

Social Security Advisory Board

Verbatim Forum Transcript

Informing Policy: A Review of Social Security's MINT Microsimulation Model

Thursday, May 30, 2019 | National Union Building



Selected Abbreviations

SSA	Social Security Administration
SSI	Supplemental Security Income
AIME	Average Indexed Monthly Earnings
PIA	Primary Insurance Amount
SIPP	Security Income and Program Participation
Numident	Numerical Identification System
DB	Defined Benefit
DC	Defined Contribution
AGI	Adjusted Gross Income
HI	Hospital Insurance
OCACT	Office of the Chief Actuary
ACS	American Community Survey
ORES	Office of Research, Evaluation and Statistics (component of DCP)
BPC	Bipartisan Policy Center
PII	Personally Identifiable Information
SAS	Statistical Analysis System
PV	Present Value
WEP/GPO	Windfall Elimination Provision/Government Pension Offset
CRS	Congressional Research Service
SOI	Statistics of Income
DDP	Delivered Duty Paid
OMB	Office of Management and Budget
CEA	Council of Economic Advisors
CBO	Congressional Budget Office
PSID	Panel Study of Income Dynamics
HISIM	Health Insurance Simulation Model
QMB	Qualified Medicare Beneficiary
SLMB	Specified Low Income Medicare Beneficiary
HRS	Health and Retirement Study
CDC	Center of Disease Control
CHS	Community Health Systems
CPS	Current Population Survey

Opening Remarks

Kim Hildred, Chair, Social Security Advisory Board

Kim Hildred >> Good morning. My name is Kim Hildred and I chair the Social Security Advisory Board. On behalf of the board, I am pleased to welcome everyone who has joined us, whether that be in person or via our webcast. Today's policy forum is entitled "Informing Policy: a Review of Social Security's MINT Microsimulation Model." The Modeling Income in the Near-Term model, known as MINT, helps policymakers understand the effects of changes to benefits and payroll taxes. Much more about MINT will follow from our other presenters.

Today's event brings together outstanding panelists, including MINT users and microsimulation experts, to discuss challenges and opportunities facing MINT and other such models and provide recommendations on MINT's future development. Following opening remarks, we will essentially have five sessions today, two this morning and three after lunch. This morning our first session, moderated by Board member Nancy Altman, will include an overview of the MINT model. Our second session, moderated by me, will cover the perspectives of users of MINT. After lunch we will hear from other Social Security model experts in a session moderated by Board member Henry Aaron. We will then hear from other microsimulation experts in a session moderated by Board member Jagadeesh Gokhale, who will also moderate our final session, where our experts will discuss their recommendations regarding MINT.

Our presenters will be seated at this table. Joining them on one side, are our Board members, Henry Aaron, who should be here momentarily, Jagadeesh Gokhale, Nancy Altman, and myself. We are also joined by our newest board member, Bob Joondeph, who is participating via webcast. On the other side of the table we have representatives from the Social Security Administration, Mark Warshawsky, Deputy Commissioner for Retirement and Disability Policy; Jason Brown, Associate Commissioner, Office of Research, Evaluation and Statistics; Mark Sarney, Acting Director, Office of Research, Office of Research, Evaluation, and Statistics; and Steve Robinson, Senior Policy Advisor, Office of Retirement and Disability Policy.

During each session, our moderators will briefly introduce the panel topic and presenters. Following the panelist presentations, the moderator will first ask questions of the panel, and then take questions from our Board members and Social Security representatives at the table, and then take questions from our in-person and online audience.

Turning to a few logistics, we have copies of the agenda, presenter and board member bios located over to the side of the room. Restrooms are on the ground floor and up one flight on the fifth floor. With respect to WiFi, there is an open network, it's called the national union building

guest, and if you log into that network, you'll be asked to enter your email. We are also hearing, though, that the coverage may not be as strong as we hoped. I offer that up as a caution. There will be a one hour and 15-minute lunch break where lunch will be on your own. We are located within the Penn Quarter, which has several eatery options within walking distance. Please see our staff if you have any questions.

We will be tweeting throughout the day, using #MINTforum, our handle is @SSAbgov. A photographer will be taking pictures throughout the day. Following our event, the day's transcript, along with the PowerPoint presentations, will be posted to the Board's website. To ensure those here and online are able to hear the questions that are asked, we ask our online audience to email your questions to events@SSAb.gov. For those who are here, we ask that you write your questions on note cards with your name and affiliation. Please raise your card, and staff will pick up the questions throughout the day. Questions will be delivered to the moderator who will then address as many questions as time allows.

It is now my pleasure to introduce Mark Warshawsky, Social Security's Deputy Commissioner for Retirement and Disability Policy, for his remarks.

Opening Remarks

Mark Warshawsky, Deputy Commissioner for Retirement and Disability Policy,
Social Security Administration

Mark Warshawsky >> Thank you very much, Kim, and good morning to everybody. I'd like to take this opportunity to thank the Social Security Advisory Board for convening this forum on SSA's microsimulation model, which we call modeling income in the near term, or MINT. Near term is sort of misleading, because it actually is a long-range projection.

I would also like to thank the distinguished group of panelists we have here with us here today. It's very generous of you to share your time and your knowledge with us. We look forward to hearing all of your thoughts and recommendations. But before we move to our distinguished panelists, I'd like to provide a little bit of background about MINT. MINT was first released 20 years ago in 1999. And actually, Howard Iams is here in the audience, and he's one of the fathers of MINT. Since that time, it has been an invaluable tool to SSA and others in examining future retirement conditions such as income and poverty. It has also helped gauge the effects of policy changes on future Social Security beneficiaries.

MINT is based on Social Security's administrative records, which are then matched to the Census Bureau's Survey of Income and Program Participation, also known as SIPP. Using these data,

and the economic, demographic, and programmatic assumptions in the trustees' reports, MINT projects 21st century retirement income, marital trends, Social Security benefits, income, and poverty.

Over the past 20 years, SSA, along with its contracting partners -- most recently the Urban Institute -- has further developed the MINT model. MINT's methodology and projections have been refined, improved, and expanded to include additional features, data, and capabilities. Indeed, this morning's session will include an overview of the model and its development, and Mark Sarney will lead that.

So how do we use MINT, both we at SSA, and others? Researchers and analysts in SSA's office of research, evaluation, and statistics -- what we call **ORES**-- have relied on MINT to analyze proposed changes to the Social Security program and to respond to requests for distributional results of proposals both from internal and external clients, and we're very pleased that many of them will be here with us today who have used MINT recently. An example is the Bipartisan Policy Center's commission on retirement security and personal savings proposal, which we will also learn more about in the first session as an example of what MINT can do.

In addition to such proposals, researchers have used MINT over the years to write papers, policy briefs, and reports on many areas of Social Security and retirement policy, which they have then presented at numerous conferences. Some of these papers have been published in the Social Security bulletin and in external journals.

You can find SSA's MINT projections regarding the effects of various retirement policy proposals by going to SSA's research, statistics, and policy analysis website, located at www.SSA.gov/policy. The website also has other MINT-based analyses, focusing on profiles of different populations, as well as links to ORES publications. Today, we have a great opportunity to take a step back, think about the MINT model, and see what we can learn from those who use it and other microsimulation models, some of which we will discuss in this afternoon's session.

We recently updated the model, we'll be putting out MINT8. But the 20th anniversary is a very good opportunity to see if we need to make small changes, larger changes, fundamental changes. And, in fact, these are the questions that we're interested in pursuing today, through today's forum.

- Are there better ways for SSA to present MINT distributional results to requesters and the public?
- Are there minor tweaks or fundamental changes needed for the model and in how SSA employs it?

- And how should we adapt MINT so that it is relevant and valuable for the next 20 years, in particular, as new types of proposals and new issues come along?

Thank you again, and I am looking forward to today's exciting program. And again, I want to thank Kim and the Advisory Board.

Session I: MINT Model Overview

Nancy Altman, Member, Social Security Advisory Board (*moderator*)

Nancy Altman >> Hi. Let me add my welcome to Kim's and Mark's. I'm Nancy Altman, a member of the Board. As Kim said, this is an overview of the MINT model. I don't think we could have found two better people on the planet to lead off today's forum. Both Mark Sarney and Karen Smith have been involved with MINT from the beginning. Mark has been a public servant with the Social Security Administration for over two decades. He currently is Social Security Administration's acting director of the office of research. Karen is the perfect complement to Mark's expertise. She's an internationally recognized expert in microsimulation. She's senior fellow at the Income and Benefits Policy Center at the Urban Institute.

Mark and Karen, we could not be more grateful to the two of you, and nor could we think of a better way to start off this program. As Kim said, Karen and Mark will each speak for around 15 minutes. We'll then have questions and comments from those at the table and conclude with questions from the audience. Karen?

Karen E. Smith, Senior Fellow, Income and Benefits Policy Center, Urban Institute

Karen Smith >> Thank you for inviting me to be here. It's really a pleasure to be here. Just as an outline in my talk today, I'm going to first describe the purpose of the model, the model history, the model structure, and give you an overview of the projections and how MINT has been used.

MINT is a dynamic microsimulation model designed to allow SSA to analyze the distributional effects of policy proposals that would affect Social Security beneficiaries and recipients of **SSI**. It has evolved from a short-term model with the initial projections going out to 2020, hence the name near term, and it was limited to the 1926 to 1965 cohorts. It's now evolved after eight versions to a model that projects out to 2099 and has virtually the entire U.S. population born from 1905 to 2068. 1905 is the oldest person in the SIPP data.

MINT captures important demographic and economic changes, including changes in female labor force participation and earnings, changes in marriage patterns, changes in pension type,

changes in health care, including expansion of the Affordable Care Act, changes in tax laws, including the Tax Cut Jobs Act. MINT projects distributional estimates of all major sources of income and MINT allows users to ask "what if" questions. What if I change taxes? What if I change the number of years in the **AIME** formula? What if I change the **PIA** bend points? What if I change the taxation of benefits?

MINT has a very long history. We're now in its 20th year. It's the brainchild of Howard Iams and Steven Sandell at the Social Security Administration. They recognized the power of the **SIPP** data matched to the Social Security earnings record for projecting future retirement benefits.

Its first version, MINT1, was designed to provide initial projections of the baby boomers in retirement. I want to recognize Stan Panis who is in the audience. Stan was responsible for doing the demographic projections for MINT1 and the Urban Institute was responsible for the earnings and economic projections. The results from MINT1 were successful, but we really needed to expand the model's ability. I remember Allen Gustman saying, "you have a retirement model without a model for retirement."

The main goal of MINT3 was to develop the model of retirement. Then MINT4, MINT5, MINT6, and MINT7 have been to improve and expand what the model can do. This is not a small effort. It's involved a huge team of people, including members from the Urban Institute, people from Brookings Institution, and RAND Corporation, and it's very much been a collaborative effort with the Social Security Administration and could not have been done without Howard Iams. He's been instrumental in its development.

As MINT has moved along, we've gone from MINT1 that was based on the 1990 to '93 SIPP panels that projected out to 2020, to the most recent version, MINT8, that uses the 2004 and 2008 SIPP panels and projects birth cohorts 1905 to 2065 out to 2099. MINT2014 is similar to MINT8, but MINT2014 uses the 2014 SIPP panel.

Each version of MINT improves on the prior version by using more recent data, projecting additional outcomes, adding additional cohorts, expanding the projection horizon, and improving the methods. MINT begins with the SIPP—core and topical module data. These include pension coverage, marriage history, fertility history, disability history, and retirement account, financial and housing assets. We get a huge amount of demographic and economic information from the starting SIPP data that's matched to Social Security Administration data including Social Security Earnings Records that begin in 1951 and Social Security benefits. It also includes information from the **numident** file that has information on mortality and immigration.

MINT projects earnings from the SIPP interview date to age 55. It does demographic projections, and it projects pensions and savings. MINT first projects from the SIPP interview date to age 55

using an earnings splicing method, which I can describe more if you have questions about it. Then at age 55, individuals enter a retirement hazard. The retirement hazard says, given my characteristics, my earnings, my health status, do I work or do I retire? And if I work, MINT assigns the earnings of workers. If I retire, MINT assigns the earnings of retirees. This includes some people who don't work and some people who have partial retirement earnings. Then given those earnings, MINT calculates the pensions, savings, and Social Security benefits for each individual in the sample. The retirement model determines, given what I would get in my pension and savings if I retired today, or given what I get in Social Security if I retire today, do I retire? This is an iterative loop that goes from age 55 until age 70.

The final process merges the demographics, earnings, pensions, savings, and Social Security benefits to generate individual and family income every year until death. MINT also calculates assets, income, and taxes.

The output of the MINT model includes a longitudinal person record and a person year file with a vast amount of information, particularly about taxes and benefits, that are used for analysis. MINT's annual income projections include earnings, Social Security benefits, SSI, pension income, asset income. It projects health insurance coverage, health insurance premiums, out-of-pocket medical spending, work expenses, child care expenses, and child support payments. It calculates co-resident income (this is income of non-spousal family members). It calculates means tested and non-means tested benefits and noncash benefits. It calculates family income, and family poverty.

We project pension assets that are built up through a series of jobs. For jobs that offer **DB** pensions, we calculate the pension benefits. We also have people who have **DC** plans, and we project their DC assets. We calculate home ownership and home equity and financial assets. We project a number of measures of asset income, including interest, dividends, rental income, capital gains, and retirement account withdrawals. We also calculate an annuity measure of asset income. If you take someone with a DC plan and you want to say how well they're doing relative to someone who has a DB plan, we want to put those two people in the same metric. We use the annuity measure of asset income for this comparison. And finally, we calculate an imputed rental value for homeowners to say that homeowners are better off than renters. For taxes, we calculate annual tax filing status, the **AGI**, taxable income, federal and state income tax. We calculate payroll tax, Medicare surtax, tier1 and tier2 taxes. The tier1 taxes are the federal income taxes on Social Security income that go into the Social Security trust fund, and tier2 taxes are the taxes on Social Security benefits that go into the **HI** trust fund.

MINT8 starts with data from the 2004 and 2008 SIPP data. This includes about 48,000 observations from the 2004 SIPP data, and about 39,000 observations from the 2008 SIPP data. We then add about 210,000 people from the POLISIM model (the Social Security model that the

actuaries run) to generate the future population. We grab the POLISIM data at age 31. We also we add immigrants projected to arrive after age 31. The total number of observations in MINT8 is about 336,000 individuals. The final MINT file contains over 11,000 variables, and there are many additional variables available on the pension, Social Security, and tax files.

We try to use the best data available for doing the behavioral estimates. Many of the behavioral equations are estimated on SIPP data matched to administrative data and the Health and Retirement Study data matched to administrative data. We also use the Panel Study of Income Dynamics for the wealth and home equity estimates. We use the Medical Expenditure Panel Survey for health insurance coverage and health and medical spending estimates. We use the National Longitudinal Survey of Youth for fertility estimates. And we use the Survey of Consumer Finances for asset validation and asset allocation and investment risk behaviors.

MINT uses minimal alignment. We do align the mortality and disability rates by age, sex and year, and we align the gross immigration numbers using the 2018 Social Security trustees' assumptions. We also use the trustees' annual wage and price growth assumptions. The future population is generated from the POLISIM model. The POLISIM model uses 2018 trustees' assumptions, as well.

We extensively validated MINT. These include comparisons of MINT with the Current Population Survey, the American Community Survey, and the DYNASIM model.

The DYNASIM model is the Urban Institute's version of MINT. We also compare MINT projections with the OCACT projections. There are many subgroups OCACT does not project, so if we don't have it, we don't validate it. But when we have it, we use it in the validation. We validate the population, educational attainment, marital status, nativity, employment, earnings, DB pensions, Social Security benefits, means-tested benefits, noncash benefits, family income, and poverty. We also validate the assets with the Survey of Consumer Finances, SIPP, the Health and Retirement Study, the Panel Study of Income Dynamics, and DYNASIM projections. And we also extensively validate the projections with the tables provided in the Social Security Statistical Supplement. The Statistical Supplement provides very detailed information about employment, earnings, Social Security, and SSI benefits. Lastly, we validate the tax data with published Statistics of Income data and published information on W-2 data. The W-2 data gives information about DC contributions. We compare distributions of MINT earnings with earnings distributions from the Current Population Survey, the American Community Survey, DYNASIM, and published SSA data. We compare MINT Social Security claiming with SIPP data matched to the master beneficiary record. And finally, we compare the DI prevalence rates, DI awards and number of DI beneficiaries in current pay status with SSA projections.

What can MINT do? MINT can do descriptive analysis of the current and future population and analysis of economic resources. This includes distributions income and benefit changes under a variety of Social Security reforms, tax reforms, pension reforms, changes in future demographic and economic assumptions. It can also do combination forms. It can do lifetime and cross-sectional analysis. It can estimate the number and characteristics of winners and losers. It allows analysis of important subgroups, including age, sex, race, marital status, immigration status. It can do projections by lifetime earnings percentiles, including just your own earnings, or including your spouse's earnings. It can do family income percentiles, and it can do numbers by poverty status.

MINT can estimate replacement rates, lifetime benefits, lifetime taxes, and tax-to-benefit ratios. MINT has been used for over 53 published reports. These are listed in citations in the MINT8 primer. There are also numerous distributional tables available on the SSA website, including analysis of population projections and analysis of prominent provisions to change Social Security.

What are the limitations? While MINT provides historic earnings and benefits before the SIPP interview, MINT does not include the population that died or emigrated before 2004, which really means that if you want to say something about the 1930 population at age 65, you're limited to people in that cohort who survived to the SIPP interview date. MINT can't really do early life events of older cohorts.

MINT projects complex living arrangements using a stylized approach. Children in MINT are attributes to individuals (there are no linkages between parents and their children), and non-spouse co-residents are imputed using a statistical match, so the model cannot calculate intergenerational transfers. MINT does not include individuals who emigrate or die before age 31. This is a relatively minor omission in the population, but it does not have the full U.S. population. The earnings splicing method (the method we use to project earnings through age 55) is fairly sensitive to recent employment and earnings trends. This particular method is not well suited for doing DI policy reform projections. The only DI projection lever in MINT is disability prevalence rate. MINT has a limited capacity to account for policy changes that are included in the estimation model. We usually accommodate this limitation by making some assumptions and generating high, medium, and low forecasts to recognize the uncertainty. Users must be careful about analysis of small subgroups. Importantly, there's uncertainty in any projection model. The uncertainty is greater the farther out you go.

Finally, the projections don't account for unanticipated policy changes. But I want to emphasize that MINT is a powerful model that can do distributional analysis and it really complements the model that the actuaries use. Thank you.

Mark Sarney, Acting Director, Office of Research, Office of Research, Evaluation and Statistics, Social Security Administration

Mark Sarney >> Hi. I'm Mark Sarney. I'm the acting director of the Office of Research in **ORES**. I'm going to start dropping acronyms that other people have introduced, because otherwise this will go on way too long and you'll all fall asleep.

What I'm presenting is a MINT case study of modeling the Bipartisan Policy Center proposal. I crunched the whole overview into my first slide here. I'm going to focus on five things: why we model it and how we model it; how we tested it; how we analyze it; and, finally, a little bit of a preview of what the results look like, but not too much. I know that the crowd is full of policy wonks who would love to get into details, and I'm going to tease you a lot, but I'm not going to pay it off. You're just going to have to live with it. There's just not enough time and you're going to get swamped in the details just from the presentation. So to start with, why we modeled it.

For those who don't know, the **BPC** proposal was released in June of 2016, so we had received a bunch of requests to model it. As soon as the Actuaries' memo came out, as soon as the BPC released the press release, we knew it was coming. This is pretty typical. We get requests for confidential modeling on a fairly regular basis, especially for major proposals. So we had already planned on modeling it as soon as we knew what it was. We usually try to anticipate these requests, because someone is going to ask, even for lesser known, smaller profile proposals, we'll get a request eventually. We will try to get started on them as soon as possible. This final bullet is sort of the theme of the presentation. It took us a year to model BPC, and the presentation will explain why. So how we modeled it. This is a brief summary of the specifications for BPC. I am not going to spend a lot of time on these, I'm not going to try to explain all the acronyms, but I just want to make a couple of key points.

For those of you who don't know BPC, it has a lot of provisions. It affects a lot of the current law rules in Social Security, especially the **PIA**, which is the primary benefit formula, the primary benefit that gets calculated. It does a lot of things to spousal and survivor benefits. It changes the cost of living adjustment. But the important things to remember are lots of provisions, lots affecting the same rules. Most of these don't affect disabled workers, which for the first time means a bunch of rules now have to be split into two versions, one for disability and one for non-disability. And finally, the lower part of the table shows the provisions we couldn't model. So those three, two of them are income tax provisions, and the reason why is kind of gets into how we use MINT inside SSA, which is not rerunning the entire model, including the really in-depth and accurate income tax model. We don't model income tax changes. We also don't know once a child reaches 18 if they're going to continue on to college or be in some sort of educational program. The extension of student benefits we just couldn't identify who would be affected, so we skipped that, as well. And finally, in addition to these being complicated provisions, but

check out the phase-in periods. They are all over the place, and several of these are overlapping for the same current law benefit rule.

Whenever we model something, we follow the actuaries' specifications for the proposal as closely as we can. We are avid readers of all OCACT memos on proposals and provisions, and we parse them like people used to parse Allen Greenspan's utterings from the Federal Reserve Board. For a complex proposal like BPC we go back to the Actuaries and we actually spend a lot of time discussing with them how the individual provisions work and how they interact with each other. This required a lot of back and forth. In fact, I think we were still going back and forth with the Actuaries even six months after the proposal was released.

When we model a proposal that has multiple provisions, we model each provision standalone, and we do this for a number of different reasons. One, to make sure we're doing it correctly on its own. We also want to know what it's doing distributionally without any other interactions. Believe it or not, a lot of these proposals' pieces will be broken up and reused by other people. So having them done individually pays off in the long run, especially if it's tweaked or updated over time. And finally, we only model the full proposal after the individual pieces are done. I think the BPC proposal had something like 11 different provisions that we could model, or 10 provisions we could model, plus the entire proposal, so it ended up being the same as modeling 11 proposals. We modeled each piece individually and then we modeled the whole proposal. That also played into why it took so long.

Karen had mentioned the MINT output file, and this is actually a key point in how we use MINT inside the agency. The red box is the MINT7 model. We're still using MINT-7. We were still definitely using it in 2016, 2017. It's built on the 2012 trustees' assumptions. It uses all scheduled benefits for everything having to do with retirement planning, savings behavior, marriage, divorce, everything. It was last run in summer of 2014, and it produced what I call the MINT7 projections, same thing as the MINT output file that Karen detailed. That is essentially a fixed data set, and we've been running all of our proposals against that since summer of 2014.

Although the model itself is very dynamic, we're not rerunning the model for every proposal. All we're doing is recalculating the Social Security benefits and the things that flow from Social Security benefits, such as income, poverty status, that kind of thing. Karen already sort of specified some of the size of the MINT7 projections. It is pretty massive. It's got over -- well, MINT8 has 11,000, but MINT7 had over 10,000 economic and demographic variables and over 300,000 people in it. And this means that since we're using it statically, there's no behavioral feedback effects, there's no macroeconomic feedback effects in any of the proposals that we model. People cannot adjust their retirement claiming. There's no adjustments of wages if the tax max goes up. It's strictly whatever their behavior was in the model continues on into all the projections that we're running as far as how benefit levels and beneficiaries are affected.

I'm going to spend some time talking about how we tested this, because it's a question that comes up a lot. We test each provision a number of different ways, mainly to try to reduce the probability that an error is going to slide through. We read the code, it's all **SAS** code, and check the logs. SAS is a little fuzzy sometimes about telling you when there's something wrong. We review the output that we run automatically whenever we run the benefit calculations. We analyze it to find what we call the X-Files, which are the really weird cases. That could mean that there's something that's gone wrong or something went right that we don't understand. And then we actually spend time analyzing those X-Files, to make sure that there either is an error and we fix it, or that we have a really interesting story to tell when someone's proposal is increasing benefits when they expected it to decrease benefits.

We also can trace the calculations of the benefits and the taxes at every step of the way for individual cases. And we do this twice. The person who programmed it and spent a lot of time living with a provision tests it and then our team has someone else test it from scratch who hasn't looked at it, so they are a fresh set of eyes. And the full proposal gets tested again, which implicitly tests all the individual provisions a third time. And sometimes we have to go back and fix things when we see the full proposal being tested. This last bullet is kind of key. We spent six months testing the full proposal. So of the 12 months it took us, half that time was spent on testing the full proposal, which was sort of a complicated beast of a proposal to model. Why was it so complicated? Because the transition periods overlapped. Because of the provisions that only affected retirees but not disabled workers. And, of course, there's disabled workers who receive aged benefits at the same time they're getting disabled benefits, and you have families where there's spouses and survivors and children and it can get really involved and complicated. Also, because certain rules were affected by multiple provisions, so the PIA changes all affect the formula and calculation of it.

The bottom part of the slide goes through some of the chicken or egg situations that happened. And if you can imagine, if you know the technical stuff, that's great. But if you don't, just imagine a rabbit hole that leads to another rabbit hole that leads to a third rabbit hole. So I'll just highlight one of them that the changes to the PIA formula, factor changes affected the **WEP/GPO** changes (Windfall Elimination Provision / Government Pension Offset), which had spawned their own set of PIA calculations for both covered and non-covered work, and they affect the working year PIA changes, or the lingo is mini PIA which was done for the earnings year.

All of those phases all had to be figured out as those things all interacted with each other. There was a lot of discussions about how to make that work. So how do we analyze the proposal results? I'm going to use a simple science fair example. Current law is the control. The proposal is the experiment. In this case, current law is there's two controls. The chart there is just illustrative of what that looks like, scheduled benefits, say \$100, payable benefits is about \$25

lower, 25% lower in around 2050, and BPC was around an 8% reduction from scheduled. So one thing that will probably become important for the rest of the day is when we model these proposals, all we're modeling is specified provisions.

There's no other adjustments made for how solvent or insolvent they are. We just model them as is. This can lead to some sort of maybe not apples to oranges, but different species of apple comparisons. So, for example, Scheduled is not solvent, but we're comparing BPC to Scheduled. BPC is solvent a certain way, we're comparing it to Payable as well, which is solvent in a very different way. We're going to get into some of the output. At first I was just going to tease you all by saying here's the output, show you this screen shot of the PDF files and Excel charts, but I'll go a little further. But when we run the calculator, which takes us about 20 minutes, it produces all these files of output. This is for comparison purposes.

The reasons we don't run the whole model is the model takes about three days to run, about 24 hours on a server crunching numbers. So that's one of the reasons we're not just rerunning the model every time we model a proposal. Most of our clients want the overview tables for 2030, 2050, and 2070, because the deep dive, it's like drinking out of a firehose.

The table shown here we produce automatically, and some of the lesser-known ones include lifetime and time series results. And for those of you who have seen our tables and are wondering why you get a table that only shows you total and maybe sex or age, nothing else, and a lot of white space, we build in our disclosure proofing, because this is Census data, including our administrative data, so that it will knock out subgroups that are too small to be shown. Sometimes we get an entire PDF that is completely empty because some of the subpopulations we run for a proposal or provision that only increases benefits, there's no one who gets a decreased benefit, so we don't show that.

Here are the actual tables from BPC. I know you're all going to be tempted to look at the actual numbers, but I'm focusing on the row headers and column headers. The first thing is when we produce these tables, we put a watermark on them because they're draft. We can provide them for publication, but nine times out of ten they're just draft tables because people are just trying to educate themselves on how proposals function distributionally. The whole point of this is to do distributional analysis. And you can see the subgroups that we've got listed, sex, age, ethnicity/race, born here or in another country, marital status, education, household income quintile, benefit type, and official poverty status.

This is really where MINT's value add comes in because it is hard to get the data you need to do these types of subgroups. This is where having the SIPP match the administrative data is really important. The two tables I'm showing are the first table, which is just a median percent change in individual benefits, and the second table is poverty effects. And these are probably the two

most popular pieces of MINT analyses that clients are interested in. The first table shows the median percent change of benefits. It also shows what proportion of the aged 60 and older beneficiaries are receiving lower benefits or higher benefits. It's not always 100 and zero as far as that breakout. In this case, for BPC almost three-quarters would be getting reduced benefits from Scheduled and 20% would be increased in 2050. Poverty numbers show the poverty rates and then the change in the number in poverty for the same subgroups.

I'm not going to go through the results too much, but there's a lot of differences by the different subgroups. The higher educated or higher income are affected different than lower educated. There's differences by marital status. And this is really the kind of information that people are looking for in our distributional analyses. I'm just going to quickly jump through this, so that we cover a couple of different populations. It's usually current law beneficiaries 60 and older. We also look at those who are just affected by the proposal in our tables, because sometimes a proposal may affect very few people, but you want to know how those few people are affected.

Our policy measures include poverty information, percent change in individual benefits and household income. And we also look at the composition and sample size for the population because that tells you a lot about why the results are happening the way they are. If you want to go deeper than the overview, we do have in-depth results where we go into four populations. There's all current law beneficiaries 60 and older, those that are affected; those who are having increased benefits; and those who are having decreased benefits. And we do about ten tables on each population with a second page of breakouts that covers more programmatic-type subgroups, like quarters of coverage claiming age, supplemental poverty measure, quintiles, in addition to household income. And for those who are really into Social Security benefit calculations, a step-by-step through the benefit calculation so you can see when the changes happen and how much change is happening at every step for worker and auxiliary benefits.

We also produce lifetime measures, replacement rates are done three different ways, and the ratio of lifetime benefit Present Value (PV) to lifetime tax PV is our money's worth measure. It is also done three different ways. The three different ratio methods generally tell you the same story. The benefit/tax ratios are just different in how they attribute the auxiliary benefits that are either paid or received by beneficiaries. We also have a system progressivity measure, which is similar to the Gini coefficient, which is much less popular than the replacement rate, or the benefit/tax PV ratios. We can expand those -- and we actually plan to -- in MINT8.

You're probably wondering, well, wow, where can I get these BPC results? You can't get them online. We don't publish full proposal results. We only do individual provisions, and the reason why is we think it's just too sensitive to publish proposal results, because individual provisions generally don't have the same kind of authorship as a full proposal, a full package that somebody has said, 'this is my proposal and we're showing a lot of the good and bad of any proposal.' We

just thought it would possibly cause too much trouble and undermine our ability to put up results at all. But we'd be happy to provide BPC results to anybody who wants them. You just need to request them. And we do publish -- I think Mark (Warshawsky) and Karen have mentioned -- we publish a lot of MINT results online for individual provisions and for populations of interest, and they call them the "who are they" tables for like spouse-only beneficiaries and low earners, and we also have explainers of current law rules like the minimum PIA, which I think is our most popular of the explainers, and other techy stuff like that.

I want to thank the Board again for letting us come and geek out over MINT all day. Thanks to all of you for being here online, as well as here in person. There's a lot of folks, you know, who have been involved with MINT from way back, including Stan Panis and Howard Iams and a lot of the SSA staff, and Steve Goss, and a lot of our clients, and it's really nice to get everyone together at the 20-year mark to see what the future of a future-projecting model is. Thank you.

Nancy Altman >> Thank you both. That was very informative. We have about 20 minutes or so for questions. As you're thinking about your questions, let me start actually with a question for you, Mark, and then Karen, you can jump in as well. You both have given us a really good sense of the strengths and limitations of the model, a conceptual understanding of it. Can you give us a little bit of a feel for on the ground?

Mark, you mentioned that it can take several days to run a proposal model. How many people work on it? What happens when the proposal comes in? That kind of thing.

Mark Sarney >> Well, first of all, I don't know if we mentioned this, but the whole model and all of our analyses are all done in SAS code, SAS datasets, that sort of thing. Everybody that works on it has got some ability to program in SAS. When a proposal comes in we've got a team of about four people, not counting me, since I'm in management and unreliable as far as time goes, who can work on proposals, and it really depends on the type of proposal. Some things can be done really quickly, or we already have done; other things can take a long time, like BPC. So usually we will get the specifications and make sure we understand them first.

If there's an Actuary memo, we follow that as closely as we can, and sometimes we've got some or a lot of questions about how they did various things. We usually have somebody who knows that area, who can repurpose another provision we've done and model the new one. And then we have someone else, like I said, test it, and that can go on for a while or can be very quick. I think some of the BPC provisions we have done within a week or two, and others we were still grinding away on for five, six months before we figured them out.

Sometimes the testing can go on for a long time because we see a lot of strange things, and partly because we want to be able to explain those when the results come out, but we also want to make

sure they're correct. So it can vary quite a bit. And, of course, the four people who are working on MINT stuff aren't just doing that full time, they have other projects to do, as well. So, you know, there's a lot of wild card variables in estimating how long it may take to model a provision comes in or a proposal, depending on what's involved.

Nancy Altman >> One more question, and then I'll turn it over to the board and SSA for comments. Karen, how is MINT like and how is it different from other microsimulation models, and how is microsimulation different from other modeling? What was done before MINT, what is done to supplement it, and so forth?

Karen Smith>> I think before MINT, most of the analysis was done using case studies. So take an example worker and project the change in benefit and examine what that means for selected outcomes. What that analysis does is miss a lot of interactions. MINT allows users to alter policy provisions and see how they percolate over time to get a different output.

I'm co-director of the DYNASIM modeling effort at the Urban Institute. And DYNASIM and MINT have grown up in tandem and they share a lot of knowledge. The human capital behind these two models are very similar. The pension and wealth provisions in DYNASIM and MINT are identical. On the other hand, DYNASIM and MINT have very different labor market methodology, very different demographic methodology, different benefit calculators. In doing the validation, I validated DYNASIM compared with MINT. I'm very reassured when both models give the same answer. Two models, different labor market methods, and different methods for doing the projections, and they get the same answer. I feel very comfortable with that, so that's been a very good exercise. My sense is that microsimulation is the only way you can get some answers because the interactions are so complicated. And I'm extremely proud of both DYNASIM and MINT.

Nancy Altman >> Thank you. And let me now turn to my fellow board members to see if there are questions. Kim.

Kim Hildred >> Thank you both. That was a very, very thorough presentation. I worked on the Hill for a number of years and was a MINT user often, and I know that we'll be getting feedback from our next panel of users. Could talk a little bit about how you've thought about and how you've changed the presentation of MINT results?

Mark Sarney >> They've changed a fair amount. The ultimate end users for MINT are people who think about and make Social Security policy. And there's kind of a spectrum there from being very technical and in the weeds to not being technical. And I think our results are probably still hard for new users to understand. Distributional analysis in general is hard to understand, because it's not a hypothetical example. There are outliers at both ends of the distribution, and

especially for policy changes, weird things can happen. We've been trying to follow what users are interested in, and so our results have morphed over time. Like I said, the overview results are for people who are less technical but still want to cover all the bases, and then there's some people who want to go very deep into the weeds. And so it's a little bit like 500 channels of cable. And that's why we output so much stuff and we still end up doing ad hoc analyses. But the biggest hurdle, I think, is trying to make them presentable in a way that are intuitive but they're also accurate and easy to explain.

Jagadeesh Gokhale >> I'm going to be on three panels later, and I don't want to use up my ammunition right here. I have five different counts on which I could ask questions. I'll just stick to one. Which is your communications methodology? How do you communicate your results? Looking at the tables you put up, I would have a hard time even motivating myself to look at those tables and analyze them. A lot of information is packed on one page. Have you thought about what it takes to communicate the complexity and draw your audience or your clients to get them to a place where they have confidence in your results? Because I think the main hurdle here is going to be how to communicate and how you convince that these results are not just -- your models are not just random number generators, to put it very bluntly. So that would be my question. What comment do you have?

Mark Sarney >> I think you nailed the issue right on the head. I mean, the communication of the results is a key hurdle, because there's so much that anybody could focus on. Trying to come up with something that will meet most clients' needs is difficult. We don't intend to demotivate people when they see our tables, but I can understand how that can happen. And even with the overview tables, that's still a lot. So that might -- it's actually a good question for the user panel. It's something we've been working on I think the whole time we've been doing this. And as we get either people internally or externally, we get different takes on the results and how confusing they are and how they can be improved. So the tables have sort of evolved over time. We've figured out the subgroups people are mostly interested in. We focus on medians rather than means, for example, was one small technical change that happened over time. And we focused more on the population than really complicated measures, because often the nature of the population is the driving factor in the results.

If you think about current law beneficiaries 60 and older, that is a group that is made up in a very different way than, say, workers age 25 to 65. And we've added and subtracted tables over time and tried to make them easier. We even added color coding to the results to try to make the general trends more understandable at a glance. Making the results understandable is something that is very much at the forefront of our mind. And trying to improve that as much as we can, that's one of the reasons we'd like to hear feedback today.

Nancy Altman >> And before we go to the audience, let me turn to Social Security Administration, Mark (Warshawsky), and if you have other points you want to make, comments.

Mark Warshawsky >> Sure. Actually, I wanted to dig into one point. I think, Karen, when you were describing building up the populations, start with the SIPP but then you go to POLISIM. I'm wondering if you could explain that a little bit more, what POLISIM is and what population you get from it. And then to even be a little more specific, you say that does limit whether you can use the model for modeling the disability program, could you explain why that is the limitation

Karen Smith>> >> The POLISIM model started with DYNASIM at the Urban Institute. It moved to Cornell with Steve Caldwell and became CORSIM. The Social Security Administration hired Steve Caldwell at Cornell to develop POLISIM. POLISIM is a microsimulation model that takes the population (I believe it starts in 1980) and projects the population forward. It does earnings projections as well. I actually worked for a while developing the POLISIM model. The actuaries still use it and it is housed in the Actuary's office.

The POLISIM model doesn't predict all economic components, like pensions, wealth, and means tested and non-means tested benefits. Those are things that MINT projects. POLISIM generates the composition of population by age, sex, race, and educational attainment. We grab them at age 31. A lot of the philosophy of MINT is that you need to get people past their education and into their career job in order to project their future. When we grab people at age 31, we're basically getting them in their career job after they've completed their education. We impute a past earnings record for these POLISIM people and then MINT projects their earnings, pension accruals, health status, disability status, marriage, divorce, death, Social Security benefits, and so on every year until death.

The POLISIM model gives MINT the right size population at age 31 including their completed education. The main limitation with respect to disability is a limitation of the earning splicing method used to project earnings from the SIPP interview to age 55. What the splicing method does is replicate the earnings and disability status of somebody who is five years older than you with similar characteristics as you. I can take somebody who looks like me at age 30, and grab them five years ago, and then they give me information from age 31, 32, 33, 34, and 35. I just splice that person's earnings, disability status, and mortality status to me to generate my future projection. The advantage of that method is it captures the joint correlation of earnings and disability and mortality. The MINT1 version failed to get that correlation right, so we moved to the earnings splicing algorithm. We align the projections to the Social Security Administration's projections on disability and mortality.

In the splicing method, we pick two potential donor records, both of which are good, reasonable projections. Then we can swap which donor record to use to hit the target mortality and disability rates. But that projection method doesn't have much steerage other than disability and mortality prevalence. For example, if we get disability prevalence from the actuary that says that 15% of the population in this age group is going to be disabled, I look for 15% of the population in this age group to be disabled. I can't change the disability rules and affect behavior. All I can do is change the prevalence rate. That's the only mechanism I have for doing that steerage.

Henry Aaron >> I'd like to step in for two things. First, publicly to thank Mark (Sarney) for providing help 18 months ago or so in developing an excel spreadsheet program with which it was possible to estimate the impact on 75-year balance of any change in the PIA formula or earnings max that you wanted. We calibrated it against estimates for specific changes that the actuaries have done, and the results were pretty much dead on. It provided a tool that I found very convenient, and I think others might find convenient similar exercises. It's a very different kind of exercise than the one you've been describing, but it was possible only because we were able to pick, in effect, a representative year somewhere between now and the 75th year and have an earnings distribution that produced results very similar to those that the actuary generates.

I tell you, it's a lot of fun to sit there and figure out what the impact on balance is of kicking the 90% bend point up or down, changing the replacement rate to 92 or 95, and getting your results instantly and with a reasonable degree of accuracy. So this is one non-mainline use of MINT that I found extremely useful.

I have a question for you, Mark (Sarney), something I think I didn't understand fully. I heard you to say that you are more comfortable, or as a matter of policy, produce results for provisions, but not for packages. Did I hear you correctly on that? And if so, really? [laughter]

Nancy Altman >> And let me ask you both to answer very quickly, because I want to turn to the audience and we're almost out of time.

Mark Sarney >> Well, the answer really quickly is we model full proposals, but we don't post those results on our website. And the reason why is, our sense from talking with clients and others is that if, you know senator so-and-so has a proposal, and we publish the results when the actuary's memo comes out, there might be a lot of blowback from that senator about our results. That could lead to us maybe not being able to publish any MINT results. So we feel like we're kind of walking a tightrope between publishing everything and nothing. We've kind of roughly drawn the line at publishing individual provisions proposed by either the SSAB (Social Security Advisory Board) or other nonpartisan groups that are commonly discussed, like an NRA

increase. So it's a more limited subset of kind of the top ten hits of provisions, but not entire proposals.

Nancy Altman >> We're being live streamed, so we may have some online questions, as well as questions in the room. I was supposed to get notes, but I guess, for the sake of time, Diane, maybe just hand them a microphone. That might be easier.

Jack Smalligan >> Could you talk a little bit more about the capability of MINT8 and specifically like what sorts of policies you could model with MINT8 that you just weren't able to model with MINT7?

Karen Smith>> The main difference between MINT7 and MINT8 is that MINT8 has expanded the set of variables that it calculates in order to better model supplemental poverty. We added child care expenses, work expenses, and child support payments. We added home ownership with a mortgage and homeowner without a mortgage that are used in determining the supplemental poverty thresholds. Work expenses, child care expenses, and child support payments are things that retirees don't generally face and were not a priority in earlier versions of MINT. But I think SSA is interested now in looking more at characteristics at younger ages, and so we've expanded the set of variables needed for calculating supplemental poverty.

We also specifically modeled covered and uncovered pensions so that the Social Security Administration can better model the WEP and the GPO. We updated the tax calculator. MINT7 used the 2001 Statistics of Income data for doing its tax projections. We now use the 2009 **SOI** and we've modeled the provisions in the Tax Cut Jobs Act. We updated behavioral estimates using more recent data. Other than having these additional projected outcomes, the capabilities of MINT7 and MINT8 are basically the same.

Mark Sarney>> One thing I'd add to that is that, to piggyback on what Henry Aaron mentioned, while we can use MINT for a variety of kind of off-brand ways, I guess, it really is just a complement to what the actuaries produce on the cost side. And I like to say that we're the flip side of the coin. They can tell you the cost of something, we can tell you the value to beneficiaries. So Henry had us doing stuff that we've never done before. But it was an interesting exercise. MINT8 is not production yet to model proposals but we're getting there. We're working on it. We have the static MINT8 projections in-house, the model is finished, and now we're figuring out how to use the projections and get ready to use it to model policy changes.

Nancy Altman >> Thank you. Well, I'm sorry that we ran out of time for more questions. But the good news is that Mark and Karen are going to be here all day and we're about to take a break, so you can ask them questions now. Let's have a round of applause, both to thank them for their presentations, but also for your work.

Kim Hildred >> A quick reminder in advance about questions. We want to tell our online audience if you have questions to please email them to events@SSAb.gov. For our guest here, please grab a card. We have people in the front that help with our sound and people with mics so the question may be heard.

Session II: Perspectives of MINT Consumers

Kim Hildred >> Welcome to our second session of the day where we have a distinguished panel of experts. Kathryn Olson, Staff Director, Social Security Subcommittee, Committee on Ways and Means in the House of Representatives. Next to Kathryn is Amy Shuart, the Republican Staff Director, Social Security Subcommittee, Committee on Ways and Means. Next to Amy is Tom Klouda, Senior Domestic Policy Advisor, Committee on Finance in the Senate. Next to Tom is Kathleen Romig, who is the Senior Policy Analyst, Center on Budget and Policy Priorities. Next to Kathleen is Andrew Biggs, Resident Scholar, American Enterprise Institute. Next to Andrew is Laura Haltzel, she is Research Manager, Income Security, at the Congressional Research Service. And next to Laura is Jack Smalligan, Senior Policy Fellow, Income and Benefits Policy Center at the Urban Institute. So, welcome one and all, again, we really appreciate it. And we will start with Kathryn.

Kathryn Olson, Majority Staff Director, Subcommittee on Social Security, Committee on Ways & Means, U.S. House of Representatives

Kathryn Olson >> Thank you for having me here. I have already found the first panel very interesting and lots of food for thought and notes to myself to call up the MINT people after this conversation and ask them some questions.

First of all, I am glad to be here, and I am speaking for myself; not any member of Congress. I have three main points and then just a couple of thoughts moving forward. First, it is very useful to have distributional analysis as an element of the policymaking process. Distributional analysis is critical especially from our perspective. It goes to the very purpose of Social Security, which after all, is not about having a particular cost rate or trust fund ratio, but is instead about providing a level of basic economic security for all workers and their families. That is the purpose of Social Security so that is why distributional analysis is so fundamental. I also see distributional analysis as a complement to stylized examples of the kinds that the actuaries and others develop. The stylized examples are very straightforward. Distributional analysis doesn't replace that. Both are useful.

Second, models are incredibly complex and some can be more of a black box than others. So to evaluate them I look at things like -- and I am, of course, not a modeler -- what is the

transparency? What is the accountability? What information is publicly available about the model?

What about the assumptions that are in the model? What kind of oversight is there of the model? Is there accountability? Another question is what is the level of the experience of those who built, maintained, and developed the model? Is it a broad team like MINT, or is it concentrated in one or two individuals? What is their level of knowledge about Social Security benefit rules, which are very complex, and numerous, and have all kinds of interactions as the BPC illustration illustrated. And also what is their knowledge about the populations being studied?

Have the modelers demonstrated their neutrality over time? Is the model subject to manipulation to achieve desired results? And finally, does it link survey and administrative data, which is, of course, very powerful. And on these points I would say MINT is not at all a black box. It is well-documented. It is linked to administrative data. And it has an extensive track record and a very large team working on it. I think MINT is particularly useful for distributional analysis, as I said, which is not available elsewhere.

I would not look to MINT as a congressional user for cost estimates or impact of plans. And I also appreciate that MINT can examine income of households in light of their total income, which is another useful perspective, not just beneficiary income or beneficiary's benefits. By contrast, I had an experience recently with another model that was not as positive. It was looking at the impact of **DDP**, a bill that's before the committee and one round of preliminary analysis was released, but just a few weeks later the modelers realized that they misunderstand a provision of the bill again due to the complexity of Social Security proposals. And also apparently changed the whole underlying model itself, which is sort of remarkable, and they published revised results and both the size and order of magnitude changed in that revision, so this doesn't give a lot of confidence about at least the sort of well-developed state of that model. And so that was sort of a more negative experience about modeling.

Three, going back to MINT, despite the many strengths of MINT, I am going to confess that I don't use MINT directly for our policymaking at the committee because of the lack of confidentiality because even though there is a certain kind of confidentiality, apparently, that results are not published necessarily. I know that within the administration it's not confidential and this is not with respect to who is in the White House or not, it is really a concern. We in Congress value our ability to, with discretion, develop policies without even our own team knowing what is going on. So that is not about who is in the White House.

So the lack of confidentiality. And then, again, I don't know if there is interference by political appointees. I don't want to allege that, but again it's about how much of a firewall is there between MINT and other parts of the administration, so regardless of who is in the White House.

But one thing I learned today is that apparently if we call the MINT folks they can give us individual provisions which I think is kind of a nice way to preserve some of the confidentiality of people who make proposals, but if there are provisions that are widely -- you know, and many proposals are offered by neutral outside entities, that is helpful and promising. So expect my call. So, third, going forward, I just want to reiterate that MINT is very useful and important work. They have a strong track record. But I recommend that SSA develop protocols for who has access and the confidentiality, if I could write the rules myself.

**Amy Shuart, Republican Staff Director, Subcommittee on Social Security,
Committee on Ways and Means, U.S. House of Representatives**

Amy Shuart >> Good morning. My name is Amy Shuart, as Kim said, and I am the Republican Staff Director for the Social Security Subcommittee. Like Kathryn, my views are my own and do not reflect necessarily any of the members that I work for.

Policy development is a critical part of the legislative process and members and staff rely on the experts at SSA to help inform their thinking. We rely on many tools, briefings on program policy, research, agency statistics, and stakeholder feedback. We also rely on MINT. Whether something informs our thinking depends on three key things and these are really where I want to focus most of my discussion today: whether something provides insight into a policy question; whether it is easy to understand and to explain; and whether it is timely.

MINT can be a really important tool to help inform policy development. It provides insights that are not available from other sources, like the Office of the Actuary. I think the point was made quite well this morning, that MINT is a complement to the Office of the Actuary and the information they provide.

The program explainers and the explainers on SSA's website are really valuable tools and certainly provide a way to get an insight into some of the power of MINT. In the past, we have used MINT to develop targeted benefit options because it really does give you a great way to look at how a specific policy affects different cohorts of folks and different types of individuals. So, if you're trying to target a certain group of people you are going to get a better sense of that cohort by using MINT than by using stylized workers from the actuary.

Unfortunately, available analysis also can drive policy designs. One of the nice things about MINT is it gives you information cut in different ways and helps you think more broadly and not necessarily in certain stovepipes which can often happen. That said, when considering how to improve MINT, I think the big overarching question is: what is the ideal use of the model? Because if the model is supposed to be informing academic research, then maybe the barriers to entry aren't as big of a deal. But if the point of the model is to inform policy development, then

we really need to be talking about the timeliness of MINT and how stakeholders are able to use it. As you're going to hear, I think, on this panel, different stakeholders will have different needs and unfortunately often these needs are going to be mutually exclusive. And I think that is a challenge for MINT over the coming 20 years.

As I see it, the three challenges facing MINT are: confidentiality; timeliness; and barrier to entry. With confidentiality, as Kathryn alluded to, technical assistance at this time from the MINT team, who really are fantastic pros, is coordinated through legislative affairs. And several years ago, that wasn't actually the case. I was able to go to what was the Office of Retirement Policy directly. I did not have to talk to legislative affairs before I did that. That's a problem for me because often when we are thinking through things, as Kathryn said, we are not always talking to our own team in the Administration.

VIDEO CUTS OUT

and we know that that's a confidential process. Certainly, are not going to be talking to political appointees about an idea that's not ready for prime time.

MINT is a really important tool that helps us think about what ideas should be ready for prime time and so the confidentiality and who has access to the requests really matter. This is very different in its current treatment than the Office of the Actuary, where we can go to Steve and his team directly - they are able to answer our questions, and we know that that's a confidential process.

VIDEO RESUMES

With timeliness, turnaround time really matters for the policy development process because decisions can move very quickly. We heard from Mark that the BPC proposal, while very complex, took a year to model. A year is a lifetime on Capitol Hill and certainly in the house where members have a two-year cycle. I cannot wait for an entire year to get full analysis of a plan so if the point of MINT is for policymakers to use it, we need to go quickly. And then finally, the barrier to entry of MINT is high. Mark showed you the volume of tables. Reading the tables is a skill and it's not a skill you learn quickly, in part because MINT isn't as well-known as some of the other tools. I think it's interesting if you look down this table most of the folks on this panel actually worked on MINT at one point. I am one of those people. That's concerning when you are thinking about whether you are an actual user. And then finally just finding the information because it is not necessarily easy to find. Thank you.

Tom Klouda, Senior Domestic Policy Advisor, Committee on Finance, U.S. Senate

Tom Klouda >> Yes, thanks for the introduction. Thank you for inviting me. Same disclaimers, that my views are my own. It is curious as to why I was invited. My guess is because I did not

have a good MINT experience and they probably wanted me to talk about it. It did have a happy ending, though. We used MINT last year. My boss wanted to improve Social Security and also alleviate poverty among elderly single women. Our team developed a proposal called the Elderly Poverty Relief Act. It was a very simple bill. At least you would think it would be simple. The bill created a new benefit which was about \$85 a month and that benefit grew every year by the growth in wages, so it went up more than the COLA. The benefit went to: Social Security beneficiaries over the age of 82 or SSI recipients at the full retirement age-Social Security and SSI recipients who had received benefits for 20 years, and Social Security beneficiaries at the full retirement age, but who had very low benefits. For you technical people, that would be a benefit below the first bend point, which is about \$950 a month. We used MINT because we wanted to see what the impact on poverty would be.

According to MINT, the bill would reduce poverty among seniors who received benefit by about 25% in 2030 and that was lifting about 420,000 seniors out of poverty. Now, this is a good statement about the bill and a good policy outcome, but it took about a month of back-and-forth emails between me and Mark and his team to nail it down and document that statement. I won't go in to the nitty-gritty details, but some of the issues with MINT we encountered: there was some confusion between the MINT results and what we had from the actuary.

I understood that MINT does not cover the entire "Social Security area" so that may have caused a little bit of disconnect.

VIDEO CUTS OUT Also, since we were doing both Social Security beneficiaries and SSI, my sense was there was sort of an either/or: so that under the bill we could say 50% of Social Security beneficiaries would be lifted out of poverty or we could say 25% of all elderly would be lifted out of poverty. We decided to go with all elderly since that seemed more relevant. We

VIDEO RESUMES tried to eliminate the confusion and frustration. And Mark's team wrote me saying, "we are concerned that we have totally confused you and you are under your desk with a bottle of grain alcohol or perhaps turpentine." [laughter] I saved all those emails, of course. [laughter] And although that wasn't true, I appreciated the sentiment because that's about how I felt. But I would say, and certainly to echo some of the other experiences, MINT could be improved to make it more user friendly.

I did not work on the MINT team and I found the output difficult to work with. The two things I think that would be most helpful - timeliness is important of course -, but a "top line summary" about the bill saying the bill did x, y and z, in plain English right up front would be helpful. Or if the MINT team noticed something that is very interesting in the output, the MINT team should summarize that up front. Those are my two comments. I hope that is helpful. I look forward to the rest of the discussion.

Kathleen Romig, Senior Policy Analyst, Center on Budget and Policy Priorities

Kathleen Romig >> Hi, I am Kathleen Romig. I am, I confess, another MINT alumnus on this panel. And I have used MINT in my work for about over 15 years now, both at SSA, when I was at **CRS** assisting congressional clients on in their proposals, here at the Social Security Advisory Board, and now I'm outside of the government the Center on Budget and Policy Priorities. So I have perspective on MINT from a number of different vantage points. And across all of those vantage points MINT has been a really invaluable tool. And I am excited to talk more about that today. So, first, I just want to say why I think MINT is so valuable for policymakers and the public and then I want to share a couple of recommendations which may sound familiar to you from the other panelists.

First of all, it is obvious, and it bears repeating, nearly every American is a Social Security beneficiary or will become a Social Security beneficiary in the future. When we make changes to this vital program, we have to understand how it affects people. Not just trust funds, not just federal budgets, but people. This is really important if you have voters to answer to or if you are a beneficiary or future beneficiary yourself. And as you've heard, MINT is designed to do precisely that.

I won't belabor some of the points that you heard earlier this morning, but it is incredibly rich data at the individual level based on multiple datasets, thousands of real people. So this allows for very detailed breakouts of subgroups that are of interest to policymakers like, for example, aged widows, or poor seniors, or African-Americans. So that's one major advantage. The richness of the data allows for these really policy-relevant breakouts.

A second advantage, especially over more stylized models, is it can model families. So not just the worker, but the auxiliaries like spouses and survivors. And again that's something that's a real advantage of microsimulation. You can see all of those complex family situations and how the auxiliary benefits flow and what those interactions look like.

And then my final point about why MINT is so important is as you have heard it has a stunning number of Social Security rules incorporated. And basically everything a policymaker would care about when it comes to older Americans. So as we all know, Social Security rules are really complicated and they interact in often unexpected ways. In my experience using MINT, often requesters were intending to do one thing. For example, help a particular group, or protect a particular group, and that is not what happened. Not after all the rules were enacted. So it is important to really reflect the complexity of the program in the model that you are using and explain that complexity so people can understand, and tweak and iterate and fix their proposals. So that's why I think MINT is so important.

Now, because I think it's so important, I have two recommendations to increase access to MINT for policymakers, researchers, and the public. First of all, I would love to see MINT results published on SSA's website exactly the way that the Office of the Actuary publishes their analyses on their website. Every single public plan, whether it's a bill, or a commission proposal, or something from a think tank would be featured on the website. Every provision within those plans would also be broken out.

Now, I want to respond directly to what Mark Sarney said about the sensitivity of doing this - I strongly disagree. You drop a bill, you get a score. Period. That is how it works with actuaries. You don't get to opt out of a CBO score. You should not get to opt out of taxpayers' understanding how a proposal would affect them, not when they have been paying for this research. That is unacceptable to me.

Right now what the SSA website has is just a small selection of that taxpayer financed research. There are no published analysis of bills or comprehensive proposals and the provisions that they have are largely generic options that honestly are pretty out of date. They're taken from years' old reports by the SSAB and by the National Academy of Social Insurance so they're missing a lot of the proposals that are a part of the policy conversation right now.

The second thing, it's no coincidence, as Amy mentioned, that there are so many MINT alumni on this panel because the only way a person can access those key MINT results beyond the few things that are on this site is by basically being an insider. You have to know to request it. And I don't think that is okay. I think that this should be publicly accessible to all policymakers, researchers, and the public.

My other recommendation very much echoes my friends from Capitol Hill that you need to keep the policy deliberation process truly confidential and that means that political appointees in the Office of Legislative Affairs should not be involved in requests of results before a proposal becomes public. Often policymakers and researchers really need to iterate those plans, deliberate, and they're not sharing it widely. And in order to do that with confidence they need to have that level of confidentiality. Again, this is standard operating procedure at the Office of the Actuary and we should extend it to MINT.

So just to sum up, MINT is really valuable. We really want to encourage evidence-based policymakers. It is an extremely important tool that taxpayers have funded, and we have the right to know what results the model has produced. Thank you.

Andrew Biggs, Resident Scholar, American Enterprise Institute

Andrew Biggs >> Thanks very much for having me today. As always, I agree with almost everything Kathleen says, which re-scrambles the things I was going to say. I was a great consumer of MINT work in the mid-2000s when working on Social Security reform, but I think it is even more relevant today. There are widespread concerns for the adequacy of Americans' retirement savings. A majority believe there's a retirement crisis. Congress is working on various legislation to increase retirement savings. States are working on automatic IRA plans. At the same time there is legislation in Congress with I think even more co-sponsors than I have ever seen to address Social Security insolvency. MINT is the tool you need to analyze these things.

Here I am going to switch into being a little bit more critical to a certain degree. In a broader discussion of retirement income adequacy, MINT which is the best tool available for analyzing, has been kind of missing in action. There are very few MINT projections available. Little analysis of provisions, or at least recently.

The last available baseline projection for MINT of future retirement income was from MINT6 going back to 2012 from a paper from Howard Iams, Karen Smith, and Barbara Butrica. I have relied on that. To give you an idea of how much that has influence, I am not a great believer that we have a retirement crisis, and one reason for that is I looked at the MINT projections. If you look at them, they include a measure of replacement rates for total retirement income, where total income at 67 as compared to the inflation adjusted average of pre-retirement earnings, which is a pretty reasonable way of doing it. What they project is that, replacement rate--total income retirement replacement rates for generation X --will be slightly higher than they were for the depression era. This sort of result, when we see this in contrast to other studies, and say, well, 55% of Americans are under-saving, 90% of Americans are under-saving, is a very, very different result. And it strikes me as troubling that at a broad range that we put probably millions of dollars into this model, it is the best available model for projecting these kinds of things, and I am about the only guy who knows about it. And, I mean, well, maybe outside of this room. [laughter] It's extremely important, at the very least to put out the baselines to give some idea what is our best guess of where we're going to be in retirement if we don't do anything, and then we can start thinking about where we want to go from there.

It is also a crucial tool for looking at Social Security reform. Here I am going to reference the Social Security 2100 Act because I think it has elements that MINT can play into. If you look at the actuarial analysis of that plan, it projects very large benefit increases for low-earning workers. For the very lower stylized earners, it's a roughly 40% benefit increase. For the low earner who makes about half the average wages it's a 23% benefit increase. But we also know that many of those very low earners are getting SSI. If you increase your Social Security benefit, their SSI benefit reduces dollar for dollar. We also know that many of them are receiving auxiliary benefits. If you increase their retired worker benefit, their auxiliary benefit declines by a dollar. So they only receive a benefit increase if their higher earning spouse receives one.

These are the sort of interactions that MINT has made for us to figure out. So it really is crucial to understand this.

At the same time going back to some of the work I did with Glenn Springstead, where we projected retirement incomes going on to the future, MINT projects that very low earners already have very high replacement rates. So, we have got a situation where we don't have a good grasp of our baseline. We don't have a particularly good grasp of how our policies might change the baseline. We have a tool which is perfectly designed for helping us in this. All we need to do is use it more. So I would urge

VIDEO CUTS OUT

just in the short time I have left in terms of I know we're very concerned about how we produce the outcome, how to make it understandable. At the very least, I would do what Urban does with the DYNASIM where they have essentially a data dump and they have everything you could possibly want in it. You can still think about producing it for policymakers but get the data out. This **SCS**, whatever, doesn't think about how we present our data. They get the data out there so researchers could use it, so at the very least I would go with that. Thanks very much.

Laura Haltzel, Research Manager, Income Security, Congressional Research Center

Laura Haltzel >> Good morning. My name is Laura Haltzel. I'm with the Congressional Research Service. And for those of you in the room that may not be familiar with CRS, I want to share a little bit because it is the lens through which most of my comments will make sense. Our only client is Congress. Our mission is to contribute to an educated national legislature.

VIDEO RESUMES

For that purpose, our analysis is required to be timely, it is required to be authoritative, non-partisan, objective, and authoritative.

Congress right now has access to information on the -- cost of a Social Security proposal either through CBO output or through the Chief Actuaries' publications. But a truly informed national legislature also requires information on how the proposed legislation is going to affect the taxpayers that contributed to the system as well as the beneficiaries of the system today and in the future. Now, CRS, unfortunately, does not have the capacity in-house to produce anything along the lines of what the MINT model can do. I'm not going to reiterate the examples that have been given down the line, but we are limited to what has been discussed as using the case simulation approach based on the stylized earners. We can provide examples and that is about it.

In terms of what would make MINT more useful to us at the Congressional Research Service, and I am going to go in order of what I think is most to least important, and from what I hope is most feasible, to maybe more pie-in-the-sky ideas for MINT in the future. The first one is to, as you will hear down the line, increase the access to and the timeliness of the results. I think this is a rare situation for consensus in the nation's capital. And so I just want to make a note of that. The other suggestions I'm going to get into are a little more technical. MINT has a lot of capabilities already and I'm extremely excited to hear about MINT 8, but one of the areas where I would like to see some additional capability is interactions with other means-tested programs.

Currently, the effects of many income support programs, I'm just going to use TANF as an example, are static. So if Social Security benefits are increased under a proposal, there's nothing in the calculator right now as it's run at SSA that would account for a decline in TANF benefits with an increase in Social Security. Trying to replicate all of the relevant program rules would be almost impossible. There is a model at Urban Institute called TRIM that does a lot of this kind of work. Given that Urban has also put a lot into the MINT model, I would hope maybe there's a chance for taking advantage of some of that knowledge already existing to try and build capacity.

Another element that I am eager to see is additional behavioral response for claiming behavior and work. The reality is that a lot of goals of legislation are to encourage or to discourage individual behavior. So it would be really helpful if MINT were able to incorporate behavioral feedback - how a policy change might lead to a behavioral change that would actually affect the outcome of the distributional results. An example would be increasing the payroll tax, how that might affect the incentive to work and thus the payroll tax incidence, for example. I know that this is a very challenging piece to take on given the massive underlying piece that is done at Urban versus SSA, but maybe we can start with claims and work behavior.

The last additional capacity I would like to see built into the calculator at SSA is the ability to replicate the taxation of Social Security benefits. The inclusion of that is critical to a true distributional outcome of a full policy proposal. You might increase benefits, but if that same individual on the backside has some of those benefits reduced through the federal income tax on Social Security benefits, the net outcome isn't necessarily what you are going to be seeing through the MINT results. So those are my points for what I would like to see for the future of MINT. And I will pass the mic over to Jack Smalligan. –

Jack Smalligan, Senior Policy Fellow, Income and Benefits Policy Center, Urban Institute

Jack Smalligan >> Thank you. I'm actually here not to speak from an Urban Institute capacity but from working at the Office of Management and Budget for 27 years. And most of those years Social Security was part of my portfolio, and so I worked on Social Security solvency issues

over the last four administrations. And it wouldn't be surprising, I think, given just the nature of OMB's role of providing kind of policy advice to the National Economic Council and the CEA, that we don't have the confidentiality issue that other panelists described. I want to say that having worked for multiple administrations, different parties, we never had an issue where we asked for MINT results and the particular plan that the administration was thinking about somehow became more widely known because of the MINT request. So confidentiality was not an issue for us and we really kind of take pride in SSA's research office for having been able to maintain that. I'm very troubled by the comments the other panelists made in terms of difficulties congressional offices have on that, and I certainly think that is an area to explore. And I don't want to throw out a proposal spur of the moment,

VIDEO CUTS OUT

but I can't imagine doing what OMB did over the years if we had not had that confidentiality confidence with the SSA staff.

In terms of challenges with MINT, I think on the timeliness, again, the White House doesn't tend to have that problem, and will be the first customer that SSA will serve, and OMB's role was kind of build up the knowledge base so we could simulate as many scenarios as possible. So when a particular request came from the White House [inaudible] -- to respond to those as possible.

VIDEO RESUMES

And then we also were very lucky in having a former ORES staffer on our team, Pat Vinkenens. And Pat was able to absorb the data and distill it down for those of us, like myself, who've never personally worked on MINT, unlike the other panelists. And so without having somebody like Pat on the team, I can't really imagine us being able to produce some of the things that we were able to do. And so I think there's a question there for SSA that sort of, as they look ahead, I mean, the demands for Social Security solvency work are so volatile, they can be very quiet for years, and then suddenly a new president or new congressional leadership can be asking for a dozen things.

A variety of people have written about the negotiations between Clinton and Gingrich during the late 1990s, so I think that's in the public domain, and so easier for me to talk about. But the intensity of the work that went on at that time is just hard to describe, I mean, you had both parties actively looking at a potential solution, and making sure that SSA has the capability to respond, even though it might not be happening today, but in a year or two years. So I think it would be really interesting this afternoon to hear a little bit better about the overall staffing capacities that SSA has in this area and how they compare to what they had say ten years ago, and how that impacts their ability to be timely on this. Because it's not credible to have every congressional office have a former MINT person like Pat. [laughter] Some of the capacity has

to be at SSA and the people have to have the time to be able to work with the congressional office or OMB, like in our current case Pat is retired, we don't have a person on board. So it's, I think, an ability for them to be flexible in that kind of dynamic environment.

In terms of building further capacities, given that I spend a lot of time thinking about disability policy, it wouldn't be surprising that I think it would really be useful. But I recognize the technical challenges to being able to model more interactions on the disability side.

But thinking ahead, just on Social Security solvency scenarios, we have to be mindful of the interaction of disability and retirement when we're thinking about solvency. And then we also have to be, as others have noted, be thinking beyond just SSA. So overall pension policy, MINT is the ability for the whole administration to think about overall retirement security to modeling out IRA's and tax issues is very important, so we have a comprehensive role. So it's important for SSA not to just think about their programs but to think about the overall priorities of the administration.

Kim Hildred >> What we're going to do first is direct a few questions to the panel before I turn to the board and our SSA colleagues and then the audience. So first of all, thank you so much. Great perspectives, some great feedback, some common themes. But I wanted to first give you the opportunity to comment -- knowing how respectful you all are of each other. Does anyone have any comments or questions they'd like to ask another panelist based on what they have to say? Any agreements or disagreements? Kathryn?

Kathryn Olson >> I would say that I'm not aware of MINT people leaking. But the confidentiality is, let's say I proposed a 50% cut in Social Security benefits and wanted to see what that resulted in. That would be very explosive coming from actually the

VIDEO CUTS OUT

democratic or the republican side, I think. So I wouldn't want anybody to get a whiff that I was willing to think about some nontraditional options. So that's where the confidentiality -- [inaudible] it's about the ability to explore things freely without anybody knowing about it besides my team and the internal folks. So it's not about leaking, it's about who knows what I'm interested in.

VIDEO RESUMES Kim Hildred >> Yes. And for Mark, that legislative affairs screen is tough. I've had in my past many conversations with legislative affairs about that. So as you know, I was blessed with my staff, as Amy is a MINT graduate and Sophia Wright on my staff is a MINT graduate. MINT reports would arrive that were 200, 300 pages, and thank goodness I had staff experts because the data was very complex. Any other comments to each other? Anything that each other has said? It might be helpful for the Social Security folks to hear examples of your

own. I think Kathleen did a great job of really talking about the individual person impact. I know there were oftentimes when, as we were thinking about policy options, we would get the MINT results and see it and say, 'Oh, gosh, no, we can't pursue this policy, forget that.' But are you able to just share some examples of how MINT really helped you in terms of your thinking about any particular policy?

Kathleen Romig >> Thank you. I can pick up on another point of agreement between Andrew and me, which is MINT is not only useful for analyzing prospective policies, but also just for looking at the current population. And one type of request that I sometimes make is what the MINT team has dubbed the "who are they" request. Just trying to get a handle on what the baseline population looks like. So, for example, I'm not very happy with the most common minimum benefit proposal that's proposed, because it doesn't target well enough and it doesn't move the needle on poverty well enough. And so to try to investigate why, I need to really get a handle on who are the elderly poor. So I requested just baseline tables of who are the elderly poor. And so just to give an example, you know, that most common proposal is geared to improving benefits for a long-term low earner. Well, when you look at who is poor in old age, those people tend to have intermittent work histories, and that shows up in the MINT data where you see, you know, half of those people have work histories of 20 years or fewer, if I'm remembering correctly offhand. And so just trying to get a better handle, well, if you're trying to move the needle on poverty, then maybe you should not frame your policy around a long-term low earner when you know that poor people tend to have had interruptions in their work for various reasons. So that's just one example of how, you know, just digging in, like Andrea said, into those baseline results can really help you reconceptualize where you want to go with your policy proposals. And then, of course, testing out what those policy proposals would do is really important, as well.

Jack Smalligan >> Yes, definitely. I think we're a whole lot smarter on minimum benefits than they were 10 or 15 years ago, and definitely agree with that. But I think it also points to a broader aspect of the distributional analysis is, I mean, a policy can be fine for 90% of the universe. And if it's really bad for that 10%, I mean, that is a serious problem. And so I think we spent a lot of time not making sure that a policy was overall good for most people, but just kind of focusing on those smaller groups and then understanding, and understanding is it an anomaly with the runs, or is it actually truly a flaw in the policy. And I think that level of analysis takes a lot of time and MINT can provide you the output, and that's why you need so much output. But it's so vital because 10% is still, I mean, millions and millions of people.

Kim Hildred >> Great. Anyone else? All right. My next question, Laura, this going to be directed towards you. I know that when you're on the committee staff you're very focused on the committee members and their needs, and your Chairman's needs, so that leaves a whole lot of Members of Congress out there without necessarily the help and support they might otherwise

get, or maybe ideally would like to receive. And so I know that often puts CRS on the front lines. Could you talk a little bit about, how you work with Members' offices and maybe how you use MINT to help illustrate certain scenarios to help them to understand the impacts of what they may be thinking about?

Laura Haltzell >> So -- sure. So that happens rather frequently. Obviously, we work both with the committee staff and with individual members' offices. From the MINT perspective, we have not yet done something along those lines. But we used to have access to the DYNASIM model way back when at **CRS** which came up, Karen mentioned earlier, which is the sister to the MINT model. And we were able to produce full reports that outlined a series of very common proposals that would achieve solvency and to demonstrate the distributional results. We put that out in anticipation because a lot of what we do is responsive, but a lot of what we try to do is anticipatory. And if there are going to be questions on a particular subject, we want to have a report out on that before the questions come in. So for me if I'm working with MINT right now, what I do is I put in a request in anticipation that there's going to be questions coming on a particular subject. So if bills are being discussed, I'm going to be asking the MINT team to share results on that so that we can be prepared, not just for the reports that we put out, but a lot of the work that we do is confidential. And so we will produce memorandums that are going to be one-on-one with either the committee staff or the members' offices that are not shared on the CRS website.

VIDEO CUTS OUT

That's where a lot of that very valuable work takes place. [technical difficulties] access to MINT result for that really drives a lot of what I think about when I'm trying to decide which way to answer is, a lot of times it's through MINT. I am fortunate that at CRS, we work for all of Congress. So when we send a request to SSA for MINT output

VIDEO RESUMES

it is not going to be clear who is asking for that result. So right now I would say we are in a very good position to help anyone who is concerned about confidentiality, because you have hundreds of members and we work for every single one of them.

Kim Hildred >> One last question for all of you. A common theme has been the presentation of the material. You know, I think Tom had the suggestion for a top line overview of the results. Clearly that's an area that I know Mark talked about in terms of needing some suggestions. I hope our other panelists will also be challenged to think about that and perhaps offer some advice in their panels to SSA. Are there other suggestions from users for SSA to help make the reports more easy to understand, more impactful, more easy to read, bottom line.

Laura Haltzell >> Just to follow on that. One of the -- and I have to confess, I, too, am a former user of the model. I have been through **SAS** code and all of that fun with the output. So I'm an educated consumer of output. And I still will periodically find it challenging to discern whether the output I'm seeing is because the policy change is working or if there's an error or how to maybe even interpret some of the tables that are coming out.

One thing that I think would help is for SSA to produce a consumer user guide to their standard set of tables so that people can begin to parse out for themselves what the meanings really are. Let me give just one example. The beneficiary status table, for example, it may not be clear to some if people are moving between categories because they have gone from being a dual beneficiary to being just a worker beneficiary, or if there are actually genuine increases or decreases in benefits occurring. So having something that would for each of the key tables provide a translator for what these results could or could not mean I think would be very helpful.

I do want to share that the provision-by-provision breakouts are particularly helpful. A lot has been shared about how we try to use policy to target certain populations. While the overall tables are very helpful for the full model, full plan, I think sharing the provision-by-provision tables in their complexity would be very, very helpful. Also, the new present value tables that are available both on the benefit side and on the tax side have been particularly illuminating at showing the burden of both tax increases and benefit increases and losses, both by, I would say, cohort and particularly by variations in AIME quintile. So just a few thoughts there on presentation.

Andrew Biggs >> When I spoke, I spoke in favor of the data dump approach. And I think Laura's comment made me think that I'm right in the sense that, for me I didn't think the provision-by-provision analysis was very helpful. Laura just said it is, in fact, very helpful, which shows more broadly, there's just a lot of different -- the analysis has to suit the purpose. And I get the idea you want to have a standard output because you need something to pump it out. We produce a lot of information, you can highlight things.

I think it's also useful -- and I think this may be something for the MINT team to think about -- of not just churning out numbers, but also playing that hand-holding role of talking to the consumer saying, what is it you're trying to do here? Not just your provisions, but what are your goals here in English? Then go through the output and say, do we think you've met your goals? How can we illustrate it to you? Just handing people numbers without -- I know SSA doesn't like to do commentary because it's a very risk-averse kind of agency, but you need to work with your client and talk to them about it. So it's give them the data dump, but then also customize based on the project.

Kathleen Romig >> I agree with that 100%. Yeah. And I think, you know, one of the major advantages of microsimulation is that there's so much richness and complexity. So we don't want to summarize it in a one-page table. That would defeat the purpose. So I agree with Andrew about the data dump. But it can't just be tables and tables of numbers where only people like us are going to be able to interpret them.

Tom mentioned maybe putting some things more into words, not just into numbers. And I think that would be helpful -- the actuaries' memos, they're not maybe the easiest things to read, but they do have words as well as numbers. There is a narrative that explains what's going on. And I think it is neutral. I think there is a lot of risk aversion at SSA about putting any words in anything because it's seen as commentary or opinions. I think you can actually put numbers and sentences and describe them, right, Laura, in a neutral manner, right? And so I think we should have some words out there.

And I also think, to Andrew's point, we need people. This is complicated stuff, and we want people to appreciate the complexity, so we need to teach people about it, and for that you need staff. I would not be doing my job if I didn't point out that SSA's staff overall has really dwindled over the last ten years. And because SSA has had to invest so much in their front line staff and they're really afraid to lose more people on the front line, rightly, and really concerned about their backlog, the research staff at SSA has really, really dwindled over the past ten years. And so we just don't have enough people who are able to fulfill that really important role to teach people about this program and to walk them through the results. I think SSA should be investing more heavily in research, and for them to be able to do that, appropriators need to invest more heavily in SSA.

Jack Smalligan >> Plenty of opinions on this one. I think it would be fascinating if we could have at a staff level, a bipartisan bicameral agreement on a summary set of output that the MINT SSA staff could produce. If it was possible to get that agreement. I frankly just don't think it's possible. Because the summary information just has a policy filter to it. I mean, and for one who's worked on various efforts to try to update the Social Security statement, I mean, there's a reason the Social Security statement has not changed in a very long time, because it's a really hard document to update in a way that satisfies all the affected parties. And believe me, SSA and OMB has tried and we've failed. And so summary information would be wonderful if it was at all possible to do it in a way that would actually satisfy the key players, and I'd welcome the hill staff giving it a try. I certainly -- anyway, give them the best of luck on that.

Kim Hildred >> Okay, great. Thank you, everyone. What I'd like to do now is turn, first of all, to our Board members for their questions. Henry, did you have a question?

Henry Aaron >> Just before Kathleen spoke, I wrote down, "Actuary publishes results for proposals, also works confidentially with groups in planned development. Why not the same rules for MINT?" So that's my question. I think it would be desirable if by the end of the day we either had a good answer to that question as to why it's not feasible, or an expression of commitment to make it reality in the future. So I'm putting that down as a challenge to Mark, and, well, to the two Marks on this.

The second is also a comment – and maybe Mark Sarney could react to this. My reaction is that in this age of enormous computer power and increasing model sophistication, I find that very extended timelines for producing analytical results on specific proposals isn't really understandable, defensible, or acceptable. I have put the point bluntly. Maybe I'm wrong. I do understand that a model may be so enormous in complexity and size that understanding all the ways in which various things interact with one another is intellectually challenging and that's what takes time; or are you saying that the sheer computational time is so lengthy that that running the model is an obstacle to doing analyses.

I remember decades ago when at Brookings the full tax model took a day to run. But in that case, the people who are running it had a stripped-down model that would run in 15 minutes. So that leads me to ask a question here. Why isn't it possible to respond to Amy on a more timely basis about distributional effects and to do so on a confidential basis before the thing is officially proposed? And wouldn't it be possible to do sort of a rough and dirty estimate before you give final numbers that's approximately right? Somebody who is intimately familiar -- as I'm sure you are -- with the model itself, can sometimes give qualitative reactions quickly, as I know Steve often does, before there's an official number. Why can't we speed the whole thing up?

Mark Sarney >> Yes, I can answer that. Things are speeding up. So after the last panel, Karen and I were both mentioning that with server upgrades and that sort of thing, the computation time has shrunk. I think back when Andrew was at Social Security it took us about four hours to run the benefit calculations that now take about 15 minutes. And so we have -- and if it involved a personal account, it took eight hours. And it was a real bummer when you got to the seventh hour and there was an error at one of the last observations, and you had to figure out what that was, fix it, and run it. So computationally it's not as much of an issue as it once was. It's really trying to figure out what a proposal is doing and are we modeling it correctly. And, in general, I agree completely, both with you, Henry, and the whole panel. I would like it to be faster and more presentable. Some of those goals are contradictory. So keep in mind that adding words adds time. Summaries also means more follow-ups. And I would much rather be operating on kind of the same turnaround time that OCACT does.

We do informally give people estimates ahead of time. So if someone suggests a particular provision we've modeled before, we can say, oh, we've done something very similar, these are

the results. And usually that is part of the back and forth as something gets developed or as something gets tweaked. But I wish there was a better answer. But essentially it comes down to resources, number of people and to some extent the complexity of the proposal. So the proposal that Tom mentioned was so complicated for us to figure out because of the SSI and Social Security pieces that we also wanted to crawl under a table. [laughter] so, you know, that can really drag things out. And some of the BPC pieces we had done in a week, or a few days, because we had done them before, but other parts, especially the PIA changes, took forever. So, I agree with everything, but there are some contradictions in what we could do to speed things up that would slow things down otherwise or make it less presentable.

Jagadeesh Gokhale >> Thank you. I think this panel has provided some very interesting comments and suggestions to the two Marks at SSA. One thing that I just thought of, which is a parallel to what **CBO** has recently done, you know, you have mentioned several barriers to using confidentiality, timeliness, barriers to understand the output, so on and so forth. Well, those things are all valid. But there are many microsimulation models in operation today and being run and are competitive, I would say, with the MINT model. One direct question would be to what extent do you go outside SSA to do your analyses of proposals. Again, confidentiality may or may not be an issue there. It could be dealt with in different ways by, you know, a contract, so on and so forth. That is one question.

The other is a suggestion to Mark. CBO just engaged in a competitive modeling exercise for a very simple Social Security experiment. It invited seven different modelers, modeling groups, some of which do both microsimulation and macroeconomic models in an integrated way to provide results and looks at variability, everybody comes out to explore model based on certainty. I mean, so why not, one way of validating what you guys are doing is to compare it with outside folks who are also modeling both micro and macro aspects of the U.S. economy and how it works. So I would suggest this would be a very useful precedent that CBO set to follow for you folks, to look at different microsimulation models, compare them, run a simple experiment as a way and have a conference to explore and discuss where the differences and similarities are and how things could be improved for MINT and other modelers would learn

VIDEO CUTS OUT

from that as well. So that's something I throw out as a suggestion for you guys to consider. The question stands. Do you guys look at other sources for your analyses?

Laura Haltzell>> I can talk a little bit to that. So when I first joined the Congressional Research Service back in 2000, we were in a similar position as we are today. We didn't have the capacity in-house to do microsimulation. We did in fact do a lot of research into various

VIDEO RESUMES

microsimulation models and we ultimately entered into a contract with the Urban Institute to bring the DYNASIM model inhouse.

The problem that we faced with that approach was that over time the small number of us that had really learned the model, knew the SAS code, could replicate results, all moved on and the individuals who remained, I mean, CRS along with SSA with other agencies, our budgets wax and wane, and replacing folks isn't always a possibility so you end up in a position where your brain drain has occurred. At that point the model is no longer useful and so for us bringing something in is no longer a viable option; we've tried that. So now we are looking at all the other options that are outside. So certainly the other ones that you're mentioning are a good place for us to start. Right now MINT, because of the use of the SSA administrative data, tends to be viewed positively by our clients as something that is reliable and so we tend to default to that as our initial go-to.

Jack Smalligan>> From the **OMB** perspective we simply didn't have the budget to like go to DYNASIM or something. I think it is interesting to note that for the BPC effort, they did have foundation funding to use DYNASIM during their deliberative process, so Karen could speak to that. So there really was kind of a two-step process, first using DYNASIM and then MINT modeled it after the deliberations were done. And so that's an interesting case, but obviously that requires either you have foundation funding or other sources of funding.

Andrew Biggs >> I guess I would just throw out that until relatively recently I believe Canada had a microsim called lifepaths which had outputs similar to MINT and I believe that was accessible to the public. And, you know, it's not like Joe six-pack is going to go dial it up, but researchers could get on there and do work. I think part of the problem, and I think this gets to the complexity issues of the use of restricted data, is that only people within SSA could use it. There would be an advantage I think to making a fully synthetic model such that it could then become public use or open source for researchers. At that point you are getting, you know, it is a huge force multiplier in terms of the use of the model to the extent that researchers from all over could examine a whole range of different issues rather than having the occasional member of Congress, or to be very frank, the very occasional research paper coming out from SSA. It's a huge investment of money and time. And the pay off, I will just say, could be way, way, way bigger than it is if you partner with people, just do something to get more research done with the model.

Nancy Altman >> I am really echoing and picking up the theme that Henry said so there isn't that much more to be said about it except that I think obviously everybody agrees the confidentiality is key to having the best use, the most robust use of MINT. And some of it, though, I think, is, and a lot of us have talked about the Office of the Actuary as being the model where you can go with complete confidence and go back and forth about proposals and if you

jettison them that is fine, or if you modify them, that is fine, you have a real sense of control over the output.

Some of that I think is culture. I mean, the Office of the Actuary has from the beginning been seen as independent and serving Congress as well as the executive branch. I mean, Bob Meyers, the longest serving actuary, really pushed that and it is really sort of equivalent to the solicitor general in the justice department that it's seen as an officer of the court, not just working for the executive branch. And I don't know whether that's, some of that is having a civil servant at the top who stays there through all of the administrations. And some of it, as I say, as kind of an understanding of the community, including people within the executive branch, that this is one where they're not reporting to, if it comes in from outside, that's who the client is. Not the inside. So I don't know how, you know, that kind of cultural change comes about, but it strikes me that it's an interesting recommendation that is coming out of this, and it has been a very, very -- I applaud you all -- it has been an excellent panel.

Kim Hildred >> So now I will turn to my SSA colleagues for your questions and comments. Mark?

Mark Warshawsky >> Thank you. This has been enormously insightful and helpful. And I have learned a lot. And there is a lot that we have to take back. Unfortunately, I am not going to be able to meet Henry's expectations. [laughter] But I really was very impressed with Jack's suggestion of a bicameral, bipartisan agreement on what's presented because that really relates to something that I thought was important in terms of the impartiality of the results, you want the results to be presented in an impartial way. Obviously there are unique aspects of any proposal which may not be met by the standard presentation. But at the minimum, I think you do want the standard presentation so that you are not accused of biasing the presentation one way or the other. So, that may be something that we may come back to our colleagues here about.

Similarly with the issue of the confidentiality/automatic publication, sounded to me like there may be a bit of tension with both desires and I'm wondering if we could work that out. I mean, the Office of the Actuary seems to have a good model. I have to think about and we would certainly have to have more discussions internally in terms of what would work in terms of publishing on a proposal that has been put out. My colleagues, I don't know if they have any questions.

Steve Robinson >> This may be a little too far down in the weeds, but one of the issues that Mark raised when his presentation on the Bipartisan Policy Center is a question between scheduled and payable benefits. Under current law, we are not able to pay all of the benefits that are scheduled, and so one of the arguments is that for comparison purposes we should have a scheduled benefit baseline. That's what is scheduled under current law, but we should also have a

payable benefit baseline so that you could compare a proposal to what actually could be paid under current law; not just what's scheduled. And so the argument for the different baselines is based on trust fund solvency.

The dilemma that we face with the MINT model is that it is not an actuarial model so that if you change benefits or change taxes, you can reflect that in the microsimulation, but it doesn't add up to equal the actuarial values, so you don't know when trust funds become insolvent or not. And so to do a payable comparison you'd be comparing payable under current law, but you can't compare payable under the proposal because the MINT model can't calculate what is or isn't payable. And so there is a bit of a disconnect. So, my question, I guess to the users is, while conceptually the payable baseline in the current law makes sense because we know what it is, is there an understandable computable payable proposal that the MINT model can actually do that? And I'm not sure at this point that the answer is yes. But even if it could, you know, is that something that is desirable so that you can compare payable under current law versus what is payable under the proposal? Is that a useful goal for the MINT model in terms of the future?

Andrew Biggs >> I was on the staff of President Bush's 2001 commission and I remember spending a lot of painful and fruitless time thinking about that because you have a payable baseline, you have a scheduled baseline. If you have a plan that is not fully funded in perpetuity you have some sort of adjusted baseline. The reality, I think, is for people that's going to generate more confusion. I think people understand you have a tax baseline and a benefit baseline that are inconsistent with each other. And when we talk about a benefit cut, people understand that's benefits relative to the promised scheduled level. If we talk about tax increase, it is a tax increase relative to the schedule level. I think people are capable of understanding and holding these two inconsistencies and adjusting relative to them. My fear is -- and we have done this, when I was at SSA, we would have a payable and a scheduled -- it was just data overload. This, I guess, rebuts my data dump argument, [laughter] but I never looked at the payable. I always just looked at the scheduled. Because I could understand it. Higher taxes versus scheduled, lower benefits versus scheduled.

Kathryn Olson >> I would underscore that. You raised some really good questions about payable versus scheduled. But scheduled is known. It's well defined. It is the baseline everybody uses. Payable, who knows what would happen. It's never happened before. In 2035 I don't know what Congress would say. We're just going to cut everybody, current beneficiaries by the same percentage, so it just adds incredible complexity. The other perspective I have on it is, again, what is the purpose of the program? It is to provide income to people. So illustrating what the income results are of a change in terms of beneficiary income and household income. And then a separate question is, so what does this cost, and how would we even be pay for that, and do we like how we pay for that? Are we willing to pay for that? But not trying to mix it all up together, but sort of separate out what do the benefits cost that we like? What does that impose on people

or the economy, those costs? And do we like that? Do we want to change it around? So not mashing it all together but having this sort of scheduled benefit baseline seems the simplest way to present it but also lets you recognize that it is just a baseline. It's just a benchmark.

Kathleen Romig >> Yeah, I strongly agree with both of your points and then I would just add it is analytically really complex to think about how to do payable in a meaningful, understandable way. Because like Andrew said, what if the proposal is not sustainable for 75 years, then you're going down this rabbit hole of payable within payable. The payable scenario is pretty simple. People understand it. You know, about three quarters of benefits are paid after the date of trust fund reserve depletion, they can hold that in their head while they are looking at these tables that are about benefits. But once you go down that rabbit hole of payable within payable, and what does this all mean? It just adds confusion and then it multiplies all these voluminous tables by two. And I don't think anybody needs that.

VIDEO CUTS OUT

Amy Shuart >> I want to add that given that MINT is not updated with each Trustees Report, you're going to dig yourself a further hole over time. Given we're on MINT7, I know you have MINT8, but it's not live and that's the 2018 Trustees' Report and by the time that you guys are ready to go with MINT8 it will be 2020. Just for the investment, is something to think about, too.

VIDEO RESUMES

Tom Klouda >> Just to clarify, I think MINT7 was based on the 2012 trustees' report so there is a bit of a lag there.

Mark Sarney >> I should probably mention that MINT8 we are going to start updating the Trustees' assumptions on an annual basis. I know this is a shock to the client panel. [laughter] So MINT8 is based on 2018, and like I said, we have got that data set in-house now. The first thing we are going to do with it internally is run it on the 2019 trustees' assumptions and we expect we can update it every year to the latest Trustees' assumptions. And there is a little bit of a lag. You guys probably noticed that when OCACT, after the Trustees' report comes out, there is a little bit of a lag before all of their projections of provisions are updated. It does take them some time and I figure we will probably have a similar kind of lag. But the idea is to track more so that you don't have to go, oh, geez, it is nearly ten years out of date.

Kim Hildred >> Any other questions or comments?

Jason Brown >> So with the risk of adding more tables to the output, I would be interested in hearing from the panel, are there outcomes that you would like to see MINT produce beyond the ones that are already produced or, you know, measures of well-being, or subgroups of the population you would like to see more, sort of more detail on?

Jack Smalligan >> MINT, again, produces it already, so it is not a deficiency in MINT, but I think just to be sensitive about how we're comparing to poverty when we're doing these forecasts out 70, 80 years and poverty as an index to inflation. I mean, if we get back to a point of actual wage growth, I mean, those comparisons become eroded over time. And so I think in thinking through the best ways to compare with poverty given that we now have new measures, it's more of a policy discussion

VIDEO CUTS OUT

to serve outside of MINT but I think just comparing the way they traditionally have may not really be as useful for the long-term.

Andrew Biggs >> Also by incorporating family size adjustments into some of the MINT benefit numbers, I think DYNASIM does this already, but the poverty measure has that. I mean, the two issues are, if a couple shifts into a single household as one dies off you lose couples living together. But a second is, couples with children, couples with a lot of their preretirement income raising the kids, that's not income they have to replace in retirement in order to maintain their standard of living. I've been doing some work with the **PSID** and you really see expenditures declining by households with kids as the kids leave home. So I think it would be an interesting kind of supplemental measure.

VIDEO RESUMES

Kim Hildred >> Anything else? Okay. Are there any questions from the audience? Steve?

Steve Goss >> Yeah, just a quick question, just in case there are one or two wonks in the audience or on the line that actually are interested in payable benefits, as Mark Sarney indicated earlier,

VIDEO CUTS OUT

in general, when MINT is applied to a legislative proposal where the provisions are a package proposal, like the example BPC 2016, these are estimates where we probably contemporaneously or maybe even preceding developed estimates of what the implications would be for Social Security cost in the future. So we would be able to inform as to what the change in payable benefits is on a

VIDEO RESUMES

year-by-year basis. That is right in our letters to the members of Congress. So if there were any that were interested in what the change in payable benefits would be and wanted to make that comparison, that would be available, not from internally within MINT, but again by the collaboration of the models.

Anya Olsen >> This is for Mark Sarney. You mentioned that there were four members of the MINT team and a lot of the panel has mentioned that they previously were on the MINT team, so could you maybe talk about how the MINT team has changed over time and how those changes may have affected some of the issues that were brought up such as timeliness of results.

Mark Sarney >> Sure. That question is from Anya Olsen who also used to be on the MINT team and who now is in management[laughter] so I guess one point of reference would be in 2010 we modeled the Bowles-Simpson or Simpson-Bowles, depending on how you want to call it, proposal.

The Obama Administration had put together that commission to look at all sorts of budget proposals, we modeled their Social Security proposal in conjunction with the Actuaries and with the commission staff. And at the time, including interns, we had ten people working on MINT, including Anya, Amy, Kathleen, and Laura. So right now we have got four people, and maybe four and a half if you count me. To some extent I still do work on MINT and manage it, but not like the four that are working on it more than half-time, but not entirely full time. So this is where the resource question comes in, meaning human resources. We have had a shrinking staff over multiple administrations and commissioners and that sort of thing. And honestly my plan is by the time I retire and end up in the audience hopefully for some MINT 40th anniversary panel, I am planning for worst-case scenario where I am the only one left, so I hand the keys to somebody else. So anything better than one person that is great. But, I think other models operate with one or two people, and I think Jagadeesh kind of hinted at that, but having fewer people damages credibility as Kathryn mentioned. The servers are going in the opposite direction. We can process things faster. But when someone leaves with all that institutional knowledge because they get their dream job somewhere else, you know, it becomes a problem. So we are continually trying to train people and that sort of thing, but the numbers have dwindled.

Karen Smith>> My question is for Kathryn. You said that you are worried about confidentiality and that is an issue. So my question is how do you get your answer if you are not using MINT?

VIDEO CUTS OUT

Kathryn Olson>> So the answer is I don't get answers necessarily. Sometimes I can get them through other sources like the Center on Budget Analysis, there is CRS, or other –

VIDEO RESUMES

somebody said something about, 'It's only used for academic research,' and I thought, 'Oh, no, that's really important.' Academic research is important because it's, again, another way to kind of at arms length, and it lets ideas be percolated around and then come to Congress once they've

been a little bit tested by the policy-making community outside Congress. So I think that is really useful and then we can exchange cards.

Kim Hildred >> First of all, I want to thank the panel. Great job, everyone. This has been very, very helpful. Thank you to the audience. And thank you to the SSA folks and the Board. We will now take our lunch break and we will be reconvening at 1:15. And again if anyone needs any directions or places to point to in terms of where to have lunch, please talk to any of our staff, as I think they are a little familiar with the area and they might be able to give you some ideas. So, we'll see you back here at 1:15. Thank you.

Session III: Perspectives of Microsimulation Experts

Part I: Other Social Security Microsimulation Models

Henry Aaron >>

We have four panelists, each of whom is quite knowledgeable about different models because they are principles involved in developing and applying them. And I'm not going to do any introductions at all because you have the biographies in your packet. But in the wisdom of those who set up the seating order, I think we're going to just move down the aisle, starting with Julie Topoleski, going through Steve Goss,

VIDEO CUTS OUT

then Brian Tysinger.

Henry Aaron >> And Jagadeesh Gokhale is a member of the Social Security Advisory Board. Jagadeesh will be speaking on the Wharton model, or a component of the Wharton model anyway. I have been instructed also to tell everybody who's using a microphone to come as close as you can to eating it so that your voice will be heard fully. Julie, you lead off.

Julie Topoleski, Chief, Long-Term Analysis Unit, Health, Retirement, and Long-Term Analysis Division, Congressional Budget Office

Julie Topoleski >> So first, thank you very much for inviting me to participate today. My name is Julie Topoleski and I'm with the long-term analysis unit of the Congressional Budget Office. We are the group -- and we've got some founding members of the group in the audience who have a longer institutional knowledge than I do who have developed and used and maintained CBO's long-term communication model. It's a model that was built to model the Social Security system. Over the years, it's been [inaudible] in different ways. So now it's acting

as the long-term [inaudible]. It has a macro model that lives within it. But it was originally developed as a Social Security model, and that's still where the meat of the model is.

I was sort of encouraged to see a lot of the similarities between CBOLT and MINT. There are a lot of things in there that we approach in very much the same way. There are things that we approach differently. For example, in the MINT model, they run policy simulations. They're not re-running the whole model. So they're not sort of recreating a base policy.

VIDEO RESUMES

The way CBOLT is built is that when you run a policy simulation, you're automatically sort of recalculating the baseline so that not only is your distribution results consistent with the new policy but you also get the underlying revenues and benefit streams and aggregate that are consistent with that. I haven't thought fully through the implications of sort of not doing that, but that's the way it was built. And that way, things are sort of all adding up and being consistent. I should note that CBOLT is just one of CBO's microsimulation models. I did work on the others, but I feel I would be giving short drift to not mention that we do have others.

I've

VIDEO CUTS OUT

been at CBO for 16 years, and so at this point, I think I've touched most parts [inaudible]. But mostly what I'm going to talk about today is how we have used the MINT model in our work, which is what Henry asked us to talk about. I really want to highlight sort of three ways that we use MINT. [inaudible] -- for consumers with the MINT documentation.

As somebody who works on microsimulation,

VIDEO RESUMES

you always want to find the other people out there who work on the same types of issues that you do and grapple with the same types of things. So whenever we're developing new features of the model, modifying existing ones, or trying to sort of understand something that's coming out of our model, it's always really helpful to look at how others have handled that same issue. We always approach an issue independently. I think it's best to sort of try and get clean thinking about the issue. But we always learn a lot by looking at what others have done. I think that probably the people in this room and maybe a few others are the ones who sort of read the documentation cover to cover and we're part of that group. There's definitely a certain level of comfort in learning that others have taken a similar approach to a problem or maybe taken a different approach to a problem or maybe taken a different approach but have come up with similar results. And that, always, as Karen said earlier, looking at sort of the differences between DYNASIM and MINT, that when they come up with -- have a different approach, but come up with a similar answer, there's definitely a level of comfort that you develop in learning that. Also,

people at CBO talk about sort of being in the middle of the distribution when we're looking at sort of different outcomes and different things and MINT is certainly part of that distribution.

Second, we're consumers of the research done using MINT. A colleague was just telling me this week that he's been looking at ways of improving our modeling of reference to education. And one of the papers he read while doing that was a -- I'm going to read this-- 2015 paper by Pat and Howard Iams on education, earnings, and equality and future Social Security benefits. So we are certainly consumers of the research done using the MINT model. Again, it's always useful to know how the modelers think about these issues and to learn about the analysis done using their models.

On another example, and perhaps the most important of those, is we work with the analysts that have worked on MINT. This is maybe less true in recent years. But certainly, CBOLT and MINT started life at around the same time. And as we we're developing CBOLT -- and I say we, and this sort of pre-dates my time at CBO -- I know that those at CBO at the time spent a lot of time learning about what people were doing and thinking about the MINT model and talking to them. And there were some things that we made different decisions about, other things we made similar decisions about. But the people that have worked on MINT over the years

VIDEO CUTS OUT

have willingly given of their time for meetings, for conversations. I know that people at CBO have been involved in reviewing parts of MINT model. And I know that we have reached out to analysts that have worked on MINT over time to review our work. So we are very grateful for that. And with that, I've gotten my 30 second warning, and I'll stop.

Steve Goss, Chief Actuary, Social Security Administration

Steve Goss >> Thank you. Okay, well, I have some slides. I'm not even sure it's worth putting them up in this short amount of time. So I'm Chief Actuary at Social Security Administration. You've heard much of the work that we do earlier today. Thanks to Mark Sarney and Karen Smith for describing some of that, and the collaboration. I really want to emphasize, first and foremost, really the collaboration that's been going on a long time now between the MINT development and what we do. Our purview is, of course, to do cost estimates under current law and for changes in the law. And the work of MINT and many other micro models is to look at the distributional aspects. And these are really not one in the same, but it's important to have them coordinated.

VIDEO RESUMES

So we're really glad that, in the models that we have -- I would just say that our models are a combination of micro and macro models. What we seek to do in the development of our models is, for whatever the intended purpose is that we need to accomplish--

we've found that, whatever the purpose is, if you want to do a cost estimate for what's going to happen with the trust funds versus have distributional analysis of the type that MINT is just primo on, for things like what's going to happen with things like how many people will be in poverty, you really need to have different models and different aspects.

We do have a pure microsimulation model called POLISIM which was mentioned by Karen and Mark Sarney earlier. We've been developing this for a lot of years. I guess probably pretty similar to the amount of time that MINT has actually been in operation. We don't use it straight up

VIDEO CUTS OUT

for our basic cost estimates, but we do use it as a side model for many purposes.

I think some other people have mentioned that things like certain things about family structures - - when you're doing estimates on things like child care dropout years or childcare credit years, family structures are really, really important. And we need the POLISIM model for that purpose. As you heard also, that our POLISIM model --

VIDEO RESUMES

probably one of its greatest strengths -- and if you want to really know all the details about it, you should talk to Karen Glen out of our office who's been the one largely running this in the very recent years is really on the demographics. I was glad to hear that MINT is still using, strongly, the demographics that we have, because we have built, very carefully, into our main model and into our POLISIM what's happening with immigration, fertility, mortality, et cetera, et cetera. So these models really speak to each other.

But I would just suggest that, as I think you heard this morning, all models basically are hybrid models. There's no such thing as a pure microsimulation model. For instance, in MINT -- I was glad to hear from Karen Smith, I had heard it before but a reminder that, for instance, in projecting earnings histories into the future, that's a really tough part, getting the texture of what earnings histories look like. We have an approach for doing that for our average model. You could do it by pure, every single year, having transition probabilities for each individual, where they're going to go, given their recent experience. Bootstrapping is a different approach. Bootstrapping takes advantage of past knowledge of what the texture and the tendencies of benefits are for people across time. It's really critical that we do that. In our models, we do much the same, to an even greater extent of sort of like cohort by cohort bootstrapping. But then, for

each succeeding generation, you've got to get in and pay attention to what you're assuming about the propensity for people to have been employed, age by age, as you move from generation to generation, what the earnings levels are, and build all of that in. So basically, I would say that we have the POLISIM model, which is very useful as a pure microsimulation model. It speaks to our main model and vice versa. Our main model is really used for developing the population, the economics, everything about Social Security, and it's used really then for aligning our POLISIM model to make sure that what's happening at a micro level is really following along.

I think you've already heard MINT, I think Urban is probably the case too, basic numbers that we've developed for the trustees in the annual trustees reports where our main model is used for alignment in many other areas. So we have a lot of side models and I still

VIDEO CUTS OUT

haven't seen the one-minute thing yet. So we have a lot of side models. We have many, many side models that are variously microsim or not microsim, again depending on what the answer is that we need for the question. I would just mention, with respect to MINT, distributional insights that come up with MINT are really important.

There was a lot of discussion this morning

VIDEO RESUMES

about confidentiality. The way that we have always worked, in terms of confidentiality, is, as many people here on the prior panel know, when we work with the Hill about something, we say, as we're developing the proposal, in our working with you, it's us and you. That's it. The only time anything that we do in terms of developing a proposal goes out public is when -- and this even includes our commissioner and everybody else in the executive office of the president. The only time anything goes public is when that proposal is finalized and the member of Congress or other is ready to go out public with what they've got and they'll be using our numbers. The one little exception to that is, if a member develops estimates and they don't want them out there but another member of Congress who might have to vote on that proposal wants them, we will put out the numbers in that case because it's important that we're informing the people that are going to have to make the decisions. With that, I'll pass the baton.

Brian Tysinger, Director, Health Policy Microsimulation, University of Southern California Schaeffer Center

Brian Tysinger >> Hi, I'm Brian Tysinger from the University of Southern California. I direct the health policy microsimulation group there. And the reason health is prominent in that title is because we care very much about health. So let me walk through the two models that we host. Okay. So why do we use microsimulation? The reason is there's a lot of heterogeneity in the

population and we really care about disparities based on income, disparities based on health, disparities based on education. We use our models to predict the future disease burden in the population, the future of disparities, future healthcare costs, and implications for federal programs.

This is difficult because there's, as I said, a lot of heterogeneity. It's complicated. So we rely on microsimulation where we can control for lots of differences in individuals at the individual level and forecast into the future. So we have two microsimulation models. The first is called the future elderly model. The second is called the future adult model. This future elderly model dates back to about, I think, the year 2000 at the RAND corporation. I worked with people who worked on it back then who still work on it now. I've been working on these models for about 7 or 8 years. And then I love the development of the future adult model, which is our younger model, both in terms of timeline and in terms of ages.

The future elderly model is based on the health and retirement study. We modeled people 51 and older and so we focus on things like the retirement decision, functional limitations, mortality, and Social Security benefits. We have a fairly built out Social Security benefit module, but that's not really the focus of what we're doing. That's more of an outcome that we also care about. So we'll change something with health and then see what happens in that module, but it's not necessarily the main thing that's driving what we're doing. Both models we have at the U.S. national level -- we've adapted them for California and also for Los Angeles county, where the University of Southern California is.

The future adult model is based on the Panel Study of Income Dynamics, and that allows us to model of people 25 and older, so we can capture the full life course, so we can capture taxes paid throughout the life course and think about interventions that might happen earlier in life. We have a broad set of applications. Most of what we do these days is more on the health side than on the benefits side.

Okay, so the team. This came up a lot in the MINT talk, and I wanted to talk about this a bit because our team is large. There are four full-time mathematicians and programmers on our team that are working on this. We're at a university, so we also have an able-bodied set of graduate students who can work on this and incorporate it into their dissertations. And then I counted up, and there's 15 faculty members who are actively working with the model, asking their own research questions. That lets us ask a very broad set of things that we work on. Additionally, we have a lot of external collaborators that we work with from OECD to many universities around the world. One of the benefits of the health and retirement study is that it has sister surveys around the world. So we've built future elderly models for Mexico, Europe, England, Ireland, Japan, et cetera, et cetera, et cetera. A lot of the alums who've worked on the model in the past

have gone to build their own model. The Canadian version, the Japanese version, those are built by people who worked on the U.S. model previously.

Okay, changing pace, I wanted to talk a little about validation because it's really important. It came up earlier, about building the confidence of potential users and consumers of what we do. We spent a lot of time working on internal validity, external validity, and then this piece that I call external collaboration, where I look to other people who do micro or macro models and I compare my projections to what they're projecting. If someone else is the gold standard, I see how I do in comparison to it. So I definitely rely on MINT for that purpose. And then within validation, I often reach out to like Melissa at Urban and ask her, 'What are you guys doing, how can we compare our results to what you're doing?' So it's a small community of people who do microsimulation, but I think we do talk to one another and we see what the best practices are.

So then on technical challenges, there's a lot of technical challenges. It's hard to use these models. You need to be a talented computer programmer, a talented mathematician, a talented quantitative social scientist in order to be able to use these models. And consequently, it's hard to find the right people to work on these models. So when you find a good team, you need to keep them around. The second piece is that it's hard for just a person off the street -- someone said Joe six pack earlier. I mean, that's very true. Someone not familiar says, 'I want to use your model.' I have to say, 'Okay, well, you want to use our model but now you're going to need access to restricted data.' So you're going to have to contact all these agencies and get permission to use it. It's a burden. The final piece is that there are consumers who don't want to do the technical stuff and so we will populate datasets so that they can access it through like a tableau visualization. So instead of having a spreadsheet that they pour over or, you know, dozens of PDFs or whatever, sometimes a tableau visualization well thought out can be a good way to present results to individuals.

And then finally, where am I going next? We want a model younger and younger. I want a model more in California. We do a lot in Alzheimer's disease and related dementias, very hot topic these days. And then we recently did a paper on mental health. Thank you very much.

Jagadeesh Gokhale, Director of Special Projects, Penn Wharton Budget Model

Jagadeesh Gokhale >> Thank you, everybody. My name is Jagadeesh Gokhale. I'm speaking on behalf of the Penn Wharton budget model. I also have some slides. I'm not sure whether they're worth going through. I will just perfunctorily. But before I start on the slides, I'd like to give you a background about the Penn Wharton budget model. The budget model was started by Kent Smithers who runs the institute. Initially, when we got together, I said that if we want to do policy analysis, we should include the entire budget. I had developed a microsimulation model before I went to Wharton. That model was focused exclusively on Social Security. We wanted to

expand it and cover the entire budget. This conference is about Social Security, so I'll reserve my comments to just the use of that model for Social Security. We've grown from three people initially to 25 now, including graduate students, interns, and so on because we need that support. Building a microsimulation of the U.S. economy is really a very complicated, complex, time-consuming exercise. It involves a lot of digging into data, considering of different micro sources for calibrating the model and so on. So we've done all that work. We've done all that work, and we have now a microsimulation that is a much bigger version of what I had developed initially. So that's a story with the Penn Wharton budget model.

So when we started, we wanted to do policy analysis.

VIDEO CUTS OUT

Usually, when one thinks about macro policy analysis, one thinks about representative-agent macro models with overlapping generations. But, to us, those seemed very limited because the population has become very heterogenous. We know that the baby boomers are retiring. They're the experienced part of the work force. They'll be replaced by a similarly experienced, but a

VIDEO RESUMES

smaller cohort of their successors. Women's representation in the labor force is changing considerably. Disability is changing. Mortality rates are undergoing changes, and not uniformly across the entire population but differently for different segments. And we need to be able to capture all of that and represent the entire set of features of the economy in the computer to have any hope for not just doing macro level analysis, but also digging into the distributional consequences of policy changes that we want to analyze.

We also know that when constructing a microsimulation, we can only calibrate it using historical micro data surveys. You can look at those surveys. You can dig into them, break them down, estimate relationships, and estimate markov processes for transition of people in different subgroups into the next period and so on. When you reach the terminal year of information (the most recent data available), where do you go after that? How do you jump off into the future? It is crucial to decide how people will be making decisions in the future. In the morning's session someone mentioned how one could splice information from older cohorts and so on. Take my birth cohort and consider the cohort that is ten years older than me. They lived in not necessarily the same economic environment that I'm living in now. The economic environment, both macro and micro, in terms of distributions of their sorts of skills that are in the economy, the various demands for skills, etc. is always changing and dynamic. So the economic environment for macro and micro are changing. So to impute information and relationships and decisions that have been observed in the past and captured in some functional forms, and used to impute future behavior in the same way as in the past seems limited.

The estimation must somehow include future behavioral changes relative to past observed behavior, which requires an aggregation and feedback step. So, a microsimulation alone is not sufficient to make projections into the future. In order to make macro projections one needs a consistent organizing framework. We could use models such as the Cobb-Douglas production function used in growth models. We can alternative ways to derive aggregate labor and aggregate capital.

Well, how do you get from the micro level, labor, supplies that individuals provide under different family structures with different ethnic backgrounds, levels of education, and so on. Individual labor supplies must be aggregated without losing the effects arising from population heterogeneity. So some part of the effects of heterogeneity can be captured through functional relationships, others would be missed. R-squareds are always less than one. If for a key projection element, such as wages, person characteristics explain 70% of the variation, (the R-squared is 0.7, 30% of the variability in wages, is unexplained.

However, capturing that 70 percent of behavioral regularities associated with population characteristics and projecting those based on projected changes in those characteristics is better than simply assuming that the future will be just like the past for everyone, no matter what types of people they are. Finally, one needs to return from the macro to the micro: We need more than a microsimulation model to project into the future by adding a layer of dynamic effects from the macro model. A dynamic macro model that can tell us about behavioral feedback effects people might engage in based on information on elasticities from the literature that others have estimated. How does labor supply change for workers with different characteristics when taxes change and so on? So the Penn Wharton model is microsimulation integrated with a dynamic equilibrium macro model. Thank you.

Henry Aaron >> Thank you very much. There are two people in the audience who were involved with the creation of two of the models here, Howard Iams who was involved with MINT and John Sabelhaus with CBOLT, or its predecessor. You've been here for half a day so far. I would like to ask you in turn -- age before beauty -- Howard, you may lead.

VIDEO CUTS OUT

Any reactions, thoughts about steps you think -- perhaps mistakes that were made or, more constructively, additional actions that can be taken in the future that would improve the product whose initiation you are responsible for? Can we get a microphone for Howard? I'm assuming he will not be completely silent.

Howard Iams >> The MINT was designed to go out about 20 or 30 years. It was to take the people in a representative cross section and project what they would look like in retirement in about 20, 30 years. At the time we were doing this -- we started working on this after the

[inaudible] trustees review. And it was clear we needed to be able to do [inaudible] pensions. [inaudible] and I came up with matching SIPP ii integrative data and using that to project what people who are 20 years from retirement would look like when they retired because [inaudible] was all based on future retirees. I [inaudible] with

VIDEO RESUMES

making MINT into milt was not my decision.

Mark and his group were drawing up conclusions 40, 50 years out through using simplified assumptions on the original MINT data so I asked Irving to come up with a way to try and do it more systematically. I really don't have a lot of confidence in the data or the projections, after you take the last observed person in the sample and they reach retirement. Using POLISIM is one way of dealing with it, but I don't particularly care for the POLISIM model. I guess that would be my one caution, that I don't think MINT is -- I think it's quite solid going in what are retirees going to look like in 10 or 20 years when Social Security is going to have real funding problems, but I do not think it's good for 2050 and 2070. I just don't think the data supports that.

I guess the other thing I would comment on is there's a number of bells and whistles in MINT were forced on me by whoever was running the Office of Policy because they wanted to add, for example, I don't know, SSI and means tested programs. I worked in a means tested program, AFDC, for about seven or eight years. It's very, very picky, picky stuff. And MINT -- and actually SIPP doesn't really have the information to replicate the picky, picky stuff of SSI, food stamps, and AFDC. So I lament somewhat adding that. I think we did the best job possible. It's okay to add on. It's okay to do it, but I don't have any confidence at all in, say, the estimates of SSI.

There is an issue on disability. Why don't we model disability? People -- back when I started working for Social Security in 1976 -- had spent years fielding surveys trying to model disability and they gave up. So I would never try and model who's going to become disabled. That's why we basically use this splicing thing of pick up people who represent five years older than what we're trying to do and we pick up disability. We originally had a disability model that Stan Panis built off survey data and it was goofy because it had disability being an absorbing state but it wasn't. You are limited in the amount of work you can do. People go in and out of that. We did not spend a lot of time trying to model disability.

Henry Aaron >> Okay, thank you. I'm going to come back to one of the observations here because I think it raises an important question that I'm going to ask the panelists to address.

John Sabelhaus >> Thanks very much, Henry. Listening to Julie, my main reaction is that it's just great to see that **CBOLT** is in great hands. Julie is the third person running the unit that runs

CBOLT. What I would do with just a minute here is to draw a connection between two things that we've heard talked about. One is what Julie mentioned, which is sort of possibility that because of the way that MINT interacts with the OCACT forecast that you can end up with inconsistencies sometimes when you're doing policy simulations. And that's one of the things which we, because we're building CBO from scratch, we were able to make sure that the micro and the macro were always consistent. One answer I would say to Mark is that when you think about when you're running an alternative simulation, it seems like you could recalibrate the model to an OCACT baseline which sort of gives you at least a first order of proximity. If OCACT has run their models on a policy proposal, if you use that as a new baseline--I think Steve was eluding to this earlier as well--it takes you in the direction where CBOLT's going. It's not fully iterated, but it's in that direction.

The second thing I would do is draw a line between that and what Jagadeesh was just talking about how at Penn Wharton, they've sort of had this insight that it's the micro and the macro, each independently can't give you everything you want. So it's actually making the two work together that, I think, does give you important information. But I would raise a caution with that, which is you have to be very careful with what you do on the macro side of things. Sometimes these long run models -- again, I think this ties in a little bit about what Howard was saying about long run uncertainty and where the economy is headed. You know, we have macro models now which still attribute most of economic growth to something called total factor productivity. We don't know a lot about that. We do know about labor force participation. That's true. But in terms of why economies grow over 30, 40, 50 years, there's still a lot of uncertainty about that. I think we have to be very cautious in terms of the interpretations we get about policy affects when we're not sure how things like government deficits, international trade deficits, things like that, affect the long run growth of the economy. So it's important to bring the micro and the macro together, you do have to keep an eye on what the macro part is doing, in terms of these projections.

Henry Aaron >> Thank you very much. I would like to get back to an issue that, I think, Howard's comments raise. These models can be used for two rather different purposes. I realize that particularly with respect to the actuarial projections, they're quite clear that they're not forecasting the future. They're providing estimates of the impact of certain initial assumptions on the course of events. But one purpose to which models are in fact put is to forecast the future, to get an idea of where the economy or where distribution, whatever, is going in the future. That's one purpose. Another purpose is to analyze the impact of changes in policy. What difference does it make if you adopt policy x to replace policy y?

Henry Aaron >> Speaking just for myself, I feel a lot more comfortable with the models that I think we're discussing here for addressing problem two, the impact of policy change, than I do about them as projections of the future. And this is not to criticize anybody's track record to date.

I would appreciate reactions from the panelists about that observation, which may be dead wrong. But I hope it will initiate discussion. Steve?

VIDEO CUTS OUT

Steve Goss >> Your bifurcation of the kinds of projections that one might make, I think, is exactly right. I would take issue though with the idea of referring to any prior projections [inaudible]. I'm not sure if Julie and others would agree with this. We do not call those forecasts or predictions. We call them projections or estimates of the future, because what is extremely critical about making a

VIDEO RESUMES

projection of the future is that you be extremely explicit about what the underlying assumptions are.

VIDEO CUTS OUT

We can look at what the productivity birthrate has been in the past, what birthrates have been in the past, how fast have [inaudible] have dropped in the past. Then we have to make a judgment about how rapidly we think any of these things will be [inaudible] in the future. These are the basic ingredients that go into cooking the cake. The cake comes out, but unless you know exactly what the ingredients are, you're not going to be in a

VIDEO RESUMES

good position to assess it. Now, on Hank's second point about the notion of when we have a policy alternative to the current law, that really becomes important. We actually had a member -- I won't name them, who it was -- of a technical panel in the not too distant past from a particular university who said some of the projections we were making for changes in law were smaller than the potential range of possibilities under current law. That's exactly right. But so what?

The real bottom line is, when the Congress -- prior panel comes in and they say, we want to consider eliminating the earning test or raising the normal retirement age, whatever, they want to know what kind of a delta, what kind of a change that will have. Not only in the cost, but also its effect on individuals

VIDEO CUTS OUT

. [technical difficulties] and that is -- by the way, I have one example I didn't have time to throw into the [inaudible], but let me toss it in here if I may now. What I think was very early -- I didn't realize how early it was in the MINT period. But back in the year 2000, many of you will know that there was a policy alternative. It wasn't enacted, two of them,

VIDEO RESUMES

the retirement earnings test for Social Security over age 65. Well, that was the normal retirement age at the time. There was consideration of eliminating it under 65, but the people in the executive office of the president feigned away from that because estimates from, I believe, MINT, indicated how many additional people, widows in particular, would be in poverty at a later time if we encouraged, allowed people to take their benefits while they're still working at age 62, have a much lower monthly benefit forever. So the value of MINT at that time was extremely important in guiding policy decisions. I don't know that it got a lot of credit, that it was out there up on the web, but it was very important in guiding decisions for making that particular policy change.

Julie Topoleski >> Just to go back to Henry's point, many of you in this room may know that CBO's projections of long-term Social Security finances and those of the trustees are pretty different, if you look at baseline levels. But one of the things we've learned over the years is that, if you look at policy analysis under either our model or their model, you often get very similar deltas. And so I think that reinforces the point that Henry was making.

Henry Aaron >> Incidentally, I do think, Howard, that response is, to some degree not complete. But to some degree, it should be comforting to you about the value of the model for purposes of analyzing policy effects, even if you feel a lack of confidence about the level projections. Jagadeesh?

Jagadeesh Gokhale >> Howard also mentioned that he had little confidence going beyond 20 or 30 years. Is this the same related point of projections versus forecast? "Little confidence" judged against what? If you think about all of these as conditional projections, that is, conditional on the assumptions that are made, going beyond 20, 30, even 50, 70 years is not invalid as long as it is interpreted as a conditional projection based on the model's construction assumptions and whatever calibrations are used. If it is divorced from these considerations, one might think of it as an unconditional forecast and, sure, you don't know how good that is.

Howard Iams >> I'd like to say, I'm a demographer. What I see MINT doing is taking an observed population with a history that's fairly rich and aging them. Basically, the equations of MINT are equations to age people based on longitudinal datasets about how people age over a period of time, ten years or whatever. So part of my comment is that, when you get to the point where you no longer are aging someone, when you're making them up -- which is traditionally how microsimulation is worked -- I have no confidence in that really, because I have no idea what the characteristics of those people would be. But MINT takes basically observing people in 2004 and 2008 who are 31 years or older and they lived part of their life. And we have their administrative earnings records through -- now it's 2018. 2015? So we have observed earnings for at least half of their work life. So I think the process of predicting how people age is different than predicting and making up what the population is going to look like.

Henry Aaron >> I want to try and get Brian into the discussion and get an interchange between you and Steve. Brian, you're in the business, you were saying, of projecting health, presumably life expectancy and mortality rates. Steve is in that business too. To what degree do your projections differ? Do you talk with one another? Let's you and him fight. [laughter]

Brian Tysinger >> Okay, I'm not looking for a fight. [laughter] But yes, I definitely do compare my life expectancy forecast to Social Security published, like Bell and Miller 2005, is something I look at regularly. I look at period life tables to understand what my **HRS** population's mortality looks like in my model compared to what they were forecasting back in 2004 and 2005. When I take the approach where I consider -- sorry, I know there are demographers in the room. I'll pretend I'm a naive demographer and I'm just going to have a period life table of age and sex. I do that within my model, I'm within a half year, roughly, for life expectancy at the individual level. For a lot of the things that I do, that's perfectly fine. For other things, it's not what you'd want to do. When I start incorporating health effects, the results change quite a bit. And I think it's up for debate whether or not incorporating all the health effects that I'm doing is the inappropriate thing to do. It might be over counting. It might be inappropriate.

Steve Goss >> No arguments here. In fact, one of the great things about symposia like this is meeting people that you haven't worked with

VIDEO CUTS OUT

in the past but, hopefully, will in the future. [inaudible] mortality projections, population projections in general. And we deal with lots of people, including other California universities like Berkeley. We deal a lot with them. We deal with the **CDC , CHS**, Cancer Institute, to get all their [inaudible]. We actually had some great

VIDEO RESUMES

work put together by the person who is the current chair of our technical panel, a guy named Bob Burlin. When he was at AIG, he actually put together a lot of work done over at Johns Hopkins University with clinicians and researchers and medical science there. We got to make sure you've got this information, Brian, if you haven't seen it so far, with their expectations over the next 30 years of what they think will be the trends on death rates for different causes of death. We do project into the future cause of death-based death rates. We found that to be very useful. You can break down the trends and not just look at an aggregate.

One of the great stories I think we had was back in the mid-80s when HIV/AIDS came around, when we were wondering what was going on. I've told the story too many times. We called up the CDC in Atlanta, talked to a guy named Meade Morgan, and Meade said, 'Hmm, this looks like a pandemic that's going to peak around the year 2000 and then drop down to about one third

of the level.’ So we said, ‘Okay, fine.’ So we projected it would peak in the year 2000 and drop back down. What happened after 1985? It peaked in 1999. Meade was off by a year. But going to people who really know and understand has really been a positive. So anyway, we’re looking forward to working more with Brian and many other people to inform our estimates.

Henry Aaron >> Okay. I would like to turn now to the SSAB. Kim or Nancy, do you have any questions or issues you would like to raise?

Kim Hildred >> Yes. Thank you, Henry. Because we had a pretty robust discussion about the importance of presentation of MINT results, Julie, I’m wondering if you can share a little bit more about the process. CBO did a lot of introspection on particularly your long-term model results and lots of other results as well. Could you talk about the process and how you worked to try to gain consensus in terms of what’s the best way to present the data?

Julie Topoleski >> Yes. So we have worked really hard over the years to think carefully about how to make our results

VIDEO CUTS OUTS

understandable. And our primary clients are Hill staff and congressmen and senators. It’s been a while since we put out a comprehensive Social Security analysis. But when we do put results out, we really want to try to make -- a few things. One, we want to keep take-aways clear. I saw someone in the earlier session mention sort of having a [inaudible] up front.

VIDEO RESUMES

That’s one of the things we’ve worked really hard on. It also means that, in our main presentations of results, you’re not going to have the level of detail that you get in all of the millions of MINT tables, partly because I think that’s overwhelming to the average user of our results.

We’ve thought really hard about what are the key takeaways and we’ve tried to come up with sort of a standardized way of presenting those results. We wanted to make sure that we can’t be accused of sort of doctoring the results or skewing to one side or the other. So when we do a policy analysis, we want to make sure we have the exact same set of tables, regardless of what the proposal is or who has asked for it. That does mean that sometimes, depending on the nature of the work that we’re doing, that you may have a display and a package that isn’t particularly useful for that particular proposal because it’s not affecting something in that way. But at least that way you’ve got sort of consistency across them. We also spend a lot of time talking to people about our results and getting feedback on what it is that we’ve put out there. This has been more of the case in recent years with things like the long-term budget outlook and other reports that we

put out on a regular basis. But the bottom line is, if our users can't take the information and we can't make it salient, and they're not finding it useful, then we're not doing our jobs.

Nancy Altman >> I have a question. Again, sort of a similar perspective. And that is -- obviously, being able to evaluate the output depends on the assumptions underlying it. To what extent -- I would be interested to hear each of you. To what extent is your model transparent? And what efforts do you take to make sure they're generally transparent to people who want to use them?

Brian Tysinger >> We make our model publicly available. That's actually part of our agreement with the National Department on Aging. We have publicly available documentation with validation results so people can get into it themselves and see what's going on. That said, being an intelligent consumer of these models is challenging and so to really understand what's going on, it requires a lot of introspection and deep thought.

Henry Aaron >> Time is getting short. So please try and make your responses brief.

Jagadeesh Gokhale >> This is the page for what the Penn Wharton model does for validations. We make sure that the models are calibrated, how well do they do when we process aggressions and look at how well they capture what the micro data is telling us, verify how the simulations perform, that is, we do source target matching after transitioning model individuals from year to year. We validate historically, including both parametric and non-parametric tests to make sure that the simulated attribute, distributions historically match the micro data survey distributions. And the last bullet point notes conditional attribute distributions and conditional attribute prevalence and incidence rates. The Penn-Wharton model is only four-years-old and validating out of sample is something we will launch into in the future. That's also an important part of validating model outcomes.

Henry Aaron >> Thank you.

Jagadeesh Gokhale >> In terms of communication -- well, I could go on.

Henry Aaron >> Please don't. [laughter]

Jagadeesh Gokhale >> One way to communicate is just to have discussions with your clients. Another way is to put out a whole slew of pre-run regression results or simulation results. And people can pick and choose what policy settings they want for their controls. So you give them a wide variety of possible policy scenarios that they can check against each other to figure out what happens under different settings -- that's another way of presenting or validating. If a result under some policy doesn't seem right, you would receive feedback from the public. That's

another way of communicating and maintaining constant updates and information for maintaining the model.

Steve Goss >> I would say, for our modeling and for the projections we put out, especially for illustrations and our results, we have constant feedback from people on the Hill, from Kim, from other people who are currently on the Hill, Tom Klouda, others who were here on the prior panel, about what they would like to see. The key is to put it out in a straightforward fashion where it will be the same for every proposal you deal with to abstract comparability. In terms of openness, we are lucky enough because our long-term projections into the future also go into the Social Security financial audited statement, that we get a full scope audit by an accounting firm with actuaries, economists, everybody else, every single year. They go through in detail. Thanks to our Social Security Advisory Board, every four years, we have a technical panel of actuaries, economists, and statisticians and demographers come and look at every aspect. And in addition to what we have on the web, which is very detailed documentation of all of our models and our assumptions, when these panels come in, they can come and they can sit down with us any time and look at any level of detail that they desire. As far as track record, the estimates have been made. Actually Bob Meyers mentioned earlier, back to 1935 and before estimates were made. Trustee reports have only been coming out every year since 1941.

Julie Topoleski >> Hello. Just to add quickly, we put out a lot of data. Every year, when we put out our long-term budget outlook or detailed Social Security projections, we put out a whole host of data to share with the world. So we're not putting our model out there, but we are giving people the opportunity to look at lots and lots of detail of what we do. And I should add that Steve and his staff are very generous in sharing their data with us. And we spend a lot of time running our model using their assumptions so we understand the similarities and differences between the two.

Henry Aaron >> And batting cleanup, Mark Warshawsky.

Mark Warshawsky >> Thank you. This is really a question to Brian. Despite what Howard said about the picky stuff, in terms of some very difficult modeling for SSI and other means tested programs, I would imagine if it's like the rest of Social Security, disability is a very complex eligibility criteria. And I'm wondering -- you know, we still have an interest in disability. If you have any advice for us, if we were to go down that route. It's not parametric, the rest of Social Security is. Disability has very complex eligibility criteria. From your experience, in modeling the health benefits, what is your advice.

Brian Tysinger >> That's a great question. Disability is challenging because of the spells issue, and particularly because I'm reliant on the Health and Retirement Study or the Panel Study of Income Dynamics. And that's individuals every two years reporting their disability status. And

people aren't really that good at it. So if you have linked administrative records, you're in a better position than I'm in. But even so, it's a challenging problem, just because of the timing and how long these things last. On the health expenditure side, that's a little bit easier because there are good data on those that you can compare against. But yeah. Your skepticism, I think, is fair on these programs like that.

Henry Aaron >> Time's almost up. I just want to put a plug in for SSAB's Technical Advisory Board. Every four years, the Social Security Advisory Board appoints such a panel. This year's is chaired by Bob Beuerlein who actually sent in a question that I haven't used so far. But I think one -- perhaps the distinguishing feature of the discussion so far, by this technical panel, has been the amount of attention paid to presentation.

There have been four or five people from outside, including leading journalists, who have come in. They're consumers of the trustee's report. The panel itself has designated a couple of members to focus on writing recommendations regarding presentation. This has been an important part of the discussion here this morning. And as luck or skill would have it, the panel this year is going to be writing, probably fairly extensively, on this particular topic.

If any of you have suggestions about how Steve and his team could improve the way in which the information that is presented in the various reports for which OCACT is responsible, please communicate them to Bob Beuerlein who is the chair of this panel,

VIDEO CUTS OUT

because we really are anxious to try and be constructive and practical. It's easy to come up with pie in the sky ideas about how the presentation should be changed. But for one reason or another, including, I have been told, such seemingly unrelated matters as the Americans with Disability Act on how material is or is not accessible by those with various physical impairments,

VIDEO RESUMES

those considerations are relevant as well. So I repeat, if you have suggestions, please communicate them to Bob Beuerlein, or to Kim, our chair on the Social Security Advisory Board.

Steve Goss >> Hank, could I ask that people not hesitate to also share them with us?

Henry Aaron >> Oh no, no, no, no. [laughter]

Steve Goss >> Sorry.

Henry Aaron >> Thank you very much. It's 2:15. The next session is going to take over. Don't leave your chairs.

Kim Hildred>> We're going to begin the next session.

Session III: Perspectives of Microsimulation Experts

Part II: Microsimulation Experts

Jagadeesh Gokhale >> Okay. Welcome to session three, part one. Part two, sorry. I'm not good with numbers. I'm just kidding. [laughter] we have a distinguished panel, we have a distinguished group of panelists here, who are going to focus on their modeling using microsimulation techniques in their respective areas. This is now not concerned with Social Security directly or with MINT. We have with us Richard Evans, who is senior lecturer for the masters of computational science program, and the director of the Open Source Macroeconomics Laboratory, and a fellow of the Becker Friedman Institute at the University of Chicago. We have Dr. Eugene Steuerle, institute fellow, and Richard B. Fisher chair, co-founder, Urban-Brookings Tax Policy Center, center on non-profits and philanthropy, and opportunity and ownership initiative, and the co-director of the tax policy and charities initiative Urban Institute. And we also have Rocky Feuer who is the branch chief in the surveillance research program within the division of cancer control and population sciences at the National Cancer Institute. Dr. Feuer received his M.S. and Ph.D. in biostatistics from the University of North Carolina, Chapel Hill. So I expect this is going to be more of a matter of discussion regarding microsimulation do's and don'ts, best practices, what the experience was, how to convince your clients that what you are doing is really valuable and should be the basis of whatever policy changes, protocol changes, and so on that you are contemplating. So I think again we will start in the same order as they are seated. Let me request Eugene to go first.

Eugene (Gene) Steuerle >> Okay.

Jagadeesh Gokhale >> And I think we have, yes, a clicker for the slides. We have ten minutes, except Rocky requested twelve minutes. So we are going to adhere to ten minutes for Gene and Rick and then twelve for Rocky. Thank you.

Eugene (Gene) Steuerle, Institute Fellow and Richard B. Fisher Chair, Urban Institute

Gene Steuerle>> Thank you. So, I have number of slides here, but I'm not going to go through all of them. In part because much has been said already and I don't want to repeat it. But I do

want to emphasize a few items and just by way of background. I know Jagadeesh introduced me, but I was involved in bringing the retirement project to the Urban Institute early and recruiting Melissa Favreault when we were debating with some of the people at the Urban Institute about whether we'd even bring on a DYNASIM project. One of the things that really got us going was then we got this contract to help create MINT 1 at which point we were lucky enough to get the fabulous and unbelievable Karen Smith to work with us as well.

Also, I cofounded the Tax Policy Center. We do a lot of microsimulation there on tax, tax bills. The one difference is we don't really do very long run projections. We do do annual numbers and try to project out ten years through ten sets of annual numbers. But it's not as complicated as happens with DYNASIM. And then when I was back in treasury many years ago, I headed up the Office of Tax Analysis, and some of my comments actually reflect my feeling about how that office operated versus some of the offices in Social Security. Because we had both the estimating and the analytic function under my auspices at various times.

So the first slide I want to talk about is just microsimulation in a policy context. And this has been said in some ways, but I want to emphasize it, is that in a policy context, the main output that you get from these microsimulation models, yes, we do try to project where current law is going. But within a policy context, the policy debate, the main parameters that people look for are sort of total spending, total cost to Social Security. They also try to see actuarial balance between the two and the distribution and often those numbers will drive the process. They'll drive the political process at times even more than you would like, because often various issues of efficiency, or target efficiency, other stuff like that, gets set by the wayside.

I'm going to talk about that a bit more because progressivity is extremely important, but it's often not conveyed in enough detail to make sure the provisions are very target efficient. I also want to emphasize just how important are the rare fiscal institutions and independent score keepers we really have. We have a lot of offices in government, but I could probably name on one or two hands the number that I really rely upon day after day, year after year to provide really solid, analytic, apolitical estimates. Not only to provide them, but be in a position to provide them to the public. The actuary's office certainly stands out in that regard. I think ORES is getting close, but I think as came up in a previous panel, I think sometimes they are a bit more limited in part because they just don't have the long history of the actuary office and that's important in this context as well.

VIDEO CUTS OUT

The next thing I want to talk about is how important packages are versus separate provisions. And this is also a fundamental problem in most policy process usage.

VIDEO RESUMES

I would count Social Security tax bills, budget bills, almost every bill I have worked on as having a fundamental problem--people, and I'm including economists and even researchers or even groups like the Bipartisan Policy Committee, they like to go in and debate these issues one at a time. You know, like what do I want to do about the retirement age? What do I want to do about the rate schedule within, within Social Security? What do I want to do about minimum benefits? And then they like to go through these and debate them one at a time whether they like it or not. And a point of fact, what you really want to care about at the end of the day, is what the package looks like. And this leads to all sorts of misleading analysis. So, for instance, you often get provisions and there's many of them in existing law. I still think the data are not out there adequately that are very target inefficient.

One of my pet peeves, I worked on for 25 years at least, is the extraordinary inefficiency of the spousal and survivor benefit

VIDEO CUTS OUT

, which basically requires everybody to pay for it. But because it's a freebie in Social Security, unlike in a 401-k plan where you would take an actuarial offset, we're requiring single heads of households, for instance, who are among the poorest people in the population and whom were probably one of the targeted groups for spouse and survivor benefits, and since they were raising children, they pay for spousal and survivor benefit and they get nothing out of it. It creates all sorts of target inefficiency, but you could argue that if you just

VIDEO RESUMES

eliminated the spouse and survivor benefit or made some simple changes, that you would be left with a lot of poor people. So it is progressive by some measures, but it's horribly target inefficient. A microsimulation model allows you take these types of provisions, re-estimate them, put them in a package, and hit your target efficiency through more targeted means such as the rates structure or a well-designed minimum benefit.

We got into a discussion how you design a minimum benefit as well, which is microsimulation allows you to do that type of thing. It often requires a lot of interactions. And in the policy process, this gets to the speed at which we often can produce results under MINT or even the actuary's office, if it's slow, it's often not possible for these groups now trying to go back and redo what they thought they had done and then not like the results. It's very hard for them often to go back and really deal with those types, those types of complex issues. But, again, I want to just make very clear how important it is, that you can solve a lot of efficiency in what I'm going to call horizontal equity issues, equal treatment of equal issues, as I meant to the spouse's and survivor benefits, you could solve those. They could be very regressive to begin with as long as you come back with another provision to deal with it. That effects not just progressivity, but a lot

of other issues here. And microsimulation gives us the tools, but it has got to be timely and it has got to be there for you.

Several people have mentioned this so I'm going to briefly mention this. Microsimulation is mainly an "if then." it's not a predictive model, everybody has said that. But I again, and here I might disagree a little bit with Howard. Is even when you are projecting out 50 years, it allows you to say a lot of things, about under those conditions. Jagadeesh called them conditional conditions, under those "if" conditions, you can actually then see what the change in the policy is and how it plays out. And microsimulation is very, very good at that. And often, Henry said this, Henry Aaron, the change in what you are seeing may be accurate, even at the level, but what you get out of it is not.

On macro outcomes, I agree with the past panels. DYNASIM, POLISIM, MINT, whatever, they are not really good at macro outcomes. If you are really worried about the macro outcomes like the long-term direction of earnings or the GDP or something else, you are better off going to much simpler models. Because they are generally driven by certain things, like political forces, culture, hard instinct, and so on and so forth. But once again, microsimulation is good at showing the budgetary distribution and other effects of your microeconomic assumptions. And I should say also that sometimes, MINT and other models, they give you the smoking gun for things you theoretically might know. So, for instance, the biggest past error in projections, if you wanted to consider these as projections, is they didn't deal with the decline in the birth rate and all of the implications that we are going to have for Social Security. But if you had had a MINT or DYNASIM model available, then you could have actually shown the consequences of different birth rates and how well that system had evolved. And you can actually interject that into policy process and you can index in Social Security to deal with issues like changes in the birth rate, and changes in mortality. You can show that in this type of dynamic model. It wasn't done in the past in part, because the evidence was never actually made available. I'm going to get prepared to tell you this. I want to make sure everybody gets enough time.

I am going to go over my list which is issues that relate to how you set up and design an office. So it is very important to create expectations for what will be produced and when. And this is kind of what several panels have made, but if something is expected to be put out and on a particular basis, then people actually not only know what to do and how to react to it because of their own planning process, but just keep the political forces at bay. The political forces working against offices, even the Office of the Chief Actuary, which at times people have tried to politicize. If people know that outfit is coming, then their office is much more protected in terms of being able to put it out, and in fact in informing the public about what's expected to come out. We need to think about a way to do that which has been resolved, when they come out three weeks later or five weeks later or one week later, we need to know exactly when it is going to come out. Once they're expected to come out, they will help to direct the policy process.

Sometimes it may be too much. It's almost like teaching to the test. So like this is every major tax bill now has a different estimate produced by the joint committee on taxation. No political force can stop that from happening. You need those types of distributions results as well

VIDEO CUTS OUT

, for government offices. Finally again, I encourage you to really think about how to keep these government offices, and make sure they are timely, make sure they assimilate dollars for a lot of resources, not just actuarially, but every other one. You want this to be the story

VIDEO RESUMES

for policy purposes, you want to integrate the estimation at building staff which has been a problem through Social Security. I am still not sure why that's not been done more. But I think that's really, really important, that builders move in that direction. And you got to worry about adding staffing. A lot of the projections to what MINT has not or been able to do or not do has to do with that. So with that, I am done with my time.

**Eric J. (Rocky) Feuer, Chief, Statistical Research and Application Branch,
Division of Cancer Control and Population Sciences, National Cancer Institute**

Rocky Feuer >> Okay. I'm talking as a statistician surrounded by economists, and I'm talking about something completely different. That's why I asked for an extra two minutes. So good luck. Thanks for the opportunity to talk about the area of simulation modeling and cancer control planning, and more specifically what we call, CISNET, the Cancer Intervention and Surveillance and Modeling Consortium. It's the National Cancer Institute's consortium for population modeling to guide public health research and priorities. It is a collaborative consortium. What is called a U01 grant mechanism of modelers, in six cancer sites, breast, prostate, colorectal, lung, esophagus, and cervical cancers. It was formed in 2000, when many of your models started. But some of the cancer sites joined more recently. These are competing competitive grants that are renewed every five years. The overall purpose is to synthesize trial, epidemiologic, and surveillance data using simulation modeling to guide public health research and priorities, and to help address what might be called the formidable and growing gap between the rapid pace of innovation and cancer research and our ability to efficiently harness it to improve population health.

I want to talk briefly about a basic principle of CISNET and that is what we call comparative model and the consortium organization we developed to facilitate it. In many instances in the past, individual modeling efforts yielded results that were very difficult to reconcile. And here's an example that occurred. It wasn't part of CISNET, but right around the time CISNET started, and it was about the cost effectiveness of CT screening for lung cancer, published in 4 separate articles in major journals at the around same time. The results went from \$2,500 per quality

adjusted life year saved all the way up \$154,000 per quality adjusted life year saved. When people in CISNET actually took some time to try and reconcile these results by looking carefully at these articles, it was impossible because there is so many different assumptions, inputs, and things like that. This really hurts the credibility of modeling this type. People feel you get any result you want from modeling if you change the inputs appropriately. The approach innovated by CISNET is a systematic comparative modeling approach, with questions to be addressed by groups collaboratively with a common set of inputs and outputs. The reproducibility across models adds credibility to the results, and differences point out areas for further study in a very systematic way. And we have designed an administrative structure to facilitate collaborative modeling with the very close working relationships that are necessary to accomplish this. We have six multiple PI grants, one for each of the six cancer sites in CISNET, each with a coordinating center and between three and six independent modeling groups at different institutions. And we really like this approach, because rather than building a single “super model,” because by building separate models we get an idea of what we call the structural uncertainty among the models.

The structure for CISNET starts with common input simulators, that's risk factor trends, screening behavior, diffusion of new treatments. Next, the heart of CISNET is the individual cancer simulation models. And because we're simulating, we could model parallel lives with and without the interventions. It starts with demographics in early life, the sex, race, year of birth, the development of risk factors. Then it moves into preclinical cancer, precancerous lesions, and tumor initiation and growth. The pre-clinical natural history of disease isn't observed directly and has to be indirectly inferred from autopsy, screening, and other studies. We only get indirect glimpses of the natural history disease, and this is really one of the major challenges of CISNET modeling. Next is post clinical cancer, the diagnosis of cancer treatment, and then death from cancer and other causes. Then we sum together the individual histories, again with and without the intervention to model metrics such as the harms and benefits, mortality, quality adjusted life years, overdiagnosis and direct medical costs. And then we add calendar time by modeling multiple birth cohorts to represent the real U.S. or other populations over time.

As an example of one of the input simulators is the lung cancer group smoking history generator, which was developed based on national surveys of smoking behaviors since the 1960s. Given a person's sex and year of birth we simulate their age at smoking initiation, how many cigarettes per day they smoked, and their age at cessation if they quit. Rather than everybody modeling this separately and in different ways, this is an example of a common shared input developed jointly and used by all of the models.

I want to briefly talk about the impact of CISNET. An overarching goal of CISNET has been outreach and collaboration with many groups asking important questions including various NCI consortia. In the upper left I highlight work in support of the U.S. Preventative Services Task

Force (USPSTF). This has been among the most important of these collaborations. USPSTF is supported by the Agency for Health Care, Research, and Quality who sponsor the USPSTF as an independent panel who make evidence based recommendations about clinical preventative services. CISNET supported the task force in developing screening guidelines in colorectal, breast, lung, and cervical cancers. Prior to CISNET the task force relied solely on reviews of direct evidence and rarely used models because of the credibility issues I mentioned, which had been ameliorated by CISNET's comparative modeling approach. Working on important issues has resulted over the years in over 450 publications for CISNET, not just in specialized or technical modeling journals, but in high profile and influential medical journals such as the new England Journal of Medicine, the Journal of the American Medical Association, and the Annals of Internal Medicine. It is interesting that one ironic way that we realized that we were making a difference was being engaged and critiqued by high level and influential decision makers in medicine and public policy, but we also got some praise as well, which was good to hear.

I want to talk briefly about what type of questions the CISNET models can answer. One type of question revolves around understanding. The impact of cancer control interventions, that's screening treatment and prevention on current and future trends in incidence and mortality. For example, why did breast cancer mortality decline 24% in the U.S. between 1989 and 2000 and then another 16% between 2000 and 2012? A second question related to extrapolating evidence from randomized controlled trials, epidemiologic and observational studies determine the most efficient and cost effective strategies for implementing technologies in the population. For example, at what age should women start and stop having mammograms and how often should they have them? A third type of question relates to being responsive to challenges due to the increased pace of technology by helping to determine which new technologies are the most promising when scaled up to the population level. For example, a CT colonography, sometimes called virtual colonoscopy, a good technology to potentially replace the usual optical colonoscopy. In the area of state, local, and international cancer control planning, what is the best strategy for colon cancer planning screening in my area?

I want to talk about the kind of iterative approach that is used for model calibration and validation using lung cancer screening as an example. There's been two major trials of lung cancer screening. The first is called the PLCO trial. That's an older trial of no screening versus chest x-ray and it included smokers and non-smokers and ended up showing no benefit for x-ray. A newer trial is the National Lung Screening Trial. This was a newer trial of chest x-ray versus low dose helical CT screening. It included only heavy smokers, those with 30 or more pack years. It showed a 20% mortality reduction for low dose CT screening. CISNET was asked by the U.S. Preventative Services Task Force to extrapolate from the National Lung Screening Trial, for example, to people who smoked less than 30 pack year, to assist them in making screening recommendations for the nation. The modelers first calibrated their models to the National Lung Screening Trial. Then they validated those models against those in the PLCO trial who met the

NLST eligibility criteria. In other words, the heavier smokers, and it fit reasonably well for most models. But then they went and tried to validate their models against the full PLCO trial including never and lighter smokers, and the fit was not as good for some of the models because this was an extrapolation. The models that didn't fit as well were recalibrated the lighter smokers in PLCO to allow them to move on to do the work with the U.S. Preventative Services Task Force.

I want to talk about two challenges to CISNET. And the first is that models tend to agree on relative ordering of strategies, but not on absolute levels. And I think this relates to a comment people said earlier about differences between strategies, which is a more robust outcome across models than absolute outcomes. As an example, the first time the CISNET breast group worked with the U.S. Preventative Services Task Force, they modeled 40 different strategies with different periodicity, starting and stopping ages of screening. This graph is for one particular model to show how we pick efficient scenarios. We graph each scenario with percent mortality reduction on the y axis, and the number or mammograms measure of health care utilization on the x axis. If you have two strategies that require equal number of mammograms, and one produced a larger decline in mortality, that strategy is considered more efficient. Therefore, scenarios to the upper left of the graph are considered more efficient than the ones to the lower right. We draw a curve called the efficiency frontier for the scenarios, which are considered most efficient for a specified number of mammograms. We plotted only those scenarios on or near the frontier.

When we display all six breast models, we don't just have one efficiency frontier, but six. There is quite a bit of consistency for those strategies that are considered efficient, but the level and shape of the efficiency frontiers differ. For example, if we select the most resource intensive scenario, that is annual screening from age 40 to 85, the models vary from just over 30% to over 55% mortality reduction relative to no screening. This reflects the inherent variability of the underlying trials of mammography screening done in different settings, which are used to inform the models. However, in estimating strategies relative to each other, i.e. which ones are on the frontier, which is usually the most important outcome, a lot of these settings specific characteristics cancel and the results are more consistent.

A second challenge revolves around confusion about where to place modeling results in the scientific evidence hierarchy. The hierarchy starts with the highest level of evidence that is randomized trials and systematic reviews of randomized trials, and goes all the way down to editorials and experts opinion. The question was where do modeling results fit in this hierarchy? So we wrote a paper arguing that modeling doesn't fit any one place in the evidence hierarchy, it's a tool to summarize information. Where the results belong in the hierarchy depends on the amount of extrapolation necessary from primary data that was used to inform the model. In that paper we produced the illustration shown here where questions such as using modeling to decide

between annual versus biennial mammography for women age 50 to 69 is high in the evidence hierarchy, because even though there is no direct trial data testing these head to head, there's really a large amount of trial data to inform the model about this question. If you look all the way to the far right in contrast, the impact of incidental medical findings that you find when you do a virtual colonoscopy, where the evidence about long-term benefits, costs, and harms of following up on these incidental findings is largely unknown, indicates that modeling results in this area deserve to be lower in the hierarchy. Now quantifying this for any particular question is, is difficult, but at least we set up what we felt is a structure for where to place modeling in this hierarchy. So thanks a lot. And I'm glad to answer questions about CISNET or anything else.

Richard Evans, Director, Open Source Economics Laboratory, University of Chicago

Rich Evans >> All right. Well, while they are pulling up my slides, I will just give you a little bit of my background of why I am here. I think it's because I am a core maintainer of an open source macroeconomics model that is used in the United States. It's used in Europe and starting to be used in India. I'm also a consumer and light contributor to open source microsimulation models, and in particular how to incorporate those microsimulation models into these bigger, dynamic generally macro models. I'm fascinated by Rocky's presentation about how medical models are different from economic models and how they are similar. So I'm going to try to use his modeling framework in my definition of a microsimulation model. I mean, this is something we have all seen. We know they are these models that take a lot of heterogeneous individuals of some type and show outcome from those individuals. But the key point in this slide is the three components to a microsimulation model. And the one we usually think of as number two, which is the institutional detail. In a tax model that is all the "if then" statements about the tax code. In a medical model, like what Rocky was showing, these cancer institute models, that has got to be the physics and biology behind how a disease advances. And then the other two components of a microsimulation model are the data that gets input into those and then behavioral assumptions. In a medical model I am not sure what that would be. It's like how would somebody's response to treatment influence the biology. In a tax model, we've already talked about what that might be. If you increase the payroll tax, people may work less. That's going to influence the revenue aspect balance that you get.

So the models that I'm used to dealing with are, in microsimulation terms, these models of tax policy, tax calculator, does the personal income tax side. And that is open source model. And then cost of capital calculator is one that does the business taxation side, and the heterogeneous incentives that businesses face. The data we use for those, there's private data sources from the IRS that limits what we can do in terms of our open sourced-ness, and so we have created a CPS match that we also use that makes it so we can have it be public as well. Then we've got all of the federal tax logic and then on top of that there is some behavioral assumptions that a user can

choose. Do you want people's elasticity in response to higher taxes to be higher or lower? That is a source of debate. Just let people choose what they think it is. So in doing these models, my presentation is going to be a little bit different, in that I want to focus on some lessons that we have learned in this open source environment that I think apply to any models whether they're open source or not. And they have to do with the modularity of these models. And that really has to do with your ability to Frankenstein a bunch of them together or have them work together or plug them into other bigger models.

In addition, with that public private data issue, I want to just make one note about synthetic data and the data input that might allow us to use administrative data more broadly. And then finally I want to talk about workflow and some collaborative methods that I think, they're certainly best practices in the open source software world. And I have found in my own research and in this open source policy model environment, those are great. All right. Some of the organizations that I work with, I'm the director of the open source economics laboratory, open source policy center is located here in Washington, D.C. out of the American Enterprise Institute. Policy simulation library is a resource that has a repository of lots of open source economic models. I think Bryan, is yours open source, the health model? You said it's public. That's fantastic.

Basically, when I hear about an open source model, I just want to go use it. I don't want to reinvent the wheel. And I want to contribute to it. And then QuantEcon. One of these things we have learned about these microsimulation models, is we need to make them modular and that involves standardizing how the inputs are passed to the models and how outputs are passed out of the models. Because you've got, MINT is a big SAS model. I think you said your stuff was SAS and STATA, you've got Python, you've got R, you've got MATLAB. If you harmonize how the inputs are passed to models and how the outputs are distributed, then it doesn't matter what platform you are on. And lots of other people can use those. Also there are a lot of tools and modern software that allow you to harmonize the environment on a given computational platform like Mac versus PC, or AWS cloud servers versus Google cloud. And in Python, that is CONDA environment. And you have got Docker and some best practices to allow common platforms, so you get the same answers even though you have the same code, you might get different answers if you are on a different computational platform, unless you use these types of standardizing environments. When your code is modular, you can link things up.

So three examples. One, we're currently trying to do static squares for the U.S. economy, that involve more than just the personal tax code. And so we've linked our tax calculator microsimulation model with this business taxation microsimulation model that captures all the business and personal incentives. The only reason we can do that is because we've made them both modular and easy to talk to each other. One that I have been heavily involved with was incorporating microsimulation tax data into a general equilibrium micro or macroeconomic model. And then lastly, this was discussed a little bit in our previous session, but you can make

these models usable to non-experts by creating web applications. And for web applications, somebody goes to a webpage, submits some parameters, and hits go. And then that needs to be sent off somewhere in the cloud to run the model and bring the answers back. That to me is kind of the holy grail of these models is when my grandmother can run them, you can crowd source policy. Data and privacy, so this one, I just wanted to highlight again, that data component of microsimulation models. When we have administrative data, that's often one of the main hurdles or obstacles we have to making our models public. And there are a couple of projects going on right now. I am just going to jump to the bottom of this slide. Len Burman at the Urban Institute is doing a project on synthesizing data, where you give a synthetic dataset that has the same statistical properties and then anybody can use it, and it doesn't jeopardize the privacy of that data. There is another project with a group at AEI, and then I am working on one as well.

Lastly, this is a collaborative workflow slide. This is a screen shot, because you never want to do a live web demonstration in a presentation. This is a screen shot of a GitHub repository. I would submit that any good project needs some careful version control, and git and GitHub are just one way of doing that. There are other methods, but I want to show you some of the virtues of this. And not just for open source. If you have a closed source model, you can use this in enterprise version on your own systems. But this is a screen shot of a repository for our microsimulation model, it just shows some of the files. You don't care about that. Also, there is a page that shows the contributors to this model. One nice thing here, is I can see who has contributed the most and I can go click on any one of these people and see exactly what they have contributed, what lines of code. And there is a much longer list to these contributors here.

And then this next slide is the hardest one. This is a workflow that I don't want you to take anything more from this slide, than to look in the upper right-hand corner and see that the main code base is walled off by two dotted lines. This is important because most of the models that we have talked about today are represented by thousands of lines of code and to go in and change any of them is a very dangerous endeavor. And so this version control workflow allows you, the vertical line separates the main code from anything else in the cloud, and the horizontal line separates the cloud from things on people's hard drives. And the only way to get thing, get any code into the upper right-hand corner is to submit some kind of a request that people have to look at and evaluate.

The last thing I want to show you is issue tracking and discussion on these types of projects. So there are two that are my favorite. I will show you some screen shots of one of those. So this is a screen shot of a discussion on a model that I was working on, and it has to do with having different interest rates that the government pays on debt versus what private companies pays on debt. That happened to be important if you are trying to forecast how the debt-to-GDP ratio is going to evolve over time. And the first line is a post by Jason Debacker saying, "Hey, we should think about this." The second one, is a post by my friend at the Congressional Budget Office,

Kerk Phillips, who is in the front row, saying, “Well, here's how we do it at CBO.” We have some discussion, I put up some equations of how I think we should do it, a pencil drawing. And finally, we get it implemented in the code and somebody writes down what we did, and says, ‘Okay, this issue is closed. It was solved. It was merged in, in pr1 whatever.’ So the nice thing about this, is this becomes part of the permanent record of this code base of this model. And six months later, when we have this in the model and I don't remember why we did it, you can always go back. These are not conversations that took place over email. These are conversations that are directly linked to the model. I just submit that this workflow, collaborative workflow is great for research. I’m an academic. I do this in my papers. It's great for open source projects and it's great for any modeling projects that are private as well. Thank you.

Jagadeesh Gokhale >> Thank you. Thank you to all of you for providing all of your presentations and comments. I think it was a very good set of presentations and information. I have one question though which has been nagging me for the last ten years. And this has to do with making distributional consequences of Social Security public. I have heard arguments on both sides of this issue. The MINT model is basically about implementing distributional analyses under existing Social Security policies and under proposed policy changes. And I have gotten comments saying you don't want the distributional consequences to be made public, because that will only polarize the discussion even more. On the other hand, there is the arguments that policy-making should proceed under full information. I’d like to ask the panelists. This was mentioned by Gene especially. So which view do you favor?

Gene Steuerle >> I think it is obvious. The goal, the primary goal of Social Security is to redistribute to people in need at some level, need because of retirement or need because of you want to correlate this with other poverty and other needs in society. And if that's your goal, then you need to measure it. I mean, there was a similar issue about whether you project long-term or not. As long as the Congress decides it is going to write in the law something that's infinitely there for the future, you know, if you had a business that signed a contract that was going to buy plants for the next 100 years, it would estimate it. It might be a bad estimate. But they are still going to estimate it. If you decide your goal of Social Security policy is going to be to redistribute, then you should measure it. In fact, one of my concerns is in point of fact, is I think that because we often think of, I talked about particular policies like spousal and survival benefits considered in isolation. I even have problems with the issue of Social Security considered in isolation.

So my budget work, my broader budget work says, all of the growth in government spending, all of it is going to three things: health, Social Security, and interest on the debt. Everything else is going in a tail spin. So sometimes I want to think about Social Security reform, that might say, ‘Well, how about Social Security reform and Medicare?’ And somebody earlier, I think maybe it was John Sabelhaus, and private retirement. So I want to know what the distributional

consequences of those things are, so I can meet that objective. I don't measure that type of thing, and across policies, then I'm not going to meet the objective very well and I'm going to spend a lot of resources of society in a very inefficient, in a target inefficient way.

Rich Evans >> I think the distributional analysis, like you said, Gene, is super important, and I think sometimes people don't want to show it because that might be the part of our models sometimes that we're least confident in. That's probably not a good reason to not show it. We probably need to show it, and take whatever heat we need to take, and fix the models. But it is also very hard to display those results. Like the distribution, we are looking at age, we are looking at income groups, we are looking by time period, maybe marital status. It is really hard to make a five-dimensional figure and show it to somebody. There are ways to do that. I think one thing that I have and we're still experimenting with this. We have these distributional results from our macroeconomic model, and we have one way of showing those results. I'm not sure it is the best way though. There was discussion earlier with the SSAB folks about whether to just do a data dump and let other people digest the information, or to write your own summary figures and reports. Maybe that's something to crowd source. It's a hard thing to present. I don't know what the right way is. Let other people decide what the right way is. So give all the information and let other people create what they think is a good distributional table or figure.

Jagadeesh Gokhale >> Okay. Another question, especially for Rich Evans, who has shown a lot of the modern technology-based methods for collaborative project development. You know, you were talking about harmonizing inputs and outputs, versioning, workflow controls, automated validations, and so on. There is a whole laundry list of things that we can do in today's computing environment compared to how things were done even ten years ago. From what I learned, the MINT model is a SAS based model. To what extent -- so this would be a question for the MINT team as well -- to what extent have you adopted these modern workflow control and versioning and collaborative project development techniques? And my question to you, Rich, who is experienced using these techniques, how much does it cost?

Rich Evans >> It's costly. Just think back to that figure, with the yellow boxes that I put up. I remember the first time I started trying to use git, it's so unintuitive the first time. Until you start running into the problems of protecting your code source. There's a government organization that's major and it will remain unnamed that I was talking to somebody there about one of their main models, and it was old legacy SAS code and they said that there are probably 20,000 lines of code in this model that they we have no idea what they do. And you get this and it is from turn-over. It is from people who knew what the code did who leave and then or they get out of the business or go somewhere else. And there were not methods and best practices put in place to keep that alive. I mean, even things as simple as documentation. There was no clear documentation in human language words about what was behind what this code was meant to do. And no, documentation is not the fun part. I'm not an economist because I wanted to write

documentation. And I hate it when I'm writing it. But it's, it's something that is totally necessary for code that I write today to be, for me to understand it a year from now. So it is costly, but I think it is a lot less costly than the alternative of building up something really big that you can't really add to anymore because it's too complicated.

Jagadeesh Gokhale >> Okay. Thank you. Let me open it up to SSAB members. Any questions?

Nancy Altman >> I'm talking about if there are other ways to get certain kinds of information? Other kinds of information to get at those kinds of specifications, variables, and so forth.

Jagadeesh Gokhale >> You are talking about the kind of documentation that Rich was, was talking about?

Nancy Altman >> More generally. That, you know, you're experts on microsimulation. And we, in the earlier panel, we saw a lot of different models. And we talked earlier this morning about the limitations of the MINT model. So I'm just sort of curious whether there is something more general. If there's a way that you can say, well, microsimulation is great for, has these strengths, but have these weaknesses, that really can't be overcome.

Rocky Feuer >> Oh, yeah. Well, we talked about you know single, a single super model, versus having a variety of models. And we found that you know a real strength in, in the type of modeling that we have done. We always have a constant debate. When I talked about the evidence hierarchy, you know, in cancer and in health, we have clinical trials, and that is usually viewed as the highest form of evidence. And trials tend to be what I called empiricists. It's pretty direct. It's pretty direct evidence. And there is a big debate about, you know, when we, when we have models with assumptions and difficulty in transparency. And there is kind of an active debate is

VIDEO CUTS OUT

to how, you know, how much you have to depend on trials. We can't have trials for every single situation. And so we have to depend on models for many, many things. But it's a distinction between the two, between detecting empirical evidence and modeled evidence which is, and as I put up in that chart extrapolation.

Rich Evans >> Well, I like that you don't have one harmonized model of the whole human body. Right. In economics we don't have that either. We've got lots of different models for finance, for personal finance, for the macroeconomy. It would be so great to have the economic super model that includes the entire world and any other worlds that we have interstellar trade with. But that's just a really--there's this tradeoff between tractability and the question that you're interested in. And just sometimes it's better to, all models are wrong, and some of them are useful.

Rocky Feuer >> Let me just say, and we found as we first got modelers, they are all

VIDEO RESUMES

say breast cancer modelers. They all thought when we first sat down that their model was right, and they are all geniuses, and then they start comparing, and they start comparing their models. And I think it gives them an ounce of modesty, and then they really work together to try and understand their models, and what are the strengths and weaknesses of each. So that's been a very good experience.

Gene Steuerle >> I'm interested, I think so many questions here in my mind., I always think of maximization subject to constraints. I would guess Rocky, quite honestly when he does his work, has ten times the resources that the SSA people have. There is many in health. I have two daughters, one is a doctor, one is not. They are both equally educated. The doctor makes twice as much. So I mean, but Rocky indicated all the constraints of what he could do, you know, even in terms of how many models he can compare. And so you are always doing this, what can I do subject to a very limited resource constraint. It's something we haven't talked a lot about also, but also you were hinting at, is also the incentives. It is very important to think about the incentives, both internal to SSA in terms of what they can get out. The incentives and academia which often preclude the types of efforts Richard when he talks about coding and documenting, you don't get publications out of that. You don't get tenure out of that. You might get tenure if you are the first person to do the model that gets it out there. But what does that mean about 20 years from now if you really want this model to be around? So we are trying to do a lot subject to a lot of constraints. But a lot of the discussions I have seen around this table, have sort of centered on that. How do I balance all of these worthwhile objectives given the limited resources I have? And also given the incentives within the systems to do it, which, of which I think you are raising just one, one piece.

Jagadeesh Gokhale >> Okay, that gives me a good segue to turn to the SSA folks. Any questions, comments?

Mark Sarney >> One comment about Richard's presentation on the GitHub. Some of the staff and I were exchanging glances at each other, because we have actually tried to use not just GitHub, but Python. We did bring in a web developer to try making both modeling for ourselves and the presentation of results easier. And we went through a succession of people with computer science backgrounds, and inevitably they would say very overtly, I'm only one person, we really need a team to do this work. So we also found there is a big chasm between the social science folks and the computer science folks as far as of what kind of software tools you could absorb and use effectively while doing other things. So I have a question, as far as the open source stuff, what kind of infrastructure, as far as support, moderating the discussions, and

managing the GitHub workflows and all of that kind of thing goes into something like that? Because to answer Jagadeesh's question, we have a very crude version of all of that at SSA, for MINT, which is we use SharePoint for a lot of the modeling discussions. We inevitably go back to those hundreds of times to try to remember what it was we did when we last modeled something similar. We also have kind of a way to keep the code secure, so we don't accidentally mess it up. We have a lot of those things built, but they are very much sort of a poor man's version of a more sophisticated software approach.

Jagadeesh Gokhale >> Okay. Do we have anything else?

Mark Warshawsky >> Yes. I actually have, this is a question for professor Feuer. It was not clear to me as to, there seems to be, you used six different models, and I don't know if that is literally six or that is just an example. And there is some standardization. But at the same time there are different models and then you aggregate them in some way. I'm a little unclear as to what actually is going on. I was wondering if you could just explain it just a little more. Because, I mean, I don't know that it will be directly useful

VIDEO CUTS OUT

for us. But one idea which did occur to me just from the conversation that you had, at Social Security we actually do have something like a field experiment. We do demonstration projects and that is the highest level of knowledge, we have not integrated all of our demonstration development. It's expensive, very expensive and they are hard to do. But anyways, if you could expand on what you are doing.

Rocky Feuer >> Yeah. So we have each of the cancer sites, j

VIDEO RESUMES

Just the state of modeling, you know, who won awards. There is between three to six different independent models that were developed prior to entering CISNET. So then, and in the early days of CISNET, every model came in as an independent, as an independent grant. And then it was difficult because we tried to put them together and collaborate. Now, they come in as a single grant and they kind of coordinate what they are going to work on. And what that means is, these are independently developed models, and it's really, the biggest difference is how they model the preclinical natural history of disease. Because that's unobserved and you could model it very different ways. And you have to make an indirect inference about what that is. How fast do polyps grow? Do some polyps stop growing, or do some polyps just grow very slowly? How do risk factors affect the growth rate of polyps? Do different risk factors affect the initiation of polyps or the growth rate of polyps? Now, there is other aspects of how they differ, even a post clinical diagnosis. So some, some of the models, some of the models just model survival and

then they, then when there is a new treatment comes out, they go to the clinical trial literature and they say this treatment has a better hazard ratio, a slightly different better survival.

But some model, do it a little more biologically, they model a cure, no cure model. So that, you know, what is the time of the first fatal metastases that somebody has biologically in their body? And if they have that, they are destined to die of their disease if they don't die of other causes. So that's the aspect at which they differ. And then some of the more demographic inputs like life tables from other causes or the smoking history generator, or we have a PSA screening generator, so it generates the U.S. history of men in PSA screening. We feel those are shared and there is a big constant debate within CISNET about, for in any particular cancer site, what should we harmonize and what should we let be independent. If we harmonize too much, we don't have six models, we have one model. If we let them, but it is nice to take things that can be modeled relatively easily and try to have those as a single common input. So kind of the, you know, so I guess I should say an economic ceteris paribus, that is an economic expression, you know all other things being equal.

So we try to see you know step-by-step, you know, what's the differences between the models. And at first the modelers didn't understand each other very well. And, you know, there were differences between the models. And we have our cervical group is fairly new. And they were asked by the U.S. preventative services task force to go in and try to do an analysis. And they said no, because they weren't ready. They didn't really understand the differences between their models. Now that they've worked together long enough, most of the modeling groups when they see a difference, they understand why there is the difference. And is the difference because somebody made a mistake in their model, or is the difference related to some underlying, the underlying lack of basic information that feeds the model. And then it reads to, what studies can we do to narrow those gaps? And that, I think that formula could be used in other aspects, in other areas. Is that clear?

Mark Warshawsky >> Yes. That was very helpful. Thank you.

Jagadeesh Gokhale >> Okay. We just have a few more minutes left. Anybody from the audience with questions?

Ted Goldman >> Hi, I'm Ted Goldman, the Director of Policy Research Analysis at the Pension Benefit Guaranty Corporation. At PBGC, we have our microsimulation model called PIMS, Pension Insurance Modeling System. We have a unique opportunity to re-modernize our system which has all of the things you said, of lines of code that nobody knows what it means anymore and so forth. What one piece of advice would you give us as we start down this path of opportunity to really take a blank piece of paper and modernize a new system?

Rich Evans >> That's really expensive, like you are going to redo the whole thing? So, one, it's really expensive to translate code that's in one programming language to another programming language. There's a whole lot of testing that has to happen to make sure that the outputs from the new code do exactly the same things as the outputs from the old code. And maybe you find some mistakes that you want to change. But testing is a big issue, translation is a big issue. This goes back to something that Mark said a second ago, that when you are constrained on staff and resources, one way to do it is you can kind of outsource some of this programming. So imagine if you could break up your model into a set of sub-functions where you know what the inputs are, and you know what the outputs should be, and you know what some of the edge cases answers should be, you could put that as kind of a help wanted thing in the public domain. There are tons of computer programmers that care about policy, policy people who care about computer programs, that might in their evening, this is what I do for fun. There are other people like this, who would say, I know what this function is. I'd love to contribute this. And then in the end they don't need to know the whole code, basically they have just contributed one piece of it. But you can outsource some of it. That is one way to get around some of the resource constraints that we have.

Henry Aaron >> Let me just throw in here, that that example was the method used for rolling out the Affordable Care Act. It was not a conspicuous success. And the reason was that the separate modules were developed independently, they didn't talk well to one other another, and they produced interactions when combined that produced bad results. This is with respect to the federal exchange, so there may be situations in which this works well. There are also situations in which it doesn't. And I would be cautious about the general approach.

Rich Evans >> And I mean the Affordable Care Act was something totally new. Like these were just hypotheses about how things would work that had never been combined before. I'm a little less skeptical about those models not being perfect.

Jagadeesh Gokhale >> Okay. Yes. Steve.

Steve Goss >> Thank you. And it works. A quick question for Gene. Gene, in your early comments, I was little bit surprised, you said something about concerns about projecting fertility. That is birth rates. If you look historically back in 1982, projections we made for the 1983 amendments, our projections of birth rates at that time and of mortality rates at the time are a stunning success. We were really good on the demographics. What we weren't doing so well at the time was projecting disability incidence rates

VIDEO CUTS OUT

and also labor productivity growth rates. So I'm surprised you mentioned fertility and the necessity for micro, because if you know the age distribution of your female population, and you

have age specific birth rates and you make assumptions about what they are going to be in the future, that is pretty straightforward.

Gene Steuerle >> I wasn't actually referring to '83, I was referring going all the way back. I mean, I was basically saying that and it was partly related, it was partly estimating which is policy, it was basically going back to the early creation of Social Security. It never decided to index at all for various factors. And it could have, and I have said this a number of times in other meetings as well, this is a technical panel, if it had been indexed for some of these risk factors, then that is something you can show through a microsimulation. Or shown through a microsimulation how if you index for these factors, not projected but index for them, you can show through microsimulation how you can remove and reduce that risk quite substantially within the system. So that was my main point. I'm not answering to whether the '83 estimates were good or not.

Steve Goss >> Understood.

VIDEO RESUMES

But if I can just respond to that. We have actually had experience with the people on the Hill, I am not sure how many are still here. Back at the time of the 1983 amendments when we had increasing normal retirement age, reflective of increasing longevity, and we said we could index this based on what longevity actually turns out to be. And we were told by people on the Hill, and maybe they'd say something different now, they say, we want to have in the law exactly what the age is going to be in the future. Not some index to tell our constituents exactly what they are going to face when they approach retirement. Not something that was indexed that they can't speak to directly, so that's, that's why we ended up with what we have and don't have it indexed.

Jagadeesh Gokhale >> All right. Thank you very much. Let's give a big round of applause to this panel. We have got about a 15-minute break before our last panel. And if our folks could all come to the front. Thank you.

Session III: Perspectives of Microsimulation Experts

Part III: Merged Expert Discussion Regarding MINT

Jagadeesh Gokhale >> This is the last session of the day -- this time I'm going to get it right -- session three, part three, is to be a general open discussion about the issues we've heard in the morning. So just to recap, we heard presentations from people who developed MINT and are operating MINT; from the Urban Institute, Karen Smith; and SSA, Mark Sarney. We then heard from the users of MINT output from various institutions, particularly from Capitol Hill. And then

we heard discussions on other microsimulation models pertaining to Social Security and then finally, models that are independent of Social Security.

In this Panel, we will try and collect some of the issues that people may still have lingering questions about, want debated, provide another chance to think about the issues and exchange ideas. So one question that has come to us from the online source pertains to the microsimulation models. 'Key to any model is the assumptions that are used.' So this questioner would like us to talk a little about the process for setting assumptions. So maybe we can start with this question.

Mark Sarney>> The process for setting assumptions. Well, for MINT, the starting assumptions are essentially the Trustee's Report assumptions from the Office of the Chief Actuary. And a lot of the inputs to MINT are based on research on a whole bunch of other data sets, mainly SIPP, but also the others that Karen mentioned. So as far as setting up the model, the basic assumptions, the sort of macro fixed ones come from the actuaries. A lot of the projection pieces are coming from research done on various data sets.

But as far as assumptions about, you know, when a particular policy is modeled or something like that, there's a lot that comes from specifications. And some of it comes from, okay, discussing now what do we do because specifications only go so far. How do we deal with a situation of a divorced woman who's only been divorced for one year and the marriage rules that apply? And the proposal doesn't speak to it. So there's many different types of assumptions. Sometimes you muddle through, and oftentimes you can lean on other sources like the Trustee's report or what the Actuaries have done.

Jagadeesh Gokhale >> One question has come from the chairman of the Technical Panel, Bob Beuerlein. So I thought it was especially important to bring up. It is about assumptions and is something that's bothered me as well. So I might as well throw out my concern about assumptions. Some of the assumptions that, especially the actuaries make -- and Steve -- is Steve here?

Steve Goss >> I would be happy to address that.

Jagadeesh Gokhale >> With a microsimulation, tone is at the micro level, at the individual and household and family level. There are ways of modeling how they behave based on what has been observed in the past. We use that information to make projections about how they will continue to behave, given the regularities we observe in similar circumstances. We collect this information in various types of transition rules and make projections into the future.

One of the key variables we project is labor force behavior, which is key for Social Security because, ultimately, it's all about the payroll tax base and payroll tax revenues. So labor force

participation, the distribution of earnings, compensation and so on adds up to the wage base. So if you have micro level information on labor force behavior that can be aggregated to the full economy level, we can call it total labor input. And it is a consistent method of aggregating labor input and yielding total output or GDP. Following from GDP we can estimate the compensation base, the wage base, payroll taxes, and then benefits.

However, projections based on such a microsimulation-based aggregate may not agree with the macroeconomic assumptions that the Social Security trustees have provided. In that case, my understanding is that the microsimulation models are recalibrated to agree with those trustees' assumptions. Which is the right methodology? Should the microsimulation inform us about some of these things, which are currently assumptions at the macro level? Or should the macro level assumptions that the trustees provide us be used to recalibrate the microsimulation?

There's a tension there that I have not seen resolved in a satisfactory matter. My sense would be, if you start at a macro level, it's very difficult from that to go down to the micro level. But if you start from the micro level and aggregate things up, you have everything in a consistent manner. Everything is micro founded. So which is the right way to make Social Security projections? Should the microsimulation not be the fundamental way in which we capture all the micro level data, aggregate it up, and then make projections? Or should the microsimulation be recalibrated in some sense based on assumptions that the trustees provide us? This is a question that's been nagging me, and I'd like to hear a discussion about it.

Steve Goss >> I would suggest taking a little step back here. This is a little bit to the point that Howard was making before. I think it was Yogi Berra who was said to say, 'predictions are difficult, especially about the future.' And so the thing that we have always looked at, for a long time now, in trying to develop assumptions for the future is there's a very simple principle. And that is, look at the historical picture, whether it's micro or macro. Take your pick. See what has happened but understand the conditions that contributed towards what happened in the past. Then you're left with a very simple question: do you believe the conditions in the future are going to mimic the conditions of the past? If you do believe that, then great, get out your ruler, and just fit it to the past trends, and you're all okay.

But if you believe that the conditions in the future are going to be different, this is where it gets to be tricky. This is where there's -- I don't think there's a simple answer. Go micro, go macro -- really, you have to look at all of these factors and let them speak to each other and see what seems reasonable. I mean, to Howard's earlier point about having no real confidence in projections for the future, I think really what we have to keep in mind is, what are these models really for? What questions are being raised that they have to address? That's the real fundamental question. That a projection into the future cannot and should not be expected to be perfect isn't even really the point.

The point is to just really say, under these assumptions, this is what we expect to have in the future. And I would suggest that, whether you're looking, fundamentally, first at a micro level -- as you are, I know, Jagadeesh, or if you're looking initially at a macro model, look very careful at the past trends or the past experience of your historical base. And you have to ask the question, do I really think the future is going to exactly mimic that?

I mean there was talk before about bootstrapping for earnings histories. We do something equivalent to that for earnings histories going to the future also

VIDEO CUTS OUT

and make modifications based on expected changes in the labor force participation. So that gets complicated. But I would say the best of models is maybe like with certain kinds of cancer. Have six models and see what they all say and see what we can learn from each other on that. That, I think, has been our experience of what really seems to work best. There is no singular approach which you can hang your hat on and say, this is sure to be the best.

VIDEO RESUMES

Gene Steuerle>> If I can jump in here, because I worked on this a little bit from a different angle, and it's somewhat parallel to your response, Steve. I think you take all sources of information. On the macro questions, I don't think any model, actuarial model, microsimulation model, econometric model, general equilibrium model/theory, in and of itself, is adequate. And you really have to look at what information you get from every source. The reason I mention this in this particular case is because I've looked at the labor supply of projections for Social Security over time. And in the late80s, they accepted the notion that somehow or another retirement was a leisure good and we'd always demand more of it. And so they projected constantly declining labor force participation, for instance, with 65 – 70-year-olds. And then gradually over time, through the labor supply curve, groups have been going up, and gradually the actuaries started projecting, well maybe it's going to level out. And they were always behind what was happening.

But the models, the DYNASIM models, I'm guessing -- not because they followed the actuaries, but they followed the same thing. And just to give you another source of information, most people are estimating a labor supply curve. In technical terms, what you actually see over time, however, is not the labor supply curve, what people would do under their conditions, but the intersection of the labor supply and the labor demand curves. You're actually seeing the equilibrium results.

And basically, when the birthrate falls, people who have to provide the work, more and more, become the older workers. An extreme example is the birthrate falls to 0, 65 years hence, everybody is over 65. Would the labor force participation of people 65 and older remain at 25%?

Of course it wouldn't. Well, the birthrate fell one third of the way there so it obviously has to have some adjustment. Now, maybe you don't accept that result, but that's just another source of information.

You've got to really figure out how to pull all of these things together, and they've got to inform each other. And at the end of the day then, somebody's got to take their best guess. And maybe the actuaries are too slow to accept the microsimulation results with the information they provide. And maybe the microsimulation people are too slow to figure out what the actuaries told you. You have to inform each other.

Steve Goss >> The fact is, I believe it's at least 20 years now that our projections of labor force participation have had a factor for people -- I believe it's over 45 years old, which we call the longevity factor. The longevity factor is that, as life expectancy is projected to grow from one generation to the next, we've been assuming the labor force participation rates for people over 45, 55, 65, 75, would be increasing. This has been in our model for a long time. We've got to make sure that we share this with you so you're aware.

Gene Steuerle >> First of all, you're talking about longevity and I'm talking about fertility. We don't have to discuss this here.

Steve Goss >> I thought you were talking about labor force participation at higher ages.

Gene Steuerle >> I'm saying, when the birthrate falls, the older workers have to start supplying more of the work, regardless of what's happening to their longevity and other issues affecting their supply.

Jagadeesh Gokhale >> Okay. It seems to me that at least a minimum requirement is that the projections be micro founded because the advantage of a microsimulation,

VIDEO CUTS OUT

looking at detail, granular analysis of how the demographic and economic forces, if you will, in the economy are moving forward. And these forces have a momentum that will continue. And that has implications for how productive

VIDEO RESUMES

we will be as a nation and how, eventually, the tax base and Social Security's finances will be affected.

That seems to be enforcing consistency at a granular level but also informing the assumptions that you make at the macro level. So making assumptions at the macro level seems to be more

arbitrary than building up to the level of productivity that we estimate, as an output of the model, of the process, rather than just an assumption imposed from outside. That's my view anyway. But let me throw it out to the audience. Any thoughts from anybody else who has thought about this or has anything to say, weigh in?

Audience >> It seems like the micro model and the macro derivation, those two things need to be consistent with each other, right? If you say I'm going to build up from a micro model and I'm going to get a new labor supply that may have differed from some macro assumption that I made, well, that's going to give me a new macro assumption. But then my macro feedbacks are going to tell me, oh, now I have different wage rates. And that should feedback into the micro system. It seems like, maybe in mathematical terms, there's a fixed point there somewhere in the middle. And that's where you want to be. Now, does that mean you want to iteratively run this procedure a million times until things don't change? I don't know.

Julie Topoleski >> Our model is sort of a hybrid. it is a microsimulation model. But within the microsimulation model, there is a macro model. And it's a dynamic macro model. At this point, it takes into account the effect of increasing debt and deficits on labor supply and on crowding out. So on the capital stock. So growth is affected by that. So let me think about the best way to explain this. So think labor supply. So we have micro level labor force participation equations. So for anyone of a given age, there's a probability of participating in the labor force. We can add up that labor force and supply that goes into the macro model, which you would use level GDP. And that's the level of earnings and everything else.

So you can see what the micro models say about the level of earnings in the economy and you can pair that to what the macro model says about the level of earnings given that method of estimation. In a perfect world those are the same. In reality, they're never the same. So what we will do is we'll align the micro earnings so that they're consistent with the macro earnings because that model is built on a set of economic projects that we have all agreed on in the agency. So in the first ten years, they're consistent with the ten-year economic forecast. CBO's primary mandate is to do ten-year budget and economic projections, as well as ten-year cost estimates. Beyond that, we spend a lot of time thinking about what we -- our best guess of what might happen in the future. So consistent with those ten-year projections and then consistent with, often, long-term trends and what we think will happen over the long run. Because, of course, we don't have a crystal ball. And no answer is going to be perfect. But by aligning our micro results to what the macro model is saying, you end up with a micro model that is consistent with a set of underlying macro principles. Is that helpful, Steve?

Steve Robinson >>
VIDEO CUTS OUT

Julie, just one follow up. In terms of calibrating between the macro and the micro, if the macro aggregate wages, for example, are higher or lower, is there some fixed way by which you adjust the micro? Is there an add or subtract factor, either to the earnings, micro earnings, or the labor force?

VIDEO RESUMES

What mechanically makes those two align?

Julie Topoleski >> So mechanically, it's an alignment feature on earnings itself. So if we are 2% too high or too low, we'll adjust everyone's earnings by 2%. Now, reasonable people can disagree about what the exact way we should be doing that is. We've chosen to do it in a pretty straightforward way and just do an across the board adjustment of that amount. You might think that, if you were missing the top of the income distribution or the bottom, you could adjust it in different ways across the distribution, but we've chosen to keep it simple.

Rich Evans >> So in our macro model, what we do is we adjust something at the basic microeconomic level, which is just the disutility of labor. And we get people to work more or less just so that we match those targets. Right? So that's a calibration exercise. But I think any good macro model is always going to be based on microeconomics. That's my school of thought.

Jagadeesh Gokhale >> So let's contrast these two ways of arriving at our projections. One is we have a trustee's assumption based on historical data. The other is the microsimulation-based

VIDEO CUTS OUT

projection without informing the macro model and iterating, as was mentioned. Suppose we did that exercise. And maybe we don't have to do that entire exercise. Maybe there's a shortcut that we can employ. Maybe we can inform the macro model with inputs from microsimulation outputs, then run the macro model because even macro models have heterogeneity, parsimonious as it is. But if in some representative, some way of matching one to the other, we inform the macro model with the heterogeneity we are observing and then run the macro model and just get deltas out of that to inform or adjust our microsimulation-based models, that's a shortcut. The ideal thing would be what Kirk is suggesting, iterating the final fixed point. I understand that. Have you done this type of exercise? How much difference would you expect in a world where we expect deficits to grow even larger in the future?

VIDEO RESUMES

The microsimulation is not going to care about future deficits in a different environment. The macro model will inform about this issue. How different would the outcomes be under these two methods? If they're significantly different, then how much confidence can we have in the projections that we see and use all the time, of Social Security, of the budget, and so on and so

forth. That's a very key question I think we need to think about in great depth. Any response to that, about the difference in particular, if you have any?

VIDEO CUTS OUT

Rich Evans >> Nobody's done that, Jagadeesh. There's one group I know of in Europe that's done -- that had one iteration. They went from microsimulation to macro and then plugged the macro back into the micro assumptions.

Jagadeesh Gokhale >> And how were differences defined?

Rich Evans >> I don't remember, but they're significantly different. I think it feels safe to say that in every microsimulation and macro model, the difference is -- I think I feel pretty confident saying they're significant. And what does that mean in our confidence of the models? I have no idea. Somebody needs to do that study. Sounds like you're the guy.

VIDEO RESUMES

Steve Goss >> Jagadeesh, I think probably what we ought to be thinking about here is that, in a macro model, if, as you've described, you're going to take past patterns of individuals, past transitional probabilities, and apply them on the basis of race, ethnicity, age, gender, whatever, and you just let it rip into the future, you're assuming a certain kind consistency with historical. And then your results, going forward, will vary, perhaps, from the past, in aggregate, based on the changing distribution of your population by those characteristics. That's one way of doing it.

Now, another way of doing it is to understand that, historically, at a macro level, we've had these kinds of changes going on all the time anyway and, somehow, maybe the different characteristics of individuals by race, ethnicity, maybe even age, are not all that important. It's sort of like I think what Gene was perhaps describing before. If you have a shrinking labor supply and it tends to be older, maybe some of the older people are going to start doing some of the jobs younger people used to do. So it's really a balance between the supply and demand.

Just to assume that the characteristics of people of a particular type in the past will be replicated in the future, when we have changing distributions of our population, is probably not going to work out. On the other hand, assuming that the aggregates are going to track exactly as in the past is probably not going to work out either, at the same rates. So really, it's kind of working back and forth. And I think Julie explained it well. And I think all of us are doing that with micro and macro aspects of our model.

Jagadeesh Gokhale >> But it seems to me that you're agreeing that one or the other is not good enough. It's a combination of information exchanged from both methods, macro and micro, that

stands a chance, or at least a hope, of getting us a better projection that we have more confidence in. At least that is --

Steve Goss >> The one exception is -- well actually, talk about meteorologists trying to project the weather tomorrow. How're they doing? Usually not so great. So remember, the accuracy, if you want to look ex post at how your projections turned out, quite frankly, the model is not really the biggest part of it.

Frankly, the biggest part of how accurate your projections are going to turn out is what were your fundamental assumptions? What do you think the birthrates, the rate of improvement in mortality, disability incidence rates? All of these things -- what are you really assuming is going to happen in the future? Whether you're talking micro or macro, decisions have to be made about where you think those are going to be in the future, based on past experience and trends and your expectations of the future. That's really what's going to drive the accuracy of your projections, much more than your choice of model. Any of these models done well and reasonably are probably going to end up with fairly similar results.

Jagadeesh Gokhale >> I disagree.

Steve Goss >> If you use the same assumptions.

Julie Topoleski >> Can I add that I think one way of thinking about this is thinking about the macro model is telling you how much and the micro model is telling you who. So if you know that 100 65-year-old men are supposed to be working in a given year, you can use the macro model to inform which 100 65 year old men are going to work in any given year. That way, you've got a set of predictions that are consistent across the micro model that add up to your macro totals.

Gene Steuerle >> Julie, in CBO, don't you effectively get that? I'm not saying it's right, because nobody's willing to project a crash when it's going to occur. But don't you effectively get there by predicting long term -- deficits create some impact on interest rates, on earnings growth rates over time so that it's sort of worked in and then that feeds back into the assumptions of the model. That's the only way to do it, but don't you sort of indirectly get through that type of assumption?

VIDEO CUTS OUT

Julie Topoleski >> So two comments. One, we don't project recessions in the long run. But we do project that our sort of aggregate long-run growth is about .5% below potential because we think that, sort of on average, is where we've been over history. We do take into account the

effect of rising debt and deficits on economic growth, which affects both interest rates and, of course, GDP growth and earning growth.

VIDEO RESUMES

Jagadeesh Gokhale >> Another question from online viewers: A very simple one: Is MINT used for Medicare policy decisions?

Mark Sarney >> No. [laughter]

Gene Steuerle >> Can I disagree slightly? I don't think it's good for deciding how you do health policy. But I think if you're asking the distribution of benefits in society, say between elderly and non-elderly, it wouldn't be bad to throw in -- it might be crude, some estimate of Medicare benefits in that. For instance, one thing I would love to see in a good reform package would be if we combined Medicare and Social Security reform. One could think about ways that the Medicare cost savings, some of that could be shifted over to a cash benefit for Social Security. And I don't know how to show that other than in a microsimulation model, about how you might do it. Even though the Medicare estimates would be -- you would be taking them from the outside, for the most part.

Mark Sarney >> Sorry, I was just answering the factual question, do we use MINT for Medicare policy decisions?

[laughter] We don't. But I think doing that kind of analysis would require us to climb a learning curve about Medicare rules that would take some effort. We would have to become conversant in the rules about all the detailed Medicare rules like **QMB** and **SLMB**. And there would be a lot of different aspects there. It is definitely doable. It can definitely be put on the list of things to do. And just to give you a little bit of background about how we would expand capacity on anything we do, including something like Medicare, is, first, we would do it really badly and very simplistically. And maybe just to get at those sort of more holistic, comprehensive estimates of, say that all the Social Security and Medicare benefits that someone is receiving over a lifetime or something like that then, eventually, we would iterate toward something more program-specific. And I know like value of health insurance benefits and how out-of-pocket expenses, all of that, you know -- there's putting a toe into the Medicare world and then there's jumping in with both feet, you know. And that could be something that could be done over time that's kind of like jumping in a deep end of a different pool in some sense. But right now, we don't do it.

Steve Goss >> Mark, could I just add -- I mean, for policies that come up all the time from within the administration, whether you're regulation or laws that would have implications, not over Social Security and Medicare, so what we generally do is we take the lead on projecting

what the implications would be for Social Security. I mean, for instance, if something in the disability -- if there are going to be fewer people receiving disability benefits, therefore 29 months hence, there will be fewer people on Medicare, we simply indicate to the CMS actuaries on the other side of the beltway the implications that we're seeing for the change in the numbers of beneficiaries that would reach that 29th month of Medicare eligibility for disability. Or if we have some kind of change that will have differential death rates or birthrates, that will have different size populations, we just share that with the Medicare folks, and then they patch on the Medicare cost.

Jagadeesh Gokhale >> Another issue that I think came up when we had our pre-conference calls for setting up this event is about where do we go from here -- after this conference is over? What does SSA take away, in particular given the constraints and capacity for developing and putting all of the modernizations and changes. What would be the right approach for having SSA collaborate with outsiders, use advisory services from outsiders? What are good practices in that area? Any thoughts would, I think, help the SSA folks out considerably. What does the CBO do, for example?

Julie Topoleski >> In terms of getting outside feedback? So we do a few things. We often, we send many, many things out for external review. We turn to experts in the field to get feedback on different model developments or, I'll refer to a paper that my colleague Marina came out with earlier this year, in replacement rates. As we were developing that we sent it out to I think five different people, some of whom are in this room, for feedback on that work. We take the feedback from outside experts very seriously and try to take the recommendations on board.

While this doesn't directly affect the long-term model, CBO does have a couple of panels of different types of experts. We have a panel of economic advisors who advise primarily on macroeconomic issues. They come in twice a year to CBO and the folks in our macro division work with those people very closely. We also have a panel of health advisors who are in once a year to talk, to advise our health folks on things in the health world, which for the most part I stay out of.

We've also convened this year a special technical panel to advise our health insurance simulation team. They came out with a new model **HISIM 2** that was used for this most recent spring baseline. And they have been working closely with that panel, to get feedback on sort of very specific and very technical issues. So we really do try to turn to outside advisors. We also generally are always looking for feedback, whether it is from our colleagues on the Hill or people that just sort of want to talk to us about their work, or come in and present their work. So there is lots of collaboration with experts in the field.

Jagadeesh Gokhale >> Any other comments?

Steve Goss >> I had one that patching on to Mark Warshawsky's comment earlier. Demonstrations, we love demonstrations. We do them occasionally in Social Security. Sometimes you have a natural experiment when a policy changes and being able to look at what the effects are before and after a natural experiment or well-designed demonstration can tell you a lot about what behavioral aspects might be, what implications would be, and we have benefited greatly from many of these over the years. So Mark, if we have the money, we can do more demonstrations, that'll be great.

Mark Warshawsky >> Well, a little editorial on that. Part of our authority as you know is going to be expired in a couple of years, so I don't know where the congressional staff are.

Jagadeesh Gokhale >> Any issues that people in the audience or panelists want to raise regarding what was heard?

Steve Goss >> Could I just mention one other little item? In our experience, we have a number of micro and macro models of our own, but we also work collaboratively not only with MINT and others, but I think it has been mentioned through the Office of Tax Policy, over at Office of Tax Analysis Treasury, and also the Joint Committee on Taxation. We don't have as good an access to detailed pervasive records for income tax returns as we would like, but they do. So working collaboratively with them, we can get tremendous information back from them. So again, you know, I don't want to say a patch work quilt, but any model that tries to be everything unto itself is probably not going to do well. But having feed in from many different models turns out to work out really well.

Lih Young >> My name is Lih Young and thank you for your presentation and your effort in this area, but the whole day, it doesn't seem you expressed objectively what SSA means for a benefit to those people who are supposed to be beneficiary. There are other people who are beneficiary, but they don't receive the benefit. So if you use micro or you use macro, it doesn't make sense if they don't spend anything, instead there is a distortion of our society. Who is benefiting from the money that they are receiving from the government? The money is toward, distortion toward, somebody who is not supposed to be there. That employment, has two meanings. One is, you do some service to benefit a society. The other one, unemployment is disservice. And judging by the roles of fraudulent or corruption, and so if you give benefit to them, if you give more money, if you think that SSA is helping the elderly, it is not. Elderly, they can spread, and their assets can be robbed by those people. And these people can be dead, but they got to be dead before they receive the service but they don't receive any benefit.

Jagadeesh Gokhale >> Thank you very much.

Lih Young>> When they dead, they don't get anything either. They don't get anything either.

Jagadeesh Gokhale >> I appreciate your comments.

Lih Young >> Thank you, sir. I hope you change your direction.

Jagadeesh Gokhale >> Thank you very much.

Lih Young >> Thank you.

Jagadeesh Gokhale >> Any other issues? I have one more issue to raise which I will unless someone else has something else they want to say.

Kim Hildred >> I had one question to ask the panel. In terms of your experience regarding validation and kind of the merits of doing validation in-house versus through outside contractors. I'm curious just to know of your experiences in this area, if you have any thoughts or suggestions along those lines.

Richard Evans >> This is something we don't do enough of and we were having a discussion. I forget who with. But just we spent so much time developing the next kind of advance in our models that we don't do this enough. We do, do things like calibrations and making sure we hit some targets, but validation you are talking about, there were some really good points about validation that we sought. But that is something that I know on our part we don't do enough of. CBO does a little bit of that. It seems like one of the few places that does it really well. But I am sure we could do more.

Julie Topoleski >> I'd actually argue that validation is something that MINT does well, and actually the other Urban models partly, because they have spent an amount of effort to see how their models project sort in a historical period, so you have actual data to compare it to. I think given that we don't have sort of the counter factual in the future to compare it to, I think is the next thing. So I would say that's something that, that the MINT model does a nice job with.

Jagadeesh Gokhale >> I have a question about whether model validation could be a tool to convince clients, users of MINT output, that the model really knows what it is talking about. In my experience, information about how the model was validated and the metrics used for validating are not actually laid out in any systematic way for people to use, confirm, ask questions about and so on, on your modelers' websites.

You could actually just look at distributions of all kinds of attributes in the population and compare them. Charts are a great way to communicate what your model is producing over time.

But I don't see those types of displays. I mean, these things would be very useful to getting users to believe in your model's validity. You may be validating it, you know, squeaky clean, but because when I look at it and it is a black box, you need to tell me what exactly you did. Any responses to that?

Karen Smith >> As a model developer, I spend about five percent of my time making the model and 95% validating. One of the challenges with validation is what you validate it against. For example, the CPS is a great data set, but it doesn't match what the Social Security trustees say is the answer. For example, if you look at the earnings distribution from the CPS, and you look at the earnings distribution that is published on the SSA website, they are not the same. The median earnings of workers on the CPS is different than median earnings of workers in the SSA data. If I go back in time, I have a mortality problem. I only have the people who survived the SIPP interview. Looking 20 years back is not informative, because many of the people 20 years ago aren't alive today. I do a lot of validation.

Really, Julie is exactly right. We have the historic data, we can see, confirm that, yes, over the historic period, the model is getting the right answer. It is getting the right relative differences. Given that historic period, I have a lot more confidence in the future. I have to say that one of the nicest things is how much information Social Security puts on its website that I can use for validation. It is tremendously helpful and Steve's group has been wonderful at providing detailed distributional numbers that we use in DYNASIM and in MINT.

Steve Goss>> If I can add one thing. I hope I get this about right, and Karen or somebody correct me if not. But on the POLISIM microsimulation model we have been running, I think it starts making its projections something like 20 years in the past. And that allows an opportunity to see what the model is doing over an historical period where we really can validate it against what was reality. Now, that is all well and good. That is a wonderful thing to be able to do that. But, of course, will the next 20 years follow the same kind of patterns as the last 20 years?

And this, of course, is the challenge for any model and every model always. You can validate against the historical trends, but the thing you are always left scratching your head about and to figure out and having to think very hard about, is how might the two trends be different at both a micro and a macro level? And that's really the challenge where I hate to use the word judgement and ultimately it just has to be there.

Julie Topoleski >> One thing that we found enormously useful, and I think Steve would agree, is the exercise once every few years of taking their assumptions and putting them into our model and to see what our results will be. And we hand them the same data in their models and we always learn a lot. The short answer is, the models produce more similar results, but still not the same results. But we do explain two-thirds to three-quarters of the difference when we do that.

And through the process of doing that, we learned a lot about how CBOLT works, about how the actuary's models work. That is another way you can think about validating your models, comparing it with others.

Jagadeesh Gokhale >> So it seems to me, based on those comments that we are again in this unclear gray area between what is a projection and what is a forecast. If you're using judgments about how the next 20 years will turn out, that is most like you are trying to make a forecast rather than a conditional projection based on the assumptions of the model. And so I'm not sure what more I can say.

Steve Goss >> Well, I suggest that you would call it a forecast if you don't disclose what the assumptions are and say, 'this is my forecast.' If on the other hand, you say, 'hey, the projections we are making under these assumptions'--but you also have sensitivity analysis, which we have, I know that CBOLT has this also--you say, 'what if we varied some of the assumptions, what else would come out?' Have different scenarios, have stochastic projections about the possible things that can come out. Not just a singular set of assumptions with a failure to disclose what they are. But really to lay out exactly what your assumptions are, and then show some variations.

So if somebody comes along and says, 'I don't believe that fertility rate assumption, I think it is going to be a lot higher or a lot lower,' then we can show a scenario. And we have it readily available that will show what that result is. So you could call them a bunch of forecasts under different assumptions, but I think it is probably more appropriate to say they are a bunch of projections under well disclosed and highly identified sets of assumptions.

Rocky Feuer >> I just want to comment briefly about, I don't know if I'll speak for myself a little bit, what we do a little bit at the National Cancer Institute with respect to projection. So if we are projecting mortality, if we are projecting for three, four years, we might use just what I call a cold mathematical model. Well it is a time series model, maybe not a straight line. If we are going back, and so suppose I'm doing lung cancer mortality, if I'm going back a little further, then I want to think not about lung cancer, but I want to think about the antecedents that are causing lung cancer. So I might do the projections of smoking. But if I want to go a little further, I go further back in the antecedents and I think about the policy changes that might occur: taxation, raising the minimum age of smoking and think about those..

Because I think, and of course all of these projections are conditional on the assumptions. But the further I want to go, the further I want to go back to the antecedents, because those seem like, you could make maybe a more reasonable set of assumptions the further back you go in the antecedents of what eventually--so policies might cause smoking and smoking causes lung cancer. I think that makes a lot of sense. And you could make it is easier to make guesses about the further you go back in the antecedents to some extent.

Jagadeesh Gokhale >> Anything else at all? Because otherwise I'm going to close the session.

Mark Warshawsky >> Yeah. I have one very discrete question. And it is really in terms of when we are thinking about improving the model, you know, where should this go, if at all, in our priorities? Should SSA improve the model's behavioral responses? As we heard, they are really quite limited. I think I heard none. And really the question is, is that an area where we think it would be helpful, where you think it would be helpful? For example, claiming decisions in response to a change in the full retirement age. Question is that, would that be helpful? Would it be believable? Would it get us into a lot of controversy? I'm curious what CBO does. If the panel could comment on that.

Julie Topoleski >> So I would say, well, so let me talk a little bit about what has happened and what we have observed in data. So we know as the full retirement age has ticked up, that there's been a remarkable response in claiming at that full retirement age. If you look, the charts are actually sort of unbelievable. You see this little thing just move out two months a year. It's really remarkable. And so that's the type of thing that we at CBO try to take into account. Now, getting claiming behavior exactly right is really hard. And so we do a lot of sort of mechanical things to massage patterns in ways we think makes sense, that's sort of where we apply judgment. Because we don't have a great model to do it. But those types of things do have an impact and then we have a downstream impact. So if you are delaying in claiming benefits, then you are more likely to be in the labor force at that age. And that can have an effect on output and so I would argue that those things are important to take into consideration.

Steve Goss >> I'd agree they are important. I think they are really tractable. I know back at the time of the 1983 amendments when we were tracking, starting in year 2000 the normal retirement age was going to be going up. We had to make an assumption about what was going to be the labor response, and what was going to be the response of delayed start of receipt of benefits. We made some assumptions on that. Joyce Manchester and I think Jay Sung did some work ex post at a later point, and it turned out we had a lucky guess on that. We assumed some fraction of the ages would increase by the average age, but really an assumption you are never going to be guaranteed an assumption is going to be right. But one thing that we always strive for is to try to have no known bias. It's pretty clear that if you raise the normal retirement age and offer a lower dollar amount of benefit at any given age of retirement, probably people tend to wait a little bit longer to come on the rolls. So put something in there, we were lucky on guessing that.

Another aspect that we do and I'm sure CBO does the same thing is if we assume the tax rates are going to change, we assume that it is not going to be just transparent and what people's earnings are going to be. If we raise tax rates on a segment, we assume there will be a response.

And say there's some others have done work in this that we sort of we make use of, that there will be some response at a minimum of shifting some of your compensation out of earnings into other compensation in response to the increased marginal tax rate. There are things like this that can be done, and Mark and Mark, we'll be happy to work with you on this.

Gene Steuerle >> My recommendation there would be particularly on the one you are mentioning, the labor force participation. I think there would be a lot, if you came up with one estimate of behavior you took out of some study somewhere, a lot of controversy over the one you picked. I think the better thing to do would be to assume a lot of different responses.

As I say, Steve and I can disagree. We don't have time here. We way underestimated the labor force participation response for reasons that have little to do with things like even what is happening on the labor supply curve or the labor demand curve. But we don't have to agree on that. So what I suggest is you do want to put in some alternative behaviors. But I think what would be really helpful are what the types of -- you don't have to favor one--but what are the types of responses you might want to do for that behavior if you think that it is causing some increase in poverty or a lack of support for some groups?

The real question there is: if we want to adjust full retirement age, so for instance, we give people now 13 more years of retirement than we did in 1940, and that is becoming very expensive in a system where the birth rate has fallen and we don't have a lot of workers, I think that inevitably we need, you don't have to agree with this, that inevitably we need some response in terms of on the employment front to maintain this system.

But that is highly controversial, because if you think you're just going to retire at 62 to 65, or say age, average of 64, that's going to have a huge effect on things like poverty and other things. I would rather you sort of assume alternative behavior. Picking this behavior by the way is herd mentality and you are not going to get off the typical econometric or typical study because -- if I have 30 percent, all my New York friends are going to try and go to Florida, I still may do it. It is still going to show up particularly in a cross section estimate anyway. So give me some way to think about it. Don't worry so much about whether you have got the right behavioral response because you could feed in some simple behavior responses. But now tell me the implications of the alternatives and give me ways to think about it.

Bryan Tysinger >> I just want to say a couple of things. So one, you picked a good, an easy example, because you do have the normal retirement age that has changed over time and you can estimate that. If you have a harder one, like changing the early entitlement age, then you're going to get in an argument with all sorts of economists that have a similar model that has just one thing, and someone has one, and says some another thing. And so getting to the cancer discussion earlier, where they had different models and they had sort of, how did they put it, the

credibility of the amount of extrapolation required. I think that's a really useful thing to put on the projection you make. If you can say I'm more confident in this because I was able to estimate using real data and this one required a modeling assumption. Thanks.

Jagadeesh Gokhale >> Well, this has been a very interesting discussion. Not everybody agrees on everything but that's par for the course for these types of discussions. But I would like again for us to give the panelists a big applause. Thank you.

Kim Hildred >> I just have a couple of closing remarks that I just wanted to share. During his opening, Mark Warshawsky expressed Social Security's interest in pursuing three questions today. Are there better ways for Social Security to present MINT distributional results to requestors and the public? There was good feedback on that. Are there minor tweaks or fundamental changes needed for MINT modeling and how Social Security employs it? There was also good feedback on that as well. How should Social Security adapt MINT so that it is relevant and valuable for the next 20 years? We certainly discussed the importance of MINT today. We also learned about the resource challenges that Mark and his staff face as they work with MINT on a day-to-day basis, never mind taking on the task of updating MINT so that it remains relevant and valuable going forward.

So Godspeed, Deputy Commissioner Warshawsky. But I believe there is very good news in the fact that we have heard from many of our experts today who were truly supportive of MINT, who have made great suggestions. You have made a number of contacts in terms of people who are willing to help you move forward. And I think that is terrific, so thank you. And thank you all. I do want to thank our presenters, our moderators, and our audience for your hard work in answering these important questions as well as others. And we look forward to hearing how this feedback will be used to determine MINT's future. I also want to thank the audience and all of you too for completing a survey you will soon receive from the board to continuously improve future events.

I want to thank those who made the success of this event possible. First, I want to thank Mark and his staff. We worked in true partnership on all aspects of this event and I am deeply grateful for their collaboration with us. Also, our staff was supported by terrific communications and other experts at Social Security. And please forgive me if I don't get your names right, including Dawn Bystry, BJ Jarrett, Dennis Stehlar, Dash Adamson, Richard Alves, Jennifer Burghardt, along with Ty Brown our photographer. These folks are doing all the audio. They are not from the building, they are from the Social Security Administration. We are truly grateful for their technical support. I want to thank our Board members for their feedback along the way for serving as moderators, and especially to Jagadeesh for being so generous with his time and expertise during the planning for this event. And I also want to thank our staff director, Claire Green and our entire Board staff, for their hard work promoting the event, supporting our

presenters, and taking care of the logistical details. And lastly, thank you to the staff wherever you may be of the National Union Building for your help with the set-up details. Enjoy your evening.

MINT Forum Panelist Recommendations Summary

The following information is a list of recommendations from the MINT Forum transcript. These recommendations also reflect the reoccurring themes and topics that were discussed. For example, certain speakers reiterated recommendations in numerous ways, such as stating the same idea at varying levels of specificity, and multiple speakers shared or repeated similar recommendations. This was particularly true for the topic of confidentiality and access. Numerous panelists recommended that SSA should both develop protocols for access and confidentiality as well as recommended that policy deliberation should be kept truly confidential. Also, some items listed are not explicit recommendations but rather are presented as suggestions.

MINT Forum Panelist Recommendations

- MINT users should make the results more understandable to different audiences/clients.
 - Focus on medians rather than means.
 - Focus on the population rather than complicated measures.
- SSA should develop protocols for who has access and confidentiality.
- Keep the policy deliberation process truly confidential. The Office of Legislative Affairs should not be involved in requests of results before a proposal becomes public.
- MINT could be improved to make it more user friendly.
 - When MINT Team receives a legislative bill to review, add a “top line summary” about the bill saying the bill did x, y and z, in plain English. Or if the MINT team noticed something interesting in the output, the MINT team should summarize that up front.
- SSA should produce a consumer user guide to their standard set of tables so that people can begin to parse out for themselves what the meanings really are.

- The beneficiary status table, for example, it may not be clear to some if people are moving between categories because they have gone from being a dual beneficiary to being just a worker beneficiary. Or if there are actually genuine increases or decreases in benefits occurring. So having something that would, for each of the key tables, provide a translator for what these results could or could not mean would be very helpful.
 - The provision-by-provision breakouts are particularly helpful. A lot has been shared about policy is used to target certain populations. While the overall tables are very helpful for the full model, full plan, sharing the provision-by-provision tables in their complexity would be very, very helpful.
 - Also, the new present value tables that are available both on the benefit side and on the tax side have been particularly illuminating at showing the burden of both tax increases and benefit increases and losses, both by cohort and particularly by variations in AIME quintile.
- It's also useful -- this may be something for the MINT team to think about -- of not just churning out numbers, but also playing that hand-holding role of talking to the consumer saying, 'What is it you're trying to do here?' Not just your provisions, but what are your goals here in English? Then go through the output and say, 'Do we think you've met your goals? How can we illustrate it to you?' Just handing people numbers without -- I know SSA doesn't like to do commentary because it's a very risk-averse kind of agency. But you need to work with your client and talk to them about it. So it's give them the data dump, but then also customize based on the project.
 - Maybe putting some things more into words, not just into numbers, would be helpful -- the actuaries' memos, they're not maybe the easiest things to read, but they do have words as well as numbers. There is a narrative that explains what's going on that is neutral.
 - There are not enough people who are able to fulfill that really important role to teach people about this program and to walk them through the results. SSA should be investing more heavily in research, and for them to be able to do that, appropriators need to invest more heavily in SSA.
 - Increase MINT access to policymakers, researchers and the public.
 - Publish MINT results on SSA's website exactly the way that the Office of the Actuary publishes their analyses on their website. Feature every single public plan, whether it's a bill, or a commission proposal, or something from a think tank on the website, breaking out every provision within those plans.
 - MINT is a crucial tool for looking at Social Security reform

- MINT can play into elements of the Social Security 2100 Act. The actuarial analysis of that plan projects very large benefit increases for low-earning workers. For the very lower stylized earners, it's a roughly 40% benefit increase. For the low earner who makes about half the average wages it's a 23% benefit increase. Many of those very low earners are getting SSI. If you increase your Social Security benefit, their SSI benefit reduces dollar for dollar. Many of them are receiving auxiliary benefits. If you increase their retired worker benefit, their auxiliary benefit declines by a dollar. So they only receive a benefit increase if their higher earning spouse receives one.
- Summary information would be wonderful if it was at all possible to do it in a way that would actually satisfy the key players.
- Increase access and the timeliness of MINT results.
- Building capacity to look at interactions with other means tested programs would be helpful.
 - For instance, currently, if Social Security benefits are increased under a proposal, there's nothing in the calculator right now as it's run at SSA that would account for a decline in TANF benefits with an increase in Social Security.
- Incorporate additional behavioral response for claiming behavior and work.
 - A lot of goals of legislation are to encourage or to discourage individual behavior. It would be really helpful if MINT were able to incorporate behavioral feedback - how a policy change might lead to a behavioral change that would actually affect the outcome of the distributional results.
- It is important to take claiming behavior, such as the full retirement age up tick, into consideration.
- Replicate the taxation of Social Security benefits.
 - Add in the ability to replicate the taxation of Social Security benefits. The inclusion of that is critical to a true distributional outcome of a full policy proposal. You might increase benefits, but if that same individual on the backside has some of those benefits reduced through the federal income tax on Social Security benefits, the net outcome isn't necessarily what you are going to be seeing through the MINT results.
- It would be useful if MINT could model more interactions on the disability side.
 - It's important for SSA not to just think about their programs but to think about the overall priorities of the administration.
- There should be a lot more validation in the modeling (following the example of CBO).

- Think beyond just SSA's programs and consider overall retirement security, including pension policies, IRAs and tax issues.