

Office of the Assistant Secretary
for Planning and Evaluation

MEETING ON DEMYSTIFYING MICROSIMULATION

November 16, 2012 | Washington, DC

DEMYSTIFYING MICROSIMULATION
MEETING REPORT

OVERVIEW

MEETING ON “DEMYSTIFYING MICROSIMULATION”

*Office of the Assistant Secretary for Planning and Evaluation
U.S. Department of Health and Human Services
November 16, 2012*

On Friday, November 16, 2012, the Office of the Assistant Secretary for Planning and Evaluation (ASPE) of the U.S. Department of Health and Human Services hosted a one-day meeting at the Department titled “Demystifying Microsimulation.” The meeting consisted of a panel of experts who discussed what information consumers of microsimulation models need from modelers in order for consumers to understand and evaluate model results. The primary focus of the meeting was on educating consumers of models and helping them to understand model output and validity. Participants from within and outside the modeling community attended the meeting in person or via Webinar.

The results of that meeting are contained within this document which contains the meeting agenda, a list of panel members, a list of preparatory reading materials sent to participants, and a summary of the proceedings. Also included are the PowerPoint presentations made at the meeting.

Following the meeting, ASPE commissioned four expert papers for which the authors, who had served as moderators of sessions during the meeting, were asked to summarize the results of their sessions. Those papers are contained herein.

A complete transcript of the day’s proceedings is available on request from Dr. Joan Turek of ASPE at joan.turek@hhs.gov

This summary was prepared by BETAH Associates, Inc. under task order contract (HHSP2332012005579G) to the Office of the Assistant Secretary for Planning and Evaluation (ASPE), U.S. Department of Health and Human Services. The views expressed herein are those of the authors and do not necessarily reflect the views of the U.S. Department of Health and Human Services or ASPE. Special thanks are due to Joan Turek, Ph.D., the ASPE Project Officer, Joy Nathan, the BETAH Project Manager, and Al Staropoli, Writer/Editor, BETAH Associates, Inc.

CONTENTS



Meeting Summary	4
Expert Panel Session Summaries	35
Citro, Constance (When to Use Models in Health and Human Services)	35
Abraham, Jean Marie (How to Choose a Model and Modeling Strategy)	43
Betson, David (Assessing and Interpreting Model Results)	50
Giannarelli, Linda (Communicating Results)	57
Meeting Agenda	66
Panel Member List	67
Panel Member Biographical Sketches	68
Background Reading Materials	74
PowerPoint Presentations	76
When to Use Models in Health and Human Services	76
How to Choose a Model and Modeling Strategy	90
Assessing and Interpreting Model Results	105
Communicating Results	115
Best Practices in Modeling	130

DEMYSTIFYING MICROSIMULATION

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
OFFICE OF THE ASSISTANT SECRETARY FOR PLANNING AND EVALUATION (ASPE)
HUBERT HUMPHREY BUILDING, ROOM 505 A
WASHINGTON, D.C.

November 16, 2012

MEETING SUMMARY

PRESENT

EXPERT PANEL MEMBERS

Jean Abraham, Ph.D., University of Minnesota
Jessica Banthin, Ph.D., Congressional Budget Office
David Betson, Ph.D., University of Notre Dame
Sharmila Choudhury, Congressional Research Service, Library of Congress
Constance Citro, Ph.D., The National Academies
Steven Cohen, Ph.D., Agency for Healthcare Research and Quality/HHS
Michael Collins, U.S. Government Accountability Office
Linda Giannarelli, The Urban Institute
Robert Gillette, U.S. Department of the Treasury
Sherry Glied, Ph.D., Columbia University
Howard Iams, Ph.D., Demographer, Bethesda, MD
Susan Jekielek, Ph.D., Office of Planning, Research & Evaluation/ACF/HHS
Richard Kronick, Ph.D., ASPE/HHS
Julie Lee, Ph.D., Medicare Payment Advisory Commission
Charles Nelson, U.S. Census Bureau
Michael O'Grady, Ph.D., West Health Policy Center
Kakoli Roy, Ph.D., CDC/HHS
Arloc Sherman, Center on Budget and Policy Priorities

OTHER ATTENDEES

Megan Campbell, Office of Child Care/ACF/HHS
Yolonda Campbell, J.D., HRSA/HHS
Rose Chu, OS/HHS
Donald Cox, Ph.D., ASPE/HHS
Madeleine De Boinville, HHS
Nancy De Lew, OS/ASPE/HHS
John Drabek, Ph.D., ASPE/HHS
Kenneth Finegold, Ph.D., ASPE/HHS
Joseph Gagnier, ACF/HHS
John Haaga, Ph.D., National Institute on Aging/NIH/HHS
Kevin Haninger, Ph.D., ASPE/HHS
Susan Huan, ASPE/HHS
Jennifer King, ONC/HHS

Alana Landey, ASPE/HHS
Ben Lum, CCIIO/CMS/HHS
William Marton, ASPE/HHS
Jim Mays, Actuarial Research Corporation
Erica Meade, ASPE/HHS
Jonathan Moore, ASPE/HHS
Donald Moulds, Ph.D., ASPE/HHS
Thomas Musco, OS/ASPE/HHS
Jennifer Nooney, Ph.D., HRSA/HHS
Donald Oellerich, ASPE/HHS
John Phillips, Ph.D., National Institute on Aging/NIH/HHS
Rosa Po, ASPE/HHS
Susan Queen, Ph.D., ASPE/HHS
James Scanlon, ASPE/HHS
Fritz Scheuren, Ph.D., NORC at the University of Chicago
Jeffrey Sims, ASPE/HHS
Laura Skopec, ASPE/HHS
Andrew Sommers, Ph.D., ASPE/HHS
Joan Turek, Ph.D., ASPE/HHS
Cathleen Walsh, Dr.PH., CDC/HHS

WELCOME AND INTRODUCTION

Jim Scanlon, ASPE, HHS

Mr. Scanlon called the meeting to order at 8:41 a.m. and welcomed all participants.

He said the meeting would assist the Department of Health and Human Services (HHS) policy community – as well as the modeling community in general – to think through various factors, criteria, and information used by consumers of models. Questions to be addressed by the meeting include: How does one choose a model? When should one consider modeling for policy analysis? and How does one interpret and communicate modeling results? The meeting will conclude with a summary of lessons learned and best practices. Participants went around the room and introduced themselves. Mr. Scanlon introduced Dr. Sherry Glied who spoke next.

OVERVIEW

Sherry Glied, Ph.D., Columbia University

Dr. Glied said that HHS uses a variety of models. For example, the Centers for Disease Control and Prevention (CDC) uses models for diabetes and the Food and Drug Administration (FDA) uses models to determine the effects of policy on tobacco use.

She added that research is not always useful for policy because research is about something that has already happened while policy is about something that will happen in the future. The results of research can be informative, but one usually has to draw inferences to determine what will happen in the future. Another way in which research and modeling vary is that research often examines one thing at a time while policy analysis usually involves looking at multiple effects on different agents. The latter generally requires a set of parameters and assumptions.

One also needs to consider that modeling policy in the federal government involves use of very large samples. This can require building complex models – such as microsimulation type models or other complex models – with multiple agents.

One of the questions to be addressed today is how staff can become more educated consumers of models. In other words, how can they be assisted in selecting a model, learning what a model can and can't do, and better understand modeling results? Consumers should also be able to better understand uncertainty.

DISCUSSION

- Steve Cohen said that one can simulate the host database by statistical matching techniques and augmentation. He asked if this could be added to the scope of the participants' discussion.
- Dr. Glied agreed and added that, from the Office of the Assistant Secretary for Planning and Evaluation's (ASPE) perspective, it might be useful to know where ASPE and HHS can make investments that would make modeling better into the future – not only how one can make

models better but also what kinds of research and data collection would make the models better underneath it.

- Jessica Banthin said that academic models at times can be well documented but private models also exist that are less well documented. Sometimes one can examine the model and the data on which it's based and determine that the underlying data set is perhaps biased or incomplete.
- Michael O'Grady said that policymakers sometimes perceive modeling through a "trust factor." Therefore, it's important that modelers not be viewed as partisan or having an agenda. Trust also comes from modelers being competent and rigorous. On the Hill, trusted sources include the Government Accountability Office (GAO), Congressional Research Service (CRS), Congressional Budget Office (CBO), and MedPAC.

WHEN TO USE MODELS IN HEALTH AND HUMAN SERVICES

Connie Citro, Ph.D., The National Academies

Dr. Citro remarked that policymakers generally ask questions that are future-oriented such as:

- How much will a new policy cost in 1, 10, 75 years?
- Which geographic areas, demographic groups, and/or organizational players will benefit (or not benefit) and by how much?
- What will be the effects of a policy change on measures of health outcomes?

To answer these questions, analysts provide estimates that are sometimes generated by using models. A model can be defined as a "mathematical framework representing some aspects of reality at a sufficient level of detail to inform a clinical or policy decision." Models are communication tools that allow the complexity of a given system to be reduced to its essential elements.

Models follow a continuum from the simple to the highly complex. They can be both transparent and opaque, although good documentation can reduce the "black boxness" of a model. However, neither complexity nor transparency necessarily matches up with a model's usefulness. In other words, a simple model can omit important factors while a complex model can be needlessly complex.

Models are generally classified as either formal or ad hoc. Ad hoc models are usually developed on the fly. They include "back-of-the-envelope" (or spreadsheet) models as well as extrapolation, regression, and cell-based models. Formal models are maintained (at least for some time) and include extrapolation, regression, cell-based, static microsimulation, dynamic microsimulation, and computable general equilibrium models.

Microsimulation models use samples of individual records for people, families, and organizations. They can either be static or dynamic. Static microsimulation models project baseline sample forward for short periods by reweighting. Dynamic microsimulation projects baseline sample forward by dynamic aging (e.g., people aged 50 become 60 in year $t+10$). Microsimulations can also be grouped into arithmetic (counting) models, behavioral models, or a combination of both. In general, all but the simplest models have components.

Dr. Citro offered three questions for discussion:

- What criteria should one list for deciding when to use models – redefining the question as when to use formal, more complex models?
- What criteria should one list for deciding on when to use microsimulation instead of other formal modeling techniques?
- What best practices can/should modelers follow to document models, evaluate models, and help users understand their differences – that is, to reduce the “black box”?

DISCUSSION

- Dr. Cohen said that one of the most difficult things to convey is the underlying uncertainty in the estimates both in terms of the host data sources, the synthetic data that’s manipulated to host the data source, the underlying modeling error that’s built in, etc. A best practice could be to examine what each model does under a modest assumption, an average assumption, and an unexpected assumption. Having a feedback loop is also essential after the modeler moves forward and makes a policy estimate. For example, RAND’s COMPARE model (<http://www.rand.org/health/projects/compare.html>) will use three different data sources to make projections of how many people will likely be uninsured by 2014. By 2014 one could look back and see what happened, what has changed, and how the model could do better.
- Dr. Citro said that hopefully leadership from ASPE and other agencies can help regularize that kind of feedback. The key is that there be some mechanism, funding, or directive for modelers to cycle back as time passes. She added that it would be helpful to have new visualizing techniques to better convey uncertainty.
- Dr. O’Grady said that part of the strategy is looking forward. There are epidemiological trends that will peak, in a policy sense, in the not-too-distant future. The questions are: Do we have the right data to build the right models? What are demographics driving? What are costs driving? Are the right models in place? When a bill is going to the floor, there is no time to make a lot of mistakes, so one needs to look forward. This is a contribution that ASPE can make because it has a foot in both camps – both the policy and the analytic camps.
- Howard Iams said he wanted to discuss formal vs. ad hoc modeling. He said he could see no circumstance when the ad hoc model is worthwhile unless one is desperate for time. The formal model forces the analyst to lay out the relationships that are involved, the feedback loops, and covariance – while ad hoc models just don’t pay any attention to what the underlying structure is. The advantage of a formal model is that it forces the analyst to decide what relationships exist. One can then start arguing about how to estimate the relationships and how good are the underlying data. So the real questions are how extensive the formal model is and how many aspects of the situation it deals with.
- Linda Giannarelli agreed with Dr. Iams. She said that with an ad hoc model, it is very easy to forget to include some of the inter-relationships that exist, or to think that those relationships are not

important enough to include. In a formal model, once those relationships are built into the model, they are always included. The model never “forgets” that relationship even if an analyst might.

- Julie Lee said that, in her experience, the process of going from an ad hoc to a formal model happens organically. Usually there is a policy question and an ad hoc model is built. Questions related to the original question then come back and one discovers that the ad hoc calculation doesn't provide the needed granularity or nuances to answer these questions, so one starts adding things. At some point one discovers that the things that were laid on top of each other don't always make sense and a more systematic approach is needed. So the right tool evolves based on the questions that come up and the demand for certain kinds of analyses.
- Dr. Cohen said that one should go through both processes. Sometimes one has to actually show what the estimate would have been under the “back-of-the-envelope” approach as well as the more elaborate approach. He says he owes it to his clients to make that comparison to justify why [additional] resources would clearly be a good return on investment.
- Kakoli Roy said that oftentimes their approach has been first to do a back-of-the-envelope estimate. They also do this to conceptualize the problem. At CDC one needs to work in an interdisciplinary fashion. Oftentimes one is trying to model a disease or a condition and the issues are the same – the resource allocation decisions – but economists need to talk to epidemiologists and physicians to begin to understand the “other” core aspects (e.g. disease epidemiology, clinical issues) in conceptualizing the model. As one conceptualizes the model one can see if the data are there to model it. If the data aren't there one can make simplifying assumptions (unless it's a key driving variable). This can often help determine whether to move forward with a formal model. If one moves forward with a formal model, it can be published and improved over the years.
- Mr. Scanlon said that program directors sometimes think about possible modifications to the program, which are future oriented. He asked if there are considerations and thresholds for one to better view this before even building the model. For example, what are the things that one should consider (e.g., cost, time frame, etc.) in a short-term projection – one or two years – where there's good trend or other data?
- Dr. O'Grady said there's always a notion of return on investment. For example, will someone drop a large sum of money to develop this kind of work even though it may be used only a certain number of times? Also, sometimes for a decision-maker it's not a question of ad hoc vs. formal models but more ad hoc vs. anecdote. This may happen if the vote is tomorrow and the person needs an answer quickly. In such cases, back-of-the-envelope models are preferred to anecdotes.
- Dr. Iams said one needs to consider complexity. There are some policies that sound straightforward but – because of interactions – don't always work the way people think they do. An example would be giving credits for women taking care of children (child care credits). Most women who take care of children a lot probably are getting a wife benefit. Thus altering the payoff of taking care of kids will not affect anybody's benefit if she's getting paid as a wife off her husband's record, because it doesn't affect the husband's record. In addition to complexity one also needs to consider – when working short-term – the consequences of making a mistake in the estimate.

- David Betson said that back-of-the-envelope estimates can play a crucial part in formal modeling because they help people to start to conceptualize potential relationships that are going to be embodied in the model. These may be a prerequisite for formal modeling. Back-of-the-envelope estimates can be both a starting point and an ending point because when explaining results one may want to go back to that simple spreadsheet. In other words, it's a way to start explaining complex results to someone else. He added that he would never argue not to do back-of-the-envelope models. There are different levels of complexity in various dimensions. There's the level of complexity of our society and the diversity in our society that is present. There's the complexity of the policy interventions or policies. And there's the complexity of the questions policymakers are interested in.
- One of the things that microsimulation can do that other things can't is to answer a series of questions with regards to who are winners and who are losers; what these look like; and whether it would be at a racial, income, or geographic level. Decision-makers are very interested in that kind of information, which conceivably could be done through a cell-based approach, but probably is better taken on in a microsimulation approach. The kinds of models [used] are decided upon the dimension of complexity of the questions that policymakers are interested in addressing.
- Dr. Glied said there are various reasons why one needs formal models. There are some types of questions, such as distributional questions, that require formal models. There's also the question of uncertainty due to systematic bias (e.g., assumptions). For example, Medicaid take-up rates have been estimated in the literature from 60 to 90 percent. When providing a back-of-the-envelope estimate to a policymaker with an 80 percent take-up rate she can ask: Are you sure 80 percent is the right number to go with? But when providing an estimate in a formal model it's hard for a policymaker to say that there's a wrong underlying parameter. So the question is how one could use both the back-of-the-envelope and the formal model together as a way of communicating.
- Dr. Banthin said that at CBO they maintain various types of models – from formal models to simple spreadsheet models to microsimulation models to hybrids. She said they decide what approach to take based on the nature of the data available and the complexity of the question. Because they develop baselines and many of the questions are reiterated, they have the opportunity to periodically update the models, elaborate on them, and improve them. They also sometimes publish a description of what has been done.
- Dr. Citro said that formal models are not limited to microsimulation, they can include spreadsheets as well. She said someone raised the issue of long-term care. Perhaps a couple of relevant agencies should be thinking [along the lines of] “We are never going to have the full panoply of models on everything, but there are some key issues that are going to keep coming back and maybe that's an area to invest in – not only in the data and the research – but in a formal model that is available when policymakers ask the question.”
- Dr. Lee said it's sometimes hard to communicate that certain models are built for a specific purpose or a specific type of question. Distinguishing the level of uncertainty and the comfort level with the results is important but also tends to get lost in communication. Microsimulation, in

general, is useful when one has to do a distributional analysis or where there are complicated actions or a sequence of relationships that are all happening in response to a policy intervention.

- Cathleen Walsh from CDC raised a concern that as one responds to questions on particular policy interventions – using existing data or data that has to be somehow triangulated – the reliance upon ad hoc models is going to work against being able to put the kind of rigor that is needed in the models that have demonstrated their utility over years.
- Dr. O’Grady said he has used simple simulations using SAS code to help committee staff see a variety of interactions that in turn helped them to refine their policy. He also said there are sometimes differences in how different professionals go about modeling (e.g., health economists, actuaries, epidemiologists, etc.). While they may use different terminologies, their results can be similar. It may be helpful to come to a better understanding in this area.
- Dr. Citro agreed that communication is important and stressed the need to help policymakers get past the long-cherished terminologies that people use and instead try to establish a nomenclature that speaks to them and goes beyond the blinders of each specific discipline.

HOW TO CHOOSE A MODEL AND MODELING STRATEGY

Jean Abraham, Ph.D., University of Minnesota

Dr. Abraham explained that microsimulation models are a tool for estimating potential behavioral and economic effects of public policies on decision-making units including: individuals, households, employers, and state and federal governments.

A microsimulation model has at least four elements: core data, assumptions/parameters, methods, and outcomes. Core data includes data for population attributes, income, health status, employment, and insurance. Examples of assumptions include price sensitivity or preferences (i.e., options) while types of methods include elasticity- or utility-based. Outcomes include primary and secondary outcomes as well as distributional effects. It’s important to keep in mind that microsimulation models can include either a dynamic or a static timeframe.

When choosing a model it’s important to consider the purpose of the model or the question being asked. For example, is the purpose of the model for policy development (e.g., budgetary impact or cost-effectiveness) or policy implementation (e.g., estimates on behavioral responses)? When choosing a model one also needs to consider a perspective or the potential user of the information generated by the model (e.g., is the user the federal government, state governments, advocacy organizations, or the private sector).

One should also consider if there are any constraints or challenges that move the user to select a certain model or strategy. Constraints can include time, money, reputation (e.g., established, trustworthy, responsive, objective), and fit – that is, can the model effectively answer the question being asked by the user based on the data and assumptions?

It’s important to keep in mind that, from the user’s perspective, evaluating different models’ options and outputs can at times seem as “apples-to-oranges.” For example, the models from CBO, the Gruber

Microsimulation Model (GMSIM: <http://economics.mit.edu/files/5939>), COMPARE, The Health Benefits Simulation Model (HBSM: <http://www.lewin.com/publications/publication/357/>) and the Health Insurance Policy Simulation Model (HIPSM: <http://www.urban.org/publications/412471.html>) provided very different numerical estimates to the question “What happens if there is no individual mandate (i.e., decrease in newly insured, change in premium, behavioral response, etc.)?”

Dr. Abraham proposed the following questions for discussion:

- What specific information is needed by potential users in order to compare models?
- Is that information available in the public domain and is it up to date?
- How frequently are changes made to model data, assumptions, or methods relative to what is documented?
- Is there any way to narrow down the key drivers of differences in the output produced by different models? For example, is it a reporting phenomenon, differences in data, assumptions, approach?
- What validity checks are in place?
- How time-sensitive are the models (e.g., macroeconomic assumptions that affect longer-range estimates)?
- What value do existing models focusing on coverage and costs have after 2014? How will these models be adapted for the longer term?
- In what ways can models that are national by design reflect a particular state’s population and preferences?
- What is the lowest level of geography for which microsimulation models can estimate outcomes while still being “valid”?
- To what extent can models incorporate choices by states about Exchange functions or Medicaid eligibility thresholds that would affect premiums and coverage decisions?
- How easily can certain assumptions be relaxed in order to gauge their importance for a given state?
- To what extent can models generate information about distributional effects?
- Are there major evidence gaps regarding our understanding of the potential behavioral responses of the following groups to Affordable Care Act (ACA) provisions: uninsured population, employers, insurers, providers, state governments?
- How much do we know about modeling policy interactions?
- Are there additional data investments that ought to be considered to improve the capabilities of microsimulation models?

- Are administrative data sources being used to adjust baseline estimates and/or assumptions? Will they be used in the future?
- Is there interest or discussion in “broadening” microsimulation models to account for different but related outcomes and/or populations?

DISCUSSION

- Dr. Glied said that, with respect to the estimates presented, while the different models are well-documented, this is still not enough to understand the three-fold difference in their estimates.
- Dr. Abraham said she’s thought about this problem and part of the reason is that there’s no standardization. There’s huge variation in the level of detail that is presented in the public domain. For example, models may use the same data source but the matching process could be opaque or there might be differences in assumptions and baselines. Also, when modeling behavior, some of the evidence might be limited. In addition, documentation may not always be up-to-date.
- Dr. Glied said that when examining the question “How will people respond to the individual mandate?” the views of educated people who are not economists may be as valid as those of the modelers, since this area is totally new. Perhaps one approach is to ask people to do some test cases. For example, what would their model give if they ran Children’s Health Insurance Program (CHIP) or extended Medicaid?
- Dr. Bantlin agreed that in areas where there isn’t any experience, assumptions will probably vary significantly across models. A while ago CBO published a paper looking at the tax compliance literature and behavioral economics literature to try to draw some lessons from the limited evidence available. Another example might be employer decisions to offer coverage where the evidence is better and so one would hope to see smaller differences in the models in a different example.
- Dr. Betson asked a question about the process of reconciliation between different estimates and using that process to try to build confidence one way or another. He asked if this would mean that the modeler would have to adopt a specific assumption.
- Dr. Glied replied that modelers would not have to adopt a specific assumption.
- Dr. Betson said there are differences in models as a result of differences in assumptions and differences in using data and one can try to isolate those. But there are also inherent modeling differences as opposed to assumptions.
- Sharmila Choudhury gave an example of work they did on modeling individual accounts and Social Security. What helped in that case was to try to make explicit to those considering reforms what the differences were in the way transition costs were modeled, the way disability insurance was either ignored or partially taken into account, etc. When legislators looked at the study using such perspectives, the differences with other studies were stark and clear. At the time there was a belief that a cottage industry of modeling results was needed and that one should not rely on one

model alone. If one is going to encourage a cottage industry, the differences need to be made clear and interpretable by those who make the decisions.

- Robert Gillette said, with respect to the estimates presented by Dr. Abraham, that CBO and the Joint Committee on Taxation spent large amounts of time talking to each other trying to figure out why they disagreed. It rarely comes down to parameter variations, because one can adjust the parameters and check if the assumption made a difference. It tends to come down to people using a survey of consumer finance or Current Population Survey (CPS)¹ data instead of, for example, tax data. When it comes to the ACA, an awful lot is calculated in terms of tax return information instead of people or households. Discussing parameters is a good practice but it might not solve the fundamental problems, which are that oftentimes the law requires one to use a database source that is not available to others.
- Dr. Cohen said it would be helpful to have a reconciliation host database that modelers could use. When it's close to policymakers making a decision, it would be helpful if there were the opportunity for each modeling enterprise to use an agreed-upon best data resource and then filter in all the assumptions. This would help to explain if it was a data host issue primarily for the differentials or if it was the underlying model and parameter assumptions. The ACA has a section, 4302, which standardized a number of demographic metrics. It could potentially be fertile ground for investments to help narrow the gap.
- Ms. Giannarelli said, with respect to the estimates presented, that sometimes two models will provide a similar estimate for one question. However, this doesn't automatically mean that those same two models will provide a similar estimate for a different question. Understanding why different models provide different results will likely not just come from the technical documentation – no matter how great that documentation is – because different groups will place different information in their technical documentation, as different things might seem important to each of those groups. However, it's still possible to understand the differences in results from the models. It takes time, which means that it takes money, for the groups of modelers to talk to each other, and to each try running their models with different assumptions in order to really understand what the other is doing. It also takes a real openness on the parts of all teams to want to understand these differences.
- Dr. O'Grady said it could help if the different teams could be convened early on, before results are published or are on the record. However, this needs to be done carefully so as not to create the impression that political actors are influencing the way models are being developed.
- Dr. Abraham agreed with Dr. O'Grady but said the approach is more challenging when one considers conversations among competing private organizations rather than just within the federal government.

¹ When referring to the CPS, it is the Annual Social and Demographic Supplement, formerly known as the March Supplement.

- Dr. Iams agreed that relationships are important, but sometimes it's the data that are really at the bottom of it. The cross sectional datasets don't always have longitudinal data. It's very important to look at what is in the data set, what is being made up, and what seems to be reliably measured from the starting point (the starting target population). There's a lot of use of cross sectional relationships because most of the datasets are cross sectional. The Survey of Consumer Finances (SCF) is cross sectional as are CPS and the American Community Survey (ACS). Dr. Iams says he uses the Survey of Income and Program Participation (SIPP) for income security. They have respondents' records from the administrative database matched up to the people in the survey (from 1951 until now). Cross sectional surveys do not provide the longitudinal data, such as lifetime earnings, that are needed for analyzing many retirement issues.
- Dr. Iams said he spent years documenting where CPS is underestimating and it's not trivial. The problem with tax records and filing is that there's a group of people that don't file 1040s. So one basically has a very selective target population to start with. One needs to look at the data: What does it represent and what is measured? What isn't being measured? And how they are making up what isn't being measured?
- Dr. Roy said that even when researchers have been true to the data, or true to the evidence that lies behind it, differences may still exist in the estimates. If reported by the same agency, it's important for different parts of the agency to talk to each other to understand what's driving the differences. If there's divergence it could be the data or something more structural or just different methods. However, people do need to use different datasets to try to answer a question, sometimes to capture different perspectives, often involving different assumptions, so it's sometimes hard to come to a convergence. It's great to have a convergence, but that may be rarely there.
- Arloc Sherman agreed that data matters but added that very often it's clear which one or two behavioral assumptions are the most controversial or are the most sensitive. When possible it would be helpful to include at least in a footnote a bit of a sensitivity analysis, so that people can get a handle on one of the most important differences between models. He agreed that it's nice to perform tests on the models such as those mentioned by Dr. Glied, though there are often so many things happening at the same time – for example, the health care system is changing, the economy is changing, etc. – that retrospective analysis may tell you less than you think. He added that in the past, in projecting policy proposals' poverty effects, the Transfer Income Model (TRIM: <http://TRIM.urban.org>) carried out a couple of sensitivity analyses to their job take-up rate assumptions which was very helpful.
- Dr. Banthin said it might also be helpful to give end users a choice in choosing assumptions and perhaps even a choice in the sensitivity of the data and degree of benchmarking, or calibrating of the data. The initial analyses for the Clinton health care reform didn't benchmark to administrative data. However, it's becoming more standard that models benchmark survey data to administrative data. More and more administrative data is becoming publicly available, at least in tabular form. The income data are critical because microsimulation models can look at distributional analyses and subgroups – and key subgroups are almost always identified by income.

- However, income is hard to measure and is multidimensional – it’s asset holdings and drawing down from assets. To measure income one has to know who the individual lives with, so defining the tax filing unit might also be needed. These decisions are critical to assessing where that individual stands either in an eligibility threshold or a poverty concept. None of these sources are comprehensive. Tax data misses the nonfilers who are at the low income level. Social Security data are great, but they don’t tell us who people live with. So one really needs to combine survey data, tax data, and Social Security information. It’s not clear what the definitive best way for doing this is, as it will vary with the question. However, people have been raising the bar slowly in terms of how to measure income and individual status in models.
- Dr. Betson said he was involved in redesigning the Consumer Expenditure Survey (CES). There are different ways of going about this including direct data collection vs. still data collection where one would use imputation or modeling to complete a picture of household based upon other data. He said that at one time he used a two-by-three cell table to impute medical out-of-pocket expenses before the government actually began asking that question on its survey.
- Charles Nelson said that in the past there were different Medicaid numbers obtained through the Centers for Medicare and Medicaid Services (CMS) and CPS. For years there was discussion whether this was due to either under-reporting or misreporting. ASPE solved this question through the Medicaid Undercount Project (State Health Access Data Assistance Center (<http://www.shadac.org/content/medicaid-undercount-project>) and <http://www.census.gov/did/www/snacc/index.html>) but this involved numerous agencies, several years of research, and lots of money. It’s important to keep this in mind because there are a lot of data questions out there. Perhaps the approach should be to pick the biggest problems, throw resources at them, and then decide which of the other problems can be addressed in other ways.
- Dr. Glied said that one year from now health models will be different, because once the expansions happen all the baselines will change. So this might not be the moment to invest heavily in existing insurance models. Since policy spaces are changing, one might want to start with an ad hoc model in new policy spaces and then move towards more formal models. She asked if it would be possible to release administrative data tabulations that modelers could use for purposes of calibration. For example, although the Department of Treasury can’t release its model, they could release how many tax filing units there are by income category. In a similar way, the Social Security Administration could release certain kinds of tabulations based on data only they have. Thinking about new survey questions and adding administrative tabulations could help advance the modeling field.
- Dr. Citro read the following sentence from a report “in addition to assessing data qualities, statistical agencies need to add more value to the data series they release than is currently the practice.” She added that some organizations have not generally seen as their role the production of analytical databases or the publishing of best estimates, such as household income or poverty, that could be developed from multiple data sources. She added that it would be helpful if an agency like the Census Bureau would think of their role as not just asking the questions and putting out the data, but actually using administrative data to replace missing data (or

misreporting). A number of models do this (e.g., TRIM) but they don't have the advantage of the full array of microdata or confidential data that an agency like Census has. ASPE and other organizations could perhaps work with statistical agencies to change this paradigm.

- Joan Turek said that the income study was unique because it actually attempted to make consistent estimates across surveys (Income Data for Policy Analysis: A Comparative Assessment of Eight Surveys: <http://www.mathematica-mpr.com/publications/PDFs/incomedata.pdf>). She said that sample size also matters and wanted to hear what other participants thought about this. One needs pretty large surveys to get state-level estimates but when one gets to the ACS, one gets the least amount of data.
- Dr. Abraham agreed this is a big issue, particularly with respect to the ACA. She added that states are heterogeneous and when modeling the implementation of federal legislation one needs to think about the fact that data investments are mostly for national purposes – although the ACS and the CPS do pretty well at the state and the Census Public Use Microdata Areas (PUMA) levels. She asked what strategies exist in terms of streamlining or focusing to basically have the infrastructure there to advance understanding at lower levels of geography.
- Dr. Cohen said that with respect to tax expenditures, there is a tremendous concentration of wealth in 1 percent of the population. Also, when looking at the impact of the ACA on an individual's health, 1 percent of the population is tied to 22 percent of expenditures and 50 percent only drive 3 percent. Therefore, understanding the policy issue and what seems to be a very resourceful dataset – where all the action is in a very small quadrant – really emphasizes the uncertainty and challenge to modelers.
- Dr. O'Grady said he thought the SHADAC project, which was an attempt to come up with a consistent set of Medicaid estimates, provided a good return on investment. He added that sometimes when federal monies are spent, there are a specific set of advocates for or against that investment, which sometimes ends up creating a gap [in supplementing or expanding the surveys]. He added that if states request better data or methodologies, the taxpayers of those states should make a decision as to whether they would like to make those investments.
- Dr. Abraham said that most models are built on federal data but states may want information that is population-specific. Is reweighing sufficient to achieve this? Or could one feed information from state surveys into models to customize them?
- Dr. O'Grady said he's seen both state-level surveys as well as states who have approached a federal agency and made an investment for an expanded sample. This might be better than reweighing and is consistent with federal sampling.
- Dr. Cohen explained that going to the sub state-level census has been quite an investment in terms of the modeling enterprise where they are not direct estimates – they are indirect estimates – and contain a lot of validity in terms of the evaluations that are underpinning small-area estimation techniques. As a middle ground, there have been advancements in Bayesian techniques and other methodologies that will be gaining a lot more respectability where one can

see the error bounds and some estimates of bias for that approach. He added he wouldn't give up on a modeling perspective as a reasonable middle ground if the resources weren't available.

- Dr. Roy commented on whether the state should be the lowest level of geography. The changing policy landscape has implications for prevention and many of the changes will happen at the community level. Therefore, CDC has data sets like the Behavioral Risk Factor Surveillance System (BRFSS) which are being improved to better measure these impacts at the community level. Also, "sites" (where one works, goes to school or plays) – for example, the work site – is beginning to matter. Impacts of worksite prevention policies may vary depending on both health plan and also location (geography).
- Ms. Giannarelli said they've been able to use the ACS very effectively for some state-based modeling – not necessarily health modeling, but anti-poverty effectiveness of policy changes. The approach they've taken is not to build a new model but to adapt the data to their existing model – to leverage the code and the methodology that's already been developed. They've used this method to answer questions for some state-specific projects, rather than trying to combine two to three years of CPS data.
- Dr. Gillette said that for smaller estimates the models generally have to be more flexible about pooling data. He added that they will usually pool from data from tax information filed in the United States. Nonfilers can be created – and while this information is not as good as the actual tax return data – one has to be flexible about the data that goes into the model. The model should be designed so that one can incorporate different kinds of data into it and see what comes out.
- Mr. Sherman said they might get a request for a back-of-the-envelope state estimate when state-level numbers are not available. A state-level stakeholder may ask for a "best guess." In such case, it's helpful to have any demographic and income distributional data from national models. What's helpful is to have results from the national model by categories that match up with other data that do go down to the state level, – for example, having one set of runs that reports results by Adjusted Gross Income (AGI) using the categories that are in the published Statistics of Income (SOI: <http://www.irs.gov/pub/irs-soi/12soisprbul.pdf>) state-by-state tables. Ideally, this would be in addition to results for married couples, families with children, single parents, race, ethnicity, etc. Anything that matches up with some other data set that goes down to the state level ends up being extraordinarily helpful as a consumer of these model results.
- Dr. Turek said that one of the challenges is the creation of synthetic data. She said she's seen examples for very small, important policy groups that have totally destroyed the validity of the data. It's important to protect privacy, but one also has to ensure that any attempts to create synthetic data don't affect the covariances.
- Dr. Iams said that one of the main questions with synthetic data is: "Do they retain the covariant structure that exists between whatever one is making up and important variables that are related to it?" For example, if one is making up synthetic earnings – Social Security earnings data – how does that co-vary with household assets or personal assets and pension income or pension rights access? Oftentimes the person creating the synthetic data is just making up that piece of synthetic data and does not retain the covariant structure.

- Dr. Lee said that, as a consumer of various data sources, she would find it helpful if the experts on the data (or the individuals who are producing the survey data) could recommended a way of calibrating for a particular type of analysis or at least recommend what not to do with that data set. Perhaps they could address if there are any general types of problems that modelers are likely to encounter and if there are recommended solutions to address them.
- Fritz Scheuren said that in Northern Europe they use administrative records a lot more. There's a slow move toward using administrative records in the census, but there has also been reluctance to change. It would be useful to use administrative records more and to continue to use surveys to interpret administrative records.
- Dr. Abraham said that slowing the growth of health care costs and changes to the supply side are going to be very important as one thinks about innovation. She asked for any insights or thoughts on the need for either research and/or data investments in terms of how to more effectively model responses to payment reform, organizational reform, or other aspects.
- Dr. Cohen said that limitations – in terms of existing data resources to carry out longitudinal analyses – are front and center when looking at a policy innovation, how it gets implemented, and both short- and long-term outcomes. Families, tax filing units, and Health Insurance Units (HIUs) are dynamic over time and understanding how this dynamism plays out with policy initiatives is critical. Surveys can be done much less expensively after they're done for two or three years. One doesn't need to include everyone, but rather oversample for the key policy groups. There are paths one could take if one prioritizes on data resource investments and this should be considered.
- Dr. O'Grady said one should consider if the model is adequate or if it's going to be ready when policymakers need to make decisions. If the methodology and the models don't match the questions that policymakers will have, then one can expect the same challenges as seen with Medicare Part D and ACA when there were different methodologies and none of them quite captured everything. He said it's better for the technical people to sit down and work out the best method before they're suddenly thrown into the middle of a policy debate.
- Dr. Roy agreed with Dr. O'Grady. She said that more and more modelers are being asked to break down who pays and who benefits. There's no point in recreating disease models, especially if they are created in siloed disease worlds, and don't address the relevant policy questions. One should try to strengthen the downstream "policy" components of the existing models targeting disease prevention and start to populate the model with the kinds of data that capture the different payer perspectives correctly. She added that the types of metrics being measured are also changing from traditional incremental cost-effectiveness ratios – to reporting health care costs averted, productivity costs averted, increased earnings, etc.
- Richard Kronick said a tremendous amount of energy has been spent building demand-side models. There will be a growing need to invest in supply-side models so that one can understand the effects of various policy levers in influencing the supply of resources in the health care system.

- Dr. Banthin said they've recently re-read the report on analyzing income measurement in various household surveys (cited earlier) and are using work from the Mike Davern study of Medicaid underreporting in CPS (SHADAC study). They will be critical to some of their upcoming analyses. She applauded ASPE for funding them.
- Dr. Turek added that Don Oellerich has developed a follow-on looking at selected income sources such as IRAs.
- Dr. O'Grady suggested that the report be posted on the ASPE website.

ASSESSING AND INTERPRETING MODEL RESULTS

David Betson, Ph.D., University of Notre Dame

Dr. Betson began his presentation by discussing two papers with the audience, *"Estimating the Effects of Proposed Legislation: The Case for Model Validation, or Why are the Numbers so Different?"* and *"Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force – 6"*. Both papers address the sources of errors that can creep into estimates. (Both are included in the list of references).

Dr. Betson also read from a 1993 report by the Committee on National Statistics, *Microsimulation Models for Social Welfare Programs: An Evaluation*, which conducted a review of social welfare policy models. The report found a "virtual absence of systematic activities to validate models and their results. Hardly ever do estimates of uncertainty or the results of sensitivity analyses accompany the cost estimates that are provided to decisions makers."

Since that report was published a good deal of effort and energy has been placed into providing estimates. Researchers have also attempted to quantify how good those estimates are. As a result it's more common today to see standard errors of estimates being provided or a sensitivity test of assumptions.

This is due in part to increased expectations of users as well as modelers reflecting increased professional standards. In recent years, increases in computing power have also made empirical techniques such as bootstrapping and empirical Bayesian approaches possible.

Best estimates can deviate from the "truth" through bias (systematic errors in estimates) and variability (random errors in estimates). Dr. Betson presented a variation of an equation presented in the papers above:

$$Z_i = \alpha + \beta X_i + \delta Y_i + \varepsilon_i$$

The equation shows the nonlinearity of the relationship between Z and causal factors X, Y and ε . These are important factors in determining whether sampling variability and parameter uncertainty will create bias. The parameters of uncertainty are represented by α , β , and δ . Uncertainty can lead to variability in the estimates or it may lead to bias. Imputation errors (bias and not maintaining the appropriate correlation with X_i) can also create errors in the estimate leading to bias and variability in the estimates.

Dr. Betson described the following challenges in modeling:

- Modeling of behavioral responses (static vs. dynamic)
- Data for the year one wants to predict not being available so the data have to be altered to reflect the future year (aging of data)
- Alternative modeling approaches that are not tested

Several strategies are used to provide confidence in modeling including validating models by using one source of data (e.g., how well does the model actually predict events captured in another data set), examining the accuracy of the forecasts, and identifying bias in a model's point estimates or systematic errors. Sensitivity tests can also be made of assumptions. In addition, one could also examine alternative modeling approaches (i.e., looking for robustness of estimates). With each set of estimates one can provide estimates of the possible extent of variability in estimates (i.e., confidence bounds for estimates).

Dr. Betson offered the following questions for discussion:

- Are there additional examples of attempts to validate model estimates or provide estimates of variability?
- Is there demand by policy makers not only for point estimates but also for estimates of the quality of the estimates?
- What can be done to "educate" policy makers about the importance of quantifying quality? Will supply create its demand?
- If measures of quality of estimates are provided to policy makers, how will these be used? Or is it more likely that they will be misused?

DISCUSSION

- Dr. Citro said that public opinion polling, such as was done for the last election, is now presenting some level of uncertainty or variability (i.e., plus or minus), so people may expect this in the future even though policymakers usually need point estimates. Perhaps the modeling community should set up standards for itself and take advantage of new graphical tools that make uncertainty easier to present.
- Dr. O'Grady said that part of enforcing quality control involves getting better models, rather than by modelers simply being more transparent. This can also be accomplished by having policymakers use trusted technical experts like CBO, CRS, and MedPAC. He added that it's up to the [modeling] community to drive the poor quality out of the market because there's still a financial incentive to produce poor quality modeling.
- Dr. Iams said the Social Security Administration has as a requirement in its evaluation contracts that the firm propose a panel of independent experts that are asked to attend the first meeting for the regular evaluation and also evaluate the final results as well as provide a technical review. He suggested it could also be an approach for microsimulation models.

- Experts can be involved in various pieces of the model. For example, the contractor or inside group could propose an approach and the experts could react to whether they think it makes sense. Another approach is to have an outside technical expert come in and evaluate the model. He said they had their model evaluated by the statistician at GAO as well. The panel is an expert authority that's independent and has no vested interest.
- Another approach is to develop two independent simulations and then compare the results. They had their model predict income outside the model and then compared it to SIPP, Health and Retirement Study (HRS: <http://hrsonline.isr.umich.edu>), CPS and the SCF. He said that two models that are independent and have different data sources can at times provide similar results. He compared results from his Modeling Income in the Near Term Model (MINT: <http://www.socialsecurity.gov/policy/docs/ssb/v66n4/v66n4p1.pdf>) to those from the KGB model² (which used a different dataset) and the results were remarkably similar.
- Dr. Gillette said that sometimes there are unknowns that are themselves unknown – variants that the model will never be able to determine. He said that forecasting economics is an act of hope, not an act of determination. Outcome accuracy comes down to a sense of coherence rather than variation. If one can be consistent and coherent over time, then one can develop a track record on answers, although one can't always track dollar outcomes because they can be dependent on things that one can't control.
- Dr. Glied said that policymakers are used to dealing with uncertainty. This was seen with hurricane Sandy where Mayor Bloomberg was informed by NOAA that a hurricane could hit the city. He had to make decisions based on a forecast that contained a certain level of uncertainty. When discussing uncertainty one needs to consider a variety of questions: What decision does the policymaker have to make? What is the timing of that decision? And how can they think about the uncertainty of a model estimate in that context?
- There are different kinds of uncertainty. For example there are unknowns that one doesn't yet know (e.g., one won't know until something happens such as Medicare Part D or ACA). But there's also uncertainty that could be better determined by collecting certain kinds of information or doing certain kinds of experiments. In the latter case one has to ask, how much is it worth to buy this information or how long will it take to create it? All of this information can be used to communicate uncertainty. This goes beyond simply communicating the standard error as uncertainty.
- Dr. O'Grady said that models sometimes have limitations and can't always generate the answers that policymakers are looking for. For example, there is a point where the modeler can tell a policymaker that if she talked to the mayors of five leading cities in her district she would probably obtain better information than if the modeler continued to "slice and dice" the data.

² Model developed in-house for ASPE and used during the Carter welfare reform initiative.

- Dr. Choudhury talked about the importance of having an oversight process. For example, SSA and the Bureau of Labor Statistics, Department of Labor (BLS) have implicit oversight processes in place via expert panels. Modelers in government agencies look at expert panels as a way to continue to improve the quality of model building and the quality of the estimates. But expert panels can also have an oversight function and their interaction with government can send the message to policymakers that oversight exists.
- Dr. Kronick said the question of how to communicate uncertainty depends a lot on what one is communicating and in what context. Sometimes Congress needs a [point estimate]. At other times, when policymakers are trying to make a decision, one can inform them about the range of uncertainty which can be helpful but difficult. He added that uncertainty is not necessarily the sampling error itself.
- Dr. Iams responded to Dr. Gillette's comment of keeping the model constant. He said that if the structure changes one needs to change the model in some way. For example, the MINT model was designed to examine the impact of Social Security reform on the Baby Boom, but the structure changed for pensions so the model had to change. Another example of a structural change is the Great Recession which could have a long-term impact. If a model is built and the relationships were based on one kind of world and the world changes, the model has to be revised to reflect the world of the future and not the world of the past.
- Dr. Turek said that years ago she was involved in a project that took a significant amount of time and used up all of her mainframe computer budget for that year. This taught her that one needs to plan ahead and determine costs because if one gets a request from a policymaker next week, one might not have the [resources] needed to answer that request.
- Mr. Scanlon asked questions about the evaluation of a model. If an agency developed a contract, where would one expect to see this evaluative information – how the model would perform historically, how it would perform recently, and some of the quality and technical dimensions? Is there a matrix or a "Consumer Reports" for this? Or is the approach that every policymaker has a technical person who is the "translator" and adviser for this? In other words, how and where would this quality information be made available?
- Dr. O'Grady said that, in the policy realm, the impression of the rigor of the model is sometimes just as important. Policymakers have trusted sources. For example, ASPE is the analytic shop that works directly with the Secretary of Health and Congress likewise has its own shop. In other words, sometimes decisions are made based on the reputation of the organization doing the modeling.
- Dr. Banthin said that CBO is somewhat unique because their models are as nonpartisan and unbiased as possible. They go through extensive literature reviews, marshal the evidence, call upon experts, and consult with a panel of health advisers and other agencies. They also develop documentation. She added that their clients often want a point estimate. However, on a recent study they did a sensitivity analysis while modeling the Exchanges and deliberately altered the parameters in both directions [of the distribution] to achieve results, so they ended reporting on four different scenarios. What happened is that policymakers instead of looking at the whole set

of scenarios chose their favorite one and said “CBO said such and such.” She followed up on Dr. Gillette’s comment about the coherence of the estimate and forecasting involved. She said that CBO develops a 10-year budget projection. They obtain new data and revise it once a year, which is part of the coherence of the model and the estimate.

- Dr. Glied said that part of the validation of the model would be to write into the contract the requirement to estimate the effect on policies that have already taken place. That is, policies for which the answers are already known.
- Dr. Gillette said that, from the Office of Tax Analysis/Treasury’s (OTA) point of view, a large part of what models are about is consistency in the estimate. He added that models can give consistency for counterfactuals that can’t ever possibly exist. It’s very important that one have the ability to construct “can’t possibly ever know” questions in a way that is coherent and that one can answer, at least comparatively. He said they don’t “live and die on the baseline” but rather on “consistently constructing deltas.” He agreed that reputation is important.
- Dr. O’Grady said that a weakness of the consistency argument is that if an agency said their estimate was X last year, they may hesitate to deviate much from that base for their next estimate. This has a stifling effect on how innovative policy can be.
- Dr. Gillette said there is a difference between economic and technical corrections. For example, an economic correction could be a revision of the budget while a technical correction would involve fixing the code, etc. He added that during the six-month period when they are preparing their budget, OTA will usually not make modeling changes. In other words, if a change is requested, they will use the same set of technicals used in the first place because consistency is important. However, models do evolve over time – they are now on version 20 of the individual model and on version 12 of the corporate model.
- Dr. O’Grady said that one of the things to keep in mind is that there is a constraint on the amount of innovation that can come into policy because sometimes if something can’t be modeled it can’t be scored and it can’t move forward. For example, the drug benefit was mostly assumption-driven and the move to generics turned out differently than in reality.
- Dr. Gillette explained that when some policies are completed and the bills have been passed, there are months or even years of details that have modeling implications. An example would be the scoring rules for bills where state actions are optional, as is the case with the ACA.
- Dr. Lee said that if it’s a score on a bill then the actual number matters, but it’s different if one is using the model to assess different policy options or the trade-offs of having a particular design parameter of a policy. In the latter context it’s more important to have a consistent model that provides more accurate trade-offs rather than hitting a final number.

COMMUNICATING RESULTS

Linda Giannarelli, The Urban Institute

Ms. Giannarelli’s presentation focused on communicating modeling results.

The following three broad questions should usually be considered when communicating research results:

- What has the research found?
- What data and methods produced that finding?
- How sure is one of the results?

These questions are harder to answer when the research is based on microsimulation modeling. To give an example, the TRIM model found that 5.05 million families were eligible for TANF during the average month in 2005. However, what data and methods were used to produce that finding?

Input data sets for microsimulation are not typically just a single data set but include augmentations made by the modelers. TRIM uses a variety of data sources to reach those results, starting from the CPS data reported by respondents but also including the imputations and edits performed at the Census Bureau. The TANF eligibility estimate is also affected by imputed immigrant status information, the way in which annual incomes are allocated across the year, and TRIM-simulated Supplemental Security Income (SSI) data. In terms of the methods, TRIM uses detailed state-by-state modeling. Some simplifications in the model may also impact results. For instance, TRIM does not model ineligibility due to the value of someone's car.

When communicating information on how a particular microsimulation estimate was obtained, one should ask a wide variety of questions.

For example, in terms of input data, what are the starting point data? Has there been any aging of the data? Have there been additions to the data through statistical matching, and if so, exactly what methods were used? Other questions include: What rule-based edits or allocations are made? For large, comprehensive models, to what extent does a particular estimate depend in part on simulated data from an earlier piece of the model?

One could also ask questions about the policy parameters for the simulation. For example, for the baseline case, where do those parameters come from? What is the level of detail and range of parameters? Is the baseline case aligned to actual caseload data for benefit programs? Do the policy parameters come from a single source or are they pieced together from multiple sources?

Big-picture method questions include: Is it a single-purpose model or a comprehensive one? Is it solely examining the household level or does it bring in other sectors? For modeling of policy changes is one looking only at the near-term impacts or longer-run impacts? Are behavioral impacts captured?

More detailed level method questions include: How is the program or the policy option being modeled in terms of eligibility, benefit, or tax computation? What simplifications or assumptions are made? For comprehensive models, what methods in other parts of the model affect this estimate?

Communicating the degree of uncertainty is particularly challenging with microsimulation modeling. In addition to uncertainty arising from sampling variability, there is uncertainty due to imputations and the impact of various assumptions. Providing a range of estimates in addition to a point estimate can be considered.

When communicating results another challenge is to decide what will be the most important information to convey, based on the research question. For example, of the thousands of pages that one could write about the methods and data for a particular model, what are the most important few pages to convey to anyone looking at those numbers in order for them to: 1) Understand that number; 2) Understand how much credence they should put on that number; and 3) Understand the characteristics of that number? In the ideal world, one should find ways to document and describe simulation models in a way that people who really want that information can find it.

Another challenge to communicating results is when estimates change. Estimates can change for various reasons, including when methods for models are improved. When estimates change they can improve the current estimate but at the same time disrupt a time series. In such cases it may be challenging to convey the reasons for a change.

A report from a task force on transparency and validation suggested that every model should have both technical and non-technical documentation. Some potential goals for technical documentation include:

- Documentation of additions to the data
- Description of methods appropriate to different users (i.e., provide ability to “drill down” to a finer level of detail)
- Availability of parameters
- Availability of “baseline” results
- Documentation being publicly accessible (e.g., on-line)

Ms. Giannarelli proposed the following questions for discussion:

- What should people know about models in general that they currently don’t know?
- Should there be improvements in the content of documentation? Accessibility of documentation? Comparability of documentation across models?
- When is a range of estimates helpful?
- How much information should be provided on the impact of changes in methods?
- How is the modeling community doing with communicating the results of its work, its methods, and uncertainty? How could this be done better?

DISCUSSION

- Dr. Turek said she has managed the TRIM model since 1976. She oversees a group called the ASPE TRIM Management Team, which has experts in each of the subject areas of the TRIM model, and they work directly with analysts at The Urban Institute who work on the module. So essentially the module is developed in cooperation with the user which makes it somewhat unique.

- Dr. Bantnin said that at CBO analysts build the model and then participate in using it. She added that it's important – when using the model to estimate policy options – to obtain clear specifications from staff on the Hill. Therefore, having close communication with staff is important. Occasionally in areas where CBO is still exploring and developing models, it will reach out to them to understand the questions they're focused on to develop a model that responds to those aspects of the policy they're interested in.
- Dr. Cohen thanked Ms. Giannarelli for providing a comprehensive list of best practices. He said one could also consider adding a statement explaining what could have been done better if some aspects were available (e.g., specific data resources or more timely policy assumptions). In other words, a statement of what more could have been done while recognizing the limitations. He added that a call for non-technical documentation is critical because heavy users of models include policy individuals who might not always have the time or background to understand [technical documentation]. For added transparency it might also be helpful to cite the funding source.
- Mr. Sherman said that TRIM is a model of transparency by providing helpful information including baseline information. This allows one to understand context and also do back-of-the-envelope calculations just by knowing who is eligible and not served under current scenarios. There is also great value in knowing key behavioral assumptions – what work response or take-up rates are built into a model. It's also great that TRIM sometimes provides microdata. Having state data available is ideal but when state data are not available other categories – such as AGI categories – can help bring some simulations to the state level.
- Having a distributional result beyond race, ethnic, or education categories is also helpful. Work by Marianne Bitler and others found that – in the case of welfare time limits in Connecticut – there were policy impacts that were not correlated to race or education or time on welfare (i.e., any of the usual predictors) and the mean impact was close to zero and yet the time limit policy was found to widen the dispersion of participants' income dramatically. Therefore, looking at dispersion of impacts can be quite important. It's also sometimes helpful to communicate uncertainty in an estimate beyond a confidence interval (e.g., assumptions, baseline, etc.).
- Dr. Roy said that when thinking about technical documentation one should also consider the audience. You may need to provide different tiers of documentation to address the different audiences. The audience includes not only one's peers (such as economists) but also policymakers who might be technocrats and who may need to know how one got the estimate as well as information on key drivers and the uncertainty around the estimate. This can help the policymaker defend the estimate – even if it's a point estimate.
- Then there's also the expert panel who might want to get into the “nuts and bolts” of the model. They might want to know where the data came from (e.g., econometric studies, RCTs, meta-analysis, expert panel review, etc.). So one needs to include information on the inputs, parameters estimates, and data. Finally, there are also one's peers or other analysts who may want to get into the technical details of the model (e.g., methods, etc.). One may also want to be able to use the model. For example, if the policy landscape is changing can the technical analysts within the

government examine the impacts of changing some of the key parameters as the policies change? In other words, provide working models.

- Ms. Giannarelli said she agreed with the need for different levels of documentation as well as for providing information for different parties. She asked what microsimulation modelers are doing well and what they are doing less well in those areas.
- Dr. Abraham said that, with respect to ACA models, what is being done very well is the fact that there are good, clear descriptions of the data sources – the population – particularly around the insurance data premiums. There’s less detail about how synthetic firms are built. It might also be useful to improve documentation in the area of effectiveness or the timing of implementation.

With more complex models one should also think about policy interactions. That is, how people respond to more than one change at a time and the interactive effects of different types of changes. It’s also important to keep in mind that many institutions are changing simultaneously. Also – when relaying back estimated parameters to what is known from the literature – in some cases one might not want to rely so much on history as one thinks about major changes. Finally, knowing the reference period (i.e., how old the documentation is) would be useful.

- Dr. Betson commented that microsimulation is sometimes graded on how minute parts of the legislation are included in the model. The challenge is how one takes a complicated and detailed model and simplifies it while still providing the idea that one has modeled all those details.
- Dr. Choudhury said there is a similar challenge in retirement preparedness. There are at least two schools of thought in this area: “the sky is falling” and “it’s not so bad.” Those who have analyzed the literature look at the differences driving the results. Perhaps documentation could shed light on certain key variables that researchers just happen to look at differently – for example, in retirement preparedness how researchers include housing equity. There’s no right or wrong, just folks looking at it differently. In other words, if someone is examining two credible studies that person may want to know what is the driving difference. Perhaps there’s a way to highlight some of the key variables that are used differently.
- Mr. Sherman said the Web is very good at helping with how much detail one presents. People can click and drill down to what they want or click on a comparison to other models (to the extent one has them). Letting that architecture guide people is helpful.
- Dr. Citro suggested that perhaps a set of certain minimum requirements could be specified in government contracts. For example, contracts could specify the minimum amount of information that needs to be provided such as the data being used, whether they are imputing for under-reported income, etc. Contract requirements could then be used to build a reference database that people could go into and also be linked to more detailed documentation.
- Dr. Turek said that TRIM is a publicly owned model and detailed information is provided to anyone who wants it through <http://trim.urban.org>. The site runs through the logic of each module; tells the user of all of the imputations and statistical matches made; provides complete data

parameters, etc. Some people think this is still not enough information while others get confused because they think it's too much. So in a sense one can't win.

- Dr. Citro said the information provided by TRIM is great. She was referring more to a simpler abstracting of information that could be used to make higher-level comparisons.
- Ms. Giannarelli said that modelers are reluctant to make comparisons with other models because the documentation is not always up-to-date. Also, modelers may be reluctant to characterize other models.
- Dr. Citro said the idea is that – if models want to be considered for HHS funding – they themselves need to provide a standard set of information which is both detailed and abstracted as well as a creation date.
- Dr. Glied said that federal policymaking bodies who contract with modelers can require beforehand that a specific set of parameters be provided along with a set of descriptive statistics. This is sometimes done in the literature, but not always done with models.
- Mr. Scanlon said that ASPE actually did some comparison and contrasting of models in 2009. ASPE put together fairly descriptive information about some models at the time, including some intramural and commercial models. However, most of the information was descriptive rather than evaluative. This information was eventually used to select three models. He asked if this approach would still be a possibility and whether it's something the federal government should do.
- Dr. Glied said that at the time a decision was made to go with multiple models because it was thought that something could be learned by taking this approach. She asked what other's thought of this strategy from a policymaker's point of view.
- Mr. Scanlon agreed it's worth it, especially when the stakes are high.
- Dr. Glied asked if it's better to do a lot of evaluative work on one model or go with a couple of models.
- Dr. Abraham agreed it would be a good idea, to the extent that the entity producing the model has the resources to do it.
- Dr. Turek said that several years ago she had contracts with both TRIM and MATH. When TRIM would make a development, she'd give it to MATH and vice versa. This kept models consistent and when they disagreed they could talk it out. However, it's unlikely that this scenario would happen today because of diminished funding.
- Mr. Scanlon asked if it's the federal government's role to evaluate private products.
- Dr. Banthin – speaking as a modeler and not representing CBO – said that this is an era when models are not perfect and it's not always fully understood why models do what they do and why they differ. It might be useful, when affordable, to compare models. Defining output statistics (e.g., baseline distribution of families across income groups) is really important. CBO's model can produce a whole slew of output statistics although it usually looks at a small set for a given run.

- Mr. Scanlon said that agencies may only have one month to pick a model and devote significant resources to it. States, nonprofits, and communities may also have to select a single model. Where does quality information or performance come from?
- Dr. O’Grady said it might be helpful if federal contracts were modified so they require proper justification and benchmarking at the outset (e.g., matching a CBO cost estimate).
- Dr. Banthin said that one can develop components or modules that do specific things very well. This provides flexibility because, as policies change, components can be applied differently in a new framework. She added that some models are better at doing some things than others. Since all models aren’t always going to be comprehensive it makes sense for the federal government to use multiple models at different times, or even at the same time.
- Dr. Roy said it’s sometimes good to see if results [from different models] can converge, especially if they take very different approaches.
- Dr. Scheuren said the discussion has centered around tax models, transfer models, and health models, but the sample designs for each of these are different. They have a degree of lesser or greater uncertainty. The tax models combined with the Federal Reserve Board surveys are very good on the top end. While the CPS models are not good at the top end, they are good at the bottom end. Health models are still developing but will be very stratified if they are focused on costs. Some of this is fixed due to the nature of the data from the Medicare system. In other words, there are three models for three data sets, each with its pluses and minuses.

MODEL PERFORMANCE AND UTILITY IN RECENT APPLICATIONS

Richard Kronick, Ph.D., ASPE, HHS

Dr. Kronick explained that ASPE has made use of microsimulation models for various purposes. For example, microsimulation models have been used to examine a variety of policies related to the ACA. A point in case is that HHS will soon be issuing a proposed rule on reinsurance fees. The ACA states that \$10 billion should be collected for a reinsurance pool – with \$2 billion going to the federal government. Microsimulation was used to estimate what the difference would be in collecting fees on a per-person vs. on a per-policy basis.

Microsimulation models have also been used to examine the reinsurance program parameters for insurers in the individual market. They have been used to estimate who is going to be in the individual market in 2014 and what their spending distribution looks like as well as what will be the effects of various attachment points and coinsurance rates. Models have also been used in the development of an actuarial value calculator.

The ACA calls for establishing methods for conducting risk adjustment. The department will soon be publishing a proposal on the methodology that the federal government will follow for risk adjustment. The ACA also calls for the Secretary to certify state approaches to converting varying current Medicaid standards to Modified Adjusted Gross Income (MAGI) standards. The goal is to try to assure that – under the converted standard – the same number of people will be eligible for Medicaid. It will also help determine eligibility. Microsimulation will be used to assist in that determination.

The ACA states that the amount of money that Medicaid pays to states for a disproportionate share of hospitals will be reduced. Microsimulation will be used to develop allocation formulas for the disproportionate share of hospital reductions. Microsimulation is also being used to establish the age curve in the individual market.

Microsimulation is being used to determine parameters for some of the features of the basic plan. In addition, microsimulation is currently being used to examine insurance outreach and enrollment efforts. HHS has been producing information to help folks who are doing this understand location and characteristics of the uninsured at a very micro level.

HHS has also made heavy use of microsimulation models in a variety of regulatory impact assessments that have accompanied various rules, such as prohibiting insurers from imposing pre-existing condition requirements on children, the requirement to allow dependents under 26 to remain on their parents' policies, and a variety of other regulations. These regulations come with impact assessments that make heavy use of microsimulation models.

Microsimulation has also been used to answer a wide variety of questions. For example, is partial year reconciliation possible? What are the effects of the ACA on premiums in the individual market? What will be the impact of the ACA on demand for primary care and public health services?

A wide variety of microsimulation models have been used to answer these questions. HHS makes heavy use of the RAND COMPARE model but also uses The Urban Institute's HIPSM and TRIM models. Other models are also used. Work from other organizations, such as CBO, is used for reference and comparison.

Dr. Kronick said one can also learn microsimulation used in other sectors. He said his daughter works for a market research firm that works with consumer product companies. They use microsimulation to determine how many units of a new product – or a variant of an existing product – will sell. He believes that for minor variants of existing products this approach seems to work better than for products that are new and very different from anything on the market. He added that more interest has also been given about how to communicate uncertainty in modeling results.

DISCUSSION

- Dr. Citro said she attended a seminar about the problems with declining survey responses. This impacts the private and public sectors differently. A big company can afford a lot of market research even if their product is not successful, because they can quickly turn around. Also, they may not have a need to care about how totally accurate the information is. However, governments are trying to inform the public as well as policymakers and therefore need to take a different approach.
- Dr. Kronick said he thought the private sector does indeed care about the accuracy of the information. They're trying to make money and also make the right decisions. He added that he didn't want to suggest that the work they do is less rigorous. However, they may have an easier job than those creating microsimulation models to estimate the effects of the ACA because a lot of this work is out-of-sample. Dr. Kronick added that there's a need to establish reinsurance attachment points and create an actuarial value calculator.

- Dr. Citro agreed that private sector work can be rigorous, but their job is simpler and also they generally have an ability to recover much more quickly than the government does from getting it wrong. She added that it's important to have several models to address underreporting of income or other aspects when engaging in public modeling.
- Dr. O'Grady explained that the private sector is also impacted by mistakes – stockholders can be quite unforgiving when a mistake is made. He said there may be some areas where the government may be further ahead because of the budget discipline that is imposed on public modelers and also because there are more evolved methodologies and more established metrics.
- Dr. Glied said there might not be a need for multiple models in the private sector when comparing two products when there is a well-established market research firm with a track record. However, it's different when one is doing something new and out-of-the-box and when there isn't an established track record. When the parameter assumptions are really at play, then getting lots of people's views on how it all fits together can be useful.
- Dr. Banthin asked if market research firms followed the same approach – calling up people and getting them to respond. She added that economists haven't yet decided whether elasticity or utility models are better. Since there's not consensus, one can get different results depending on the setup.
- Dr. Kronick replied that in market research they do have different methods they can follow and thus can come up with different results. He agreed with Dr. Glied about the importance of using multiple models when there's so much uncertainty.
- Dr. O'Grady said that one should also consider multiple databases. For example, for income one could consider CPS while for healthy spending (i.e., how many times one goes to the doctor) one could consider a different database since each was built for a different purpose and each has their own sweet spot. He added that one of the challenges is groupthink – when there is a group that has been working on the same model for a long time and may at times miss something big.
- Dr. Gillette agreed that groupthink can be a challenge. There's a need not only for multiple models but also for different modeling groups – there isn't one group that is big or coordinated enough to address all of the policy issues raised. He added that there are very different but equally valid economic approaches to these problems that unfortunately tend to give different answers when they're pushed in different ways. Therefore, what is needed is many people thinking about the issue in different ways. This means having different models, rather than one model with many different people trying to run it.

LESSONS LEARNED AND BEST PRACTICES

Jessica Banthin, Ph.D., Congressional Budget Office

Dr. Banthin spoke briefly about CBO and also summarized the day's discussion. CBO's role is to provide objective, non-partisan estimates. CBO provides projections of federal spending and revenues with models focusing on the federal budget. The time horizons for CBO projections include a 10-year budget window for baseline projects as well as other long-term projections.

CBO uses a variety of models including cell-based models, regression modules, microsimulation, or a combination of the above. Using these models facilitates consistency and replicability of estimates over time while enabling timely responses to requests for estimates.

The process of constructing and reviewing models incorporates a variety of aspects. These include reviews of research literature, reviews of historical data from federal programs and states, extensive internal reviews (e.g., review of assumptions), original research using administrative records and survey data, and analysis by the staff of the Joint Committee on Taxation. External considerations involve research organizations, government agencies, and private sector organizations. Models are validated before being used.

Income distribution can be of particular importance in modeling health care because income distribution can at times determine eligibility for health care services such as Medicaid and Exchange subsidies. Income distribution will also likely play an important role in influencing employer decisions to offer coverage after 2014.

Benchmarking or calibrating data is therefore an important best practice, especially because household surveys vary in their ability to measure income accurately. Administrative data based on tax results can provide useful benchmarks but can be limited in access because of confidentiality reasons, missed information from nonfilers, and be tabulated at the tax filing unit level.

CBO uses as a starting point SIPP 2005 data. It then constructs tax filing units from administrative data and benchmarks sources of income and unit-level income against these data. Sources of income are grown over a budget window. The distribution of income across families over time is adjusted based on CBO projections developed through CBO's tax and macro divisions. As a result, CBO goes through a very careful calibration process even before beginning to model.

Dr. Banthin proposed the following best practices for modeling income distribution:

- When possible model builders should consider matching survey data to tax records
- Attention should focus on:
 - Constructing tax-filing units
 - Calibrating the number of units
 - Accounting for major sources of income
 - Supplementing tax records with estimates of nonfilers

Dr. Banthin summarized the day's discussion.

- The meeting began with discussion of when it's appropriate to use microsimulation models.
- Microsimulation models require a large investment of resources and are used when a complex question varies with many interactions and/or possibly nonlinear relationships between key parameters.
- Microsimulation models are also used when distributional impacts really matter and also when questions come back again and again, which justifies the investment of resources to develop such models.

- Over time it's becoming cheaper to build microsimulation models, which means they are more used today than in the past.
- The group also discussed how to evaluate models as well as the pros and cons of supplementing or imputing data.
- There are more Bayesian methods and statistical methods available to microsimulation modelers as well as more/less sophisticated ways for statistical matching.
- Maintaining as much covariance as possible is ideal.
- There was less consensus on how to address uncertainty and also about how to validate models. This can be hard and time consuming, but the group generally agreed that it is important to validate models.
- Various approaches to validation were discussed: sharing or standardizing output statistics and using a reference case or replicating a previous policy to see if a model can match it.
- Funders may not always be concerned with characterizing uncertainty, but modelers think it's important and in the future might find better ways to publicize or illustrate the uncertainty of outputs and predictions. Visualization techniques may help for certain types of outputs.
- The group also discussed the importance of allowing users to choose parameters. It's important to keep in mind that models are designed differently and parameters are not always easy to isolate or convey.
- The group agreed that it's important to document models both in greater detail as well as in a summary format. It's also important to keep documentation current.

DISCUSSION

- Dr. O'Grady said that generally modelers are more interested in issues surrounding confidence/uncertainty than clients.
- Dr. Betson said he didn't believe one should use multiple models for everything. However, there are times when it's helpful to have multiple models, such as when there's true uncertainty.
- Dr. O'Grady asked who is advising the employers and who is doing the modeling for them when it comes to setting premiums [in the future]?
- Dr. Banthin said she saw, but did not attend, a notification of a webinar on how to go through those steps. The preview to the webinar discussed workers at different income levels. The general idea, she thought, was how employers may want to look at their workforce and where they may stand to gain or lose.
- Dr. Glied said there are many people advising employers and they don't all seem to be on the same page, which is probably a good thing.

DEMYSTIFYING MICROSIMULATION

OFFICE OF THE ASSISTANT SECRETARY FOR PLANNING AND EVALUATION
U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
NOVEMBER 16, 2012

Summary of Session

WHEN TO USE MODELS IN HEALTH AND HUMAN SERVICES

Constance F. Citro, Director

Committee on National Statistics, National Research Council/National Academy of Sciences

DEFINING THE QUESTION

This session was charged to address the question of when to use models in health and human services. The session moderator, Constance Citro, redefined the question as one of when to use more or less formal models. She began her argument by listing sample questions that policy makers frequently ask when making decisions:

- How much will a new policy cost next year? In 1 year? 10 years? 75 years?
- Which geographic areas and/or demographic groups and/or organizational players will benefit or not and by how much?
- What will be the effects of a policy change on, say, health outcomes? On, say, measures of educational readiness?

She pointed to the use of the future tense (“will”) in each question set, noting that when policy makers ask future-oriented, “what will,” questions that require numbers, then the policy analysts that are tasked to respond will need to provide estimates. They cannot provide an exact number because it is, by definition, not known for the future, and underneath any kind of estimate, no matter how seemingly straightforward or simplistic, is a *model*. She provided two definitions of models from the International Society for Pharmacoeconomics and Outcome Research (ISPOR/2012):

A model is a mathematical framework representing some aspects of reality at a sufficient level of detail to inform a clinical or policy decision (p. 798).

[Models are] communication tools that allow the complexity of a given system to be reduced to its essential elements (p. 796).

She discussed how even a single number-estimate requires a model, using the example of models developed to answer the question of how much of gross domestic product will go to Medicare costs 1, 10, 25, and 75 years from now. Extrapolating historical costs is a simple linear model that assumes no change in any factors that drive costs; in 2009, extrapolation would have led to an estimated 30 or more percent of GDP spent on Medicare 75 years out. The Congressional Budget Office (CBO) “bent” the cost growth curve to reach a more “reasonable” estimate of around 15 percent, but did not specify what factors would slow growth (Friedman, 2010). More complex models could also be used to provide an estimate.

There is a continuum from *simple to highly complex* models, which does not necessarily match up with the continuum from *transparent to opaque* models—good documentation can reduce the “black boxness” of even a complex model (examples are the framework in Glied, Remler, and Zivin, 2002, and the ISPOR papers; see ISPOR, 2012). Moreover, neither complexity nor transparency necessarily matches up with how useful a model is—a simple model can omit important factors, but a complex model can be needlessly so. In practice, the choice of a particular modeling approach and a particular model is driven heavily by the time and resources available, which in turn leads analysts to consider the available models on a dimension that runs from *ad hoc* (the analyst develops the model on the fly) to *formal* (the model is well developed, maintained over time, and intended for multiple uses).

DEFINITIONS OF TYPES OF MODELS

Dr. Citro indicated types of models that can be termed *ad hoc* or *formal*:

- **Ad hoc models**, often referred to as “back-of-the-envelope” models, can use extrapolation, regression, or cell-based techniques, usually implemented in spreadsheet or statistical software packages.
- **Formal models** can also use extrapolation, regression, or cell-based techniques in spreadsheets or statistical packages. In addition, formal models include static microsimulation models, dynamic microsimulation models, and computable general equilibrium (CGE) models, which generally require complex programming.

Dr. Citro also provided definitions of some model types (see Friedman, 2010; National Research Council, 1991, 1997; Zucchelli, Jones, and Rice, 2012):

- **Microsimulation models (MSM)**—Key to MSM is the use of samples of individual records for people, families, organizations, or other units. MSM may use different techniques to project the data forward to the desired year—*static* microsimulation models project a baseline sample forward for short periods by reweighting; *dynamic* microsimulation models project a baseline sample forward by dynamic aging, such that, e.g., people age 50 become age 60 in year $t+10$ and so on. Some MSM are arithmetic or accounting models; other MSM are behavioral models; most are a mixture. MSM is the most flexible among the available formal modeling techniques.
- **Cell-based models**—These models work with pre-specified groups of people, families, organizations, or other units. They may use static or dynamic transition matrix techniques to project estimates forward. They are generally simpler but less flexible than MSM.
- **Computable general equilibrium (CGE)** —These are macro models of longer run policy effects, which take account of behavioral response and feedback effects. They are quite opaque and are not suited for short-term projections.

All but the simplest models will have components, both for calibrating the baseline sample to match control totals and for estimating the effects of policies. Health care policy changes are particularly difficult to model. Some components of a health care reform model are straightforward accounting (e.g., applying a new tax credit, assuming there is 100 percent take-up of the credit by those eligible). Other components will require their own modeling. For example, Glied et al. (2002) identified four

approaches to modeling the health insurance enrollment decision by a family or individual—elasticity, discrete choice, matrix, or reservation price.

CRITERIA FOR SELECTION

Dr. Citro presented some criteria for choosing between an ad hoc model and a formal model. Assuming, for the moment, that a formal model that has been well thought through has advantages over an ad hoc approach, the criteria boil down to asking whether there is a formal model that:

- Addresses the policy question(s) of interest and the desired time horizon?
- Is available to the agency and contractor staff?
- Has up-to-date data, control totals, and baseline policy parameters?
- Is well parameterized or modularized to support modeling of various components of interest?
- Is well documented, has a good track record, and is trusted by analysts in the field?

If these conditions are not met, then the question becomes:

- Are there time/data/resources to develop a formal model?

Health care reform offers an instructive example. Policy analysts who participated in the modeling for the Clinton health care reform initiative in 1993-94 uniformly noted major difficulties due to the lack of critical data and research-based behavioral parameters (e.g., National Research Council 1997; Bilheimer and Reischauer, 1996). There was no continuing Medical Expenditure Panel Survey, and the Medicare Current Beneficiary Survey had barely begun. Much modeling, perforce, was ad hoc, and modelers used widely different assumptions in key areas, producing very different estimates.

Modelers were in a much better place for estimating the Protection and Affordable Care Act (ACA) of 2010. Much more data were available, as was the availability of formal models. There were still differences among estimates developed by different analysts, but the differences were much less pronounced (see Glied and Tilipman, 2010).

QUESTIONS FOR DISCUSSION

Dr. Citro posed the following questions for discussion by the session participants:

- 1) What criteria should be listed for deciding on when to use models—redefining the question as when to use formal, more complex models?
- 2) What criteria should be listed for deciding on when to use MSM instead of other formal modeling techniques?
- 3) What best practices can and should modelers follow to document models, evaluate models, and help the user understand their differences—i.e., to reduce the “black box”?

FORMAL VERSUS AD HOC MODELS

Views differed on the utility of ad hoc models as opposed to formally developed models. The general consensus was that both approaches may be necessary.

Howard Iams (Demographer) said he could see no circumstance in which an ad hoc model would be worthwhile unless the analyst was desperate for time. Formal modeling requires the analyst to lay out the relationships that are involved, the feedback loops, and the covariance structure. Ad hoc models pay little, if any, attention to the underlying structure. He gave an example of the kind of interaction that could well be overlooked in an ad hoc model of changes to the Social Security program--namely, that crediting women for child care would not necessarily help poor widows, given that they already benefit from their deceased husband's benefit.

Linda Giannarelli (The Urban Institute) agreed that formal models require the analyst to examine the relationships among elements of the policies and behaviors that are being simulated. In contrast, ad hoc models do not force the analyst to spell out all the assumptions, and they also make it easier to forget key underlying relationships.

Julie Lee (Medicare Payment Advisory Commission) countered that, in her experience, there is usually an organic process of going from an ad hoc to a formal model. Typically, an ad hoc model is built to respond to a specific policy question. Then, as more questions are asked about the ad hoc estimates, there is a need to provide additional nuances and granularity that lead to a more formal, systematic model. She concludes that the right tool evolves based on the questions that the policy maker keeps posing as estimates are provided.

Steven Cohen (Agency for Healthcare Research and Quality) recommended going through both an ad hoc and a formal modeling process, if possible. Sometimes one has to show what an estimate would be under a "back-of-the-envelope" approach as well as a more elaborate approach. The back-of-the-envelope calculation is needed so that the policy maker can decide if additional resources to develop or refine a formal model would be a good return on investment.

David Betson (University of Notre Dame) said there is a continuum from really dumb to better to best estimates. One wants to avoid producing a really dumb estimate, but a reasonable back-of-the-envelope estimate can both help in the development of a formal model and also help explain the results. Michael O'Grady (West Health Policy Center) agreed that a back-of-the-envelope estimate is preferable to one or more anecdotes.

Sherry Glied (Columbia University) said there are several reasons why one needs formal models. There are some types of questions, such as distributional questions, that require formal models. She also noted the value of a back-of-the-envelope model to make more transparent the output from a formal model. For example, Medicaid take-up rates have been estimated in the literature at anywhere from 60 percent to 80 percent. When a certain point estimate is used in a formal model, it may be hard for policy makers to know about and question that estimate; in contrast, having a particular estimate (e.g., 80 percent) in a back-of-the-envelope model will likely be obvious to the policy maker, who may well ask about the certainty of the estimate.

Jessica Banthin (Congressional Budget Office) said that her agency maintains formal models of various types, ranging from simple cell-based models to complex MSM. They also have hybrid models. Because they develop baselines and because many questions are repeated, they have the opportunity to periodically update, elaborate, and improve on their models.

Kakoli Roy (Centers for Disease Control and Prevention) said that CDC analysts often start with a back-of-the-envelope estimate. In modeling a disease or health condition, they need to work collaboratively with physicians and epidemiologists to obtain sufficient information on the nature and components of what is being modeled. As they conceptualize a model, they determine if data are available with which to do estimation. If not, then they make simplifying assumptions. Their conceptual work and first-cut estimates often help determine whether to move forward with a formal model. If that is done, then work with the model can be published and the model improved over the years.

Dr. O'Grady said that agencies such as the Office of the Assistant Secretary for Planning and Evaluation (ASPE) need to have a forward looking strategy. For example, there are demographic trends toward an older, frailer population that will peak, in a policy sense, in the not-too-distant future. It will be important to be able to model policy changes that may affect nursing home care and other needs of an aging population. ASPE, which has a foot in both the policy formation and policy analysis worlds, can make a real contribution by asking such questions as: Do we have the right data to build the right formal models for an aging population? What are the extent and rate of forecasted demographic changes that drive policy? What are the cost drivers? Are the right models already in place? When a bill is going to the floor, there is no time, and hurried ad hoc model development will likely lead to mistakes that undermine credibility. A strategy for ensuring that the needed data and formal models are in place is critical.

Dr. Citro agreed that relevant agencies should think along the lines of: "We are never going to have the full panoply of models to cover the waterfront of issues, but here are some key issues (e.g., long-term care) that are going to keep coming back. Perhaps these are areas to invest resources in data and models, so that formal models are available when policy makers ask for estimates."

MSM VERSUS OTHER MODELING APPROACHES

The topic of choice of model type was discussed at greater length in other sessions. In this session, Dr. Betson noted that modeling choices need to reflect the complexity involved in a particular issue. There are three dimensions of complexity: that of our society and the diversity it displays; that of the policy interventions being proposed; and that of the questions the policy makers want answered about their proposals. MSM have an attractive property in that they can answer a series of questions as to which population groups, defined by race/ethnicity, income, or other characteristics, and which geographic areas are winners or losers under one or more policy changes. Policy makers are very interested in this kind of information, which could conceivably be done through a cell-based approach, but for which MSM are very well suited.

BEST DOCUMENTATION, EVALUATION, AND COMMUNICATION PRACTICES

The need for good model documentation, systematic model evaluation, and improved methods for communicating model results and the associated uncertainty was a major theme of this and other sessions.

Dr. Cohen commented on the difficulties in conveying the underlying uncertainty in model-based policy estimates that derives from the host data source, other data that are matched into or used to calibrate the host source, the underlying error from model misspecification, and other sources. He suggested that

a best practice could be to produce the same estimate using a modest assumption, an average assumption, and an unexpected assumption to determine the sensitivity of the model results to changes in assumptions. Having a feedback loop is also essential once an initial estimate is produced so that the model can be improved. He cited with approval the RAND COMPARE model, which uses three different data sources to make projections of how many people will likely be uninsured in a future year (see <http://www.rand.org/health/projects/compare/how-it-works.html>). He hopes that RAND will look back, once the projection year becomes an actual year, to assess the accuracy of the COMPARE projection, what has changed in the interim, and how the model could be improved.

Dr. Citro encouraged ASPE to provide leadership by encouraging the kind of feedback loop outlined by Dr. Cohen. There must be some kind of mechanism or directive and funding for modelers to cycle back as time passes to assess and improve their models. She also wondered if new data visualization techniques could be used to better convey uncertainty in model estimates.

Dr. Lee noted that it is sometimes hard to communicate that certain models are built for a specific purpose or type of question. She agreed with Dr. Cohen that distinguishing the level of uncertainty in model estimates and the general comfort level of the analyst with the results is important, but tends to get lost in communication.

Dr. O'Grady said that there are often differences in how the members of different professions go about modeling (e.g., health economists, actuaries, epidemiologists), although their results can be very similar. To improve communication, it could be very useful to develop common terminologies. Dr. Citro agreed that common nomenclature to facilitate multi-disciplinary efforts would be very useful.

CONCLUDING OBSERVATIONS

The discussion in the session on “When to Use Models in Health and Human Services” made the following points:

- All policy estimates that are derived to answer “what if” questions (e.g., how much will a policy change cost in 10 years) rest on models, which may be simple or highly complex, transparent or opaque, and ad hoc (developed on the fly) or formal (well developed, maintained, and intended for multiple uses).
- Formal models offer many advantages, such as compelling the analyst to work out the likely interactions of a policy change with other policies and characteristics of families or other decision units and providing a platform on which to model a range of policy alternatives. However, ad hoc models are often necessary given time and resource constraints and in new policy areas where formal models have not been developed.
- Ad hoc models used iteratively with formal models can help to communicate results to policy makers and identify where and how important it is to improve a formal model.
- Agencies such as ASPE can usefully assume a leadership role by identifying demographic and other societal trends and their implications for likely needed data sources and the development of formal models to support the production of policy estimates when the need for new policies comes to the fore. An example is the aging of the population and the implications for data and modeling needs for estimating the effects of alternative policies for long-term care.

- Microsimulation models offer the great advantage of being able to identify who will benefit and who will lose under one or another policy change, something that is usually very salient to policy makers. MSM is the most flexible of the major modeling techniques.
- It is essential but difficult to convey the uncertainty associated with model-based estimates to policy makers and how much is due to various sources, such as the underlying data and the model specification. New data visualization techniques, as well as the use of ad hoc models for illustration, may help in this regard.
- It is also essential but difficult to develop feedback loops whereby model results are evaluated for sensitivity to assumptions and by looking back at how well the results matched with the actuality. To close the loop, modelers must be able to improve their models in response to feedback, whether it occurs during the current modeling process or through looking back.
- Agencies such as ASPE can usefully assume a leadership role by providing directives, guidance, and funding to modelers to regularly evaluate and improve their models.

REFERENCES

- Bilheimer, Linda, and Robert Reischauer (1996). Confessions of the estimators: What didn't we know and when will we know it? [including discussion by Len Nichols and Jon Sheils]. Pp. 147-165 in Henry J. Aaron, ed., *The Problem That Won't Go Away*. Washington, DC: The Brookings Institution.
- Eibner, Christine, Federico Girosi, Carter C. Price, Amado Cordova, Peter S. Hussey, Alice Beckman, Elizabeth A. McGlynn, Establishing State Health Insurance Exchanges: Implications for Health Insurance Enrollment, Spending, and Small Businesses, Appendixes A-C. RAND Technical Report TR-825-DOL, Prepared for the U.S. Department of Labor, Santa Monica, CA, 2010.
- Friedman, John N. (2010). Predicting Medicare cost growth. PP. 83-106 in Gooloo S. Wunderlich, rapporteur, *Improving Health Care Cost Projections for the Medicare Population*. Committee on National Statistics, National Research Council. Washington, DC: The National Academies Press.
- Glied, Sherry, Dahlia K. Remler, and Joshua Graff Zivin (2002). Inside the sausage factory: Improving estimates of the effects of health insurance expansion proposals. *The Milbank Quarterly* 80 (4):603-635.
- Glied, Sherry, and Nicholas Tilipman (2010). Simulation modeling of health care policy. *Annual Review of Public Health* 31:439-455.
- ISPOR [International Society for Pharmacoeconomics and Outcomes Research] Task Force (2012). Modeling good research practices—overview: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Value in Health* 15:796-805.
- National Research Council (1997). *Assessing Policies for Retirement Income—Needs for Data, Research, and Models*. Panel on Retirement Income Modeling. Constance F. Citro and Eric A. Hanushek, eds. Committee on National Statistics. Washington, DC: The National Academies Press.
- National Research Council (1991). *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling, Vol. 1, Review and Recommendations*. Panel to Evaluate

Microsimulation Models. Constance F. Citro and Eric A Hanushek, eds. Committee on National Statistics. Washington, DC: The National Academies Press.

Zucchelli, Eugenio, Andrew M. Jones, and Nigel Rice (2012). The evaluation of health policies through microsimulation methods. *International Journal of Microsimulation* 5(1):2-20.

DEMYSTIFYING MICROSIMULATION

OFFICE OF THE ASSISTANT SECRETARY FOR PLANNING AND EVALUATION
U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
NOVEMBER 16, 2012

Summary of Session

HOW TO CHOOSE A MODEL AND MODELING STRATEGY

Jean Marie Abraham, Ph.D.¹

Division of Health Policy and Management, University of Minnesota

INTRODUCTION

Microsimulation models are an important tool for estimating the potential behavioral and economic effects of public policies on decision-making units, including individuals and households, employers, suppliers of health insurance and health care services, and government. Among the broader set of models available to researchers and analysts to inform policy decisions, microsimulation models (MSMs) are distinguished from others by their capacity to analyze the policy impact at the level at which it is intended and to incorporate changing behavioral responses and institutional attributes over the time period being assessed (Citro & Hanushek, 1991; Chollet, 1990). While the following discussion focuses on the use of health-related MSMs to inform policymaking, MSMs have application to a broad set of policy contexts.

There are four basic components of any MSM: (1) the data infrastructure, (2) behavioral assumptions and parameters, (3) statistical methods, and (4) model output. As its foundation, every MSM must have a core data infrastructure. For many health-related MSMs, the core data file is built using at least one survey or administrative databases that contain detailed attributes of individual units within the population. A second component of any MSM is the set of parameters or assumptions pertaining to specific behavioral responses of individuals that would be anticipated as a result of a new policy. Typically, modelers utilize a range of estimates about behavior drawn from the scholarly literature. Modelers then use statistical methods to estimate how changes in behavior due to the policy affect designated outcomes of interest to policymakers. Finally, MSMs produce output summarizing aggregate and in many cases, distributional effects of policy scenarios for a defined population.

Recent improvements in data collection, scholarly research evidence, and computing technology over the past two decades have led to the emergence of a new generation of health-related MSMs, including those focusing on health insurance, medical care spending, and population disease burden. Models have been developed and are being used by Federal agencies (e.g., Congressional Budget Office, U.S. Department of Treasury, Centers for Disease Control and Prevention) and private sector entities. Among

¹ The author would like to thank the Robert Wood Johnson Foundation's State Health Reform Assistance Network Program for supporting related research on a comparative analysis of health policy microsimulation models, which contributed to the content of this document.

the latter, MSMs have been developed by research and policy-based organizations (e.g., Urban Institute, RAND) as well as consulting firms, universities, and individual academicians.

Throughout the policy development and implementation process of the Patient Protection and Affordable Care Act (ACA) of 2010, estimates from several health-related MSMs received considerable attention by policymakers and interested stakeholders seeking objective estimates of the potential effectiveness and economic implications of the legislation. Notably, with the passage of the ACA, many of the major health policy simulation models continue to be used to assess how key provisions may affect specific populations. While potential users or contractors of microsimulation modeling are diverse, there are likely common issues faced by each when choosing a model and modeling strategy. Additionally, as the number and types of applications grow, many experts within the microsimulation modeling community suggest that there may be added value from increased transparency of MSM development and education of potential users regarding the capabilities of existing health-related MSMs. Within this context, the remainder of this report addresses the following questions:

- 1) Who uses health-related MSMs and for what purpose?
- 2) What are the constraints and challenges faced by potential users when making decisions about models and modeling strategy?
- 3) What are some areas for improvement and investment for the development and use of health-related MSMs in future policymaking activities?

WHO USES HEALTH-RELATED MSMs AND FOR WHAT PURPOSE?

Within the Federal government, health-related MSMs are used extensively by the legislative and executive branches during the policy development phase as alternative designs are identified and assessed in terms of their effectiveness and efficiency for addressing the policy objective at hand. For example, in the early development of the ACA provisions, multiple Federal agencies sought ‘unofficial’ estimates of the number and cost per newly insured person that could be achieved under different combinations of public and private insurance expansion strategies. This was done as a way to ‘fine-tune’ policy recommendations as language was being drafted. Of course, once formal legislation was introduced in Congress in 2009, the non-partisan Congressional Budget Office and the Joint Committee on Taxation provided official estimates of cost, revenues, and impact.

While MSMs are extremely useful for evaluating the relative strengths and weaknesses of particular policy designs in the development phase, they are also widely used during implementation. For example, complex legislation may require multiple years to achieve full implementation. In order to have the most current estimates, models may incorporate new information into modeling assumptions, such as updated expectations about overall economic growth. Another rationale for using MSMs during implementation is that the administrative rule-making process often creates more certainty regarding the enactment of specific provisions. In turn, this may affect how modelers parameterize particular policies when estimating behavioral responses.

Health-related MSMs also may be used for state-specific policy development as implementation of federal policy in cases where states are given discretion. Within the ACA legislation, for example, states were given the option of planning and developing their own insurance exchanges through which individuals and small employer groups can purchase coverage. Among states that have chosen to do so,

many decision-makers have considered contracting for MSM services to help them understand the tradeoffs associated with different design features (e.g., pooling individual and small risk pools or leaving them separate). Other state governments have contracted for modeling services to estimate whether additional state investments are needed to address second-order effects of federal reform, such as whether a state has adequate provider capacity to address increased demand for medical care resulting from the coverage expansion.

Health-focused advocacy organizations as well as private organizations within the health sector also utilize MSMs. For advocacy organizations, their use is often for producing materials to influence public opinion and/or the legislative process. For health insurers and health care delivery organizations, output generated by MSMs can provide useful information for developing business strategies that will put them in strong positions given changing market conditions that are the result of a new policy.

WHAT ARE THE CONSTRAINTS AND CHALLENGES FACED BY POTENTIAL USERS WHEN MAKING DECISIONS ABOUT MODELING STRATEGIES?

Several constraints and challenges are faced by potential users of MSMs when making decisions about modeling strategies. Within the Federal government, it is not uncommon for there to be a short window of opportunity during which there is support among lawmakers to advance such initiatives in the policy process. Thus, legislative committees as well as agencies within the executive branch may be under considerable time pressure to obtain model estimates that can inform decisions about policy positions. Expediency is an important factor.

A second criteria/constraint considered by potential users is the reputation of the modeling organization, particularly when considering private-sector MSMs. Because the information created by MSMs can profoundly affect a policy design and its subsequent impact, potential users want to make sure that estimates available for public consumption utilize valid data, current research evidence, and appropriate modeling techniques. Potential users also have expressed the importance of contracting for services from modelers who have evidence of scholarship and/or experience, given the complexities of these models and the difficulty faced by non-technical experts in understanding the specific approaches used.

Financial constraints represent a third challenge. In contrast to the Federal government, states and other stakeholders within the private sector may be more limited with respect to the resources available for MSM contracting. As potential users weigh contracting decisions, it is vital to carefully consider the scope of work, including the number and types of scenarios being put forth as well as the deliverables requested by the users.

In addition to the scope of work, potential users also must consider the issue of “fit.” Many health-related MSMs are developed from national databases and designed to generate national estimates. However, many questions about impact may be state-specific or even local in nature. In the implementation of the ACA, for example, many states have contracted with MSMs to generate state-based estimates. While many MSMs have been able to make adjustments to the data infrastructure to reflect the population attributes of the particular state, there is considerably more hesitation on the part of modelers to generate estimates that are sub-state or for small population sub-groups within a state.

Finally, potential users of MSMs as well as members of the modeling community are challenged by the lack of reconciliation across competing models, that is, why different models produce different results for the same outcome of interest. Through discussion with experts, several possible reasons exist. First, there are distinct differences among models both in terms of the data sources and steps used to construct the analytic file. Another difference reflects how behavioral assumptions are incorporated into the model. While many health-related MSMs use empirical estimates from the scholarly literature, it is well known that extensive variation exists in the amount, quality, and certainty regarding the evidence for particular behavior. In the context of insurance expansion MSM applications, for example, the scholarly literature has much stronger evidence regarding the price-sensitivity of workers in their decision to take up employer-based coverage as compared to estimates of price-sensitivity for persons seeking coverage in the individual market. Finally, comparisons of estimates generated by competing models are made more challenging by the use of different reference periods, whereby one model might report the outcome of interest as though it were fully implemented today while another might report its estimate for the first year during which the policy is expected to be fully implemented.

WHAT ARE SOME AREAS FOR IMPROVEMENT AND INVESTMENT FOR THE DEVELOPMENT AND USE OF HEALTH-RELATED MSMS IN FUTURE POLICYMAKING ACTIVITIES?

Following discussions with MSM experts, several ideas have been proposed to advance the development and use of health-related MSMs in policymaking activities. Among the many ideas proposed, three themes emerged. These included: (1) improvements in the clarity and transparency in the documentation and communication of modeling methods and output; (2) investments in data and measurement; and (3) investments in research to improve certainty about behavioral assumptions used within MSMs.

IMPROVING TRANSPARENCY AND CLARITY IN THE DOCUMENTATION AND COMMUNICATION OF MSM RESULTS

Although MSMs can provide valuable information in the development and implementation of public policies, many potential users and other interested parties have difficulty understanding what is inside the “black box” of most health-related MSMs. Overall access to MSM documentation with details of data development, behavioral assumptions, baseline output, and simulated results has become more widespread in recent years. However, since the development of MSMs can include hundreds of decisions related to building the data infrastructure and specifying the models, it is often impractical to document every detail. Some modelers have noted that publicly-available documentation reflects what is deemed important to the team creating the model. Others have indicated that private-sector modelers may simply wish to not reveal too much information, lest they lose competitive advantage. As one expert noted, this is one reason why government agencies may choose not to rely exclusively on private-sector models and instead build internal capacity. The evolving nature of MSMs presents another challenge. There can be considerable gaps in time between the version of documentation available for a given MSM and the actual version of the model in use to produce estimates for a given simulation. Thus, one suggestion for improvement is to document clearly the version of the model being used and having current documentation that coincides with those estimates.

Beyond understanding any one MSM's "black box," there is the related issue of how to compare across MSMs' "black boxes" to try to identify reasons why estimates might vary so much for the same outcome. For potential users and modelers themselves, the issue at hand is whether the variation in results across models is due to difference in data, assumptions, or something else. One strategy put forth by experts was to encourage organized communications among the modeling community to really understand the "nuts and bolts" of each model. While this may provide valuable "peer-to-peer" learning among modelers, particularly for those within the Federal government, it seems less likely that modelers in the private sector would be as amenable to sharing such detailed information with their competition. Moreover, one expert noted that such activities would yield the largest return if done during the development phase rather than after each model is built and producing output that is being released into the public domain.

A second strategy to facilitate reconciliation among model estimates, as proposed by Glied, Remler, and Graff Zivin (2002), is to build a reference case, which includes a well-documented set of assumptions and parameters utilized in MSM estimation. By using a reference case, policymakers can more easily compare across models and begin to understand which differences in the results are due to behavioral assumptions and which ones may be due to data. While commended as a good idea in theory, some experts have questioned whether or not the value gained from this type of exercise would exceed the necessary time, money, and coordination costs.

INVESTMENTS IN DATA AND MEASUREMENT

Microsimulation models are only as good as the input data from which they can be developed and estimated. Most health-related MSMs utilize Federal government-sponsored population surveys (e.g., Current Population Survey, Medical Expenditure Panel Survey, Survey of Income and Program Participation, American Community Survey), employer surveys, and administrative data to develop the core analytic file. Although progress has been made with respect to the quality and scope of data collection efforts in recent decades, there are still notable limitations with respect to the types of information being collected and modelers' ability to access existing data sources that are not available in the public domain.

One concern that has been raised by the modeling community is the imbalance in investment between cross-sectional and longitudinal data sources. While cross-sectional data can provide researchers with large sample sizes and greater precision for generating distributional effects of policies, longitudinal data have the advantage of modeling both short- and long-run effects of a policy change for the same set of individuals.

Even in cross-sectional data, no one source contains all the necessary demographic, economic, and health-related information for an individual in the population. Thus, modelers typically rely on multiple data sources along with statistical matching and imputation techniques to generate a comprehensive "picture" of individuals. As some modelers have noted, the process of building the data file does not require just one match or imputation, but these methods may be used at multiple points in the analytic file development. While there have been notable advances in statistical methods and software to execute such techniques, some experts have noted that more work should be done to ensure that in the

application of matching and imputation techniques, appropriate covariance structures are maintained between the variable of interest and other key factors in the model.

A third issue recognizes that different types of data sources may be used in MSM development and that each has relative strengths and weaknesses. Many existing health-related MSMs rely primarily on survey-based data. As a result, modelers frequently must deal with both under-reporting and misreporting of key information. Within health policy models, the measurement of coverage provides an important example, whereby large disparities existed between survey-based estimates of Medicaid enrollment and administrative records. With investments by government and private foundations, researchers have been able to analyze administrative enrollment and survey data together to identify potential reasons for these differences and to propose changes to survey methods to improve the measurement of coverage.

Another context in which there may be important gains from utilizing both administrative and survey data together is in the measurement of income and assets. As noted by several MSM experts, administrative income data (e.g., IRS tax filing information) is likely to be more accurate and complete than survey responses. Yet, there are clear tradeoffs in using administrative data. The first is that tax filing units may not necessarily align with the set of individuals who comprise a household or family. To the extent that household or family income is the relevant measure of income for determining eligibility for public programs, this may be problematic. Second, tax filing data often misses non-filers who tend to be lower-income and are often more likely to qualify for many social and health-related programs. Surveys are more likely to capture these individuals. Despite these tradeoffs, some MSM experts have noted that having aggregated tabulations from administrative data sources in which the microdata are prohibited from release could still provide important value. Notably, such tabulations could provide modelers with a more accurate income distribution, enabling them to gauge the magnitude of the reporting error from surveys and to improve upon existing imputation routines for survey data.

INVESTMENTS TO INFORM BEHAVIORAL ASSUMPTIONS

After building the data infrastructure to establish the baseline, modelers are then able to estimate the behavioral response of individuals and households as well as other stakeholders to particular policy provisions. One important way in which models diverge is their approach to incorporating or estimating behavioral responses to the policy provisions of interest. Among health-related MSMs, the scholarly evidence can vary widely both in terms of published estimates and in the overall number of studies upon which to rely. For the former case in which there is extensive uncertainty about behavior, sensitivity analyses can be useful. Sensitivity analyses are conducted by changing a given assumption(s) and re-estimating the model to assess how the outcomes change as a result. Within this approach, MSMs can generate a range or an average estimate under differing assumptions. While the value of sensitivity analyses can provide greater assurance to modelers and analysts, their value to policymakers may be less, as lawmakers often prefer point to interval estimates.

As existing MSMs are refined and new ones are developed, there may be value in identifying and prioritizing areas in which research investments are needed given the lack of evidence. In some instances, the questions that need to be asked are fairly narrow. For example, there is still limited scholarly research on employer decision-making with respect to offering health insurance. However, in

2014 with subsidized Exchange-based insurance for low-income Americans without access to affordable coverage, there is an increased need to understand how workers and employers will respond to this new option. Additionally, as one expert noted, for very innovative policies, there may be almost no scholarly evidence on which to rely. Clearly, the most notable case of this within the ACA legislation is the individual mandate provision.

Health-related MSMs also have been developed to evaluate the potential effects of other types of policies, including those that focus on spending and others on prevention and treatment of disease. One area in which there is considerable need for investment in data and research evidence is the supply-side of health care markets. For example, are there particular types of policies that may be relatively more effective and efficient for increasing capacity in local markets to address demand for medical care, given reform and longer-term population shifts? Within newly created Exchanges, how might insurers respond in ways that affect the price and quality of coverage, given competitive forces? How will hospitals, physicians, and delivery-focused organizations respond to changes in payment systems (e.g., pay-for-performance, bundled payments, and global capitation)?

LOOKING AHEAD

The past two decades have seen tremendous advances in the development and use of health-related microsimulation models for policymaking activities. Looking ahead, there is still much room for identifying ways to facilitate their use and usefulness in policymaking discussions. As the number and types of applications grow, it will be important for the modeling community, government, and other entities that value such tools to make critical investments in data and research to inform model development, particularly in areas that are understudied such as the supply-side of health care markets. In addition, efforts to improve the transparency of MSMs, so as to illuminate what is inside the “black box,” will help potential users as they develop their strategy for using microsimulation models in policy development and implementation.

REFERENCES

- Chollet, Deborah. “Microsimulation Models in Health Services Research.” Employee Benefit Research Institute Center Working Paper No 91-3. Paper prepared for the Committee on National Statistics, National Academy of Sciences/National Research Council, Washington DC: The National Academies Press, October 19, 1990.
- Citro, Constance and E. Hanushek. “Microsimulation models for social welfare programs: An evaluation.” Committee on National Statistics, National Academy of Sciences/National Research Council. Washington, DC: The National Academies Press, 1991.
- Glied, Sherry, D. Remler, J. Graff Zivin. “Inside the Sausage Factory: Improving Estimates of the Effects of Health Insurance Expansion Proposals.” *The Milbank Quarterly*, 2002, 80(4): 603-635.

DEMYSTIFYING MICROSIMULATION

OFFICE OF THE ASSISTANT SECRETARY FOR PLANNING AND EVALUATION
U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
NOVEMBER 16, 2012

Summary of Session

ASSESSING MODEL RESULTS FOR POLICY ANALYSIS

David Betson

College of Arts and Letters, University of Notre Dame

INTRODUCTION

An estimate of the impact of a policy change derived from a microsimulation model, like any other point estimate from other modeling approaches such as econometric modeling, represents what can be called ‘our best informed guess’. While the use of the word ‘guess’ may be objectionable, it does capture in common vernacular the reality that the estimate may deviate from what may actually occur if the policy is implemented. However, the ‘guess’ is not generated by pulling a number out of a hat but is ‘informed’ by empirical evidence collected from survey and administrative data as well as theoretical reasoning. Finally, the modifier ‘best’ is used perhaps more as a statement of a goal for the estimate if not a statement about the actual quality of the estimate. The purpose of this essay is to examine how one approaches thinking about the quality of estimates on the impact of changes in policy derived from microsimulation models.

MICROSIMULATION MODELING

Many governmental policies are directed toward the individual. Microsimulation modeling can prove instrumental in providing estimates of the cost and number of individuals affected by these programs but also the impacts the programs will have on the individuals’ lives. Eligibility as well as the amount of transfers provided by a specific government program will depend upon the characteristics of the individual such as the amount and sources of income, their demographic characteristics, where and with whom they reside, and participation in other governmental programs. The informational demands of the programs are further compounded by the non-linearity of program rules. Aggregate or cell-based approaches are viewed as being unable to adequately capture the complexity of the program rules, trace how the individual is overall affected when changes in one program interact with other government programs, or provide sufficient detail to describe the distributional consequences of these programs. These factors have been the historical justification for the microsimulation approach that bases its modeling efforts at the individual level.

Microsimulation modeling begins with the construction of a baseline – a sample of individuals reflecting their situation given the state of the economy and government policies assuming no change in policy.

This baseline can be either a single year or a series of years.¹ For simplicity, the remaining discussion will focus upon the single year analysis framework. While survey data (for example, a common source is the Current Population Survey (CPS)) is a useful starting point, it is rarely a perfect source of information for a model. While a considerable modeling effort is required to characterize a wide range of program rules, effort will also be directed to modifying survey data to reflect the concepts used by the programs. In some instances, information needed to simulate the program rules will be missing and this data will have to be imputed to the survey. While the data used by the microsimulation model is often a nationally representative sample of individuals and households, the aggregate reported or simulated amounts and participation in government taxes and transfers programs in the baseline may significantly differ from amounts found from administrative sources. The baseline is then ‘calibrated’ or adjusted so that the baseline at the aggregate level and for selected demographic groups receiving benefits reflects the estimates of the program expenditures and participation derived from administrative sources.

To estimate the impact of a change in policy, in essence the program rules embedded in the model are modified to reflect the new set of programs. The model is rerun under the assumption that none of the underlying income or demographic information about the individual changes. It is often claimed that this modeling scenario doesn’t reflect any behavioral responses to the changes in policy. This characterization isn’t entirely true. To the extent that the policy change reflects new programs, expands eligibility or generosity in an existing program whose participation is voluntary, these models will often reflect the individual’s decision to participate or not in the program and consequently the model is reflecting a response to the program. Models that restrict themselves solely to the decision of whether to participate or not in a program are ignoring the potential ‘feedback’ effect of policy on other behaviors which could then affect the estimates of the impact on individuals as well as the aggregate cost of the policy to taxpayers. For example, an important behavior for many transfer programs is the work effort of individuals.

Whether the model reflects the full set of potential behavioral responses to the policy change or not, the intent of this second picture of the nation created by the microsimulation model is to answer the question “what would the nation look like if the proposed policy had been fully implemented?” The estimate of the impact of the policy is constructed by comparing the second constructed picture of the nation with the baseline picture.

CONCEPT OF THE QUALITY OF ESTIMATES

Any estimate, prediction, or forecast of a future event can be expected to deviate from what actually happens. In statistics, this deviation is referred to as an ‘error’. The choice of this term is unfortunate since in common usage an error is associated with the act of making a mistake, a mistake in calculation or coding of the model. While these problems can occur and should be avoided through careful implementation and checking of the model, they aren’t the focus of our present discussion. The

¹ In the past, static and dynamic models would have been used to differentiate those models that examine a single year or the time path across multiple years. However, the more recent debates about the use of models in scoring of congressional bills have appropriated these terms to refer to ignoring the responses of individuals and businesses to changes in the law (static) versus models that attempt to account for these responses (dynamic). To avoid this potential confusion, we will refrain from using these terms.

overarching presumption is that if we possessed perfect knowledge then we could predict events without error. A lack of perfect knowledge requires that we make assumptions or simplifications in the modeling process that result in estimates that are likely to deviate from what happens.

While microsimulation appears to be complex and committed to the detailed modeling of program rules, this modeling approach, just as in other approaches, is forced to adopt assumptions or utilize incomplete data or data that is subject to mis-measurement. The potential sources of error in microsimulation modeling can be grouped in the following categories:

- An incomplete picture of the nation's individuals due to use of a sample
- Lack of appropriate data needed for simulation of program rules
 - Surveys didn't collect appropriate data needed to simulate program rules
 - Collected closely related data (modification of data)
 - Didn't collect data (imputation of data or rules ignored in model)
 - Misreporting of data by survey respondents; and
- Lack of perfect knowledge of how individuals will respond.

Any specific estimate of a policy change requires the modeler to adopt a set of assumptions or procedures to confront each aspect of their imperfect knowledge. Since the appropriate assumption or procedure is not known, adopting a less than 'perfect' assumption will lead to an error in the estimate unless some miraculous event occurs and the errors cancel each other out. These modeling choices are made to reflect what the modelers believe to be the most appropriate approach to account for the lack of perfect and ideal information needed for the prediction. However even a strong belief in what is correct isn't sufficient to guarantee the choices are correct (leading to no estimate error) and there is a likelihood that alternative choices might have led to lesser error. The conclusion that we wish to draw from this discussion is that alternative modeling choices can lead to variation in the estimate error. Our lack of knowledge is likely to be reflected in a significant chance of committing a prediction error as well as an error of significant size. The concept of quality of an estimate should reflect these concerns.

The quality of the estimate can be described in terms of the first two moments of the distribution of estimate errors, the mean and standard deviation, or alternatively in terms of the concepts of bias and precision. If the average error is zero, then the estimate is said to be unbiased. If the mean error is nonzero (positive or negative), then the estimate is denoted as biased and our lack of knowledge can be expected to lead to systematic over or under estimation of the true impact of the policy change.

Bias in this context reflects upon the consequences that the lack of information plays in the estimation process. For example, consider the effect of the use of a given sample in the microsimulation. If the sample is truly a random sample of the national population then we should expect that a different sample would produce a different estimate of the impact of a policy change. From sampling theory, we should expect that across numerous samples, the average error in the estimate would be zero. An unbiased estimate doesn't mean that a specific estimate has zero error but that the factors causing variation in estimates (sampling in this case) shouldn't be expected to create a systematic error in the estimate.

While sampling should lead to unbiased estimates, variability in the estimates due to different samples will exist. A measure of the variability is the standard deviation of the estimate errors or in terms of its

inverse, the precision of the estimation. Sampling theory tells us that the variability in the estimation errors can be reduced if larger samples are used. These gains in precision of the estimates reflect that as the sample size is increased, the effect of alternative samples will have less and less of an effect on the estimates. Estimates of the standard deviation along with the assumption of unbiased estimates can be used to construct ranges for the estimates that reflect the values of the true impact of the policy change that we can state are likely at a given level of statistical confidence. As the precision of the estimates is increased, the range of these values narrows.

While sampling theory is well established, the effect of errors in variables (conceptually inappropriate data, absent data and misreported data) and our lack of knowledge of how individuals respond to policy changes is a less developed literature and there is no general theory of their effects on estimates. We know that income underreporting occurs in current surveys based upon comparisons with alternative estimates of aggregate incomes. If we are interested in the cost of a transfer program directed to low-income individuals, the underreporting of income is likely to lead a systematic overestimate of the cost of an expansion of an existing program. But how to appropriately correct for this underreporting is largely unknown because our knowledge of the determinants of misreporting is limited. While we would expect that any correction would likely lead to a lower estimate of the cost of the program, it is possible that certain methods might overcorrect for underreporting and lead to an underestimate of the cost of the program.

Our lack of knowledge of how individuals may respond to the policy change leads to a similar set of conclusions. If we assume that individuals don't respond then economic theory suggests that this assumption will lead to an underestimate of the cost of expansions in transfer programs.² While any likely response will likely lead to an increase in the program's cost, the modeling of the response could lead to overstating the true response and consequently the program's cost.

Ideally one would like to estimate the entire distribution of model errors to gain an understanding of the quality of any given estimate. While this is difficult exercise, it is possible to implement. Assuming that sampling error is independent of the other sources of modeling error, there would be at a minimum four different dimensions to explore (needed modifications to available data; imputation of missing data; misreporting of data; and behavioral responses). If only two alternatives for each dimension (each dimension would be the approach used in the estimate plus two alternative approaches) were examined, this would require 80 additional 'runs' of the microsimulation model. With today's computers, this form of validity test of the model is feasible. What makes this approach impractical is that it would have to be repeated for each different policy change since it isn't assured that the variation in model estimates is independent of the type or magnitude of the change in policy.

However, even this description understates the complexity of the problem in measuring the quality of model estimates. Take for example the dimension of data imputation. The above characterization

² When transfers become more generous, the amount of income transfer to an individual is increased. Increases in income will most likely have an effect on work resulting in lower work effort and consequently less earned income. If the benefits of the program are based upon earnings of the individual, this behavioral response will result in higher benefits to be paid out. We have limited our discussion to expansions in transfer programs. For reductions in the generosity of transfer programs, the direction of the effect would be the reverse of what is stated for expansions. Tax programs can be expected to have the opposite effects from changes in the transfer programs.

suggests that two alternative approaches be compared to the one that is actually used. Let us assume that a hot deck approach is being used in the estimate; two alternatives (for example, multiple imputation and a multivariate regression model approach) could be used. But all of these methods contain a random element to the imputation and consequently the extent to which the Monte Carlo variation in each separate imputation approach also should be examined.

This discussion assumes that individuals making policy decisions not only care about the point estimate (assumptions used) but also about the quality of the estimates (variation in estimates reflecting choices of assumptions not made). The policy maker is likely to care about the quality of the estimates in a rather asymmetric way. Policy makers will most likely care more about their chances of deciding to go forth with a policy based upon a favorable estimate of the impact of the policy when the policy change turns out to be a failure than they would care about making the error of failing to implement a policy that actually would have been a success but whose estimated impact suggested it would have been a failure. This assertion mirrors the behavioral economist's assumption of an asymmetric loss function and should be reflected in any measure of the quality of the model estimates.

RISK AND UNCERTAINTY

In the 1920s, Knight and Keynes independently made distinctions between the concepts of risk and uncertainty. Situations involving risk assume that the relevant probabilities of a random process are known while uncertainty reflects situations where the probability of the event isn't known. To illustrate the difference between risk and uncertainty, consider an urn with 90 marbles of three different colors. The following facts are known about the urn. There are 30 red marbles but the remaining 60 balls are an unknown combination of black and yellow marbles.³ If the marbles are randomly distributed in the urn and one marble is picked at random (no peeking), we know the probability of picking a red marble is $1/3$ but the probability of a black marble is unknown but can range from 0 (no black marbles) to $2/3$ (60 black marbles).

Now consider a gamble based upon a draw of a marble from the urn. If a red marble is picked, then you win \$100, otherwise you lose \$10. You go to an expert who tells you that $1/3$ of the time you will win \$100 but $2/3$ of the time you will lose \$10. You as the decision maker will need to take this information and decide whether this gamble is better taken than turning it down by taking into account how you value both the \$100 gain and the loss of \$10. This situation is considered risky and is the primary focus of decision-making under imperfect or incomplete information.

Now instead basing the gamble upon the event of picking a red marble, consider a gamble with the same monetary consequences but your winning depends upon picking a black marble. While we are still picking from the same urn, the situation is no longer one of risk but one of uncertainty – the probability of picking a black marble is unknown. Again you go the same expert who tells you that there is some chance that you might never win, while there is a chance you will at most win $2/3$ of the times you play the game. She can't tell you what the true probability of picking a black marble out of the urn. In the absence of any further information, how do you make a decision of what to do with the existing

³ This example is based upon construction that Ellsberg used to show the potential contradictions implied by the expected utility hypothesis.

information? The temptation is to convert this uncertain gamble into a risky one. You go out and find an expert to provide the historical average percentage of black marbles in similar urns and use that estimate as the probability of winning this revised game. Of course this expert doesn't have access to the urn that will be used in the gamble but only to 'similar' urns.

Utilizing the advice provided by one expert carries a good deal of risk depending upon the ability of your expert to correctly estimate the proportion of black marbles in the urn. To hedge your bets, you decide to consult some additional experts for their estimates of the proportion of black marbles. If the additional estimates closely cluster around your first estimate, you will likely gain confidence in your first expert. However, you might realize there is no guarantee that even if the dispersion of estimates is small that you have found the true proportion of black marbles.

But acquiring estimates from different experts may not only be costly, it may be impractical in terms of time available in the policy process. If you were to rely upon one estimate then how do you select which expert to use? The expert's reputation is likely a key factor in the selection. An expert's reputation can be built on the basis of their access to data that others don't have that is viewed as superior to alternative sources of data. This follows the adage; better data makes for better estimates. For example, if one agency has access to tax data that others don't have then logically that agency will have a comparative advantage over others in estimating the costs of tax reforms. One would hope that reputations are earned on the basis of providing what is perceived as accurate as well as timely estimates. Decision-makers may have trust in an expert that they will provide estimates that reflect objective analysis and aren't molded to advance any partisan agenda. Reputations are often slow to lose since all participants in the policy process recognize that the world is truly uncertain and hence errors are likely to occur. But if an expert's track record worsens over time, their reputation will and should suffer. Once an expert's reputation is lost, it will be difficult to recover since fewer people will ask for their advice.

One would think that one's reputation would reflect one's track record of predicting the impact of a policy change. However, establishing one's track record of predicting the impact of changes in policy is a difficult, and in many cases, an impossible task. The model's estimate of the impact of a policy change reflects the difference between what the nation would like if the policy were implemented to what the nation would like if the policy were not implemented. Let us assume that the policy is implemented. Then one could directly observe a picture of the nation given the policy change using survey and administrative data. But where does one acquire a picture of what the nation would look like if the policy hadn't been implemented? Does one use a picture of the nation prior to the implementation? Or does one use the model to construct the counterfactual from the data used to represent the nation given the implementation of the policy change? There truly doesn't seem to be a good solution.

PRACTICAL STRATEGIES FOR REPORTING UPON QUALITY OF MODEL ESTIMATES

Providing confidence bounds that reflect all potential sources of estimation errors is impractical to do for every estimate. Reporting confidence bounds (or standard error of the estimates) that reflect sampling error is a practical first step. While these confidence bounds will understate the true uncertainty in the estimates, they do provide a valuable lower bound.

Modeling of a policy change will necessitate making assumptions. While some of the assumptions can reflect empirical evidence, sometimes an assumption reflects the professional judgment of the modeler. We believe that it is good practice if the modeler not only identifies the more key assumptions but also then provides model estimates under alternative assumptions. This type of sensitivity test will document the robustness of the estimates to these assumptions but doesn't attempt to measure the overall quality of the model's estimates.

Both of these practices will assist both the modeler and policy maker in understanding the quality of data from a limited subset of possible sources of model error. To improve the model, all sources of potential model error need to be examined in a manner that allows estimation of the interactions between various sources of model error. While time consuming to undertake, a strategy of model validation as described in the paper by Citro and Hanushek (1993) should be periodically undertaken. The information produced by such a study will create a clearer picture of the overall quality of the model's estimates as well as the areas of opportunity where improvement to the model's quality can be achieved.

A unique aspect of microsimulation modeling is the calibration of the baseline to administrative data. While there are clearly measurement errors present in administrative sources, they are assumed to be accurate measures of the aggregate amount of benefits paid and caseloads. Modelers rarely report the extent to which they had to calibrate the model in order to meet the control totals. A simple statistic that would capture the extent to which the model has augmented the survey data would be to report the ratio of aggregates from the model prior to calibration relative to the same aggregate after the calibration has been performed. This ratio would be like an R2 statistic reported in regression analysis reflecting the percentage of the aggregate 'explained' by the model.

CONCLUSIONS

Citro and Hanushek (1993) noted that a National Academy of Science Panel on Microsimulation Modeling for Analysis of Social Welfare Programs remarked that there "was the virtual absence of systematic activities to validate models and their results. Hardly ever do estimates of uncertainty or the results of sensitivity analyses accompany the cost estimates that are provided to decision makers." While policy makers may not utilize or understand measures of the quality of the model's estimates, this is not a sufficient reason for not providing them. With rising professional standards and lower computing costs, indicators of the quality of estimates are more routinely being provided. The state of microsimulation modeling today may not be as dire as noted by Citro and Hanushek in the early 1990s.

REFERENCES

Citro and Hanushek (1993) "Estimating the Effects of Proposed Legislation: The Case for Model Validation, or Why are the Numbers so Different?", CHANCE Magazine, 1993.

DEMYSTIFYING MICROSIMULATION

OFFICE OF THE ASSISTANT SECRETARY FOR PLANNING AND EVALUATION
U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
NOVEMBER 16, 2012

Summary of Session **COMMUNICATING RESULTS**

Linda Giannarelli, Senior Fellow
The Urban Institute

INTRODUCTION

The last step of any microsimulation modeling project—after the model has been selected or fine-tuned, the estimates have been produced, and the results have been analyzed by the technical staff—is to effectively communicate the findings to policymakers and other interested analysts. Different types of users need different levels of information about an estimate. Some individuals considering a model result will want the full details about how that result was produced—the complete technical documentation of the model. Others will want a less technical version of the documentation. Policymakers who are using the results of a microsimulation model on a quick-turnaround basis may have time to examine a short version of the documentation, but would be more likely to use a document that abstracts from the documentation the information that is most key for a particular estimate.

The Panel on Demystifying Microsimulation—convened by the Office of the Assistant Secretary for Planning and Evaluation of the Department of Health and Human Services (DHHS/ASPE) on November 16, 2012—devoted one session to the challenges inherent in communicating microsimulation modeling results. This paper summarizes the issues covered in the session, including introductory comments made by the session moderator (Linda Giannarelli) and comments made by other participants. Guidelines for both model documentation and for communicating results to policymakers are proposed, and the potential role of government staff in improving communication about microsimulation modeling is discussed.

CHALLENGES IN COMMUNICATING RESULTS FROM A MICROSIMULATION MODEL

Any presentation of research results must answer the same basic questions: What has the research found? What data and methods produced that finding? What is the level of uncertainty? However, when the research is based on microsimulation modeling, communicating this information becomes more complex.

- *What has the research found?* A single run of a microsimulation model can be analyzed to produce a myriad of different results—overall results, results by demographic subgroups, information on what kinds of individuals became better-off or worse-off, and so on.

- *What data and methods produced that finding?* Datasets for microsimulation analysis may include information that goes beyond the questions asked in a particular survey. Datasets used for microsimulation often include information obtained by statistical matching – which involves matching up survey respondents with similar respondents in another survey that includes a desired variable. Microsimulation datasets may also include variables created by regression or other econometric equations. Modelers may also “age” the data to represent a future year. Thus, even if two different models are both based on the same initial survey data, imputations and adjustments may result in quite different data being used for modeling. Regarding the methods used to generate microsimulation estimates, the models themselves can involve tens of thousands of lines of computer code.
- *What is the level of uncertainty?* Like all other survey-based estimates, microsimulation estimates are affected by sampling variability—the uncertainty in an estimate due to the fact that only a portion of the population is captured by the survey. However, a microsimulation estimate is also affected by all the other details and decisions regarding the data and methods. Thus, the level of uncertainty around a particular estimate cannot be easily described by a single statistic.

As a concrete example, consider one of the program eligibility estimates produced regularly by the TRIM3 microsimulation model for HHS/ASPE. In the average month of 2005, according to the model, there were 5.05 million families eligible for cash assistance from the Temporary Assistance for Needy Families (TANF) program (HHS/ASPE 2008). In brief, that estimate could be explained as the result of applying the TRIM3 model’s TANF simulation to the calendar-year 2005 data from the Current Population Survey’s Annual Social and Economic Supplement (CPS-ASEC), using the policies in place in each state in 2005. The uncertainty in the estimate that is due to sampling variability could be computed, placing a lower bound on total uncertainty. For some users, this level of information would be sufficient, particularly if they already viewed the TRIM3 model as a standard source of TANF eligibility estimates.

However, a user interested in more fully understanding TRIM3’s TANF eligibility estimate might want some or all of the following:

- *Additional information on the findings:* A user might be interested not only in the overall estimate, but also estimates by family structure or other characteristics.
- *Additional information on the input data:* An in-depth understanding of the estimates would also require knowing the ways in which the CPS-ASEC data have been augmented. For example, the TANF eligibility estimate is affected by the imputation of immigrants’ legal status, the allocation of annual income across the months of the year, and the model’s simulation of Supplemental Security Income benefits.
- *More information on the model’s policy data:* Another type of input for TRIM3 and many microsimulation modeling efforts is information about actual policies. Users interested in fully understanding TRIM3’s TANF estimates would want to know the source of the TANF policies used in estimating eligibility, and the fact that the estimate captures state-level variation but not sub-

state variation. Even when a model is used to estimate the effects of a *change* in policies (rather than a statistic related to current policies) the information on current policies is important because it helps to create the “baseline” estimates against which the results of the alternative policies can be compared. Models may differ in the extent to which they use state or sub-state details versus more stylized policies. Establishing the baseline policies may also involve assumptions, for example if there is incomplete information about detailed policies in the year of interest. (In the case of health-related modeling, there is no single source of information about the full range of states’ current Medicaid and Children’s Health Insurance Program eligibility policies; modelers may need to piece together the information they need from different sources.)

- *More information on modeling methods:* TRIM’s TANF modeling can be described briefly as attempting to apply each state’s real-world eligibility determination policies to the families in the input data, one at a time. However, the actual implementation of that intent requires 19,000 lines of code and many dozens of individual choices about how to model policies and their interactions with other policies. For example, TRIM’s TANF eligibility estimate is affected by the assumptions and simplifications made in modeling assets tests and time limits. One set of choices involves whether a model’s baseline case is aligned to some sort of target data, and if so, the particular methods and target data that are used in that alignment. Different groups of modelers make different choices depending on their input data, the amount of time available, and the main purposes of the model. Differences in results between two models may not be due to a single obvious difference in assumptions; instead, a large overall difference could be the result of many individual differences in methods and assumptions.
- *More information about uncertainty:* As mentioned above, a standard error could be computed that treats the TANF eligibility estimate as if it was tabulated from survey-reported data, placing a lower bound on the confidence interval. However, additional uncertainty is produced by every modification to the survey data and every assumption made in modeling the TANF program. A user interested in more information might want some understanding of how the estimate was affected by different assumptions.

Possible approaches for discussing results, communicating methods, and conveying uncertainty are discussed below, followed by comments on roles for both modelers and government staff.

PRESENTING THE FINDINGS OF A SIMULATION

The first challenge for a modeler conveying microsimulation results is determining exactly what results to show. For example, a simulation of a change in the tax system could be analyzed to produce estimates of the aggregate change in tax collections, the numbers of tax units affected, the distribution of affected tax units by the magnitude of the change in tax liability, the ways in which different subgroups of families are affected, the impact of the change on families’ after-tax income, and the impact of the change on family well-being as assessed by the Supplemental Poverty Measure. Obviously, the particular results that are presented depend on the particular purpose of the simulation project. However, several types of information can be considered for inclusion in any report of microsimulation results.

- *Subgroup Estimates*: In terms of the types of results presented, session participants mentioned the importance of subgroup estimates to the extent those are feasible given the limitations of the data and methods. These could include results by state, type of family, or family income level. Subgroup estimates can help users to better understand the estimates (for example, a user may gain confidence by seeing that the pattern of results by a particular characteristic is in line with his/her expectations), and can help users draw additional inferences.
- *Winners and Losers*: To the extent that a particular policy has different impacts on different individuals, the variation in impacts can be important as well as the overall impact. For example, even if the overall impact of a policy is to increase program eligibility, it is important to convey the extent to which the simulation found families who lost eligibility. In general, the unexpected findings—interactions that were not initially anticipated, or individuals who gain benefits when only losers were anticipated—can be among the most important results from a microsimulation exercise.
- *Example Families*: Some complex results may be conveyed more easily through the use of specific examples from the microdata. For example, in the case of a policy that could cause some families to gain eligibility for a benefit while it causes others to lose eligibility, an example of each type of family could be found in the simulation’s input data and presented as part of the aggregate results.
- *“Back of the Envelope” estimates*: One session participant mentioned the possibility of using a “back of the envelope” estimate as part of a discussion of microsimulation results. For example, it might be explained that a simple “rule of thumb” would produce an estimate of X, but that the simple rule fails to account for numerous real-world complexities. The microsimulation result—which attempts to take those real-world complications into account—can then be compared to the simpler estimate.

EXPLAINING THE METHODS: TECHNICAL DOCUMENTATION

A full understanding of a microsimulation result requires understanding the details of the model that produced it. Higher-level descriptions, and descriptions geared to particular sets of model results, can then be based on the underlying technical documentation.

A group of researchers considering the transparency of microsimulation models observed: “The benefits of full technical documentation are obvious; it is the only way readers can understand how the model works.” (Eddy et al, 2012). The National Academy of Sciences’ panel that evaluated microsimulation in the early 1990s also stressed the importance of complete technical documentation with regular updating (Citro and Hanushek 1991). Technical documentation should be complete, up-to-date, and suitable for users wanting different levels of detail.

Considering recommendations in Citro and Hanushek 1991, Eddy et al 2012, and the contributions of participants in the HHS/ASPE meeting, complete written technical documentation should include all of the following:

- Overall information on the model
 - General type of model—static or dynamic¹, near-term or longer-term
 - Funding source
 - Expert panel reviews / validation of the model
 - Types of economic actors included—households, employers, etc.—and methods to develop/add those data
 - Programs and policies modeled
 - Interactions between different components of a model
 - Approach to “aligning” baseline simulations, and sources of alignment targets
 - Types of behavioral changes that may be captured
- The primary micro-level datafile(s) used as input
- Information on any additions or modifications to the data
 - “Aging” of the data
 - Statistical matching to add in variables not present in the underlying survey
 - Other imputations of variables not present in the underlying survey
 - Methods for handling missing data or under-reported data for key variables
- Policy parameters
 - Source(s)
 - Assumptions for handling incomplete policy information
- Methods for modeling each program or policy
 - Eligibility determination for benefit programs
 - Benefit or tax computation
 - For comprehensive models: interactions with other parts of the model
 - For “baseline” simulations:
 - Alignment to targets, if any: methods, source of targets
 - Validation against targets
 - For simulation of alternative policies:
 - Enrollment decisions in the case of alternative policies (In other words, if a person becomes newly eligible for a benefit, will s/he take the benefit?)
 - Other behavioral assumptions, such as labor supply effects (If a policy change affects the net benefit of working, such as through a tax credit or a child care subsidy, will a person change his/her choice of how much to work?)
 - The economic literature and other rationale underlying the assumptions

Regarding the documentation of data modifications and modeling methods, two additional points are important to note. First, each element should be explained in sufficient detail that a technical reader

¹ A “dynamic” model moves a starting-point survey data file into the future by applying a variety of demographic processes—birth, death, marriage, divorce, obtaining or losing a job, and so on. A “static” model either does not try to move a survey’s data into the future, or makes those changes through one-time alterations to the data (such as modifying sampling weights to hit population targets) rather than through year-by-year changes.

could form a judgment about how the model might function for a particular use. For example, saying that a particular piece of information was obtained via a statistical match with another dataset is useful; but without knowing more details about how the match was performed—what variables were taken into account in selecting a donor record—a reader cannot form a judgment about the appropriateness of the matched data for a particular policy application. Taking an example from health modeling, one participant commented that it would be helpful if the documentation of models of the Protection and Affordable Care Act (ACA) of 2010 provided more detail on the creation of synthetic firms. Second, documentation should include references to what is *not* included, or what could have been done better if there had been additional time, funding, or data.

In addition to having as many of the above elements as possible, model documentation also needs to be up-to-date—a perpetual challenge in the case of large microsimulation models that are updated on an ongoing basis. Papers presenting microsimulation results should point users to the version of model documentation that applies to the estimates being discussed.

Finally, model documentation should be available that is appropriate to users wanting different levels of information. This can be accomplished through different stand-alone documents or through an integrated set of on-line products that allow readers to “drill down” from less technical to more technical information.

Users can obtain greater understanding of the model’s methods through access to detailed model parameters and micro-level results. For example, once HHS/ASPE has reviewed and approved baseline simulations from the TRIM3 model (simulations of a particular year’s policies applied to that year’s survey data) the detailed simulation parameters and the micro-level model results are made available on the project’s website, trim3.urban.org. Users can check whether the model’s baseline policy parameters match their own understanding of the rules, and can download and further analyze micro-level results—either to better understand the baseline situation or to provide insights into possible alternatives.

One meeting participant commented that researchers can conceivably share working versions of their models. This is generally feasible only in the case of models in the public domain, and places heavy demands on model documentation.

EXPLAINING THE METHODS: DESCRIPTIONS GEARED TO A PARTICULAR SIMULATION EXERCISE

While a model’s technical documentation is essential, most policymakers who are examining microsimulation results require a different type of document: a summary of the aspects of the data and methods that are most important for the particular estimate being examined. This discussion will generally include:

- Input data:
 - A brief description of the underlying input data, whether it has been “aged”, and how the aging or lack of aging might affect the estimate
 - Key imputations, data edits, or statistical matching

- Information on the baseline simulation:
 - Summary of methods for modeling the current law
 - Sources of parameters for current-law rules
 - Summary of alignment or validation of the baseline results compared to external data
- Information on the policy being modeled:
 - Simplifications or assumptions made in modeling the proposed policy, and how they might affect the results
 - Some information on the degree of uncertainty in the estimates (discussed further below)

An additional item suggested by a meeting participant is possibly the most important:

- The modelers' own assessment of the most important limitations of the analysis, and the most important ways that the analysis could be improved/augmented with additional time or additional data

Meeting participants also noted that it could also be helpful to see tabulations of the population by factors most important for a particular simulation—for example, the number of families tabulated by income group.

Further, when the estimates that are being provided update or correct earlier estimates, the summary will include an explanation of the reasons for the changes.

CONVEYING UNCERTAINTY

Like the estimates from any other type of model, the estimates from a microsimulation model are only approximations of reality. As mentioned above, modelers can compute the *minimum* uncertainty by treating a simulation result as if it was computed directly from a survey datafile. For example, when HHS/ASPE releases TRIM3's state-level estimates of eligibility for federally-funded child care subsidies, a range of estimates is shown to convey how sampling variability affects the precision of the numbers (HHS/ASPE 2012). However, that type of estimate does not capture the impact of data adjustments, modeling assumptions, and simplifications in implementing policies.

Modelers can provide insights into the impact of their assumptions and methods by providing a range of estimates using different key assumptions or alternative methods—for example, different assumptions about take-up rates for a new program, different assumptions about labor supply impacts from an expanded tax credit, or two different approaches to a key imputation. One meeting participant noted that in many cases, there is no right or wrong assumption or method, but it would be useful to know the extent to which the choice of assumption / method is driving differences in results. This type of sensitivity testing cannot possibly cover all combinations of assumptions, but can focus on whatever assumptions the modelers view as most important.

In a situation when the same policy is being simulated with more than one model, a group of modelers might agree to all run their models with a single standardized set of assumptions. Of course, this is possible only to the extent that different models are built in such a way that they can accommodate a different model's assumption or method.

POSSIBLE ROLE OF GOVERNMENT STAFF

The session included some discussion of the ways in which government staff could facilitate improved communication about microsimulation models and their results.

It was suggested by multiple participants that government contracts could specify a standard set of information to be provided about any microsimulation model used as part of a government contract. This could include the data file(s) being used, key imputations being performed, and so on. It could also include specified parameters and a set of descriptive statistics. This information could then be used to build a reference database, with links to more detailed documentation.

Contracts could also include tasks to gain a better understanding of a model's predictive accuracy. For example, a prior year's input data and model code could be used to simulate a policy that was not in place at the time, but that has since been implemented, and the model's predicted changes could be compared to the actual changes.

Participants also discussed the issue of government funding for comparison of models and model results. It was noted that ASPE has funded model comparisons in the past. Another participant agreed that when funds are available, comparisons across models are useful.

CONCLUSION

Communicating results from microsimulation models is challenging—more challenging than communicating results from other types of research because of the wealth of results that can be analyzed, the complexity of the methods, and the resulting difficulty of quantifying uncertainty. Modelers can facilitate communication by making their technical documentation as complete, up-to-date, and accessible as possible. Communication of results to policymakers should highlight unexpected results and convey the impact of key assumptions. When resources are available, comparisons between models can increase our understanding of how different approaches affect simulation results. ASPE could play a role in promoting greater sharing of information about methods and assumptions.

REFERENCES

Eddy, D.M., W. Hollingworth, J. J. Caro, et al. 2012. "Model transparency and validation: A report of the ISPOR-SMDM modeling good research practices task force." *Value Health* 2012; 15:843-50. International Society for Pharmacoeconomics and Outcome Research.

Citro, Constance F., and Eric A. Hanushek, editors. 1991. Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling, Vol. 1, Review and Recommendations. Washington DC: The National Academies Press, Commission on Behavioral and Social Sciences and Education.

HHS/ASPE staff. 2008. "Indicators of Welfare Dependence. Annual Report to Congress 2008." Available on-line, <http://aspe.hhs.gov/hsp/indicators08/index.shtml>.

HHS/ASPE staff. 2012. "Estimates of Child Care Eligibility and Receipt for Fiscal Year 2009." ASPE Issue Brief, available on-line, <http://aspe.hhs.gov/hsp/12/childcareeligibility/ib.cfm>.

DEMYSTIFYING MICROSIMULATION – A MEETING FOR MODEL USERS

NOVEMBER 16, 2012, 8:30 A.M. – 5:00 P.M.

AGENDA

- 8:30 a.m. – 8:45 a.m. Welcome and Introductions
Jim Scanlon, Deputy Assistant Secretary for Science and Data Policy
Office of the Assistant Secretary for Planning and Evaluation, HHS
- 8:45 a.m. – 9:00 a.m. Overview
Sherry Glied, Professor, Department of Health Policy and Management
Mailman School of Public Health, Columbia University
- 9:00 a.m. – 10:00 a.m. When to Use Models in Health and Human Services
Moderator: Connie Citro, Executive Director
Committee on National Statistics, The National Academies
- 10:00 a.m. – 10:15 a.m. Break
- 10:15 a.m. – 12:00 p.m. How to Choose a Model and Modeling Strategy
Moderator: Jean Abraham, Associate Professor, Division of Health Policy and
Management, School of Public Health, University of Minnesota
- 12:00 p.m. – 1:00 p.m. Lunch
- 1:00 p.m. – 2:30 p.m. Assessing and Interpreting Model Results
Moderator: David Betson, Associate Professor for Economics and Public Policy
University of Notre Dame
- 2:30 p.m. – 3:30 p.m. Communicating Results
Moderator: Linda Giannarelli, Senior Staff
The Urban Institute
- 3:30 p.m. – 3:45 p.m. Break
- 3:45 p.m. – 4:15 p.m. Model Performance and Utility in Recent Applications
Richard Kronick, Deputy Assistant Secretary for Health Policy
Office of the Assistant Secretary for Planning and Evaluation, HHS
- 4:15 p.m. – 5:00 p.m. Lessons Learned and Best Practices
Moderator: Jessica Banthin, Senior Advisor
Congressional Budget Office

DEMYSTIFYING MICROSIMULATION

MEETING PANEL MEMBERS

Jean Abraham

Associate Professor

Division of Health Policy and Management
University of Minnesota, School of Public Health

Jessica Banthin

Senior Advisor

Congressional Budget Office

David Betson

Associate Professor

Economics and Public Policy
University of Notre Dame

Sharmila Choudhury

Research Director

Domestic Social Policy Division
Income Security Section
Congressional Research Service

Connie Citro

Director

Committee on National Statistics
The National Academies

Steven Cohen

Director

Center for Financing, Access and Cost Trends
Agency for Healthcare Research and Quality

Michael Collins

Assistant Director for Economics

Government Accountability Office

Linda Giannarelli

Senior Fellow

The Urban Institute

Robert Gillette

Director

Economic Modeling and Computer Division
Office of Tax Analysis

Sherry Glied

Professor

Department of Health Policy and Management
Columbia University

Howard Iams

Senior Research Advisor

Social Security Administration

Susan Jekielek

Social Scientist Research Analyst

Child Care Team
Administration for Children and Families/Office of
Planning Research & Evaluation (OPRE)
Department of Health and Human Services

Richard Kronick

Deputy Assistant Secretary for Health

Office of the Assistant Secretary for Planning
and Evaluation
Department of Health and Human Services

Julie Lee

Senior Policy Analyst

Medicare Payment Advisory Commission (MedPAC)

Charles Nelson

*Assistant Division Chief for Economic
Characteristics*

Bureau of the Census

Mike O'Grady

President

West Health Policy Center

Kakoli Roy

Lead Economist

Policy, Research, Analysis and Development Office
Centers for Disease Control and Prevention

Arloc Sherman

Senior Researcher

Center for Budget and Policy Priorities

DEMYSTIFYING MICROSIMULATION – A MEETING FOR MODEL USERS EXPERT PANEL MEMBER BIOGRAPHICAL SKETCHES

JEAN ABRAHAM

University of Minnesota

Dr. Jean Marie Abraham is an Associate Professor at the University of Minnesota School of Public Health, Division of Health Policy and Management. She holds a Ph.D. in Public Policy and Management from Carnegie Mellon University and a B.A. in Economics and Political Science from the University of Arizona. Jean's research interests include health insurance access and cost; the effectiveness of employer-based wellness programs; health insurance and hospital competition; and consumer awareness and use of provider quality information. Her research has been funded by the U.S. Department of Health and Human Services (AHRQ and ASPE), the U.S. Department of Labor, the Robert Wood Johnson Foundation, and the Commonwealth Fund. For academic year 2008-2009, Jean served as the senior economist for health issues on the President's Council of Economic Advisers in Washington, DC. In this role, she authored two chapters of the Economic Report of the President. Additionally, she had the opportunity to participate on the Obama Administration's work group to develop policy positions relating to federal health care reform, as well as to co-author the 2009 report entitled *The Economic Case for Health Care Reform*.

JESSICA BANTHIN

Congressional Budget Office

Jessica S. Bantthin was appointed Senior Advisor in the Health, Retirement, and Long Term Analysis Division at the Congressional Budget Office in 2011. In this position she directs the development and application of CBO's Health Insurance Simulation model as well as other models used to support estimates of coverage and budgetary impacts due to changes in health policy. Prior to CBO, Dr. Bantthin worked for many years at the Agency for Healthcare Research and Quality (AHRQ) where she directed the Division of Modeling and Simulation. At AHRQ she participated in the design and analysis of the Medical Expenditure Panel Survey, a set of household, provider, and employer surveys that yield nationally representative health expenditure and premium data. Her research has spanned a range of issues including trends in out-of-pocket burdens for health care, take-up of employment-based coverage, adverse selection and plan choice simulation models, Medicaid and CHIP eligibility and take-up, the crowding-out of private insurance by public programs, non-group markets, prescription drug expenditures by therapeutic categories, and the impact of policy reforms on costs, access, and the distribution of out of pocket expenditure burdens on families. Dr. Bantthin received her Ph.D. in Economics from the University of Maryland and her A.B. from Harvard University.

DAVID BETSON

University of Notre Dame

David M. Betson is an Associate Professor of Public Policy and Economics and is associated with the Hesburgh Program in Public Service at the University of Notre Dame. Previously he has been a research associate at the Institute for Research on Poverty at the University of Wisconsin and a staff economist at the Department of Health, Education, and Welfare. He has been a member of seven National Academy of Science panels. In 2004, he was elected a National Associate of the National Academy of Science. His research has dealt with the impact of tax and transfer programs on the economy, poverty and the distribution of income. A particular area of interest is child support policy in which he has written academic papers and consulted with numerous state governments on the development of their child support guidelines.

SHARMILA CHOUDHURY

Congressional Research Service

Sharmila Choudhury is the Research Manager for the Income Security Section in the Domestic Social Policy Division of the Congressional Research Service. The section provides research support to the United States Congress on Social Security, private pensions and savings, Unemployment Insurance, Disability Benefits and other income support programs. Earlier, she was at the Social Security Administration as an economist and Project Manager for the Retirement Research Consortium which funds research to centers at the University of Michigan, Boston University, and the National Bureau of Economic Research. Prior to that, she was an Associate Professor of Economics at the State University of New York-Cortland. Her research interests and publications are on retirement income and wealth, Social Security reform, income sources of the elderly, and labor market earnings. She received a B.A. and M.A. in economics from the University of Calcutta and a Ph.D. in economics from Wayne State University.

CONNIE CITRO

The National Academies

Constance F. Citro is director of the Committee on National Statistics, a position she has held since May 2004. She previously served as acting chief of staff (December 2003-April 2004) and as senior study director (1986-2003). She began her career with CNSTAT in 1984 as study director for the panel that produced *The Bicentennial Census: New Directions for Methodology in 1990*. Dr. Citro received her B.A. in political science from the University of Rochester, and M.A. and Ph.D. in political science from Yale University. Prior to joining CNSTAT, she held positions as vice president of Mathematica Policy Research, Inc., and Data Use and Access Laboratories, Inc. She was an American Statistical Association/National Science Foundation/Census research fellow in 1985-1986, and is a fellow of the American Statistical Association and an elected member of the International Statistical Institute. For CNSTAT, she directed evaluations of the 2000 census, the Survey of Income and Program Participation, microsimulation models for social welfare programs, and the NSF science and engineering personnel data system, in addition to studies on institutional review boards and social science research, estimates of poverty for small geographic areas, data and methods for retirement income modeling, and a new approach for measuring poverty. She co-edited the 2nd–4th editions of *Principles and Practices for a Federal Statistical Agency*, and contributed to studies on measuring racial discrimination, expanding access to research data, the usability of estimates from the American Community Survey, the National Children’s Study research plan, and the Census Bureau’s 2010 census program of experiments and evaluations.

STEVEN COHEN

Agency for Healthcare Research and Quality

Steven B. Cohen, Ph.D., is Director, Center for Financing, Access and Cost Trends at the Agency for Healthcare Research and Quality (AHRQ). Dr. Cohen directs a staff of approximately 50 highly trained and skilled economists, statisticians, social scientists, clinicians and support staff conducting intramural and supporting extramural research on behalf of the Agency. He also directs activities necessary to conduct and support a wide range of studies related to the cost and financing of health care services. Studies include analyses of health care use and expenditures by individuals and families for personal health care services, the sources of payment for health care, the availability and cost of health insurance, and health status, outcomes and satisfaction. Dr. Cohen also leads the Center's administration of surveys and development of large primary data sets, including the Medical Expenditure Panel Survey (MEPS), to support health care policy and behavioral research and analyses. Dr. Cohen has authored over 100 journal articles and publications in the areas of biostatistics, survey research methodology, estimation, survey design and health services research. He is co-author of the text *Methodological Issues for Health Care Surveys*. He has also served as an Associate Professor in the Department of Health Policy and Management at the Johns Hopkins University and the Department of Health Services Administration at the George Washington University. He received his Ph.D. and M.S. in Biostatistics from the University of North Carolina and his B.A. in Mathematics and History, CUNY. He is also a Fellow of the American Statistical Association.

MICHAEL COLLINS

Government Accountability Office

Michael Collins is an Assistant Director Economics at the U.S. Government Accountability Office (GAO). Michael's work at GAO has focused on retirement income security issues, including Social Security reform and pensions. He has also been involved in work looking at other social insurance programs, such as disability, Supplemental Security Income (SSI), Temporary Assistance for Needy Families (TANF), FECA, and Unemployment Insurance. Michael's research interests are in retirement income and aging issues, as well as the interactions between retirement income and other social insurance programs. He earned a Ph. D. in Economics from the University at Albany, State University of New York.

LINDA GIANNARELLI

The Urban Institute

Linda Giannarelli is a Senior Fellow at the Urban Institute whose research experience covers a broad range of federal tax and transfer programs. She is the Project Director for the maintenance and development of the TRIM3 microsimulation model, which produces the government's estimates of how many families are eligible for several key assistance programs, and which supports numerous government and foundation-funded analyses. Recently, she led the development of the Net Income Change Calculator, a web-based tool that displays the changes in taxes and benefits resulting from an increase in earnings for a user-specified family. Earlier in her career, she developed a microsimulation model based on program administrative data and worked with the Urban Institute's DYNASIM model. She has also provided expert support to National Academy of Sciences panels analyzing and using microsimulation modeling. Much of her recent work has used microsimulation modeling to estimate how potential policy changes could reduce poverty, using expanded definitions of poverty based on the Census Bureau's Supplemental Poverty Measure. She co-directed a national-level analysis of how various policies might reduce poverty for the Center for American Progress, and has analyzed anti-poverty policies for four states.

ROBERT GILLETTE

Office of Tax Analysis

Robert Gillette is the Director of Economic Modeling and Computer Applications for the Office of Tax Analysis, U.S. Department of the Treasury. He has over 25 years of experience in microsimulation modeling.

SHERRY GLIED

Columbia University

Sherry Glied is Professor of Health Policy and Management at Columbia University's Mailman School of Public Health. She was department chair from 2002-2009 and has been on faculty since 1989. Dr. Glied served as the Assistant Secretary for Planning and Evaluation at the U.S. Department of Health and Human Services from July 2010 through August 2012. In 1992-1993, she was a senior economist for health care and labor market policy on the President's Council of Economic Advisers under Presidents Bush and Clinton, and participated in the Clinton Health Care Task Force. She has been elected to the Institute of Medicine of the National Academy of Sciences, the National Academy of Social Insurance, and the Board of AcademyHealth and has been a member of the Congressional Budget Office's Panel of Health Advisers.

HOWARD IAMS

Social Security Administration

Howard Iams earned a Ph.D. at the University of Michigan in sociology. After coming to the Social Security Administration (SSA) in 1978, Dr. Iams has worked on a variety of research and evaluation activities connected with demonstrations and survey data analysis on SSA administered programs.

Since 1986 he has been conducting policy evaluations with the Census Bureau's Survey of Income and Program Participation (SIPP) matched to SSA administrative records of earnings and benefits. He designed and developed the Modeling Income in the Near Term (MINT) data system with matched SIPP data using contractor support for model estimations to project the retirement of the baby boomers. He has estimated the distributional impact of Social Security reform proposals and compared the baby boom birth cohort economic status to earlier cohorts in numerous conference papers and published papers.

SUSAN JEKIELEK

Administration on Children and Families, Department of Health and Human Services

Susan M. Jekielek, Ph.D., is a Researcher in the Division of Child and Family Development of the Office of Planning, Research, and Evaluation (OPRE) in the Administration for Children and Families (ACF). In this role, Dr. Jekielek oversees numerous research grants examining child care issues in low-income families, participates in the design and management of the National Survey of Early Care and Education, and develops funding priorities for research that can inform ACF programs. She collaborates across agencies on multiple projects, including the Interagency Forum for Child and Family Statistics. Dr. Jekielek's own research has addressed issues related to family structure and child development, indicators of child well-being, work-family issues, and the measurement of family processes and child well-being in large national datasets.

RICHARD KRONICK

Office of the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services

Richard Kronick, Ph.D., is the Deputy Assistant Secretary for Health Policy in the office of the Assistant Secretary for Policy and Evaluation at HHS. He is also a Professor of Family and Preventive Medicine at the University of California San Diego. Dr. Kronick is a nationally recognized specialist in health care policy. In 1993-94, he was a Senior Health Care Policy Advisor in the Clinton Administration, where he contributed to the development of President Clinton's proposal for health care reform. His articles have appeared in such journals as the *American Journal of Political Science*, *The New England Journal of Medicine*, and the *Journal of the American Medical Association*. He has served as Director of Policy and Reimbursement in the Massachusetts Department of Public Welfare and the Assistant Director in the Massachusetts Office of Health Policy.

JULIE LEE

Medicare Payment Advisory Commission

Julie Lee is a senior policy analyst at the Medicare Payment Advisory Commission. Prior to joining MedPAC, Julie was a principal analyst at the Congressional Budget Office. Previously, she worked as a research director at the Engelberg Center for Health Reform at the Brookings Institution, focusing on payment and delivery system reform issues. She also worked as a management consultant in Cambridge, MA. She received a Ph.D. in economics from Yale.

CHARLES NELSON

Bureau of the Census

Charles Nelson is the Assistant Division Chief for Economic Characteristics in the Social, Economic, and Housing Statistics Division, U.S. Bureau of the Census. In his thirty-five years as a Census Bureau analyst, Mr. Nelson has authored numerous Census Bureau reports and research papers on such subjects as the characteristics of households receiving noncash benefits, after-tax income, pension coverage and retirement income, health insurance coverage, alternative poverty definitions, and the effect of taxes, government transfers, and noncash benefits on the distribution of income and prevalence of poverty.

MICHAEL O'GRADY

West Health Policy Center

Michael J. O'Grady, Ph.D. is President of the West Policy Center in Washington, DC. The Policy Center is part of the West Health Initiative, an independent, one-of-a-kind initiative that provides human and financial resources for the creation and promotion of innovative solutions to lower the cost of health care for the benefit of the public. Our singular mission, charitable resources, and strategic expertise enable West Health to identify and drive changes in health care delivery that lower costs.

Prior to the West Health Policy Center, Dr. O'Grady was a Senior Fellow at the National Opinion Research Center (NORC) at the University of Chicago and Principal of O'Grady Health Policy LLC, a private health consulting firm. His research has concentrated on the interaction between scientific development and health economics, with a particular concentration on diabetes and obesity. He also serves as the Chair of the National Academies of Sciences / Institute of Medicine Panel Measuring Medical Care Risk In Conjunction With The New Supplemental Income Poverty Measure.

He is a veteran health policy expert with 24 years working in Congress and the Department of Health and Human Services (HHS). From 2003 to 2005, Dr. O'Grady was the Assistant Secretary for Planning and Evaluation at HHS, where he directed both the policy development and policy research across the full array of issues confronting the Department. During his tenure as the Assistant Secretary, he increased the quality and rigor of the Department's research and analysis significantly, providing rapid and critical analyses of legislative and regulatory proposals. Prior to his Senate confirmation as the Assistant Secretary, Dr. O'Grady served as the senior health economist on the majority staff of the Joint Economic Committee of the U.S. Congress. Previously, he held senior staff positions at the Senate Finance Committee, the Bipartisan Commission for the Future of Medicare, the Medicare Payment Advisory Commission and the Congressional Research Service at the Library of Congress. He received a Ph.D. in Political Science from the University of Rochester.

KAKOLI ROY

Centers for Disease Control and Prevention, Department of Health and Human Services

Kakoli Roy is the Lead Economist at the Policy Research, Analysis, and Development Office (PRADO) within the Office of the Associate Director for Policy at CDC. In this capacity, she directs the health economics research program and leads a team of economists in developing models and conducting policy analyses to inform efficient resource allocation in public health and prevention, to advise executive staff on evidence-based policies, and to support PRADO's mission in spearheading and coordinating policy work at CDC. Dr. Roy's recent work has focused on evaluating measures of population health, developing methods for measuring and assessing the burden of disease and health inequalities in the United States, integrating labor economics methods to inform public health workforce research, and analyzing health policy issues in low-income countries. She has authored several book chapters and numerous articles in peer-reviewed journals, including the *American Journal of Public Health*, *Journal of Health Economics*, *Journal of Labor Research*, *Health Economics*, *Health Policy*, *Health Services Research*, *Social Science and Medicine*, and *The Lancet*. She is currently leading a research initiative that includes multiple projects and collaborations within CDC and with various other partners to develop and apply microsimulation models in assessing the health, economic, and budgetary impact of recommended or promising public health interventions and policies to prevent multiple chronic diseases and related risk factors, reduce teen pregnancy, and reduce prescription drug overdose. Dr. Roy holds a B.A., M.A., and Ph.D. in economics. Before joining CDC in 2000, she was a Senior Fellow at the Center for Development Research (ZEF), University of Bonn, Germany.

ARLOC SHERMAN

Center for Budget and Policy Priorities

Arloc Sherman joined the Center as Senior Researcher in 2004. His work focuses on family income trends, income support policies, and the causes and consequences of poverty. He is a specialist in the impact of tax and benefit policies on poverty and the effect of public policy on child development. He has written extensively about parental employment and unemployment, welfare reform, barriers to employment, family structure, the depth of poverty, racial inequality, poverty measurement, and the special challenges affecting rural areas. Sherman worked for 14 years as senior research associate at the Children's Defense Fund. His book "Wasting America's Future" was nominated for the 1994 Robert F. Kennedy Book Award.

DEMYSTIFYING MICROSIMULATION- A MEETING FOR MODEL USERS

BACKGROUND PAPERS AND PREPARATORY READING

MAJOR STUDIES OF MODELING

A New Type of Socio-Economic System

Author: Guy H. Orcutt

Source: International Journal of Microsimulation (2007), 1(1), 3-9

http://ima.natsem.canberra.edu.au/IJM/V1_1/IJM_1_1_2.pdf

Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling

Editors: Constance F. Citro and Eric A. Hanushek

Vol. 1: http://www.nap.edu/catalog.php?record_id=1835 Vol. 2: http://www.nap.edu/catalog.php?record_id=1853

Microsimulation Models for Social Welfare Programs: An Evaluation

Authors: Constance F. Citro and Eric A. Hanushek

Source: Poverty Institute, Focus

Note: This is an abbreviated description of the above study.

<http://www.irp.wisc.edu/publications/focus/pdfs/foc153b.pdf>

Modeling Good Research Practices Task Force Reports

Source: International Society for Pharmacoeconomics and Outcome Research

Note: While the focus of these reports is on models to inform medical decisions and health-related resource allocation questions, many of the issues are similar to those addressed by the microsimulation models the Office of the Assistant Secretary for Planning and Evaluation has traditionally used.

Modeling Good Research Practices – Overview

http://www.ispor.org/workpaper/modeling_methods/DRAFT_Modeling_Task_Force_Report_Overview.pdf

Conceptualizing a Model

http://www.ispor.org/workpaper/Modeling_Methods/DRAFT_Modeling-Task-Force_Conceptual-Modeling-Report.pdf

State-Transition Modeling http://www.ispor.org/workpaper/Modeling_Methods/DRAFT_Modeling-Task-Force_State-Transition-Modeling-Report.pdf

Modeling Using Discrete Event Simulation

http://www.ispor.org/workpaper/modeling_methods/DRAFT-Modeling-Task-Force_Discrete-Event-Simulation-Report.pdf

Dynamic Transmission Modeling

http://www.ispor.org/workpaper/modeling_methods/DRAFT-Dynamic-Transmission-Modeling-Report.pdf

Model Parameter Estimation and Uncertainty Analysis

http://www.ispor.org/workpaper/modeling_methods/DRAFT-Modeling-Task-Force_Model-Parameter-Estimation-and-Uncertainty-Report.pdf

Model Transparency and Validation

http://www.ispor.org/workpaper/modeling_methods/DRAFT-Modeling-Task-Force_Validation-and-Transparency-Report.pdf

WHEN TO USE A MODEL AND HOW TO COMPARE AMONG THEM

Dynamic Microsimulation Models for Health Outcomes: A Review

Authors: Carolyn M. Rutter, Alan Zaslavsky, and Eric Feuer

Source: Medical Decision Making (2011), 31(1)

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3404886/>

Inside the Sausage Factory: Improving Estimates of the Effects of Health Insurance Expansion Proposals

Authors: Sherry Glied, Dahlia K. Remler, and Joshua Graff Ziven, Columbia University

Source: The Milbank Quarterly (2002), Vol. 80, No. 4

<http://www.jstor.org/discover/10.2307/3350441?uid=309165991&uid=3739704&uid=2129&uid=2&uid=70&uid=3&uid=67&uid=62&uid=3739256&uid=2457706015&sid=21101099434143>

Simulation Modeling of Health Care Policy

Authors: Sherry Glied and Nicholas Tilipman

Source: Annual Review of Public Health (2010), 31:439–55

<http://www.annualreviews.org/doi/abs/10.1146/annurev.publhealth.012809.103542?journalCode=publhealth>

Note: Fee to Download

The Evaluation of Health Policies through Dynamic Microsimulation Methods

Authors: Eugenio Zucchelli, Andrew M. Jones, and Nigel Rice

Source: International Journal of Microsimulation (2012), 5(1) 2-20

http://www.microsimulation.org/ijm/V5_1/1_IJM_5_1_spring_2012_Zuchelli_Rice_Jones.pdf

Predicting the Effects of the Affordable Care Act: A Comparative Analysis of Health Policy Microsimulation Models

Author: Jean M. Abraham, Ph.D., University of Minnesota

Source: State Health Reform Assistance Network, A Program of the Robert Wood Johnson Foundation (2012)

<http://www.rwjf.org/en/research-publications/find-rwjf-research/2012/03/predicting-the-effects-of-the-affordable-care-act.html>

The Role of Survey Data in Microsimulation Models for Social Policy Analysis

Authors: Alberto Martini and Ugo Trivellato

Source: Labour (1997), 11(1) 83-112

<http://onlinelibrary.wiley.com/doi/10.1111/1467-9914.00030/abstract>

VALIDATION

Estimating the Effects of Proposed Legislation: The Care for Model Validation, Or Why Are the Numbers So Different?

Authors: Constance F. Citro and Eric A. Hanushek

Source: For Submission to CHANCE Magazine (1993)

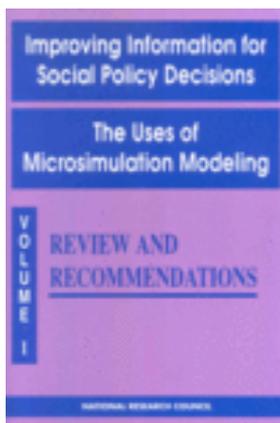
<https://www.dropbox.com/s/awrrh8mvqyos9x2/CHANCEMSM.pdf>

Demystifying Microsimulation—When to Use Models in Health and Human Services

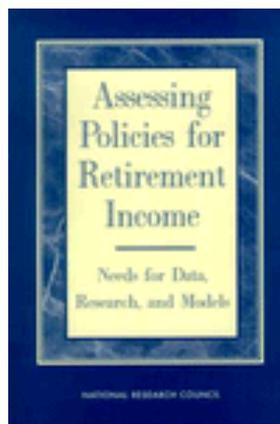
Constance F. Citro, *Director, CNSTAT*
Washington, DC – Nov. 16, 2012

CNSTAT's evaluations of policy models

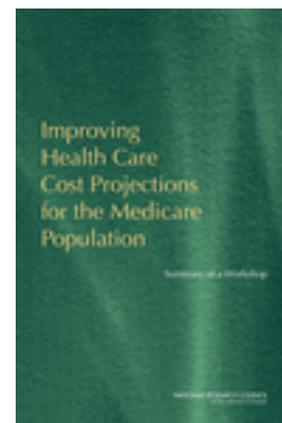
Committee on National Statistics established in 1972 at the National Academy of Sciences to improve statistical methods and information on which policy decisions are based.



1991



1997



2010

Policy maker questions—a sample

How much *will* a new policy cost next year? In 1 year? 10 years? 75 years?

Which geographic areas and/or demographic groups and/or organizational players *will* benefit or not and by how much?

What *will* be the effects of a policy change on measures of health outcomes [think not only insurance, but also treatments]? On measures of educational readiness [think Head Start]?

Underneath every estimate for policy is a model

When policy makers ask future-oriented, “*what will*” questions that require numbers, then—

Policy analysts must provide *estimates* (no one can know for certain the exact numbers), and underneath any estimate is a *model*.

“A model is a mathematical framework representing some aspects of reality at a sufficient level of detail to inform a clinical or policy decision.”

Models are “communication tools that allow the complexity of a given system to be reduced to its essential elements.”

[*ISPOR Task Force Report 1*]

Even “simple” estimates require a (simple?) model

Example—How much of GDP will go to Medicare costs 1, 10, 25, 75 years from now?

- Extrapolating historical costs is a simple model that assumes no change in any factors that drive costs; in 2009, would have led to an est. 30%+ of GDP spent on Medicare 75 years out
- CBO “bent” the cost growth curve to reach a more “reasonable” estimate of ~15%, but did not specify what factors will slow growth

Models are more or less complex and/or “black box”

There is a continuum from *simple to highly complex models*

Does not necessarily match up with continuum from *transparent to opaque*—good documentation can reduce “black boxness”

Neither complexity nor transparency necessarily matches up with how useful a model is—a simple model can omit important factors, but a complex model can be needlessly so

Ad hoc/formal dimension— practical driver of model choice

Ad hoc models (analyst develops on the fly):

- Back-of-the-envelope (or spreadsheet) models

- Extrapolation, regression, cell-based models

Formal (maintained, at least for a time) models:

- Extrapolation, regression models

- Cell-based (spreadsheet) models

- Static microsimulation models

- Dynamic microsimulation models

- Computable general equilibrium (CGE) models

Some definitions—Microsimulation (MSM)

Key to MSM is use of samples of individual records for people, families, organizations

Projection techniques—*static* microsimulation

projects baseline sample forward for short periods

by reweighting; *dynamic* microsimulation projects

baseline sample forward by dynamic aging—e.g.,

people age 50 become age 60 in year $t+10$

Another dimension—arithmetic (accounting) MSM

v. behavioral MSM; most are a mixture

Some other model types

Cell-based models

Work with pre-specified groups of people, families, organizations; may use static or dynamic transition matrix techniques to project estimates forward; generally simpler but less flexible than MSM

Computable general equilibrium (CGE)

Macro models of longer run policy effects taking account of behavioral response and feedback effects; quite opaque and don't do short term

Models within models— Example of health care

All but the simplest models will have components, both for calibrating the baseline sample to match control totals and for estimating effects of policies; health care particularly tough to model

Some components are straightforward accounting—

E.g., applying new tax credit (assuming 100% takeup)

Other components will require their own modeling

E.g., Glied et al. (2002) identified four approaches to modeling the health insurance enrollment decision—elasticity, discrete choice, matrix, reservation price

Ad hoc or formal? Some criteria for choosing

Is there a formal model that:

Addresses the question(s) and time horizon?

Is available to the analyst (and contractor staff)?

Has up-to-date data, control totals, policy parameters? Is well parameterized/modularized?

Is well documented, has a good track record, and is trusted by analysts in the field?

If not, are there time/data/resources to develop a formal model?

Example—Clinton Health Care Reform v. ACA

Clinton health care reform estimation in 1993-94—

Policy analysts uniformly noted major difficulties due to lack of critical data and research-based behavioral parameters (e.g., Bilheimer and Reischauer, 1996)

There was no Medical Expenditure Panel Survey,
Medicare Current Beneficiary Survey barely begun

Much modeling, perforce, was ad hoc, and models used widely different assumptions in key areas

Modelers in much better place for estimating ACA—

Much more data available , more use of formal models,
still differences in estimates but to a lesser extent

Questions for discussion

- (1) What criteria should we list for deciding on when to use models—redefining the question as when to use formal, more complex models?
- (2) What criteria should we list for deciding on when to use MSM instead of other formal modeling techniques?

MSM is definitely a formal, complex tool—but it is less of a “black box” than, say, CGE, and can be made less so with good documentation and interpretation; it is also the most flexible policy simulation tool available.

Questions for discussion - 2

(3) What best practices can/should modelers follow to document models, evaluate models, and help the user understand their differences—i.e., to reduce the “black box”?

The background papers for this conference show progress toward identifying and codifying best practices (e.g., ISPOR work, “reference case” idea of Glied et al., 2002), but there is still much to do.

These three questions overlap with later sessions, but let’s get started with some lists, which we can refine as we go along.

Demystifying Microsimulation

How to Choose a Model and Modeling Strategy

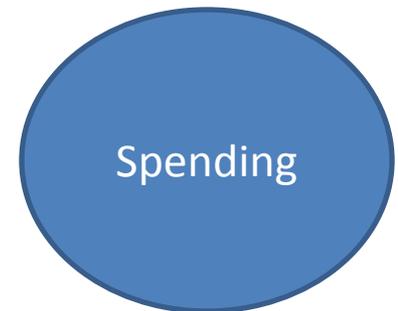
Jean Abraham, Ph.D.
Division of Health Policy & Management
University of Minnesota

November 16, 2012

Acknowledgement: Support for the work was provided by a grant from the Robert Wood Johnson Foundation's State Health Reform Assistance Network Program.

What is a microsimulation model?

- Tool for estimating potential behavioral and economic effects of public policies on decision-making units, including
 - Individuals and Households
 - Employers
 - Government
 - Federal
 - State



Elements of a Microsimulation Model



Age Group	Gender	Population	Income	Health Status	Employment	Insurance
0-14	Male	1,234,567	\$12,345	Good	0%	0%
0-14	Female	1,234,567	\$12,345	Good	0%	0%
15-24	Male	1,234,567	\$12,345	Good	10%	0%
15-24	Female	1,234,567	\$12,345	Good	10%	0%
25-34	Male	1,234,567	\$12,345	Good	20%	0%
25-34	Female	1,234,567	\$12,345	Good	20%	0%
35-44	Male	1,234,567	\$12,345	Good	30%	0%
35-44	Female	1,234,567	\$12,345	Good	30%	0%
45-54	Male	1,234,567	\$12,345	Good	40%	0%
45-54	Female	1,234,567	\$12,345	Good	40%	0%
55-64	Male	1,234,567	\$12,345	Good	50%	0%
55-64	Female	1,234,567	\$12,345	Good	50%	0%
65-74	Male	1,234,567	\$12,345	Good	60%	0%
65-74	Female	1,234,567	\$12,345	Good	60%	0%
75-84	Male	1,234,567	\$12,345	Good	70%	0%
75-84	Female	1,234,567	\$12,345	Good	70%	0%
85-94	Male	1,234,567	\$12,345	Good	80%	0%
85-94	Female	1,234,567	\$12,345	Good	80%	0%
95-104	Male	1,234,567	\$12,345	Good	90%	0%
95-104	Female	1,234,567	\$12,345	Good	90%	0%

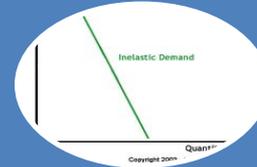
Core Data

- Population Attributes
- Income
- Health status
- Employment
- Insurance



Assumptions or Parameters

- Price-sensitivity
- Preferences re: options



Methods

- Elasticity-based
- Utility-based



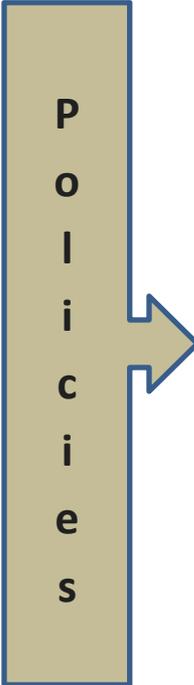
Outcomes

- Primary and secondary outcomes
- Distributional effects

Dynamic versus Static Timeframe

Modeling the ACA Coverage Expansion Provisions: Comparison of CBO, GMSIM, COMPARE (RAND), HIPSM (URBAN), and LEWIN (HBSM)

Population



Behavior

Individuals:

- Take-up of employer-sponsored insurance (ESI) and Non-group
- Public insurance enrollment given eligibility

Employers:

- Offer health insurance
- Premiums and plan generosity (employer and employee share)

Outcomes

Individuals:

- Insurance coverage distribution (ESI, Medicaid/CHIP, non-group, uninsured) overall and by other attributes

Employers:

- % offering health insurance
- ESI premiums

Government:

- Public program participation
- Public program spending

Choosing a Model and Strategy

Perspective: Who is the potential user of the information generated by the model?

Purpose: What is the question that needs to be answered?

Constraints & Challenges: Are there constraints or challenges that move a user to select a certain model/strategy?

Perspective

Federal Government

- White House
- Exec Branch Agencies
- Congress

State Government

- Human Services
- Health
- Commerce

Advocacy Organizations

Private Sector

- Insurers
- Providers
- Pharmaceuticals
- Medical Device

Purpose

Policy Development

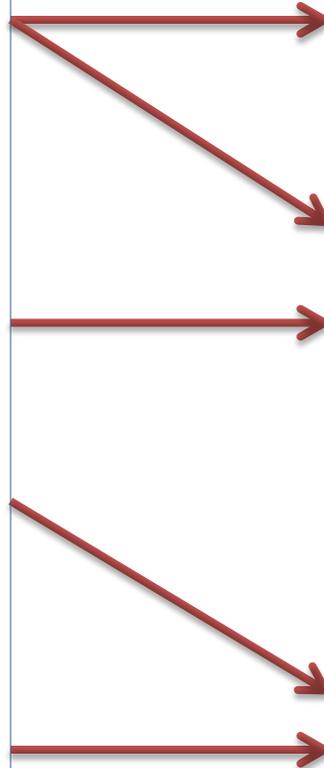
- Budgetary impact
- Cost-effectiveness

Implementation

- Estimate behavioral responses of stakeholders previously not considered in model
- Estimate potential implications of decisions to be made by other parties (e.g., States)

Strategy

Develop strategies to pursue objectives, given new market/policy environment.



Constraints and Challenges

- Time
 - How quickly does the potential user need the information?
- Money
 - How much does the user have available to spend on simulating the policy?
 - Initial cost and incremental costs of additional scenarios
 - Example: State governments
- Reputation
 - Established
 - “Trustworthy”
 - Responsive
 - Objective
- Fit
 - Can the model effectively answer the question being asked by the user?
 - Data
 - Assumptions

Strategy & Selection: Example #1

- Perspective: Federal Government (Internal)
- Purpose: Preliminary development of policy parameters (e.g., subsidy schedule, actuarial value, mandate (y/n), etc) to guide recommendations
- Constraints
 - Time
 - Reputation
- Result
 - GMSIM

Strategy & Selection: Example #2

- Perspective: State government
- Purpose: Implementation phase; Issues of health care labor supply and distribution for newly insured population
- Constraints
 - Budgetary
 - Model to estimate newly covered didn't differentiate between primary and specialty care demand
- Result
 - Used one of the major microsimulation models to estimate newly covered within regions of the state
 - Used a “niche” model to estimate number of additional visits by newly insured along with supply projections to identify potential gaps.

Questions and Issues to Consider

From the user’s perspective, evaluating different models’ options and output is very “apples to oranges.” Here is an example of answers to the question, “What if there is no individual mandate?”

	CBO	GMSIM (Gruber)	COMPARE (RAND)	HBSM (LEWIN)	HIPSM (URBAN)
Decrease in number of newly insured relative to full ACA implementation	16.0 million	24.0 million	12.5 million	7.8 million	13.4 million
Change in average Exchange premium	15-20%	27%	9.3%	12.6%	10%
Behavioral Response	Elasticity	Elasticity	Utility	Utility	Utility
Assumptions about timing	2021	2019	2016	2011	2011

Questions & Issues to Consider: Comparing Models and Results

- Several simulation models developed by private organizations and individuals exist. Each is uniquely defined by its attributes, including data sources, construction of synthetic firms, behavioral assumptions, and approaches.
- Questions
 - What specific information is needed by potential users in order to compare models?
 - Is that information available in the public domain and is it up to date?
 - How frequently are changes made to model data, assumptions, or methods relative to what is documented?
 - Is there any way to narrow down the key drivers of differences in the output produced by different models? Is it a reporting phenomenon? Differences in data? Assumptions? Approach?

Questions & Issues to Consider: Comparing Models and Results

- What validity checks are in place?
 - Are there any examples of pre-ACA policies that were implemented and could/can be used to gauge how well/poorly a microsimulation model performed?
 - What types of sensitivity analyses or checks are done?
- How time-sensitive are models (e.g., macroeconomic assumptions that affect longer-range estimates)?
- What value do existing models that focus on coverage and costs have after 2014? How will these models be adapted for the longer-term?

Questions to Consider:

Strategy and Selection – State Perspective

- If a state agency was getting ready to contract for microsimulation services to estimate insurance and/or medical care demand-related outcome(s) given the ACA, what are some issues to be considered to make an informed choice?
- Questions
 - In what ways can models that are national by design reflect a particular state’s population and preferences?
 - Data
 - Assumptions
 - What is the lowest level of geography that microsimulation models can estimate outcomes and still be “valid”?
 - To what extent can models incorporate choices by states about Exchange functions or Medicaid eligibility thresholds that would affect premiums and coverage decisions?
 - How easily can certain assumptions be relaxed in order to gauge their importance for a given state?
 - To what extent can models generate information about distributional effects (e.g., attributes of the newly insured)?

Questions to Consider: Strategy and Selection Going Forward

- Are there major evidence gaps regarding our understanding of the potential behavioral responses of the following groups to ACA provisions?
 - Uninsured population
 - Employers
 - Insurers
 - Providers
 - State governments re: Medicaid expansion and Exchanges
- How much do we know about modeling policy interactions?
- Are there additional data investments that ought to be considered to improve the capabilities of microsimulation models?
- Are administrative data sources being used to adjust baseline estimates and/or assumptions will they be used in the future?
- Is there interest or discussion in “broadening” microsimulation models to account for different, but related, outcomes and/or populations?
 - Employment and income responses to coverage policies
 - Non-elderly and Medicare populations

Assessing and Interpreting Model Results

*ASPE Workshop on Microsimulation Modeling
November 16, 2012*

*David M. Betson
University of Notre Dame*

Primary Sources

- *Citro and Hanushek (1993) “Estimating the Effects of Proposed Legislation: The Case for Model Validation, or Why are the Numbers so Different?”*
- *Briggs et al (2012) “Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force – 6”*

Brief Outline

- *A brief history of attempts at quantifying the quality of model estimates*
- *Thinking about sources of errors: a regression model analogy*
- *General Strategies for Building Confidence in Models*
- *Suggested Questions for Discussion*

State of affairs in 1993, has it changed?

“What troubled a panel of the Committee on National Statistics (CNSTAT) that conducted a review of the social welfare policy models was the virtual absence of systematic activities to validate models and their results. Hardly ever do estimates of uncertainty or the results of sensitivity analyses accompany the cost estimates that are provided to decisions makers.”

Citro and Hanushek (1993)

I think it has!

- *A good deal of effort and energy has been placed initially into getting estimates and now the focus has been directed at attempting to quantify how good are the estimates derived from the models*
- *It is more common today to see standard errors of estimates being provided or sensitivity tests of assumptions*
- *Why?*
 - *Increased expectations of users and modelers reflecting increased professional standards*
 - *Increases in computing power that have made empirical techniques such as bootstrapping and empirical Bayesian approaches possible*

Quality of Estimates

Ones “Best Guess” can deviate from “Truth” and these deviations represent an “Error”

- *Bias – Systematic Errors in Estimates*
- *‘Variability’ – Random Errors in Estimates*

$$Z_i = \alpha + \beta X_i + \delta Y_i + \varepsilon_i$$

Goal: Obtain estimate of sum of Z_i

Structure of Model: Nonlinearity of the relationship between Z and causal factors (X , Y and ε) will an important factor in determining whether sampling variability and parameter uncertainty will create bias

Parameter Uncertainty – don't know α , β , and δ but have estimates or can make assumptions about their values. This uncertainty can lead certainty to variability in the estimates but may also lead to bias.

Measured Heterogeneity – X_i and Y_i but only X_i is in data but may be measured with error. Y_i is directly observed and needs to be imputed. Imputation errors (bias and not maintaining the appropriate correlation with X_i) can create errors in the estimate leading to bias and variability in the estimates

Unobserved Heterogeneity – assumption about the distribution of ε

Additional Challenges in Modeling

- *Modeling of Behavioral Responses (Static versus Dynamic)*
 - *How will individuals respond to choices they have not confronted before? Will the initial response be different than the 'long run' response?*
 - *The effect aggregate constraints on individual choice? Or the indirect effect of policy on other causal factors (feedback effects)*
- *Data for the year you want to predict is not available so the data has to be altered to reflect the future year (aging of data)*
- *Alternative modeling approaches are often not nested*

Strategies to Provide Confidence

- *Model Validation*
 - *Using one source of data how well does the model actually predict events captured in another data set?*
 - *Accuracy of Forecasts*
 - *Reconciling the differences is helpful in identifying bias in model's point estimates or systematic errors in the estimates*
- *Sensitivity Tests of Assumptions or Alternative Modeling Approaches – Looking for Robustness of Estimates*
- *With each set of estimates provide estimates of the possible extent of variability in estimates – Confidence Bounds for Estimates*

Possible Discussion Questions

- *Are there additional examples of attempts to validate model estimates or provide estimates of variability?*
- *Is there demand by policy makers not only for point estimates but also for estimates of the quality of the estimates?*
- *What can be done to 'educate' policy makers about the importance of quantifying quality? Will supply create its demand?*
- *If measures of quality of estimates are provided to policy makers, how will they be used? Or is it more likely that they will be misused?*

Communicating Results

ASPE Workshop on Demystifying Microsimulation

November 16, 2012

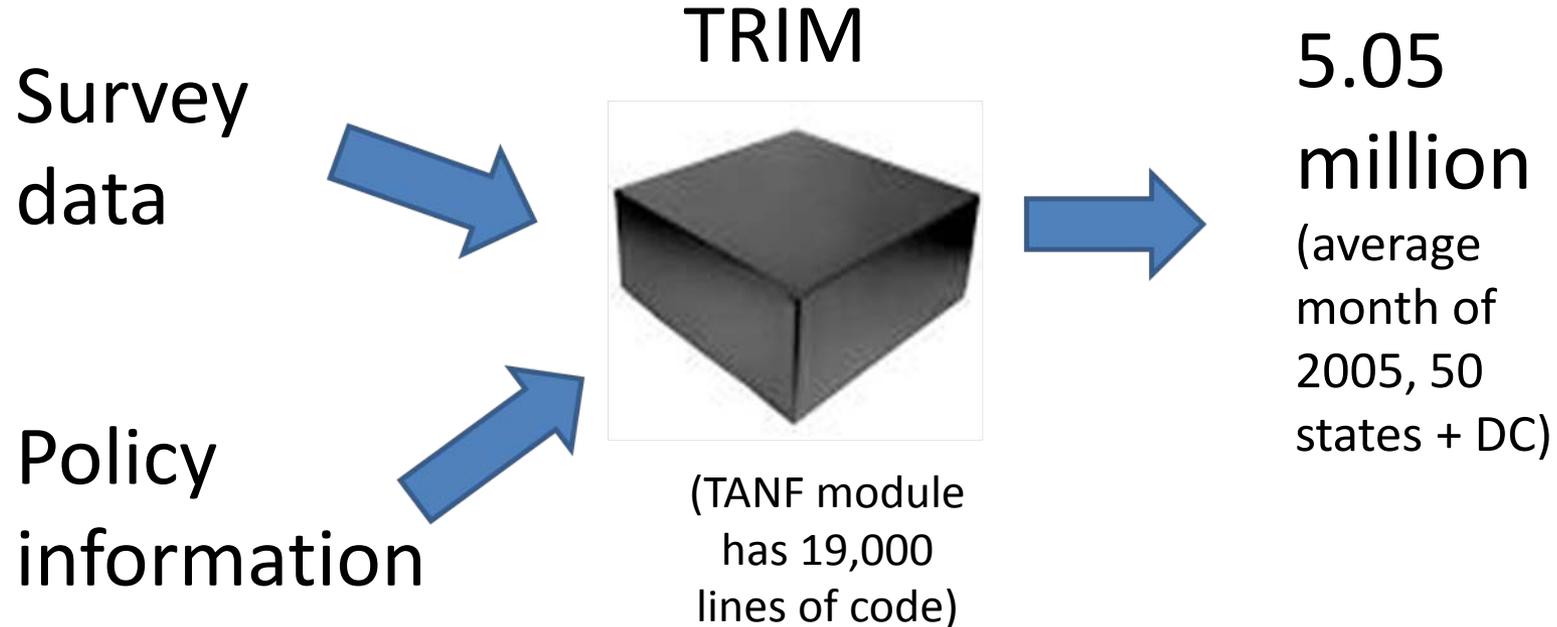
Linda Giannarelli, Urban Institute

Communicating Research Results:

- What did we find?
- What data and methods gave us that finding?
- How sure are we about the result?

Communicating this information can be harder with microsimulation than with some other types of research.

Example: How many families are eligible for TANF?



TANF eligibility estimate: Input data

- CPS-ASEC data for CY 2005
 - Reported by respondents
 - Imputed/edited by the Census Bureau
- Imputed immigrant status information
- Monthly income amounts derived by allocating annual amounts across the year
- TRIM-simulated SSI receipt (corrected for under-reporting)

TANF eligibility estimate: Methods

- Detailed state-by-state modeling
 - Policy rules from the Welfare Rules Database
 - Captures applicant vs. recipient rule differences
 - Simulated participation in one month affects eligibility in next month
- Some policies not modeled
 - Ineligibility due to value of vehicles
 - Loss of eligibility due to full-family sanctions

Input Data for Microsimulation

- Starting-point data
 - Any “aging” of the data?
- Additions to the data
 - Statistical matching
 - Imputation equations
 - Rule-based edits/allocations
 - Simulated data from an earlier step

Policy Parameters for Microsimulation

- For the “baseline” case
 - Types of parameters
 - Rules: eligibility, benefit or tax computation
 - Caseload: size, characteristics
 - Source or sources?
 - Is year of policy data same as year of the input data?
 - State-level variation?

Methods: Big Picture

- Single-purpose model or comprehensive?
- Solely household-level or also other sectors?

For modeling of policy changes:

- Does the model estimate the near-term impacts of a change or longer-run impacts?
- Are behavioral impacts captured?
- Is the “baseline” case aligned to control totals?

Methods: Details

- Step by step, how is the program / policy modeled?
 - Eligibility determination?
 - Benefit or tax computation?
 - Enrollment decision?
- What simplifications are made?
- What assumptions are made?
- For comprehensive models: What methods in other part of the model affect this estimate?

Uncertainty

- Sampling variability
- Impact of imputations
- Impact of assumptions

- Issue of range of estimates vs. point estimates

Quick Turnaround Work

- By definition, no time for a full report
- Requires a short, focused report of results plus key methods

Challenge: Choosing the most important information to convey, depending on the research question

When the Estimates Change...

- For most models, methods are periodically improved
- Changes improve the current estimate but can disrupt a time series

Challenge: Conveying the reasons for the change

Recommendation of the Task Force report on Transparency and Validation

(Eddy et al)

Every model should have:

- Non-technical documentation
- Technical documentation: “The benefits of full technical documentation are obvious; it is the only way readers can understand how the model works.”

Possible goals for technical documentation

- Publicly accessible, on-line
- Documentation of additions to the data
- Description of methods appropriate to different users
 - Ability to “drill down” to finer level of detail
- Availability of parameters
- Availability of “baseline” results

Some Questions

- What should people know about models in general that you don't think they currently know?
- Should there be improvements in:
 - Content of documentation?
 - Accessibility of documentation?
 - Comparability of documentation across models?
- When is a range of estimates helpful?
- How much information should be provided on the impact of changes in methods?

An Example of Best Practices in Modeling: The Importance of the Income Distribution

**Presentation for “Demystifying Microsimulation”
U.S. Department of Health and Human Services**

**Jessica S. Banthin, PhD
Senior Advisor
Health, Retirement, and Long-Term Analysis Division
Congressional Budget Office**

Outline

- CBO's work
- Importance of the income distribution, using the Affordable Care Act as an example

CBO's Role

- Provide objective, non-partisan, timely analyses to facilitate economic and budgetary decisions by the Congress
- Make no policy recommendations

Types of CBO Projections

- Federal spending and revenues under current law
 - Federal spending and revenues under current policies
- Effects on the federal budget of legislation under consideration
- Economic and budgetary effects of policy options

Time Horizons for CBO Projections

- 10-year budget window for baseline projections and formal cost estimates (sometimes with an indication of effects on the deficit in subsequent decades)
- Longer-term projections for the whole budget with more detailed projections for some population-based programs: Social Security, health care programs

CBO's Use of Models

- Enables CBO to use existing evidence to make future projections
- Facilitates consistency and replicability of estimates over time while enabling timely responses to requests for estimates
- Incorporate behavioral responses (if feasible)
 - Households
 - Businesses
 - Federal agencies
 - State and local governments

Types of Models Used in CBO's Estimates and Analyses

- Cell-based models using spreadsheets
- Regression models
- Microsimulation models
 - Health Insurance Simulation Model (HISIM)
 - CBO Long-Term Model (CBOLT)
- Combinations of the above

Model Construction and Review

Inputs

- Reviews of research literature
- Historical data from federal programs and states
- Original research using administrative records and survey data
- Analysis by the staff of the Joint Committee on Taxation
- Extensive internal review

External Consultations

- Research organizations
- Government agencies (federal, state, and local)
- Private-sector organizations and associations
- Subject matter experts (academia, private sector, and government)
- CBO's Panels of Economic Advisers and Health Advisers

Importance of the Income Distribution

- For estimates of the impact of the ACA, the projected income distribution determines what percent of the population will be eligible for Medicaid and for exchange subsidies
- The income distribution also plays an important role influencing employers' decisions to offer coverage after 2014

Importance of the Income Distribution

- Employers will weigh the value of the tax exclusion for employment-based coverage against the value of Medicaid, CHIP benefits, and exchange subsidies to all of their workers
- These subsidies vary significantly across income levels
- Modeling this decision depends on getting the workers' income distributions right – both within and between firms

The Income Distribution – Measurement Challenges

Challenges:

- Household surveys vary in ability to measure income accurately
- Administrative data based on tax returns provide useful benchmarks, but
 - Tabulated at tax-filing unit level
 - Misses information from non-filers
 - Limited access due to confidentiality

The Income Distribution – Measurement Challenges

Under the ACA, modified adjusted gross income (MAGI) has become the standard income definition

- Models need to define tax-filing units, which are different from household, family, and health insurance units
- Models require information about various sources of income to calculate MAGI

The Income Distribution- CBO's Calibration

Calibrating income in CBO's HISIM:

- SIPP (2005) data is starting point
- Tax-filing units constructed
- Sources of income benchmarked against administrative data
- Sources of income grown over budget window and income distribution adjusted according to CBO projections

The Income Distribution & Employer Offer Decisions

Will employers continue to offer insurance after 2014?

- CBO projects that 3 million to 5 million fewer people, on net, will obtain ESI coverage each year from 2019 to 2022.
- Why not a larger reduction in ESI coverage?
- Our projections of the income distribution show that large tax subsidies for employer coverage will remain for middle- and high- income workers.

Best Practices for Modeling the Income Distribution

When possible model builders should consider matching survey data to tax records

Attention should focus on:

- Constructing tax-filing units
- Calibrating number of units
- Accounting for major sources of income
- Supplementing tax records with estimates of non-filers