Mt Hood III Report

Diabetes has many complications that take decades to develop. Scores of preventive treatments are available. Yet most traditional randomized controlled trials evaluate just one or two treatments, for a relatively brief time. Patients vary widely in their risks and histories of complications. No clinician can calculate the long-term relative risk, benefits, and costs of each and every treatment combination, for every individual patient. How, then, can we choose among the many treatment options patients face?

Computer simulation is gaining prominence as the preferred, and perhaps the only valid method to harmonize and personalize the burgeoning mass of evidence about the long-term effects of diabetes and its treatments. Simulation harnesses the power of digital computers to combine data from many studies and to answer what-if questions about treatment effects. More and more models and model-derived predictions are appearing. Some health systems and governments already use models to set priorities and make regulatory decisions for populations with diabetes. Soon, models will be used to provide personalized advice to clinicians and patients in the normal course of clinical care, and on the world-wide web.

To avoid doing harming rather than helping, these new diabetes models must be accurate and valid. The Mount Hood Challenge meetings were organized to promote validity and reliability in diabetes modeling. Mount Hood I, held in August, 2000 compared just two models.[ref] Six models competed in 2002 at the Mount Hood Challenge II in San Francisco. At Mount Hood III, hosted by Oxford University on August 29-31, 2003, 12 models presented data and 10 models formally competed, presenting predictions about future complications, costs, and lifespan for five pre-specified patients and five standardized treatments. Dr. Richard Stevens of the Oxford Diabetes Trials Unit organized and chaired the meeting.

At Oxford, the field was divided into “clinical risk models,” which made predictions about a limited number of cardiovascular endpoints, one at a time, and “health economics models” that presented data on the joint occurrence of a wide range of endpoints, plus survival, quality of life, and cost. The three models that competed under the clinical risk model banner were the Archimedes Model, a complex biologically-based discrete event simulator developed by the Care Management Institute of the Kaiser Permanente Medical Care Program headquartered in Oakland, USA;[refs] the statistical risk-prediction models estimated by the US Framingham Heart Study;[refs] and the statistical models disseminated by the United Kingdom Prospective Diabetes Study as the UKPDS Risk Engine.[refs] The seven health economics competitors were the Cardiff Model (Cardiff Research Consortium, Wales, UK); [refs] the CORE Diabetes Model (Centre for Outcomes Research, Basel, Switzerland); [refs] DiDACT (York Health Economics Consortium, UK); [refs] EAGLE (Institute for Medical Outcomes Research, xxxxx, Germany); [refs] the Global Diabetes Model/GDM (Kaiser Permanente Center for Health Research, Portland, Oregon, USA); [refs] the ScHARR model (University of Sheffield, UK); [refs] the UKPDS Diabetes Outcomes Model (UKPDS Group, Oxford, UK); [refs] and the University of Michigan Model (Lansing, Michigan, USA). [refs] A variety of

model structures was represented, including modified-Markov state-transition models (CORE, EAGLE, Michigan, ScHARR), a fully continuous model (UKPDS), a partly deterministic model (DiDACT); and a mixed model (GDM).

Following Mount Hood II, several models made major changes in their structure and assumptions, for example by substituting microvascular results from the UKPDS study for older data, which resulted in higher estimates of renal, retinal, and neuropathic risk. Modelers now recognize that increased survival due to increasingly aggressive cardiovascular disease prevention will increase the proportion of persons with diabetes who will live to experience end-stage microvascular events.

Overall, there appeared to be some convergence in model predictions compared to 12 months earlier. The models became more sophisticated and complex, and able to accommodate very detailed assumptions about patient history, future treatment, and treatment effects. Differences in these numerous assumptions led to considerable differences in competitive results--and to a decision to publish much more explicit and detailed standardization of competition assumptions prior to future meetings. Among the causes of divergence was change over time in both the definition of diabetes and in the speed with which physicians make the diagnosis. The early studies from which modelers derived hazard rate functions and other parameters, such as the Wisconsin Epidemiologic Study of Diabetes Retinopathy, occurred at a time when the type 2 diabetes diagnosis appears to have been made three or more years after the actual onset of disease, [ref Harris and Klein; ref Brown, Pedula, Summers] even using the conservative diagnostic standards of the time. Now, in some settings, the diagnosis is made in most patients almost immediately. [ref]

Differing definitions of CVD events emerged as another barrier to model validation and use. For example, the Framingham Heart Study—popular in many models because it provides an almost untreated "natural-history" description of CVD—captured early, difficult-to-ascertain events such as angina but did not capture modern events, such as angiography, angioplasty, and coronary artery surgery. Similarly, some datasets, such as Framingham and the UKPDS, contain the results of hand-reviews of mortality records to ascertain instances of CVD-death—but the databases that many modelers use do not. Finally, some models attempted to track post-event occurrences such as mortality after an MI directly, while others do this indirectly.

A new technical challenge that emerged at Mount Hood III is the modeling of time-varying and cumulative risk burdens, including those related to hyperglycemia. The UKPDS follow-on report had been presented at the International Diabetes Federation meeting in Paris, just prior to Mt. Hood III, and was fresh in many participants' minds. The UKPDS had reported that past glycemic control affects complication events well into the future, whereas the blood pressure treatments used in UKPDS do not. To date, most modelers have been able to ignore time-varying and cumulative risk burdens by asking what-if questions that assume constant or monotonically increasing long-term differences in risk levels. Such questions are quite important. However, as models start to be used for treatment planning for specific patients, and as health systems use models for real-

world guideline development and purchasing decisions, more realistic assumptions will have to be made. The natural development path seems to be to introduce that ability to program detailed step-care treatment algorithms, which can interact with growth in risk factor levels to generate time-varying levels of current and cumulative risks. This is done, for example, in the CORE model.

At least two of the Mount Hood III models have gone further and created treatment algorithms that are triggered by both risk factor levels and clinical events. For example, use of an ACE-inhibitor could be triggered either by an increase in blood pressure, by the need for treatment after an MI, or by the onset of microalbuminuria; and ACEI could be required to be withdrawn if end-stage renal disease occurs. In addition, more models may begin incorporating data about the trial-based effects on outcomes of specific agents. (Recent trials have shown that individual compounds and classes of compounds exert independent effects on outcomes and treatment persistence, apart from their effects on their risk factor targets, such as blood pressure or LDL or HbA1c.) The Global Diabetes Model does this now for treatments with antihypertensive action and for antithrombotics. The Archimedes model, one of the new models at Mount Hood III, does this now for several kinds of treatments.

The models presented at Mount Hood III used a variety of assumptions and sources of data regarding costs of medical care and utilities. (Utilities account for the quality of life decrements associated with diabetes and its complications.) All models used direct estimates of the costs of some specific treatments, such as a hospitalization for a heart attack. Most models also used this direct, "bottom-up" to compute all other costs. However, some models used a "top-down" approach. Top-down (hedonic) cost-estimation draws on data systems that capture the entirety of direct medical care costs expended on patients' behalves, and then uses case-control methods or regression models to assign costs to diabetes and its complications. The authors of top-down models note that as many as half the costs induced by diabetes are not for the treatment of specific diabetes risks or recognized complications.[refs?] Only a top-down approach can capture these induced costs. However, bottom-up modelers like the flexibility and portability that bottom-up, item-specific costing gives, and question whether all those top-down induced costs will actually be saved when complication events are prevented. Medical economics is a social process as well as a physiological one. Future model versions may need to take the process of cost-generation and cost-reduction more explicitly into account.

Regardless of the costing method used, the Mt Hood III models usually predicted that risk-lowering (tighter glucose control, smoking cessation, blood pressure control, lipid-lowering) will be highly cost-effective and will often saved money in the long run, even when distant benefits are discounted at the agreed-upon rate of 3.0% per annum. In other words, despite increases in life expectancy of a year or more, the cost of additional hospitalizations and treatment during that period was often more than offset by reductions in the lifetime number of costly adverse medical events. In some cases, intensified treatment allowed patients to die of non-diabetes-related causes rather than diabetes complications like heart attack, renal failure, or stroke. Some attendees wondered whether the potentially high cost of these other causes of death should somehow be

incorporated into the calculations. This would reduce the predicted value of diabetes treatment. Another attendee proposed that models should include allowances for the difference between medical care cost inflation and inflation in the general economy, which is much lower. Doing so would increase estimates of savings from diabetes treatment.

Cost-effectiveness assessments were expressed at Mt Hood III in terms of cumulative total medical care costs divided by cumulative total quality-adjusted life years. It emerged that the methods used to adjust for quality of life had a very large impact on this ratio. One of three methods was used. Some models assigned a utility score between zero and one to each possible complication and multiplied them. Other models added up reductions in scores. One model based a person's quality of life on his or her lowest applicable score. Models also used different starting points. One model assumed that a person with uncomplicated diabetes should have a utility of 1.0, but most assumed at starting point of 0.70 to 0.80, less than the life-quality of a healthy young adult. Another debate involved the 'point of view' to use in assessing utilities. When interviewed, persons without a disease or a complication tend to assign a much lower quality of life to having it than persons who have actually experienced the disease. Nevertheless, persons without disease may better represent the views of taxpayers and society at large. To improve comparability, the Mt Hood III attendees agreed that they would adopt a standard utility schedule and method for future competitions.

Perhaps the fundamental long-term design question in diabetes modeling is, "What mathematical approach and architecture will prove most valid and useful in the long run?" All the models at Mt Hood III were stochastic micro-simulators, at least in part, which means that they simulated the life courses of thousands of individual patients to obtain average results, using random number generators to build in risk and uncertainty. The Mt Hood III teams used one or a combination of three distinct mathematical techniques to achieve microsimulation: modified Markov modeling, continuous statistical modeling, and discrete event simulation. Markov models divide a person's potential medical future into thousands of mutually exclusive health states and calculate the probabilities of transitioning among these states. Continuous models use a linked web of equations involving both continuous and categorical variables, typically derived from a large research database. Discrete event simulators create a world of interacting objects, such as pancreases, livers, muscle, and drugs—each operating on its own timescale. Each technique has its strengths and weaknesses. Modelers can be passionate about their choices. Mt Hood III showed that useful and potentially valid models can probably be constructed using any of the major techniques.

Of course, different uses demand different attributes in models. Real-time clinical decision support or web-based dissemination, for example, requires speed and patient-by-patient precision. Guideline development and scholarly writing can perhaps get by using slower models, that are only accurate in populations. The mean run-times of the models presented at Mount Hood III ranged from seconds to days.

A related fundamental debate involves the reproduction of causality in the models, which affects the ability of models to analyze a broad range of possible scenarios and situations. In principle, a model that fully reproduces true underlying causal reality can simulate the full spectrum of possible genetic make-ups, environments, treatments, and decisions. This is a project of enormous complexity, however. (As one attendee observed, life itself may be its own most efficient causal model.) The task of modeling is therefore the design of useful and valid simplifications. The task is complicated by fact that we do not actually or at least fully understand the world that we are trying to simplify. The whole vast enterprise of modern science is working to gain this understanding. In the meantime, diabetes modelers have drawn on accumulating knowledge in biochemistry, physiology, observational epidemiology, economics, mathematics, and clinical-trial research.

Unfortunately, epidemiologists and other scientists usually do not package their findings in forms that match the causal structures and estimation techniques that modelers want and need to use. One solution is to calculate new equations from existing databases, such as the UKPDS, the San Antonio Heart Study, or the electronic clinical records of large health care systems, such as Kaiser Permanente. This is what the continuous statistical models do. The problem with this approach is that the resulting model structures may be more predictive than causal or, if causal, may be valid only for persons and populations very much like the ones that generated the original data. This is a matter not only of the genetics and ethnicity of the populations contributing the data, but of local treatment practices, diagnostic criteria, environmental factors and health behaviors generally.

By the end of the Mount Hood III meeting, it was clear that the method of testing models against each other rather than against an independent source of gold-standard data had started to outlive its usefulness. The group agreed to look for an independent source of data from a clinical trial to test against next time. The use of trial data, rather than a purely observational dataset, may be important because, in trials, treatments (or at least some of them) are closely controlled. This tells modelers how algorithms governing future treatment should be programmed.

The next Mount Hood Challenge, Mount Hood IV, will be held in Basel, Switzerland September 2-4, 2004, prior to the annual congress of the European Association for the Study of Diabetes. Dr. Andrew Palmer will chair the local organizing committee. The meeting will be longer in 2004, allowing time for topic sessions as well as for comparisons among models. Presenters will be asked to provide yet more specific details about their models. (In fact, one of the important agreements emerging from Mount Hood III was a promise from each modeling group to publish all the equations on which their models are based.) Plans are for modelers to submit results for the gold-standard trial population, plus results for one or two standard patients, which will be specified more realistically and in greater detail for 2004. Persons interested in attending or participating should contact Dr. Palmer (ap@thecenter.ch), or Dr. Alastair Gray (alastair.gray@ihs.ox.ac.uk), who has agreed to chair the Mount Hood Steering Committee for 2003-2004.

--Jonathan Brown