Archimedes

A trial-validated model of diabetes

DAVID M. EDDY, MD, PHD[1]
LEONARD SCHLESSINGER, PHD[2]

OBJECTIVE — To build a mathematical model of the anatomy, pathophysiology, tests, treatments, and outcomes pertaining to diabetes that could be applied to a wide variety of clinical and administrative problems and that could be validated.

RESEARCH DESIGN AND METHODS — We used an object-oriented approach, differential equations, and a construct we call "features." The level of detail and realism was determined by what clinicians considered important, by the need to distinguish clinically relevant variables, and by the level of detail used in the conduct of clinical trials.

RESULTS — The model includes the pertinent organ systems, more than 50 continuously interacting biological variables, and the major symptoms, tests, treatments, and outcomes. The level of detail corresponds to that found in general medical textbooks, patient charts, clinical practice guidelines, and designs of clinical trials. The model is continuous in time and represents biological variables continuously. As demonstrated in a companion article, the equations can simulate a variety of clinical trials and reproduce their results with good accuracy.

CONCLUSIONS — It is possible to build a mathematical model that replicates the pathophysiology of diabetes at a high level of biological and clinical detail and that can be tested by simulating clinical trials.

Diabetes Care 26:3093–3101, 2003

The use of mathematical models in clinical diabetes is well known (1–9). This article describes a new type of model of diabetes called Archimedes. This model is different from other models in several important ways. It is a person-by-person, object-by-object simulation. It is broad, spanning from biological details to the care processes, logistics, resources, and costs of health care systems. It is written at a deep level of biological, clinical, and administrative detail. It is continuous in time; there are no discrete time steps, and any event can occur at any time. Biological variables that are continuous in reality are represented continuously in the model; there are no discrete "states" or "strata." It includes many diseases simultaneously and interactively in a single integrated physiology, enabling it to address comorbidities, syndromes, and treatments with multiple effects. Currently, the model includes diabetes and its complications, coronary artery disease, congestive heart failure, and asthma. Other diseases are b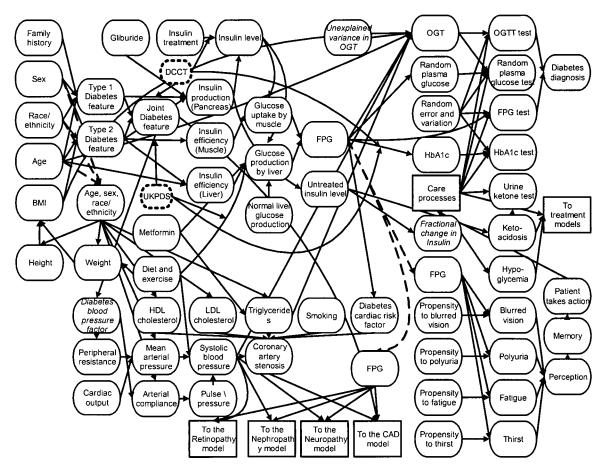eing added. Finally, as described in a companion article (10) in this issue of *Diabetes Care*, the model has been validated by simulating 18 clinical trials.

The model is written in differential equations, using object-oriented programming and a construct we call "features." The mathematical foundations have been described elsewhere (11). The entire model is too large to be described in a single article. Furthermore, the equations themselves can be written at different levels of detail, depending on the proposed applications. This article is a clinical, nonmathematical description of the part of the model that addresses the pathophysiology of diabetes at the level of detail pertinent to the clinical management of the disease. The actual equations, assumptions, and sources for this part of the model are summarized in an online appendix (available at http://care.diabetesjournals.org), which also includes an English translation for each equation. The equations presented are sufficient to address the great majority of questions relating to the progression and management of the disease, and they are the equations used in all of the validation exercises reported in the companion article (10).

Overview of the model

The full Archimedes model is designed to be comprehensive and includes not only individual people (patients) but also other important aspects of a health care system, such as health care personnel, facilities, equipment, supplies, policies and procedures, regulations, utilities, and costs. The design objective is to simulate what happens in a real health care system at a realistic and natural level of detail. Thus for a typical application of the model, there will be thousands of simulated patients, each with a simulated anatomy and physiology, who can get simulated diseases, can seek care at simulated health care facilities, will be seen by simulated health care personnel in simulated facilities, will be given simulated tests and treatments, and will have simulated outcomes. The simulated tests and treat-

See accompanying editorial, p. 3182.

Figure 1—*Diagram of variables in diabetes metabolism model. Circles represent variables. Lines represent relationships. In general, the arrows coming into any variable represent one equation. Squares represent other parts of the model that are too complicated to be shown here; they have their own diagrams, typically with dozens of variables and relationships. "FPG" and "Age, sex, race/ethnicity" are repeated to improve the visibility of the figure. Information from the UKPDS and DCCT trials were used to help build the part of the model illustrated. They are shown as the circles with dashed borders, and the parts of the model they affected are shown by the arrows emanating from those circles. No other trials used to validate the model (10) were used to build any of the parts of the model illustrated.*

ments will use simulated equipment and supplies, may require admissions to simulated hospitals, and so forth. As in reality, each of the simulated patients is different, with different characteristics, physiologies, behaviors, and responses to treatments, all designed to match the individual variations seen in reality. The simulated personnel are also different, reflecting the variation in skills and behaviors seen in reality.

The diabetes model

We describe the model in parts: the anatomy, the main variables (which we call "features") that determine the course of the disease; risk factors, incidence, and progression of the disease; glucose metabolism; signs and tests; diagnosis; symptoms; health outcomes of glucose metabolism; treatments; complications; deaths

from diabetes and its complications; deaths from other causes; care processes; and system resources. The variables pertinent to the development and progression of diabetes and their relationships are illustrated in Fig. 1. This figure also indicates the places where data from trials were used to help build the model.

Anatomy. In the model, all the simulated people/patients have organs, such as hearts, livers, pancreata, gastrointestinal tracts, fat, muscles, kidneys, eyes, limbs, circulatory systems, brains, skin, and peripheral nervous systems. Each of these organ systems in turn has the necessary parts and subparts. For example, all the people's hearts have four coronary arteries, atria, ventricles, myocardium, and sino-atrial nodes; all of the coronary arteries have lumens; and the lumens can have plaque or thrombi at any point. Pancreata

have β-cells, kidneys have glomeruli, and so forth.

In the model, as in real organ systems, all of the organs and their parts have functions. For example, a function of the β-cells is to produce insulin, a function of coronary arteries is to carry blood to the myocardium, and a function of the myocardium is to pump blood and maintain cardiac output. Furthermore, the functions of any part can change or become abnormal, as in real diseases. For example, in the model, the uptake of glucose by the simulated muscle cells can fail to respond to insulin. When the functions of organs become abnormal, the functioning of other organs is in turn affected. For example, a change in insulin levels affects the production of glucose by the liver.

Features. We conceptualize the physiology of a person as a collection of contin-

uously interacting objects that we call "features." Features can represent real physical phenomena (e.g., the number of milligrams of glucose in a deciliter of plasma), behavioral phenomena (e.g., ability to read a Snelling chart), or conceptual phenomena (e.g., the "progression" of a cancer). The full model—which currently includes diabetes and its complications and also nondiabetic coronary artery disease, congestive heart disease, and asthma—contains hundreds of features.

We represent the causes of diabetes as two features, which we call "type 1 diabetes feature" (DF_1) and "type 2 diabetes feature" (DF_2). At present, these two diabetes features are mathematical constructs, placeholders for the real determinants of diabetes when they are better understood. The appropriateness of this formulation is confirmed in the validations (10).

Three other features in the model are especially important factors in the cause and manifestations of diabetes: the insulin amount (I), the efficiency of insulin use (E), and the effects of insulin resistance (H). In the model, type 1 diabetes is characterized by an inability of pancreatic β-cells to produce appropriate amounts of insulin. Thus in the model, type 1 diabetes primarily affects the value of I through the type 1 diabetes feature. Type 2 diabetes is the result of a complicated set of interactions involving all five features introduced in this section (DF_1, DF_2, I, E, and H) and described below.

Risk factors, incidence, and progression. Equations that predict the progression of diabetes over time and cause people to develop diabetes at rates that match the observed rates were derived from data on the cumulative incidence rates for various populations (12–14). For type 1 diabetes, the feature DF_1 is a function of age, sex, family history, and race/ethnicity (African American [male and female], Hispanic American [male and female], Asian American [male and female], and white [male and female]) (12,13). For type 2 diabetes, the feature DF_2 is a function of age, sex, race/ethnicity (Hispanic American [male and female], African American [male and female], Native American, Asian American, and white) (13,14), BMI (13,15,16), and a factor that registers the effect of glucose intolerance (17). See Eqs. 1–4 in the online appendix.

Glucose metabolism. In the diabetes model, the main biological variables are fasting plasma glucose (FPG), HbA_{1c}, oral

glucose tolerance (OGT), random plasma glucose, and blood pressure. Many other biological variables that are related to the complications of diabetes are in the model but are not described here.

FPG. The progression of diabetes, the development of signs, symptoms and complications, and the response to treatments are determined primarily by the steady-state level of glucose, which can be represented by either the FPG or HbA_{1c}. In the model, the FPG in a person with diabetes is determined by six variables that represent *1*) the average FPG in people who do not have diabetes, *2*) hepatic glucose production, *3*) the effect of insulin resistance on hepatic glucose production, *4*) the insulin amount (I), *5*) the efficiency with which the body (liver, muscle, and fat) uses insulin (E), and *6*) the two diabetes features (DF_1 and DF_2). The conceptualization is as follows. In people who develop type 2 diabetes, the simulated liver cells develop a resistance to the effects of insulin. This causes the simulated liver to produce too much glucose. In response, the simulated β-cells produce more insulin. Over time, this compensatory mechanism begins to fail through a combination of decreased insulin production (e.g., "β-cell fatigue" [18]) and increasing resistance to insulin by the liver. In addition, the uptake of glucose by the simulated muscles and fat gradually decreases due to insulin resistance affecting those organs. Taken together, these factors create a relative deficiency of insulin with resulting increases in glucose. These relationships are addressed by Eqs. 5–10 in the online appendix. As indicated in the online appendix, three main sources provide the data needed to estimate these equations (19–21).

These equations are the steady-state solution of a more detailed equation for the instantaneous glucose level that takes into account the sources and sinks of glucose and the appropriate time constants. They are similar to the steady-state solution of the homeostasis model (22), the main difference being that our equations also include the progression of the disease over time and the effects of a patient's characteristics (age, sex, race/ethnicity, family history, and BMI).

HbA_{1c}. HbA_{1c} is related to FPG (23), as described in Eq. 11. We use two equations, one for people with type 2 diabetes, which is estimated from data in the U.K. Prospective Diabetes Study (UKPDS)

(24), and one for people with type 1 diabetes, which is estimated from data in the Diabetes Control and Complications Trial (DCCT) (20). We note that the measurement of HbA_{1c} varies considerably across laboratories. When measurement methods are standardized (25), the equations may have to be rewritten.

Random plasma glucose. In the model, the random plasma glucose is a function of the FPG and random factors (23) (Eq. 12).

Oral glucose tolerance. In the model, the 2-h OGT test is affected by many biological variables. To speed the calculations for simulating long-term clinical trials, we use a modification of an existing regression equation (26) to estimate the true tolerance to an oral glucose load (OGT). Our modification adds a residual variance, the variance not explained by the variables in the previously published regression equation (27). Our regression equation (Eq. 13) calculates the OGT as a function of the FPG, BMI, systolic blood pressure, and triglycerides.

Blood pressure. People with diabetes typically have higher blood pressures than people who do not have diabetes. We model this by multiplying the patient's peripheral resistance by a factor, which we call the "diabetes blood pressure factor" (*DiabBP*). The factor *DiabBP* is a function of the diabetes features and is therefore higher for people with more severe diabetes. A formula for the factor (Eq. 14) was estimated from data in Diabetes in America (19). The Archimedes model of blood pressure includes cardiac output, arterial compliance, pulse pressure, peripheral resistance, mean arterial pressure, diastolic pressure, and systolic pressure.

Signs and tests. The model currently includes tests for four biological variables relating to the pathophysiology of diabetes: FPG, OGT, HbA_{1c}, and random plasma glucose. In each case, the result of the test is determined by the patient's true value of the biological variable, as calculated elsewhere in the model, and a random variable that reflects the observed variability and errors in test results (23,28). See Eq. 15 for an example. Many other signs and tests that are pertinent to the complications of diabetes are in the model but not discussed here.

Diagnosis. There is no clear biological line that defines the disease we call diabetes. The American Diabetes Association

(ADA) defines a person to have diabetes if either he or she has symptoms plus a casual plasma glucose >199 mg/dl or a fasting plasma glucose >125 mg/dl or an OGT test >199 mg/dl (29). Impaired glucose tolerance is defined as an OGT test between 140 and 200 mg/dl. Impaired fasting glucose is defined as FPG between 110 and 126 mg/dl. Because the model is written at the level of the underlying biological variables, it can accommodate any definition. More specifically, the diabetes features do not determine the progression of a patient to a "state" called "diabetes." Rather, the features determine the progression of the underlying biological phenomena, which in turn determine a person's glucose level at any time, as described in Eqs. 1–10.

Symptoms. The model currently includes four symptoms relating to glucose metabolism: thirst, polyuria, fatigue, and blurred vision. The approach to each symptom is similar. Using thirst as an example, in the model there is a feature that represents the magnitude of a patient's thirst due to diabetes at any time. It is a function of the person's FPG and a randomly assigned factor for each person that represents the variation in thirst experienced by different individuals. (Consider it the person's "thirst propensity.") In the model, when a patient first experiences the symptom of thirst, a message is sent to the person's perception and is stored in the person's memory. The person's perception multiplies the number of symptoms of that type by the intensity of the symptom. The person's perception does this for each type of symptom, adds them together, and then compares that value to a "symptom threshold," which is unique for each patient. If the sum of all the symptoms multiplied by their intensities exceeds the symptom threshold, the person will seek care. The distribution for the intensity of the symptom as a function of FPG is estimated from data on the proportion of people who are diagnosed as having diabetes because of symptoms, at various levels of FPG (30). Equations for thirst are shown as an example in the online appendix (Eqs. 16–18). The actions the person can take and the responses they trigger in the health care system are included in the model but not described here.

Health outcomes of glucose metabolism. The model includes two main acute health outcomes associated with diabetes metabolism: ketoacidosis and hypoglycemia.

Ketoacidosis. In the model, ketoacidosis occurs when intracellular glucose levels are low, the liver attempts to correct for this by metabolizing fat into glucose, and ketones are produced as a byproduct. It occurs almost exclusively in type 1 diabetes. The equations are estimated from the observed times between episodes of diabetic ketoacidosis (20,31) (Eq. 19).

Hypoglycemia. In the model, hypoglycemia can occur when a person's insulin amount is artificially raised, either by taking insulin or by taking an oral medication to enhance natural insulin production. We model the frequencies of moderate and severe hypoglycemia as a function of the fractional change in the insulin amount caused by treatment, which in turn is related to the fractional change in FPG. The equations for type 1 and type 2 diabetes were estimated from data in the DCCT (20) and the UKPDS (24), respectively (Eq. 20).

Hyperglycemia is included in the model in the sense that it affects signs (e.g., glucosuria), symptoms (e.g., polydipsia), and the complications of diabetes. However, the current version of the model does not include any acute outcomes caused by elevated glucose levels per se.

Treatments. The model currently includes three main types of treatments: insulin, oral drugs, and lifestyle (diet and exercise) relevant to the pathophysiology of diabetes.

Insulin. Using results from the DCCT (20) we estimated a factor, which we call the "insulin factor," that determines the change in the insulin amount (I) caused by 1 unit of insulin per kilogram per day. To represent individual variations in response to insulin, the insulin factor for each person is drawn from a distribution that reflects the degree of variation in the population.

Oral drugs. Archimedes includes a variety of drugs, which can have different mechanisms of action. To illustrate how drugs are represented in the model, we describe two that have different mechanisms of action: glyburide and metformin. Ultimately, both of these drugs affect FPG, although they appear to do so in different ways (32–34). Because glyburide causes a person to produce more insulin, we model its effect as causing the β-cells to increase the insulin amount by a factor, which we call the "glyburide factor." The generic form of a treatment factor is given in Eq. 21. In the case of glyburide, it becomes a multiplicative term applied to Eq. 9. Because metformin causes the liver to produce less glucose, we model its effect as causing hepatic cells to decrease the production of glucose by a factor (the "metformin factor"), which in turn affects a person's reference FPG (FPG_0). It is a multiplicative term applied to Eq. 6. In addition to their effects on plasma glucose, both of these drugs affect other variables. The effect of glyburide on weight was estimated from data in the UKPDS (21). The effects of metformin, used alone or with sulfonylurea, on plasma triglyceride and LDL cholesterol levels are also included in the model (33). The effect of metformin on weight was not included in the version of the model reported here. However, it has since been added to the model. (We report the earlier version here because it was the version used to predict the results of the Diabetes Prevention Program reported in the companion article.) Equations have also been written for combinations of drugs and doses (32–37).

Diet and exercise. Changes in lifestyle, such as diet and exercise, affect many parameters in the model, as in reality. One is a direct effect on FPG, which we model through the hepatic production of glucose (Eq. 6). Diet and exercise also change lipid levels (38), blood pressure (39), and weight (39). Equations for those effects are in the model but not shown here.

Complications. The model contains more than 100 other biological variables, symptoms, tests, treatments, and outcomes relating to the complications of diabetes and their management. A complete description is not possible here. Briefly, coronary artery disease is modeled through two primary features called "slow occlusion" and "fast occlusion." They correspond to the gradual formation of atherosclerotic plaque in coronary arteries and to the sudden occlusion of a coronary artery due to rupture of plaque and/or development of an occlusive thrombus, respectively. In the model, either of these types of occlusion can occur at any point in any of the four coronary arteries, with appropriate implications for the amount of the distal myocardium that is affected, myocardial contractility, cardiac output,

and so forth. The equations that describe the time course of these and related features are derived from population-based datasets, such as the Framingham study (40), and include many variables such as age, sex, HDL cholesterol, total cholesterol, smoking, systolic or diastolic blood pressure, and enlargement of the heart (left ventricular hypertrophy), as well as FPG. However, the Archimedes model handles these variables quite differently than, for example, the Framingham heart-risk equation or similar regression models. Most importantly, in Archimedes the equations are not calculating the risk of an outcome such as a myocardial infarction, but are rather modeling the occlusion of specific coronary arteries in specific locations. Other important differences are that they include FPG as a continuous variable, and they incorporate not only the degree of elevation in FPG but also the duration of time that the FPG has been elevated to different degrees. The other variables needed for these equations are calculated in other parts of the model. The model also includes equations for the chain of biological events that occur with an infarction (e.g., myocardial damage, decrease in myocardial contractility, and decrease in cardiac output), as well as the treatments for those events (e.g., reversibility of myocardial damage with thrombolytics as a function of time since infarction). Strokes are handled in a similar fashion.

The primary feature for nephropathy is the progressive loss of glomerular function. The amount and type of protein leaked into the urine and the later manifestations of diabetic nephropathy are functions of the value of this feature. This feature is in turn a function of the person's FPG, blood pressure, and a variable we call the "glycemic load." The latter represents not only the degree of elevation in FPG but also the duration of time that the FPG has been elevated to different degrees.

A feature called "retinopathy" determines the course of that complication. The clinical manifestations (e.g., micro aneurysms, retinal hemorrhages, and hard and soft exudates) are functions of the value of the retinopathy feature. These manifestations in turn define the steps that are used to measure the progression of retinopathy. Thus retinopathy progresses continuously through the steps that clinicians have defined according to the signs of the disease, and the model can track the step a patient is in at any time. Like nephropathy, the progression of the retinopathy feature is a function of a person's FPG, blood pressure, and glycemic load.

Diabetic neuropathy manifests itself in several ways. In the current version of the model, the main clinical manifestation is loss of sensation. We model the occurrence and progression of this complication by defining a primary feature called "neuropathy," which is a function of a person's FPG, blood pressure, and glycemic load. In the model, complications such as foot ulcers and the diabetic foot are functions of the neuropathy feature.

Deaths from diabetes and its complications. In the model, people can die of the complications of diabetes by mechanisms that correspond to the real causes of death. For example, a person in the model will die if a coronary artery is occluded and the subsequent infarction reduces their myocardial function to the point that cardiac output and blood pressure cannot be maintained.

Deaths from other causes. The model includes deaths from other causes, but only deaths due to coronary artery disease, congestive heart failure, and stroke are calculated using physiology-based equations like those described here. Although the model will accurately calculate changes in life-years and life expectancy due to coronary artery disease, congestive heart failure, asthma, and diabetes, it will not predict any unexpected effects of diabetes treatments on other diseases.

Care processes. The model contains detailed descriptions of the processes of care, written in the form of algorithms. They describe what providers do in specific circumstances. For example, an algorithm for the control of cholesterol in a patient with diabetes might say: "If the patient's LDL cholesterol is >180 and their creatinine is <2, then give lovastatin 80 mg. At 2 months, have the patient get a lipid panel and creatinine test. At that time if the LDL is not <130 and the creatinine is still <2, then switch to simvastatin 80 mg . . ." and so forth. The full set of care processes includes more than one hundred algorithms like this. Care processes can vary from setting to setting and even from physician to physician. The algorithms can also include variations in practice styles, uncertainty, and random factors; can depend on the type of provider (e.g., specialist versus primary care physician); and can depend on other factors (e.g., attendance of a particular continuing medical education course, or access to a clinical information system with reminders). The current version of the model includes care processes derived from one of the medical centers at Kaiser Permanente. Care processes can vary considerably across different settings. Archimedes accommodates this by enabling users to modify the care processes to fit particular settings.

System resources. Finally, the model includes system resources such as personnel, facilities, equipment, and supplies needed to deliver care. These are all included at a high level of detail. For example, in the current version there are 37 different types of office visits. Use of these resources is triggered whenever patients encounter the health care system or an intervention is applied. The model tracks every use of every resource and its associated time and cost for every patient. As with care processes, system resources and costs are currently modeled after Kaiser Permanente, but the descriptions and values can be modified for other settings.

Methodological notes
The process for building the model consists of five main steps. The first is to develop a nonquantitative or conceptual description of the pertinent biology and pathology—the variables and relationships—as best they are understood with current information. For this step, we consult experts and basic texts. The result of this step can usually be described in a figure like Fig. 1. The second step is to identify studies that pertain to the variables and relationships. Typically, these are the basic, epidemiological, and clinical studies that experts identify as the foundations of their own understanding of the disease. The third step is to use the information in those studies to develop equations that relate the variables. The development of any particular equation involves finding the form and coefficients that best fit the available information about the variables. The types of research results and specific methods for using data to build the model are described elsewhere (11). The next step is to program the equations in Smalltalk. This is followed by a series of exercises in which the parts of the model are tested and debugged—first one at a time and then in

appropriate combinations, using inputs that have known outputs. The next step is to use the entire model to simulate a complex trial. This tests not only the individual parts but also the connections between all the parts. It is also a rigorous test for any remaining bugs in the software. Finally, we do this for a broad range of studies that span different populations, organ systems, and treatments. The calculations are done using distributed computing techniques. Simulation of a clinical trial currently takes about 10 min using 250 PCs.

Demonstration of the pathophysiology model
The operation of the model is demonstrated in the companion article (10).

CONCLUSIONS
The Archimedes model differs in many ways from other clinical models. In addition to including behaviors, care processes, logistics, resources, and costs, which are not described in this article, the most obvious differences are that the Archimedes model is written at a fairly deep level of biology. It is continuous in time, and it preserves the continuous nature and simultaneous interactions of biological variables. Structurally, it is written with differential equations and is programmed in an object-oriented language called Smalltalk. Another difference is the fairly extensive comparisons to empirical studies, which are described in the companion article.

These differences were driven by our objectives, which in turn were driven by years of hands-on work with clinicians. First, we wanted the model to be able to analyze guidelines, performance measures, the what-to-do parts of disease management programs, clinical priorities, medical necessity, and coverage policies at the level of detail at which they are written and at which clinicians debate these issues. This meant that the model had to be built at a fairly high level of biological detail, preserve the continuous nature of biological variables, and include their interactions and feedback loops (homeostatic mechanisms). Second, we wanted to be able to address issues of timing, such as how long to try one treatment before switching to another. A related objective was that the model should be able to address problems that range in pace from minute-to-minute or hour-to-hour (e.g.,

the events after a myocardial infarction) to year-to-year (e.g., the effect of lifestyle on long-term complications). Both of these required that the model be continuous in time. Third, we wanted to address problems relating to care processes, such as continuous quality improvement projects, the how-to-do-it parts of guidelines and disease management programs and variations in practice patterns. This required that the model include care processes at the level of detail at which these projects are conducted and evaluated. Fourth, we wanted to be able to analyze the effects of interventions on logistics, use of resources, costs, and cost-effectiveness. This required inclusion of those variables, again at the level of detail at which people plan and make decisions. Fifth, the model should be able to address the interactions between diseases and comorbidities. This meant that there had to be a single integrated model of biology from which all the diseases in the model arise, so that the important interactions can be realistically represented. This design criterion was also needed to address interventions such a diet and exercise that affect multiple biological variables and multiple conditions, interactions between treatments, and syndromes that affect multiple organ systems. Sixth, to help set priorities and strategic goals, the model had to be able to span a wide range of interventions and a wide range of diseases. Seventh, we wanted to be able to simulate clinical trials and other clinical experiences to check the model and build credibility. This required that the model be able to handle all of the important biological, clinical, and procedural factors that are part of the design of a trial, such as inclusion criteria, treatment and testing protocols, biological outcomes, and health outcomes, at the level of detail at which they are actually defined in trials. Finally, we wanted to build a model that could be used over and over again to address a broad range of problems—not as a special-purpose model created to address a single problem and then be retired or rewritten. This requirement meant that the model had to be fairly complete and natural, it could not skip over variables just because they could be finessed for a particular question.

Although we have built several Markov models in the past (41), in our opinion Markov-type models did not have virtually any of the properties we needed. In a sense, we wanted to build a

model to address what happens underneath the clinical states, between the annual jumps, and inside the transition probabilities. Although we could have pushed the Markov approach in the direction we wanted, we believed it would never be able to handle our requirements. Furthermore, there are far more straightforward ways to accomplish our objectives. We wrote the Archimedes model using differential equations because that is the most natural way to model variables that are continuously valued and that interact continuously over time. We chose an object-oriented approach because of its ability to handle great complexity and because of the ease with which it can be expanded and updated. With the addition of the concept of features, we were able to build the model described here.

A crucial decision in the design of any model is the level of detail at which the model should be built. We used three main criteria. The first is that we wanted the model to include all the aspects of the disease or its management that clinicians or administrators consider important for addressing questions of interest. The second is that we wanted to build the model at the level of detail that distinguishes clinically important aspects of a disease and its management. For example, a protocol for reducing HbA_{1c} to a specified goal might begin with diet and lifestyle, then add glyburide 10 mg, then increase the dose of glyburide to 20 mg, then switch to metformin, then try a combination of glyburide 20 mg and metformin, and finally go to insulin. To represent the effects of sequences and combinations like this, we needed to build the model down to the level of detail that distinguished the separate effects of each treatment (e.g., glyburide affects the production of insulin by the pancreas, metformin affects hepatic production of glucose, and so forth). For another example, if an analysis is to address an issue like the timing of administration of a thrombolytic, the model needs to include the viability of myocardial cells as a function of time. The third criterion is that we wanted to include the level of detail researchers consider important in the designs of clinical trials. We needed this to be able to simulate the trials to check the model, as is described in greater detail in the companion article. These criteria forced the inclusion of a relatively high level of biology compared with other clin-

ical models. However, they did not require that we model the underlying biochemical reactions. Neither did they require that the underlying causes of diabetes be understood at the deepest levels.

Our intention is that at any time, the model should represent the best available knowledge about the included diseases (in the case of this article, diabetes) at the level of detail we desired. However, there are many remaining points of uncertainty about diabetes. Just as expert knowledge is subject to these uncertainties, Archimedes will be as well. We approach this in two main ways. First, although we have based the model on current knowledge of the disease, we make no claims that our representation of diabetes is perfectly "accurate" in the sense that it describes precisely how the disease works. Indeed, we know that that is not the case because it contains abstractions like "insulin resistance." Instead, we ask whether the model is "realistic" in the sense that it produces results that match real results at a clinical level, as observed over a wide range of trials. This builds confidence that the model is producing accurate clinically relevant results, even if there are gaps in the current understanding of the underlying biology of the disease. As the underlying biology becomes better understood and as that information becomes clinical relevant, it can be included in the model.

A very important feature of the model as we have formulated it is that it is easy to expand and update. For example, as our understanding increases about such things as genomics, β-cell fatigue, or the relative importance of FPG versus HbA$_{1c}$ for specific populations, it will be relatively easy to include them in the model. From our point of view, the barrier to the accuracy and completeness of the model is the performance of the research, not the ability of the model to incorporate the research.

As with any model or any other method for making decisions, including expert judgments and clinical trials, there are potential sources of error or bias. Uncertainties about the underlying biology have already been mentioned. We try to address this through the extensive validation exercises (10). But that approach itself introduces a source of error or bias. Trials are subject to both random and systematic errors. The effects of random variation are addressed through the statistical analyses of each trial's results. The main source of systematic bias is that what happens in a clinical trial may not represent what happens in more realistic settings. We discuss this further in the companion article.

An important issue with any model is its transparency. Traditionally, transparency is achieved by presenting the assumptions, equations, and sources. Our assumptions, equations, and sources are described in the online appendix to this article. However, the mathematics used in the Archimedes model is considerably more advanced than the mathematics of a Markov model or decision tree, and we recognize that many readers will not have the mathematical background needed to evaluate the equations. Although this is true of any complex model in any field, it is still desirable to enable those who use the model to gain some confidence that it will work the way they expect. To serve this need, we have taken three approaches. One is to describe the variables and relationships that are included in the model, so that experts can judge whether they correspond to their understanding of the disease within the bounds of current uncertainties and controversies. That is the purpose of this article. The second approach is to test the model against empirical observations that experts trust. The idea is that if the model can be shown to simulate a wide range of important studies, picked and reviewed by an independent and respected group of experts, people can gain confidence that the model is reasonably realistic. The companion article describes the methods and results of the validation of the parts of the model relevant to diabetes and its complications.

The third approach to transparency is to make the model available for others to use. To that end, we intend to make the Archimedes model widely available over the World Wide Web, through a very user-friendly interface on a nonprofit basis. This will enable clinicians, administrators, and modelers to view not only the remainder of the assumptions, equations, and sources, but also to explore how the model functions and come to understand it at whatever level of detail they desire. They will also be able to tailor and apply it to analyze their own problems.

We make no claims that this model can "replace reality." If it is possible to answer a question with a well-designed empirical study, then that approach is always preferable. Our goal is to provide a trial-validated method that can be used to address problems that cannot feasibly be addressed through empirical studies due to such factors as high cost, long follow-up times, large sample size, unwillingness of providers or patients to participate, large number of options, or rapid pace of technological change. In the way that a flight simulator provides valuable experience, shortens the time needed in real planes, and simulates experiences that are too dangerous or rare to attempt in reality (like severe wind shear), the Archimedes diabetes model should be a useful tool for sharpening our understanding of diabetes and its management.

References
1. Eastman RC, Javitt JC, Herman WH, Dasbach EJ, Zbrozek AS, Dong F: Model of complications of NIDDM: I. Model construction and assumptions. *Diabetes Care* 20:725–734, 1997
2. Eastman RC, Javitt JC, Herman WH, Dasbach, Copley-Merriman C, Maier W: Model of complications of NIDDM: II. Analysis of the health benefits and cost-effectiveness of treating NIDDM with a goal of normoglycemia. *Diabetes Care* 20:735–744, 1997
3. Vijan S, Hofer TP, Hayward RA: Estimated benefits of glycemic control in microvascular complications in type 2 diabetes. *Ann Int Med* 127:788–795, 1997
4. Centers for Disease Control Diabetes Cost-Effectiveness Study Group: The cost-effectiveness of screening for type 2 diabetes. *JAMA* 280:1757–1763, 1998
5. Vijan S, Hofer TP, Hayward RA: Cost-effectiveness analysis of screening intervals for diabetic retinopathy in patients with type 2 diabetes mellitus. *JAMA* 283:889–896, 2000
6. Brown JB, Russell A, Chan W, Pedula K, Aickin M: The global diabetes model. *Diabetes Res Clin Pract* 50 (Suppl. 3):S15–S46, 2000
7. de Lissovoy G, Ganoczy DA, Ray NF: Relationship of hemoglobin A1C, age of diabetes diagnosis, and ethnicity to clinical outcomes and medical costs in a computer-simulated cohort of persons with type 2 diabetes. *Am J Manag Care* 6:573–584,

2000

8. Palmer AJ, Brabdt A, Gozzoli Y, Weiss C, Stock H, Wenzel H: Outline of a diabetes disease management model: principles and applications. *Diabetes Res Clin Pract* 50 (Suppl. 3):S47–S56, 2000

9. The CDC Diabetes Cost-effectiveness Group: Cost-effectiveness of intensive glycemic control, intensified hypertension control, and serum cholesterol level reduction for type 2 diabetes. *JAMA* 287: 2542–2551, 2002

10. Eddy DM, Schlessinger L: Validation of the Archimedes diabetes model. *Diabetes Care* 26:3102–3110, 2003

11. Schlessinger L, Eddy DM: Archimedes: a new model for simulating health care systems: the mathematical formulation. *J Biomedical Informatics* 35:37–50, 2002

12. Dorman JS, McCarthy BJ, O'Leary LA, Koehler AN: Risk factors for insulin-dependent diabetes. In *Diabetes in America*. Harris MI, Cowie CC, Stern MP, Boyko EJ, Reiber GE, Bennett PH, Eds. Bethesda, MD, National Diabetes Data Group, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, NIH publication no. 95–1468, 1995, p. 165–178

13. LaPorte RE, Matsushima M, and Chang YF: Prevalence and incidence of insulin dependent diabetes: In *Diabetes in America*. 2nd ed. Harris MI, Cowie CC, Stern MP, Boyko EJ, Reiber GE, Bennett PH, Eds. Bethesda, MD, National Diabetes Data Group, National Institute of Diabetes and Digestive and Kidney Diseases, NIH publication no. 95–1468, 1995, p. 37–46

14. Harris MI, Flega KM, Cowie CC, Eberhardt MS, Goldstein DE, Little RR, Wiedmeyer HM, Byrd-Holt DD: Prevalence of diabetes, impaired fasting glucose, and impaired glucose tolerance in U.S. adults: the Third National Health and Nutrition Examination Survey, 1988–1994. *Diabetes Care* 21:518–524, 1998

15. Rewers M, Hamman RF: Risk factor for non-insulin dependent diabetes. In *Diabetes in America*. Harris MI, Cowie CC, Stern MP, Boyko EJ, Reiber GE, Bennett PH, Eds. Bethesda, MD, National Diabetes Data Group, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, NIH publication no. 95–1468, 1995, p. 179–220

16. Schwartz LM, Woloshin S: Changing disease definitions: implications for disease prevalence: analysis of the Third National Health and Nutrition Examination survey, 1988–1994. *Eff Clin Pract* 2:76–85, 1999

17. Edelstein SL, Knowler WC, Bain RP, Andres R, Barrett-Connor EL, Dowse GK, Haffner SM, Pettitt DJ, Sorkin JD, Muller DC, Collins VR, Hamman RF: Predictors of progression from impaired glucose tolerance to NIDDM: an analysis of six prospective studies. *Diabetes* 46:701–710, 1997

18. Kahn SE: The importance of beta-cell failure in the development and progression of type 2 diabetes. *J Clin Endocrinol Metab* 86:4047–4058, 2001

19. Cowie CC, Harris MI: Physical and metabolic characteristics of persons with diabetes. In *Diabetes in America*. 2nd ed. Harris MI, Cowie CC, Stern MP, Boyko EJ, Reiber GE, Bennett PH, Eds. Bethesda, MD, National Diabetes Data Group, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, NIH publication no. 95–1468, 1995, p. 117–164

20. DCCT Research Group: The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 329:977–986, 1993

21. UK Prospective Diabetes Study Group: Relative efficacy of randomly allocated diet, sulpholyurea, insulin, or metformin in patients with newly diagnosed non-insulin dependent diabetes followed for three years (UKPDS 13). *BMJ* 310:83–88, 1995

22. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC: Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose an insulin concentrations in man. *Diabetologia* 28:412–419, 1985

23. Rohlfing CL, Wiedmeyer HM, Little RR, England JD, Tennill A, Goldstein DE: Defining the relationship between plasma glucose and HbA_{1c}: analysis of glucose profiles and HbA_{1c} in the Diabetes Control and Complications Trial. *Diabetes Care* 25:275–278, 2000

24. UK Prospective Diabetes Study Group: Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 352:837–852, 1998

25. US National Glycohemoglobin Standardization Program [article online]. Available at http://www.missouri.edu/~diabetes/ngsp.html. Accessed 19 September 2003

26. Bando J, Toya D, Ushiogi Y, Tanaka N, Okafuji K, Fujisawa M: The relationship of fasting plasma glucose and other variables to 2-h postload plasma glucose in Japanese subjects. *Diabetes Care* 24:1156–1160, 2001

27. Perry RC, Shankar RR, Fineberg N, McGill J, Baron AD: HbA_{1c} measurement improves the detection of type 2 diabetes in high-risk individuals with nondiagnostic levels of fasting plasma glucose: the Early Diabetes Intervention Program (EDIP). *Diabetes Care* 24:465–471, 2001

28. Mooy JM, Grootenhuis PA, de Vries H, Kostense PJ, Popp-Snijders C, Bouter LM, et al: Intra-individual variation of glucose, specific insulin and proinsulin concentrations measured by two oral glucose tolerance tests in a general caucasian population: the Hoorn study. *Diabetologia* 39: 298–305, 1996

29. The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus: Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus (Position Statement). *Diabetes Care* 20:1183–1197, 1997

30. Harris MI: Classification, diagnostic criteria, and screening for diabetes. In *Diabetes in America*. 2nd ed. Harris MI, Cowie CC, Stern MP, Boyko EJ, Reiber GE, Bennett PH, Eds. Bethesda, MD, National Diabetes Data Group, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, NIH publication no. 95–1468, 1995, p. 15–36

31. Fishbein H, Palumbo PJ: Acute metabolic complications in diabetes. In *Diabetes in America*. 2nd ed. Harris MI, Cowie CC, Stern MP, Boyko EJ, Reiber GE, Bennett PH, Eds. Bethesda, MD, National Diabetes Data Group, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, NIH publication no. 95–1468, 1995, p. 283–292

32. DeFonzo RA: Pharmacologic therapy for type 2 diabetes mellitus. *Ann Int Med* 131: 281–303, 1999

33. DeFronzo RA, Goodman AM, for the Multicenter Metformin Study Group: Efficacy of metformin in patients with non-insulin dependent diabetes mellitus. *N Engl J Med* 333:541–549, 1995

34. Stumvoll M, Nurjhan N, Perriella G, Dailey G Gerich J: Metabolic effects of metformin in non-insulin-dependent diabetes mellitus. *N Engl J Med* 333:550–554, 1995

35. White JR, Campbell RK: Recent developments in the pharmacological reduction of blood glucose in patients with type 2 diabetes. *Clinical Diabetes* 9:153–159, 2001

36. UK Prospective Diabetes Study Group: Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). *Lancet* 352:854–865, 1998

37. Hollander PA, Schwartz SL, Gatlin MR, Haas S, Hongjie Zheng, Foley JE, Dunning BE: Importance of early insulin secretion. *Diabetes Care* 24:983–988, 2001

38. Frick MH, Elo O, Haapa K, Heinonen OP,

Heinsalme P, Helo P, Huttunen JK, Kaitaniemi P. Koskinen P, Manninen X, et al: Helsinki Heart Study: primary-prevention trial with gemfibrozil in middle-aged men with dyslipidemia: safety of treatment, changes in risk factors, and incidence o coronary heart disease. *N Engl J Med* 317:1237–1245, 1987

39. Blumenthal JA, Sherwood A, Gullette ECD, Babyak M, Waugh R, Georgiades A, Craighead LW, Tweety D, Feinglos M, Applebaum M, Hayano J, Hinderliter A: Exercise and weight loss reduce blood pressure in men and women with mild hypertension; effects on cardiovascular, metabolic, and hemodynamic functioning. *Arch Intern Med* 160:1947–1957, 2000

40. Anderson KM, Odell PM, PW, Wilson PWF, Kannell WB: Cardiovascular disease risk profiles. *Am Heart J* 121:293–298, 1991

41. http://davidmeddy.com [online]. Accessed 19 September 2003