Linking Electronic Medical Records To Large-Scale Simulation Models: Can We Put Rapid Learning On Turbo?

The Archimedes model offers one example of how mathematical modeling can assist in medical decision making.

by David M. Eddy

ABSTRACT: One method for rapid learning is to use data from electronic medical records (EMRs) to help build and validate large-scale, physiology-based simulation models. These models can then be used to help answer questions that cannot be addressed directly from the EMR data. Their potential uses include analyses of physiological pathways; simulation and design of clinical trials; and analyses of clinical management tools such as guidelines, performance measures, priority setting, and cost-effectiveness. Linking the models to EMR data also facilitates tailoring analyses to specific populations. The models’ power and accuracy can be improved by linkage to comprehensive, person-specific, longitudinal data from EMRs. [Health Affairs 26, no. 2 (2007): w125–w136 (published online 26 January 2007; 10.1377/hlthaff.26.2.w125)]

A central component of rapid learning is the use of data from electronic medical records (EMRs) to help address clinical and administrative questions that would otherwise go unanswered. There are two main ways EMRs can do this. The first is simply to use the data in the EMR to look up the answers. The other is to use the data to build mathematical models and then use the models to answer the questions. The models can vary from a single equation to deeper, more comprehensive models that use hundreds of equations to, in essence, try to simulate everything important that happens in a health care system, from the underlying physiology to logistics and costs. The latter type of a large-scale simulation model obviously involves much more work than simply looking up an answer in the database or building a simpler model. But once built, such a model can help answer a considerably broader range of questions. This paper describes how large-scale, physiology-based simulation models can increase the use...
of EMRs for rapid learning, and it illustrates their use with a particular model called Archimedes.

The Need For Rapid Learning From EMRs

The need for better methods to answer questions is obvious. The modern practice of medicine is bristling with questions, ranging from the medical director of a managed care organization wanting to know the effects of improving control of blood glucose in diabetics, to the head of medical benefits at a large corporation wanting to know why the company’s health costs went up in double digits three years in a row, to Medicare wanting to design a pay-for-performance (P4P) program and needing to know how much to pay for how much performance on which indicators, to Jane Smith wanting to know how much her risk of a stroke will go down if she loses ten pounds.

In theory, all of these questions could be answered if we could conduct enough evaluation studies and clinical trials. But the feasibility of that approach is severely limited by several facts: the long times required to observe long-term outcomes; the large sample sizes needed to sort out real results from random “noise”; the large number of options that need to be evaluated; the need to disrupt current practices to implement each of the options; the reluctance of physicians and patients to participate; the narrowness and artificiality that trial designs often require; the fact that the answer one gets in one setting does not necessarily apply to other settings; the high cost of the studies (typically in the tens of millions of dollars, and often in the hundreds of millions); the rapid changes in technologies and practices that make answers obsolete almost as soon as they come in; and the fact that even if we can set a study in motion, we will still need to know what to do while we wait for the results. If it is feasible to answer a question with evaluation studies and clinical trials, then by all means we should do that. But the unfortunate fact is that empirical studies are not feasible for the great majority of decisions that need to be made.

The Appeal Of EMRs

From this list of limitations, it is easy to see the attractiveness of EMRs. The data are already collected; it would seem that all we have to do is sort through them and pull out the answers. For a simple example, to determine the proportion of people with diagnosed diabetes who have poorly controlled blood glucose (hemoglobin A1c [HbA1c] levels greater than 7 percent), a medical director could go into the database, identify everyone with a diagnosis of diabetes, look up their HbA1c levels, and calculate the proportion with HbA1c levels above 7 percent. Such an approach is fast and inexpensive. It is also real; the data in EMRs represent real patients getting real treatments and having real outcomes in real practice settings. The hope for EMRs and rapid learning is well summarized in the commonly expressed idea that EMRs should provide opportunities for “natural ex-
periments.” Properly used and interpreted, EMR data used in this way can help answer many important questions. Several good examples are described elsewhere in this collection of Health Affairs papers.1

The Limitations Of EMRs For Rapid Learning

Unfortunately, with the simplicity, low expense, and speed come some limitations and pitfalls. For an example, suppose that after identifying people with diabetes and HbA1c levels above 7 percent, the medical director now wants to know whether treating them with a combination of an oral drug, metformin, and a new injectable drug called exenatide has any affect on the incidence of myocardial infarction (MI) in this group. Ideally, the medical director would like to have a randomized controlled trial (RCT) in which people with HbA1c levels greater than 7 percent are randomized to receive metformin and exenatide or not, and then followed for ten years to observe MI rates in both groups. The randomization would ensure that the two groups were statistically identical in all ways except the treatment, so any difference in MI rates could be attributed to the treatment. There is no clinical trial that does this. Could we use the natural experiment of the EMR to reproduce such a trial?

Unfortunately, we cannot. Consider the steps. We would have to go back in the EMR records ten years and try to identify two groups of people: a group that had HbA1c levels above 7 percent and were put on metformin and exenatide, and another group that had HbA1c levels greater than 7 percent but were not put on any drug. At this point we encounter the first problem: We can do this only for treatments that have been in use for at least ten years, and exenatide is new. This approach is useless for new or even recent treatments. But set that aside and imagine that we can find the two groups of people we need. Now the second problem arises: We need to ensure that those put on metformin and exenatide had the same pretreatment HbA1c levels as those not so treated. This is very unlikely, because people are put on drugs for the very reason that they have abnormal HbA1c values. Suppose we set that problem aside as well and imagine that the two groups we found had the same initial levels of HbA1c. We immediately encounter a third problem: The two groups also need to be the same in all other ways that might affect their chances of having an MI. That is, they must have the same age distribution, sex proportion, race/ethnicity mix, weight, proportion of smokers, blood pressure, high-density lipoprotein (HDL) level, glucose level, use of aspirin, and so forth. This is extremely unlikely. After all, whatever caused them to be put on metformin and exenatide (an alert physician?) would also likely cause them to be treated for hypertension, hyperglycemia, obesity, tobacco addiction, or anything else they might have. If they had been treated for any one of these, we would not be able to determine if any difference in MIs was due to the use of metformin and exenatide versus any of the other risk factors. Thus, the approach of trying to “find” the desired clinical trial in the data set will not work. We might try a differ-
ent approach, such as trying to find people who are matched in all important ways except the use of the drugs. But by the time we list all of the important variables that can affect heart attack rates and try to find people who are matched on all of them, we would end up with exceedingly few matched pairs, if any. If we reduce the list of the variables to those we need to match, we would find more pairs, but we would never know how the unmatched variables biased the results. There are other approaches we might try, but they all suffer from some bias or other.

In addition to the impossibility of analyzing new treatments, the confounding effects of the unmatched variables, the difficulty of finding a sufficient number of people, and the need for long follow-up times, the look-it-up approach has other problems. They include cohort effects (people who were being observed ten years ago are different from people being observed today); variations in and gradual drifting of practice patterns and performance levels; changes in technologies such as the introduction of new tests and treatments; changes in definitions of diseases and outcomes; turnover in the population; and difficulty of transferring results from one setting to another. In general, the pure look-it-up approach is good for observing what happened in the past and what is happening now, but it is not good for determining what will happen in the future, or what would happen if we did something differently, or the merits of different options.

The Role Of Models

To answer the latter set of questions, we need to turn to the second way of using EMR data to achieve rapid learning: use the data to build models, and then use the models to answer the questions. The concept is illustrated with an analogy. Imagine that an investigator is interested in the time it takes to get from New York to Boston. Imagine a data set that tracks cars as they pass tollbooths and other checkpoints, among other things. The investigator can answer the question by going to the data set, identifying all of the cars that traveled from New York to Boston, noting the times of departure and arrival, subtracting, and averaging the results. To go further, the investigator might report the results separately for cars that left in the daytime (between 6 a.m. and 6 p.m.) and those that left at night. That would answer the question; with the look-it-up method, that would likely be the end of it. But we can squeeze much more out of the data set if we use it to build a model. For example, if we were careful to collect the right data, we could start with the model “distance = rate × time” and then go on to add the effects of number of lanes, time of day, weather, congestion, accidents, tollbooths, radar, direction signs, curves, road surface, and so on. That model could then be used to answer not only the original question but also any number of other questions, such as the
effects of adding a carpool lane or designing a freeway around New Haven.

The use of models in this fashion is not new. When Boeing designs a new airplane, it faces scores of questions about the design of the wing, shape of the body, weight, engine power, and other variables. It does not build prototypes of all possible combinations, fly them, and get the pilots’ assessments. Boeing uses the principles of physics and data from wind tunnels to build mathematical representations of the planes and then flies them inside a computer. It does conduct experiments. But the purpose of the experiments is not to get an answer to one particular question; it is to collect data to build models that can answer hundreds of questions. Virtually every other sector of our economy uses mathematical models to manage complexity and uncertainty. Architects use them to design buildings, engineers use them to calculate the trajectories of satellites, UPS uses them to calculate optimal transportation routes, and Intel uses them to design computer chips. Even in medicine, computed tomography (CT) scans, radiation doses, nurses’ schedules, decisions to buy versus lease hospital beds, and the funds in which physicians invest their money are all based on models. So the idea of using data from EMRs to build models and then use the models to answer a variety of questions has a long history of success in a variety of fields, including medicine.

**Types Of Models**

Data from EMRs along with other sources can be used to help build a wide variety of models. At one end of the spectrum, in terms of simplicity, are single-equation models such as regression equations to help adjust for confounding factors (“risk adjustment”) or to help project a patient’s future costs based on the patient’s history. Both of these are common and well-accepted types of models. But models can be used for considerably more than that. Properly built and validated, they can be used to help analyze physiological processes; design guidelines, performance measures, and the “what-to-do” parts of disease management programs; design the “how-to-do-it” parts of disease management programs, case management protocols, and continuous quality improvement (CQI) projects; forecast logistics, utilization, costs, and cost-effectiveness; set clinical priorities and design strategic goals; prioritize or combine performance measures; analyze the effects of multiple diseases (comorbidities), syndromes that affect multiple organ systems, drugs that have multiple effects, and combinations of drugs; address questions of timing, such as screening, frequency of follow-up visits, or how long a medication should be tried before the dose is changed; and help design and predict clinical trials.

For these types of applications, one needs larger and deeper models. A rule of thumb is that if one wants to answer a question that involves a particular variable, then the model needs to include that variable. And if one wants to explore the effect on particular outcomes of changing variables, the model needs to include the variables that are to be changed, the outcomes of interest, and the pathways that
connect them. For example, if we want a model to help individual patients and clinicians make decisions; to help design guidelines, performance measures, and the “what-to-do” parts of disease management programs; and to be credible, the model should start at the level of physiological and clinical detail at which clinicians think. Essentially, it should encompass all of the biological variables that physicians consider to be important in the management of their patients, and it should relate them in a natural way. This level of detail is also required to help analyze the physiological processes underlying diseases and their treatments and to help design, interpret, and extend clinical trials. If we want a model to address the issues that arise in the design of the “how-to-do-it” parts of disease management programs, case management protocols, and CQI projects, the model should include care processes, logistics, and behaviors at an equally high level of detail. If we want a model to provide credible information about logistics and cost-effectiveness, it should be able to track the use of resources such as facilities, personnel, visits, admissions, equipment, and their costs. If we want a model to help set clinical priorities, design strategic goals, and prioritize or combine performance measures, the model should span all types of interventions (primary prevention, screening, diagnosis, treatment, secondary prevention, and support care) and span multiple diseases using the same methodology. A broad span is also required to address patients who have multiple diseases (comorbidities), syndromes that affect multiple organ systems, drugs that have multiple effects, and combinations of drugs. If we want a model to address questions of timing—such as screening, frequency of follow-up visits, or how long a medication should be tried before the dose is changed—the model should function in continuous time and be able to address events that can change as rapidly as minute by minute, or as slowly as year by year. Finally, if we want a model to be credible, the model should be able to simulate the most important epidemiological studies and clinical trials at the level of clinical detail at which they are designed and reported, and to match or predict their results within the appropriate confidence limits. In general, if one is interested in a broad and deep range of questions, one needs a broad and deep model.

Archimedes: An Example Of A Large-Scale, Physiology-Based Model

The feasibility of building models like this can be illustrated with the Archimedes model, which has been developed over the past ten years by a team led by Len Schlessinger and me. Briefly, the core of the model is a set of ordinary and differential equations that represent physiological pathways at a level of detail roughly
comparable to that found in general medical textbooks, clinical trials, and patient charts. The model includes pathways relating to diabetes, congestive heart failure, coronary artery disease, stroke, hypertension, obesity, and metabolic syndrome in a single integrated model; and asthma in a stand-alone model. Cancers of the lung, breast, and colon are being added. Other conditions will be added in the future.

Stemming from this core are additional equations that represent the development of signs and symptoms; patients’ behaviors in seeking care and complying with treatments; clinical events such as telephone consultations, office visits, and admissions; tests and treatments (including errors, side effects, and complications, to the extent they are known); physicians’ behaviors and performance; logistics and utilization (for example, emergency department visits, admissions); resources; costs; and measures of quality of life.

Because the model begins with differential equations, time is continuous, biological variables are continuously interacting functions of time, any event can occur at any time, and the timing of events is as condensed or drawn out as occurs in reality. Clinical outcomes are defined in terms of the underlying variables, which enables the model to incorporate different definitions and changes in definitions. The effects of interventions are modeled at the level of the underlying biology. The inclusion of multiple organ systems and diseases as part of a single physiology enables accurate representation of the effects of syndromes, comorbidities, multiple drugs, and drugs with multiple effects. The model is programmed in an object-oriented language called Smalltalk and runs on a grid of computers using distributed computing methods.

The objective is to create a virtual world that can do the things listed four paragraphs above. The virtual world of Archimedes consists of thousands of virtual people, all of whom have virtual physiologies, can get virtual diseases, and have virtual behavior. To represent real populations, Archimedes can create copies or “clones” of real people drawn from the settings in which the questions are being asked, using person-specific data from surveys and data sets such as the National Health and Nutrition Examination Surveys (NHANES), health risk appraisals, personal health records, and, of course, EMRs. It does this at the level of detail captured in the survey, including, if available, demographic characteristics, physical examination results, behavior, family history, medical conditions, biological variables, medical history, symptoms, current medications, and so forth. Archimedes uses the information provided on each person (for example, characteristics, lab values, medications, symptoms, medical history) to calculate the hidden chronic damage these factors imply, such as the degree of stenosis in coronary arteries.

The model is constructed so that it can be tailored to different settings, distinguish by specific populations (for example, age, sex, race/ethnicity, behavior, risk factors, incidence rates); clinical practice guidelines and levels of performance; and costs. Examples of possible settings are a managed care organization, the U.S. population, the employees of General Motors, Medicare, Colorado Medic-
aid, people covered by Blue Cross of Michigan, and the uninsured of Los Angeles.

Example Of An Application

The potential uses of a large-scale simulation model can be illustrated with an example. Suppose a medical director is interested in the effects of giving simvastatin 40 mg, a cholesterol-lowering drug, to men and women ages 40–80 who have had a previous MI or currently have angina, or who have had a bypass graft or angioplasty. We have already seen that this type of question, as common as it is, is beyond the capability of the look-it-up approach for EMRs. We could use a model like Archimedes to select virtual people in the virtual world who meet those criteria, as well as other characteristics we might want to specify (for example, average HDL level, proportion taking aspirin), randomize them to receive simvastatin or not (delivered by virtual physicians through virtual visits, and so forth), follow them for ten years in virtual time, let the patients’ virtual physiologies progress (for example, virtual atherosclerosis developing in virtual coronary arteries), and record the occurrences of virtual MIs. The results of such a virtual trial are shown as the dashed lines in Exhibit 1.

One can appreciate the range of potential uses of models like this by listing some of the questions the Archimedes model has been used to address recently: (1) What would be the effect of a cure for insulin resistance (insulin resistance is the major cause of diabetes and the metabolic syndrome)? How would such a cure compare with simply getting all patients and physicians to faithfully follow exist-

**EXHIBIT 1**
Comparison Of Model Results With Those Of An Actual Clinical Trial: Cumulative Probability Of Major Coronary Events In The Heart Protection Study

<table>
<thead>
<tr>
<th>Fraction of people with major coronary events</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.20</td>
</tr>
<tr>
<td>1.00</td>
</tr>
<tr>
<td>0.08</td>
</tr>
<tr>
<td>0.06</td>
</tr>
<tr>
<td>0.04</td>
</tr>
<tr>
<td>0.02</td>
</tr>
<tr>
<td>0.00</td>
</tr>
</tbody>
</table>

Time (years)


**NOTES:** Comparison of results predicted by the Archimedes model (dashed lines) versus actual results (solid lines) from the Heart Protection Study (HPS). Open circles are placebo group; solid circles are treated group.
ing guidelines for control of risk factors such as hypertension, high cholesterol, and smoking? (2) What would be the effect on biological variables and clinical outcomes of a drug that decreases weight by 5 percent? For what populations would such a drug be best indicated if the goal is to decrease heart attacks? To decrease net costs? (3) How does delivering insulin through inhalation compare with the current method of injection in people with Type I diabetes? What are the implications of the different effects on HbA1c for downstream clinical outcomes? What trial should we conduct to determine the appropriate indications? (4) We want a calculator that will tell people their risk of diabetes and its complications, taking into account not only the usual Framingham-type variables (sex, age, systolic blood pressure, total cholesterol/HDL cholesterol, whether they smoke, whether they have diabetes) but also duration and severity of diabetes, medical history (for example, previous MI), past treatments, past and current weight, current symptoms and complications, and current medications. (5) What are the relative effects of raising performance from the average to the ninetieth percentile level for each of the Health Plan Employer Data and Information Set (HEDIS) measures for cardiovascular disease, diabetes, congestive heart failure, and tobacco use? If we can focus on only three measures, which are the most important? (6) We've done a Phase II trial that showed the effects of [a particular drug] on [a particular set of biomarkers]. What would a Phase III trial show for clinical outcomes? What would happen in different populations (that is, different indications)? What is the optimal size and duration for a trial? (7) We want to combine two drugs. What will be the combination drug's effects in populations that have different initial risks, different current treatments, and different current degrees of control? (8) We are contemplating [a new approach to the treatment of diabetes]. We want to know its effects on clinical outcomes and costs. Will it save money? If so, by when?

**Validation Of Large-Scale Models**

A critical aspect of using models, large or small, is validating them. Large-scale simulation models provide a helpful method for accomplishing this; their accuracy can be checked by using the virtual world to simulate activities that have happened in the real world and then comparing the results. For example, the application just described—the effects of simvastatin 40 mg on MIs in people at high risk of cardiovascular disease—has been studied in a real clinical trial called the Heart Protection Study (HPS). The real results are shown in Exhibit 1 as solid lines. Here it is important to understand that the HPS was not used to build the Archimedes model; it is being used only to test it. The close concordance between the real and virtual results builds confidence in the model's accuracy.

This approach of comparing predicted results against real results can be repeated for a range of trials that span the types of questions the model will be used to address. For example, thus far the Archimedes model has simulated about three
dozen clinical trials relating to diabetes and coronary artery disease. Exhibit 2 presents a scatter plot showing the accuracy of the first seventy-four exercises (counting the different arms and outcomes of a trial as separate exercises). More than half of the validations are independent, which means that no information from the real trials was used to build the model. Overall, the correlation between the real and predicted results is 0.98. For the independent validations, the correlation is 0.96. Although these results certainly do not prove that every new analysis will be accurate, they do indicate that it is possible to build large-scale, physiology-based models that reflect the available evidence from clinical and epidemiological studies reasonably well.

Using Data From EMRs

The quality of a model depends directly on the quality of the data available to build it. To build a large-scale, physiology-based simulation model like Archimedes, one would ideally like to have access to several data sets that are large (from many people), person-specific (reporting the measurements and events of each person separately), comprehensive (recording of all the variables of interest), and longitudinal (tracking each person's measurements and events over time—preferably five to ten years). Until very recently, this has been extremely difficult, if not impossible, to achieve. A model like Archimedes typically has to be constructed from published data—epidemiological studies, basic physiology studies, large

EXHIBIT 2
Comparison Of Model Results With Those Of An Actual Clinical Trial For Seventy-Four Validation Exercises

<table>
<thead>
<tr>
<th>Results calculated by model</th>
<th>Results from trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.0</td>
</tr>
<tr>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>0.0</td>
<td>0.5</td>
</tr>
</tbody>
</table>

NOTES: Actual results (horizontal axis) versus calculated results (vertical axis) from seventy-four independent validation exercises involving eighteen randomized controlled trials relating to diabetes and cardiovascular disease. Forty of the validation exercises are independent; no data from the trial were used to build or calibrate the model. Each square represents a particular outcome of a particular arm of a particular clinical trial (outcomes vary from arm to arm and include morbidity, mortality, and various clinical markers). Perfect accuracy would show all of the squares on the diagonal line. Deviations from this line are attributable to sampling variations in the real clinical trials as well as any inaccuracies in the model.
surveys, and clinical trials. These may be supplemented by single-point-in-time (not longitudinal) data from a public survey like NHANES and by occasional access to person-specific data from narrowly defined studies of particular physiological pathways, but the majority of the model is built from published sources.

Published data are limited by two main facts. First, they are aggregated; they give the average results in the population without reporting the values for individual people. Second, they report what the investigators are interested in, which is not always what a modeler needs. These limitations can be severe. To appreciate this, let us revisit the example of travel times from New York to Boston. Suppose we are interested in how rainy conditions affect travel time. A published summary of the data set might report the average time it takes cars to go from one point to another and the average number of inches of rainfall. It is easy to see that this is not sufficient to answer the question. To determine the relationship between travel time and rainfall, we need data on each car—the amount of rain it was experiencing and its times at each point on the route. This illustrates only one way in which the object-specific longitudinal data can be used. There are many others. Compared with using only reported averages, this type of object-specific information is like moving from black and white TV to color, or from a car radio to high-end hi-fi.

Data from EMRs could be extremely helpful in filling this gap. First, they could be used to develop a much more accurate understanding of biological pathways—how different biological variables relate to one another in complex physiological processes. Second, such data would greatly improve our understanding of individual variations—how the values of biological variables are distributed across individuals and how they change or progress within individuals over time. Third, they can help us understand how medicine is actually practiced, not just in idealized settings like clinical trials but in real settings of patient care. Fourth, they can be used to develop person-by-person copies of specific populations. This would enable analyses to be tailored to particular settings so that the one-size-fits-all assumptions we now have to make could be discarded. A fifth use of EMR data is validation. A model like Archimedes is now validated against clinical trials. Data from EMRs would enable the validations in more realistic settings.

For these and other reasons, we can expect that access to EMR data will make large-scale models like Archimedes much more powerful and accurate. It is important to understand, however, that EMR data are not a panacea; they have their own limitations that will affect their value and building models. To list just three, they are not particularly useful for determining the effects of interventions, for reasons already given; important data may be miscoded or missing; and they typi-
cally do not include research variables that are needed to build deeper models of physiological pathways.

**Conclusions**

The practice of medicine is extraordinarily complicated. The number of questions we face far outstrips our current methods for getting answers. Clinical trials and other well-designed empirical studies are essential for helping us understand human physiology, disease, and the effects of tests and treatments. They are our fundamental anchor to reality. However, it is not feasible to answer all questions with empirical studies. For questions that cannot be answered with empirical studies, mathematical models can be useful. In particular, large-scale, physiology-based models can help answer a wide variety of questions ranging from physiological pathways; to clinical trials; to management tools such as guidelines, performance measures, priority setting, forecasting, and cost-effectiveness. The Archimedes model has demonstrated the feasibility of building these types of models. The main limitation of these types of models at this time is the quality of data available to build and validate them. Patient-specific, comprehensive, longitudinal data from EMRs can greatly improve the quality of data available for building these types of models. Together, the linkage of data from EMRs and large-scale, physiology-based models could open up a promising new way to improve the quality and efficiency of medical care.

**NOTES**

1. A listing of these papers is available at http://content.healthaffairs.org/cgi/content/full/hlthaff.26.2.w107/DC2.
4. Eddy and Schlessinger, “Validation of the Archimedes Diabetes Model.”