17

The Future of Diagnostics

From Optimizing to Satisficing

Henry Brighton

Abstract

In health care, our observations are shaped by interactions between complex biological and social systems. Practitioners seek diagnostic instruments that are both predictive and simple enough to use in their everyday decision making. Must we, as a result, seek a trade-off between the usability of a diagnostic instrument and its ability to make accurate predictions? This chapter argues that sound statistical reasons and evidence support the idea that the uncertainty underlying many problems in health care can often be better addressed with simple, easy-to-use diagnostic instruments. Put simply, satisficing methods which ignore information are not only easier to use, they can also predict with greater accuracy than more complex optimization methods.

Introduction

In the United Kingdom, 81% of costs in patient services result from hospital expenditure (Dept. of Health 2005). One way to reduce these costs is to improve health care management of people at high risk of hospital admission. Among older people, certain subgroups are at a particularly high risk. If these high-risk individuals can be identified early, and their health care managed more effectively, preventive care could mean financial savings (Wagner et al. 2006). This holds for many problems in health care. Thus, the identification of predictive diagnostic instruments relating variables to outcomes is a necessary goal, and observations will be used to guide this process.

Statistics provides tools for crafting sophisticated predictive models from observations. These observations are shaped by the interaction of complex biological and social systems. On the other hand, practitioners seek diagnostic instruments which are both predictive and simple enough to use in their everyday decision making. I will argue that this apparent dichotomy is false: There are sound statistical reasons, and evidence, supporting the idea that the complexity

Chapter from "Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020," edited by G. Gigerenzer and J.A.M. Gray. Strüngmann Forum Report, vol. 6, J. Lupp, series ed. Cambridge, MA: MIT Press. ISBN 978-0-262-01603-2

underlying the problems faced in health care can often be better addressed with simple, easy-to-use diagnostic models.

A bottom-up approach will be taken, starting with the identification of three forms of uncertainty that stand in the way of making accurate predictions. The remainder of the discussion will then contrast two broad perspectives on constructing diagnostic models. The first is optimization, which stresses the explicit attempt to conduct rationally motivated calculations over potentially complex models. Optimization attempts to find the most predictive model (or at least, approach it). The second perspective, satisficing, is less well-known, and seeks a good enough model by ignoring information. Satisficing tends to rely on simpler models and can consequently achieve greater predictive accuracy than "optimal" models.

The notion of robustness will be used to explain this counterintuitive relationship. A system is robust to the extent that its function is maintained when operating conditions change. In health care, uncertainty surrounds our models, observations, and context of use. Diagnostic instruments which attempt to "over-model" the problem can be less robust than those which ignore information. When designing and deploying diagnostic instruments, we should assume that changes in operating conditions will occur, even though their exact form may be hard or impossible to specify.

Fitting Models to Observations

By considering eight variables, such as a person's self-perceived health and their access to a caregiver, Wagner et al. (2006) considered the problem of predicting if older people will require hospital admission at some point during the following year. To develop a predictive model, they collected observations of older people. Each observation related the eight variables associated with the individual to a dependent variable detailing whether or not this individual required an overnight hospital admission at any point during a follow-up period of one year. Now, a good statistical model, one capable of improving health care management among older people, must somehow capture what is systematic in these observations. Once identified, these systematic patterns can be used to predict whether or not a previously unseen person will require hospital treatment in the future.

In health care, education, sociology, psychology, and beyond, a hugely diverse set of problems are framed in these terms. Despite this diversity, the statistical machinery employed is astonishingly uniform. The linear regression model rules in a methodological dictatorship. It has two components: an assumed linear relationship between the variables and the criterion, and the assignment of weights to these variables by the method of least squares. A raft of statistical add-ons can complement this basic machinery, but the basic properties of the machine remain the same. The methodological dictatorship

Chapter from "Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020," edited by G. Gigerenzer and J.A.M. Gray. Strüngmann Forum Report, vol. 6, J. Lupp, series ed. Cambridge, MA: MIT Press. ISBN 978-0-262-01603-2

is not the fault of the linear regression model, but the collective behavior of its users.

Often, the variables used in a linear model are referred to as "predictors." Predictors which "explain" a large amount of the variance are typically seen as informative. Similarly, when the linear model achieves a good fit to the data, the model is said to predict the data well. Additional manipulations, those which improve the fit of the model to the data further, are seen as a move in right direction, a step closer to the objective of capturing systematic patterns in the data. Stepping back from this statistical routine, it is worth considering exactly what has been predicted. The parameters of the model have been estimated from the observations. This parameterized model is then evaluated on its ability to "predict" these same observations. This is a process of postdiction, not prediction. The predictive ability of a model, under any meaningful definition of the term, refers to its ability to second guess properties on unseen data presented at some point of time in the future, after the model has been parameterized.

Faith in data fitting and abuse of the term prediction can, and often will, lead to faulty inferences from data (Pitt et al. 2002; Roberts and Pashler 2000). Because these inferences may inform policy, it is worthwhile pulling apart the conceptual distinction between post- and prediction. The following example provides a visual illustration. The temperature in London on a given day of the year is uncertain but nevertheless follows a seasonal pattern. Using the year 2000 as an example, Figure 17.1 plots London's mean daily temperature. On top of these observations, two polynomial models have been plotted; they

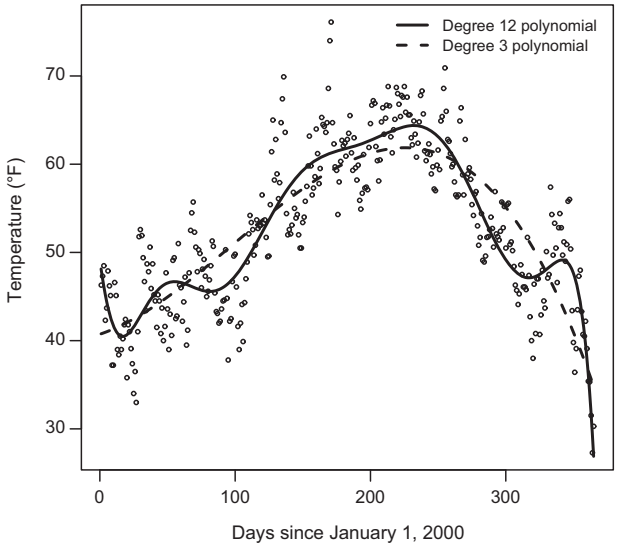

Figure 17.1 Mean daily temperature in London for the year 2000. Two polynomial models (a degree-3 and degree-12 polynomial) are fitted to this data.

Chapter from "Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020," edited by G. Gigerenzer and J.A.M. Gray. Strüngmann Forum Report, vol. 6, J. Lupp, series ed. Cambridge, MA: MIT Press. ISBN 978-0-262-01603-2

both attempt to capture what is systematic in London's temperatures. The first model is a degree-3 polynomial (a cubic equation with four parameters), and the second is a degree-12 polynomial (which has 13 parameters). Comparing these two models, we see that the degree-12 polynomial captures monthly fluctuations in temperature, whereas the degree-3 polynomial captures a simpler pattern charting a rise in temperature that peaks in the summer, followed by a slightly sharper fall.

Now, which of these two models best captures London's mean daily temperatures? Outside idealized "laboratory" settings, data can be seen as containing both systematic and accidental patterns. A model which accurately captures both will describe the data closely and achieve a good fit. In the temperature example, the degree-12 model achieves a better fit to the data than the simpler degree-3 model for precisely this reason. However, accidental patterns, by definition, will not hold true of future data, and relying on these patterns to make predictions about the future will lead to errors. The key point here is that when judging models on the basis of their goodness-of-fit, one has no way of knowing if a good fit reflects the ability of the model to capture accidental patterns accurately, its ability to capture systematic patterns, or both.

Below, it will be demonstrated that the simpler degree-3 model provides a better model of the London's daily temperature even though it achieves a poorer fit than the degree-12 model. Before doing so, it is worth pointing out that, taken to an extreme, the policy of evaluating models exclusively on their ability to fit data implies that we only need one model. The best model is one which simply memorizes the list of observations. This model guarantees a perfect fit on all problems. Among the statistically trained, many are aware of the dangers of model fitting, but far fewer act to avoid them.

Out-of-Sample Prediction

> Prediction is very difficult, especially if it's about the future.
> —Niels Bohr (1885–1962)

Because the future is uncertain, how can we possibly evaluate the ability of a model to predict accurately whether an older person will require hospital admission or not in the future? Waiting is not an option. Recall that the objective is to identify systematic patterns governing the population given only a sample of observations. Because samples are likely to contain noise and accidental patterns, they will always provide an uncertain and potentially misleading view of the population. Several model selection criteria have been developed which provide principled alternatives to measuring the goodness-of-fit of a model to the data (e.g., Pitt et al. 2002). Cross-validation is one criterion, and it works by first fitting the model to a fraction of the available observations; thereafter it uses the remaining observations to measure the predictive ability of this fitted model (Stone 1974). This process is then repeated, each time partitioning the

Chapter from "Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020," edited by G. Gigerenzer and J.A.M. Gray. Strüngmann Forum Report, vol. 6, J. Lupp, series ed. Cambridge, MA: MIT Press. ISBN 978-0-262-01603-2

data randomly. The final estimate of predictive accuracy is the mean predictive accuracy with respect to many such partitions.

For example, if we sample the temperature in London on 50 randomly selected days in the year 2000, and then fit a series of polynomial models of varying degree to this sample, we can measure two quantities. The first measure is a familiar one, the goodness-of-fit of the models to the sample. The second measure is perhaps less familiar and considers how accurately these fitted models predict the temperature on those days of the year 2000, which were not observed. As a function of the degree of the polynomial model, Figure 17.2 plots the mean of these two measurements. Focusing on the goodness-of-fit of the models to the samples, we see that the more parameters the model has, the better the fit. This relationship reflects the point made above: Achieving a good fit to the data is trivial. Achieving high predictive accuracy is not so simple. The model with the lowest mean prediction error (with respect to many such samples of size 50) is a degree-4 polynomial. More complexity is not better. If we had relied on goodness-of-fit as an indicator of model performance, we would not have chosen this model.

Cross-validation estimates the out-of-sample predictive accuracy of the models, which is their accuracy in predicting outcomes for unseen observations drawn from the population when estimated from the samples of that population. Note that access to the population was assumed in the preceding example. This need not be the case. Cross-validation is usually applied to a

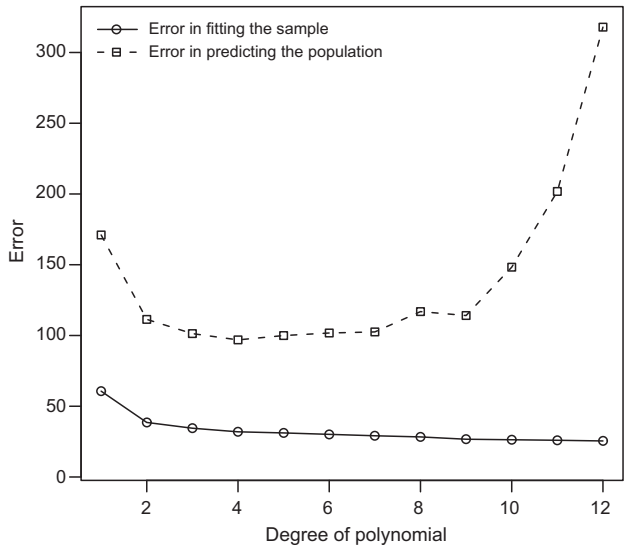

Figure 17.2 Model performance for London temperatures in 2000. For the same data, mean error in fitting the observed samples decreases as a function of polynomial degree. Mean error in predicting the whole population of the entire year's temperatures using the same polynomial models is minimized by a degree-4 polynomial.

Chapter from "Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020," edited by G. Gigerenzer and J.A.M. Gray. Strüngmann Forum Report, vol. 6, J. Lupp, series ed. Cambridge, MA: MIT Press. ISBN 978-0-262-01603-2

sample of data, because we rarely (if ever) have knowledge of the population. For example, a fraction of the available observations of people's hospital admissions can be used to fit the model parameters, and then competing models can be judged on their ability to predict the remaining observations. Although the model parameters may be estimated from fewer observations, the idea is that this is a price worth paying if we can then estimate the predictive ability of the models.

The temperature example highlights two forms of uncertainty. First, a sample of observations provides an uncertain indicator of the population from which they are drawn. If the tape of experience were replayed, then we are likely to have a different sample with different characteristics. Second, there is uncertainty surrounding which model should be chosen to capture systematic patterns. In the temperature example, we do not know the functional form of data-generating function behind London's daily temperature. Similarly, we can be sure that when using a linear model to predict if a person will be admitted to hospital in the next year, this model will be a gross abstraction of the data-generating distribution. In both cases, the model is misspecified. Many of the issues raised thus far come into sharper focus when we examine the origins of linear regression.

Toward the end of the 18th century, Adrien-Marie Legendre and Carl Friedrich Gauss were both concerned with measuring the length of the meridian quadrant, the distance from the equator to the North Pole (Stigler 1986, 1999). Using error-prone observations of subsections of the arc measured at various points between Dunkirk and Barcelona, they faced the problem of how to integrate these observations. The accuracy of their estimates was important. The meter was to be defined as 1/10,000,000 of this arc. When tackling this problem, it appears that Legendre and Gauss independently developed the method of least squares. In contrast to the problem of estimating London's mean daily temperature, or estimating the number of future hospital admissions, Legendre and Gauss had the benefit of a precise geodesic model. There was little uncertainty surrounding this model, and the observations were used to estimate precise values for its parameters.

The problems that confronted Legrende and Gauss, such as measuring the length of the meridian quadrant or predicting the trajectory of a comet, can be seen as the antithesis of those we face in health care. Indeed, relative to our temperature example, Legendre and Gauss where almost certain where on the x-axis of Figure 17.2 they should be focused, which meant that they were not searching for the functional form of the data-generating model. Now, put in these terms, it is no surprise that choosing the best fitting model will lead to poor inferences. Without knowing the gross properties of the data-generating model, data fitting becomes the statistical equivalent of groping in the dark. Thus, the method of least squares is used, now as a matter of routine, in contexts which bear little resemblance to the context of its discovery. The key

Chapter from "Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020," edited by G. Gigerenzer and J.A.M. Gray. Strüngmann Forum Report, vol. 6, J. Lupp, series ed. Cambridge, MA: MIT Press. ISBN 978-0-262-01603-2

difference is the uncertainty surrounding our knowledge of the data-generating distribution.

Out-of-Population Prediction

> In times of change learners inherit the earth; while the learned find themselves beautifully equipped to deal with a world that no longer exists.
> —Eric Hoffer (1902–1983)

Thus far, it has been assumed that a sample is drawn from a population and that the properties of this population will determine the accuracy of our predictions. This is an idealization. One assumption underlying this belief is that the population will remain stationary over time. However, this is just one idealization among many. After finding a predictive model of future hospital admission among older people, this model could become standardized and used by doctors more widely. When used in a different hospital, city, region, country, or continent, the population will almost certainly change. Perturbations arise due to one or more of an endless list of factors (e.g., differences in the measurement techniques used, demographic and cultural changes affecting the health of individuals). Couched in terms of the weather example, a more realistic test of the models estimated from a sample of measurements is to consider how well they go on to predict the mean daily temperature in London on each day of, for instance, 2001. What we are predicting now lies outside the population used to estimate the model parameters. The two populations may differ because factors operating over longer time scales come into play, such as climate change.

To illustrate, Figure 17.3 examines how well the models estimated from samples of the temperatures in London in 2000 go on to predict the temperatures in 2001 and 2002. I have also plotted out-of-sample error for 2000 as a point of comparison, and, as we should expect, the error increases when we move from the out-of-sample problem to the out-of-population problem. Although there is more uncertainty in the out-of-population setting, much of the same pattern can be observed as before: A degree-4 polynomial yields the minimum mean error. This tells us that what we learned from the out-of-sample task transfers to the out-of-population task, since a degree-4 polynomial remains a good choice of model.

An additional change to the population that we wish to predict can be introduced by imagining that we want to use the temperatures in London to predict those observed in Paris. Paris lies 212 miles southeast of London, and Figure 17.3 shows how the prediction error suffers as a result of this geographical shift; however, the finding that degree-4 polynomials predict with the least error remains. Often, we will not be so lucky. This is an example of a robust model: The degree-4 polynomial models fitted to relatively small samples of observations remain, relative to the other models, accurate. The accuracy of the

Chapter from "Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020," edited by G. Gigerenzer and J.A.M. Gray. Strüngmann Forum Report, vol. 6, J. Lupp, series ed. Cambridge, MA: MIT Press. ISBN 978-0-262-01603-2

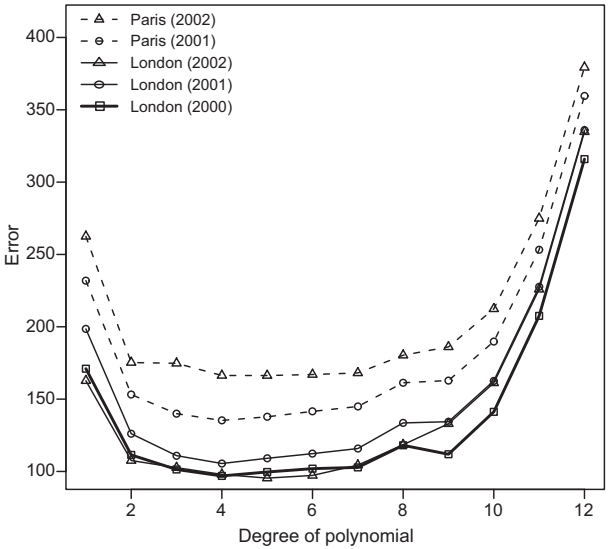

Figure 17.3 Out-of-population prediction. The models estimated from 50 samples of London's daily temperatures in 2000 can be used to predict the daily temperature for that entire year (thick line). This plot also shows how well these models go on to predict the daily temperature in London for 2001 and 2002, and in Paris for 2001 and 2002. Much the same pattern is observed across applications of the model, although the error increases due to greater uncertainty coming from changes in time and space.

model appears to be robust against several kinds of perturbation. The notion of robustness will be central to the remainder of the discussion.

Robustness to Uncertainty

Robust systems maintain their functioning when operating conditions change (Wagner 2005). The ability of diagnostic tool to make accurate predictions when the population changes is an example of a robust statistical model. The ability of the immune system to maintain a functioning organism despite continually evolving pathogens is an example of a robust biological system (Hammerstein et al. 2006). The ability of the flight control system of an aircraft to follow a course despite potentially severe atmospheric changes is an example of a robust human-engineered system (Kitano 2004). Some disciplines, such as biology, attempt to reverse-engineer robust design. Other disciplines, such as engineering, attempt to design robust systems. In diagnostics, too, we would like to engineer predictive instruments which remain predictive when operating conditions change. Which principles might guide the design of robust diagnostic instruments? Before examining this question, it is worth consolidating the three kinds of uncertainty discussed so far:

Chapter from "Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020," edited by G. Gigerenzer and J.A.M. Gray. Strüngmann Forum Report, vol. 6, J. Lupp, series ed. Cambridge, MA: MIT Press. ISBN 978-0-262-01603-2

1. Model misspecification: The complexity of the processes that determine the content of our observations is such that our knowledge of the data-generating distribution will be uncertain. Consequently, we have no way of formulating a space in which the correct model is known to exist. Short of assuming certain knowledge of putative "natural laws" governing the observations, all models are misspecified and lead to error. Put simply, there is uncertainty surrounding our ability to formulate the data-generating distribution.
2. Model underspecification: Given a parameterized space of models, misspecified or not, observations are required to estimate the parameters of the model. Observations are finite in number and may be too few to estimate these parameters reliably. For example, even if we knew we were searching in the right space, a single observation will be insufficient to reliably select a good model, and errors result. Put simply, there is uncertainty surrounding the ability of the data to select the model parameters reliably.
3. Out-of-population prediction: Even with certain knowledge of the data-generating distribution, and therefore the "optimal" model, there is uncertainty surrounding the context in which this model is used to make predictions. Put simply, data-generating distributions can change for many reasons. You may, for example, have discovered the optimal model at the time of data collection, but this data-generating distribution is likely to have been perturbed by the time the model is used to make predictions.

All approaches to diagnostics must contend with these uncertainties. The substantive issue is how to design diagnostic instruments that are robust to the realities of model misspecification, model underspecification, and the uncertainty surrounding the problem itself. With these questions in mind, the remainder of this discussion will contrast two approaches to developing diagnostic systems: optimizing and satisficing.

From Optimizing to Satisficing

Optimizing is the process of seeking the optimal solution. What is the optimal solution? Consider a tin can manufacturer who attempts to reduce costs by minimizing the surface area of the cans produced. To package 12 ounces of soup, the manufacturer has calculated the height and width of the can which minimizes the amount of tin used. No other design uses less tin to package the same soup. This is an example of an optimal solution to a problem: From the space of candidate solutions, the optimal solution is the one which cannot be improved on. In this example, the relationship between variables and solution is certain. Solid geometry and the precision of our measurements ensure a

Chapter from "Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020," edited by G. Gigerenzer and J.A.M. Gray. Strüngmann Forum Report, vol. 6, J. Lupp, series ed. Cambridge, MA: MIT Press. ISBN 978-0-262-01603-2

close fit between our model and the real world. For most problems, including diagnostics, this kind of certainty will not exist. Optimization can nevertheless still be carried out, but in practice, one moves closer to the solution that is optimal only relative to a set of assumptions.

What characterizes the process of optimization? If we lacked knowledge of solid geometry, iteratively fine-tuning the tin can dimensions until the surface area is at a minimum would be a process of optimization. This process assumes that we can the measure the effect of changes to the parameters, and this measure serves as a proxy for performance. If we had knowledge of solid geometry, then we could derive the optimal solution directly. Both are examples of optimization. Broadly speaking, optimization is any process which (a) explicitly maximizes some criterion and (b) assumes that this criterion is monotonically related to performance. As a consequence, optimization methods have a strong tendency to assume that more computation leads to greater precision.

Everyday examples include the method of least squares, which minimizes the sum squared error between the model and observations. In Bayesian statistics, the aim is to select the model with the highest posterior probability, or perform model averaging to determine the most probable prediction. Underlying this approach is the view that the optimal model (or prediction) is "out there," and optimization provides a way of approaching it. Although these approaches are often given rational justification by appealing to probability theory, exact optimization methods are largely fictitious for real-world problems. The computational demands required to conduct the search are often too great. Instead, inexact and approximate methods are used. Thus, there is no single "process of optimization," since in practice additional assumptions often need to be made.

Satisficing, an approach first proposed by the Herbert Simon, offers an alternative to optimization (e.g., Simon 1990). Instead of attempting to maximize some criterion, a satisficer might ignore several factors seen as crucial to an optimizer and seek a good enough (or better) solution. As well as ignoring information, a satisficer will typically not attempt to maximize any criterion. Examples of satisficing include restricting attention to unit weights (either –1 or 1) in a linear regression (Dawes 1979; Einhorn and Hogarth 1975; Wainer 1976), ignoring covariation between cues (Langley et al. 1992), or simply relying on a single cue to make a decision (Gigerenzer and Goldstein 1996; Holte 1993). Strategies like these raise the following question: Why should we even entertain the idea of satisficing? After all, deliberately using impoverished models, such as unit weights, and deliberately curtailing search, as opposed to performing an exhaustive search, both seem potentially foolhardy policies. If these strategies seem ill-advised, then consider the examples given below. Keep in mind that these examples essentially pit two fictional statisticians, A and B, against each other. Statistician A, an optimizer, and statistician B, a satisficer, share the same knowledge of the problem, but they differ in how they approach the problem.

Chapter from "Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020," edited by G. Gigerenzer and J.A.M. Gray. Strüngmann Forum Report, vol. 6, J. Lupp, series ed. Cambridge, MA: MIT Press. ISBN 978-0-262-01603-2

Optimizing versus Satisficing

First, consider how a retail marketing executive might distinguish between active and nonactive customers. Experienced managers tend to satisfice, by relying on a simple hiatus heuristic: Customers who have not made a purchase for nine months are considered inactive. Yet there are more sophisticated methods, such as the Pareto/NBD (negative binomial distribution) model, which considers more information and relies on more complex computations. When tested, however, these methods turned out to be less accurate in predicting inactive customers than the hiatus rule (Wübben and von Wangenheim 2008). Second, consider the problem of searching literature databases, where the task is to order a large number of articles so that the most relevant ones appear at the top of the list. In this task, a "one-reason" heuristic using limited search outperformed both a "rational" Bayesian model that considered all of the available information and PsychINFO (Lee et al. 2002).

Third, consider the problem of investing money into N funds. Harry Markowitz received the Noble Prize in economics for finding the optimal solution: the mean-variance portfolio. When he made his own retirement investments, however, he did not use his optimizing strategy, but instead relied on a simple heuristic: 1/N (i.e., allocate your money equally to each of N alternatives). Was his intuition correct? Taking seven investment problems, a study compared the 1/N rule with fourteen optimizing models, including the mean-variance portfolio and Bayesian and non-Bayesian models (DeMiguel et al. 2009). The optimizing strategies had ten years of stock data to estimate their parameters, and on that basis had to predict the next month's performance; after this, the ten-year window was moved one month ahead, and the next month had to be predicted, and so on until the data ran out. 1/N, in contrast, does not need any past information. Despite (or because) of this, 1/N ranked first (out of 15) on certainty equivalent returns, second on turnover, and fifth on the Sharpe ratio, respectively. Even with their complex estimations and computations, none of the optimization methods could consistently earn better returns than this simple heuristic.

Now, after being informed of these results, statistician A (the optimizer) might raise the following objection: These examples fail to provide an argument against optimization because, quite clearly, the "optimization" models are suboptimal. Statistician A is correct, but his argument is irrelevant. Recall that there is no single process of optimization, and it will always be possible to find an alternative optimization model, after the fact, which outperforms the satisficing method proposed before the future was observed. Obviously, this is an entirely different question unrelated to the substantive issue of designing a diagnostic instrument which predicts well in the face of uncertainty, before the future has been observed.

Why Satisfice?

There are two reasons to satisfice. First, as the examples given above illustrate, satisficing methods can predict with greater accuracy than optimizing

Chapter from "Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020," edited by G. Gigerenzer and J.A.M. Gray. Strüngmann Forum Report, vol. 6, J. Lupp, series ed. Cambridge, MA: MIT Press. ISBN 978-0-262-01603-2

methods, particularly when facing real-world uncertainties. Indeed, this discussion would be incomplete without mentioning a diverse collection of analytic results and arguments in support of satisficing. Over several decades, research into judgment and decision making has provided support for the robustness of unit weighing schemes and simple cognitive heuristics (Gigerenzer and Brighton 2009). In data mining and knowledge discovery, Domingos (1999) lists numerous cases in which methods that restrict search (a key concept in satisficing) can outperform methods that attempt to optimize formal measures of simplicity. In pattern recognition and machine learning, the naïve Bayes classifier ignores information by assuming variables are conditionally independent. Despite this, the naïve Bayes classifier often outperforms more sophisticated methods which consider these dependencies, even when these dependencies are known to exist (Domingos and Pazzani 1997). In statistics, Hand (2006) questions the assumed progress associated with evermore sophisticated optimizing methods by invoking some of the arguments presented here, as well several others. In forecasting, Makridakis and Hibon (1979) questioned the common assumption that simple models often pay the price of reduced predictive accuracy. They compared 22 forecasting models and found that a simple model which ignores observations outperformed more complex models on 111 problems. Each of these contributions supplies pieces of an emerging picture, where satisficing, and related concepts, provide principles for engineering robust diagnostic instruments.

The second reason to satisfice is that the resulting models tend to be easier to use and understand. The examples given above included a simple rule which can be used to predict inactive customers, a single criterion which can be used to search for documents, and a simple method for allocating money to a collection of funds. To take another example, when a patient arrives at hospital with acute chest pain, a doctor must decide if the patient should be admitted to coronary care unit or a regular bed. Pozen et al. (1984) developed the Heart Disease Predictive Instrument (HDPI) to help doctors make this decision. HDPI is based on a logistic regression and requires doctors to consider seven variables and then conduct a series of calculations with the aid of a chart and pocket calculator. HDPI improved decision making, but doctors found it difficult to use. Green and Mehr (1997) went on to shown that, after withdrawing the HDPI, doctors continued to make better decisions, rather than revert to their original performance levels. How can this finding be explained? Green and Mehr hypothesized that the doctors had learned which variables were relevant, and then applied a simple cognitive strategy to make decisions using this subset of these variables. Investigating further, Green and Mehr developed the simple decision tree, shown in Figure 17.4 (based on the take-the-best heuristic; Gigerenzer and Goldstein 1996), to replace the HDPI. This tree poses a short series of yes/no questions, enabling doctors to make decisions without carrying out any additional calculations or looking up numbers in a chart. Crucially, this decision

Chapter from "Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020," edited by G. Gigerenzer and J.A.M. Gray. Strüngmann Forum Report, vol. 6, J. Lupp, series ed. Cambridge, MA: MIT Press. ISBN 978-0-262-01603-2

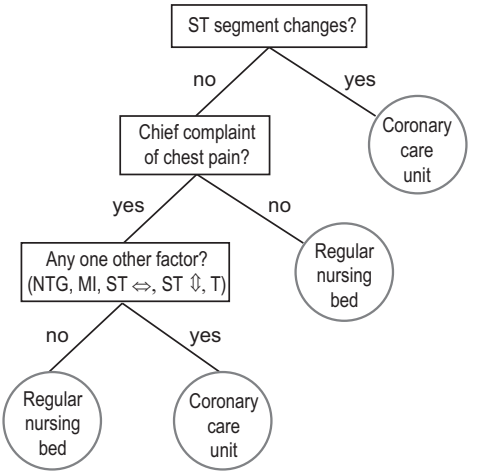

Figure 17.4 A simple "fast and frugal" decision tree used in coronary care unit allocation. ST ↔: flattening of ST segment; ST ↑↓: elevation or depression of ST segment; NTG: nitroglycerin; MI: myocardial infarction; T: T-wave.

tree proved to be more accurate in providing patients the appropriate care, and was far easier to use.

Problems in health care are shaped by immensely complex interactions between biological and social systems. Traditionally, there has been a tendency to believe that the best response is to develop statistical models and diagnostic systems which respond, in kind, by combating complexity with complexity. Satisficing offers an alternative viewpoint and suggests that the complexity of optimization is better suited to simple problems where we can be sure that we are optimizing the correct criterion. For complex problems involving high degrees of uncertainty, our assumptions will often fail to hold and operating conditions will change in unexpected ways. Satisficing can simultaneously offer a more robust response to such perturbations and simpler, easier to use, diagnostic instruments.

Chapter from "Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020," edited by G. Gigerenzer and J.A.M. Gray. Strüngmann Forum Report, vol. 6, J. Lupp, series ed. Cambridge, MA: MIT Press. ISBN 978-0-262-01603-2