INFORMS TutORials in Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Healthcare Analytics: Big Data, Little Evidence

Joris van de Klundert

To cite this entry: Joris van de Klundert. Healthcare Analytics: Big Data, Little Evidence. *In* INFORMS TutORials in Operations Research. Published online: 04 Nov 2016; 307-328.
https://doi.org/10.1287/educ.2016.0158

TUTORIALS IN
OPERATIONS RESEARCH
INFORMS 2016

 | ISBN 978-0-9843378-9-7
http://dx.doi.org/10.1287/educ.2016.0158

Healthcare Analytics: Big Data, Little Evidence

Joris van de Klundert
Institute of Health Policy and Management, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands, vandeklundert@bmg.eur.nl

Abstract While the healthcare sector contributes more than 10% of gross domestic product in most developed countries and is approaching 20% in the United States, it remains a relatively modest area in the fields of operations research, management science, and analytics. There is considerable room for a larger and more valuable contribution, especially in view of the important advancements in information technology taking place in healthcare across the globe, which are already contributing to a reduction in the global burden of disease. For analytics professionals and scientists to reach the full contribution potential of their discipline, it is beneficial to understand the dominant research paradigms and results of clinical and health sciences research. These sciences are rooted in empirical evidence, in empirical data, thus offering connection opportunities. In this tutorial we review the current position of analytics as covered in the operations research and management science literature, and outline a path for the science of analytics to enlarge its contribution to the health of populations.

Keywords healthcare; analytics; big data; evidence based

1. Introduction

Since approximately 2005, INFORMS has referred to its scientific domain as "the science of better." This phrase proudly brands the professions of operations research, management science, and analytics, where analytics was added in 2007. INFORMS defines analytics as the scientific process of transforming data into insight for making better decisions. Careful reading of this definition suggests that the purpose of analytics goes beyond improving the process of decision making itself. The definition, and in particular the *adjective* "better," doesn't just refer to the procedures or algorithms applied to reach decisions but explicitly to the decisions. Analytics is positioned as a process that increases the quality of decisions made. Then what makes a decision a better decision? If being related to features of decisions beyond the steps by which they are construed, the quality of a decision may lie in the effectuation. A decision is better than alternative decisions (or better than no decision) if implementation of the decision leads to better results than implementation of alternatives (or no implementation). A decision is then considered to be better if it has a better effect on reality, on the world. While this may not be the only possible interpretation of "making better decisions," it does form the basis for the exploration of healthcare analytics in this tutorial. It firmly grounds analytics in practice and positions analytics, at least partly, as an empirical science.

Our next question might then be, when can we qualify our world to have become better? The answer to this long-standing philosophical question lies far beyond the scope of this tutorial and, in fact, beyond the scope of the science of analytics. Nevertheless, it will be readdressed throughout this tutorial—sometimes from a methodological perspective, sometimes just pragmatically. As healthcare analytics is our topic of interest, let us narrow down the ambitions of "better" to apply to health of humans. The definition of *health* as

agreed by the World Health Organization since 1948 reads, "Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (World Health Organization [53]). The research question we consider to answer in this tutorial, and for which we provide methods as well as reflection, therefore is as follows: Can analytics improve health? Or, more elaborately, can scientific transformation of data into insights for decisions improve human health? Because of the empirical nature of this question, we seek answers to this question based on empirical evidence. Adopting a systematic approach to the existing scientific literature, we also review scientific empirical evidence that analytics can improve health.

As a result, the purpose of this tutorial is not to sharpen the technical operations research or analytics skills of the reader. The value of this tutorial hopefully lies in advancing the readers understanding of health, in providing scientific contributions that improve health, and in providing empirical evidence of such improvement. This tutorial teaches to evidently practice the science of better in the healthcare domain.

2. Health and Health Improvement

A first and basic approach to assessing health is to consider the health of a population. For instance, we may consider the health of the population of the United States of America. Alternatively, we may consider the health of a subpopulation, such as the African American population of the United States, the population of U.S. truck drivers, or the population of citizens of the United States suffering from dementia. A basic measure to express the health of a given population at a certain moment in time can now be to dichotomously count the health of a person alive as one and the health of a person who has lost life as zero. For a given time interval and a given population alive at the start of this time interval, this dichotomous measure implies that the more persons who stay alive, and the longer persons stay alive over the time interval, the larger health. As the population is given, this holds true both for total health and for average health. Put differently, the lower the mortality rate, the larger the health of the population. For the proposed measure, improvement of (expected) health of a given population then translates to improvement of the (expected) average length of life. Figure 1 shows that life expectancy has increased by almost 30 years for the population of the United States since 1900. Notice also how the difference between the general population and the population of African Americans has decreased since 1900, but it still is more than three years. Laden et al. [18] find that for a population of 54,319 U.S. truck drivers, the median age of death was 61.9, more than 10 years less than for the general U.S. male population. The relationship between dementia and life expectancy will be addressed more extensively below.

It is well researched that human valuation of well-being highly depends on health status (see, for instance, Fryback et al. [11] for instruments and validation in the United States). Hence, a longer life expectancy (LE) is not necessarily valued to imply better health. Disability-adjusted life years (DALYs) form a common measure to value the impact (loss of) health has on well-being. The DALY metric has been adopted by the World Health Organization and the Institute for Health Metrics and Evaluation (Murray [27], Murray and Lopez [28]). A DALY stands for one year in healthy life lost. More precisely, for person p in year y, $DALY_{py}$ refers to the loss of health of person p in year y. A year of perfect health corresponds to $DALY_{py} = 0$, and $DALY(py) = 1$ represents a full year of death. As considered more explicitly below, a person p may lose future years of life in year y, in which case the corresponding burden of disease is attributed to $DALY_{py}$. Hence, 1 is not an upper bound on $DALY_{py}$.

For adjusting life expectancy for loss of value due to present and future disability, it has been proposed to take into account that people value future health differently from present health. Hence, rather than simply adding up the sum of disability-adjusted life years over

FIGURE 1. Life expectancy in the United States since 1900.

a person's lifetime, health-adjusted life expectancy (HALE) may consider a discounted sum of future DALYs. Because empirical research on discount rates is contradictory (Severens and Milne [40]) on the stated discounting valuations, and discounting has also been met with ethical concerns (Murray and Lopez [28]), we disregard discounting in the remainder of this tutorial and simply define, for any human person p, health-adjusted life expectancy $HALE(p)$ as

$$HALE(p) = \sum_{y=0}^{LE(p)} (1 - DALY_{py}). \tag{1}$$

Likewise, for population P and year y,

$$DALY(P, y) = \frac{\sum_{p \in P} DALY(py)}{|P|}, \tag{2}$$

and the $HALE(P)$ is defined as

$$HALE(P) = \frac{\sum_{p \in P} HALE(p)}{|P|}. \tag{3}$$

Note that population health considers the average health of a population, rather than the sum, as is more insightful for comparison of the (average) disability among populations (see, for instance, Table 1).

The aforementioned quantitative valuations of health and disability (as well as related valuations) have been commonly accepted and adopted measures of effectiveness in evaluation frameworks by the scientific community, national governments, and international organizations. However, these effectiveness measures are commonly considered for decision making in relation to cost. A small improvement of health-adjusted life expectancy against high cost is considered less valuable than a larger improvement against the same costs, or a similar

improvement against lower cost. To further elaborate this notion of *cost-effectiveness*, let us consider a population P with a condition that can be treated by treatments t and s. Let $HALE(P,t)$ (respectively, $HALE(P,s)$) denote the health-adjusted life expectancies of population P, conditional upon receiving treatment t (respectively, treatment s). We say treatment t is *effective* in comparison to treatment s for population P if $HALE(P,t) \geq HALE(P,s)$.

Developing the notation for cost, we first define for person p in year y and treatment t the cost of treatment for person p in year y as $cost(p,t,y)$. Moreover, we define

$$cost(p,t) = \sum_{y=0}^{LE(p)} cost(p,t,y). \tag{4}$$

It is not uncommon to discount future costs, but for uniformity of definitions, we disregard discounting of costs. Next, for treatment t and population P, we define

$$cost(P,t) = \frac{\sum_{p \in P} cost(p,t)}{|P|}. \tag{5}$$

Now, a treatment t is considered *cost effective* in comparison to another treatment s (e.g., care as usual) if and only if

$$HALE(P,t) - (HALE(P,s) \geq T \times (costs(P,t) - costs(P,s)), \tag{6}$$

where T is a cost-effectiveness threshold value. In other words, treatment t has to provide at least T units of improvement in health-adjusted life expectancy per dollar over treatment s to be considered cost effective in comparison to s. Note that this definition may also consider treatment t as cost effective in comparison to treatment s if t results in poorer health outcomes yet is sufficiently less costly. In practice, cost-effectiveness analysis is, however, typically applied to assess new treatments t that are effective in comparison to present care-as-usual treatment s.

While we base our line of reasoning completely on DALYs and disregard quality-adjusted life years (QALYs) for ease of exposition and consistency of terminology in this tutorial, our notation broadly follows common definitions of cost-effectiveness and the incremental cost-effectiveness ratio (ICER) (Drummond et al. [9]). Many countries and international organizations such as the World Health Organization and the World Bank practice cost-effectiveness as a leading decision-making criterion, implying that they consider it to lead to better decisions. Cost-effectiveness also forms the basis of the often-advocated "health per dollar" measure of value-based healthcare (Porter and Teisberg [34]). Let it be noted, however, that cost effectiveness has been met with criticism and is not globally adopted. Despite favorable attitudes toward value-based healthcare in the United States, and the recommendation of the U.S. Panel on Cost-Effectiveness in Health and Medicine to use cost-effectiveness analyses as a standard metric for identifying and assigning value to health outcomes, cost-effectiveness has been explicitly barred as a criterion for Medicare and Medicaid decision making in the Affordable Care Act (see Neumann and Weinstein [31] and the references therein).

Prioritizing the reduction of the burden of disease, and hence focusing on effectiveness rather than on cost-effectiveness, we view the science of healthcare analytics as the science of transforming data into insights for decisions to improve health for individuals and populations. Moreover, in this tutorial, we define improving health as increasing the health-adjusted life expectancy of populations and individuals. Before advancing, let us consider the empirical data on life expectancy and health-related life expectancy for the United States and the world, presented in Table 1 (data extracted from Murray et al. [30]).

TABLE 1. Life expectancy and health-adjusted life expectancy for various populations in 2010.

Population	Male LE	Male HALE	Female LE	Female HALE
Global	68.8	60.6	74.3	64.1
Developed	75.5	66.0	81.8	70.0
United States	76.3	65.8	81.4	68.6
Canada	79.4	69.1	83.4	71.0
Western Europe	78.6	68.2	83.7	71.1
Australasia	79.5	68.4	83.8	70.6
Japan	80.0	71.1	86.4	75.6
Cuba	76.3	67.3	80.4	69.0
Sub-Saharan Africa	58.8	51.0	61.6	52.8

While the data presented in Table 1 are insightful by themselves, they also raise a number of questions that invite further analysis. What explains the differences between the United States and comparison countries? For instance (and more specifically), which diseases cause U.S. citizens to have a lower (health-adjusted) life expectancy than Canadians? Which diseases incur the largest loss of health, i.e., the largest burden of disease? For the purpose of further analysis to answer these questions, we further operationalize the burden of disease by $DALY(P,y)$ for a given population P in year y by

$$DALY(P,y) = YLL(P,y) + YLD(P,y), \tag{7}$$

where $YLL(y)$ refers to the health-adjusted life years lost by members of the population P in year y because of premature death, and $YLD(y)$ is a measure for the burden of disability in the population experienced in year y itself (Murray et al. [29]).

The measure $YLL(y)$ refers to the burden of mortality in year y and is slightly more complex than $YLD(y)$. The term "premature death" needs elaboration. Death is considered premature when occurring earlier than in a virtual reference population P^*. The reference population P^* is now obtained by taking, for each age group a, the population $P^*(a) = \arg\max_{P \in C} LE(P,a)$, where P is taken from the set of all relevant comparison populations C, and $LE(P,a)$ refers to the life expectancy of persons of age a in year y. Now, summing the differences $HALE(P^*,a) - HALE(P,a)$ over all age groups a in P, weighted by their relative cardinalities, we obtain the difference in life expectancy of the population P with the virtual reference population P^* in year y, which defines $YLL(P,y)$. Notice that, by definition, $YLL(P,y) \geq 0$ if $P \in C$. Taking C to be the set of country populations, the life expectancy of population P^* equals 86.0 for the year 2010 (Murray et al. [29]).

In the United States, "the diseases and injuries with the largest number of YLLs in 2010 were ischemic heart disease, lung cancer, stroke, chronic obstructive pulmonary disease, and road injury," and "the diseases with the largest number of YLDs in 2010 were low back pain, major depressive disorder, other musculoskeletal disorders, neck pain, and anxiety disorders" (U.S. Burden of Disease Collaborators [41]). Moreover, these collaborators find that "as the US population has aged, YLDs have comprised a larger share of DALYs than have YLLs" and establish dietary risks, tobacco smoking, high body mass index, high blood pressure, high fasting plasma glucose, physical inactivity, and alcohol use as the leading risk factors. Note that most of these diseases are noncommunicable diseases, and that all of the mentioned risk factors are lifestyle related.

While Table 1 may create a sense of urgency for improvement in the United States, the data on sub-Saharan Africa witness that the burden of disease does not weigh heaviest on high-income countries but on low-income countries, where financial resources, medical staff, data, and scientists are much scarcer. Most of the betterment opportunities lie where analytics professionals are fewest. For instance, the per-capita burden of disease in sub-Saharan

Africa is approximately three times higher than it is in high-income countries. The difference in per-capita burden of YLLs is even considerably higher, as may already be inferred from the lower life expectancy. The resulting younger population still carries a burden of living with disability (YLD) that is comparable to the YLD of other populations. The (mostly communicable) diseases with the largest burden of diseases in low-income countries are neonatal infections (and prebirth conditions), lower respiratory infections, diarrhea, malaria, and HIV. The leading causes of disability for low-income countries are childhood malnutrition, water/sanitation/handwashing, unsafe sex, dietary risks, air pollution, and high blood pressure, which together account for roughly half of the burden of disease in low-income countries (Institute for Health Metrics and Evaluation [15]).

Clearly, the data presented in this section may assist analytics scientists in setting priorities in selecting conditions for which to improve health through insights and better decisions. Following the definitions and purposes of analytics sketched above, the contribution of analytics is larger if the decisions contribute more to reducing the burden of disease. From a global perspective, this directs priority toward populations with a high burden of disease. Within the United States, the data also help to prioritize populations, diseases, and risk factors. Below, we continue along these lines of priority. Before doing so, we note, however, that the data used above are partly taken from the University of Washingtons Institute for Health Metrics and Evaluation, which has benefitted from a $105 million grant from the Bill and Melinda Gates Foundation (Mankowski [22]). More recently, both the University of Michigan and the Data for Health Initiative have also initiated large-scale developments for big data in health with investments of the same order of magnitude (Centers for Disease Control and Prevention [4], Michigan institute for data science [24]). Much data are already being collected and transformed into insights; huge analytics efforts have already been made and are being made and scaled up to advance global health.

3. Evidence-Based Health Analytics

The definition of analytics proposed by INFORMS can be viewed to have a sequential nature, stepping from data to insights to decisions. Above, we have seen that large-scale data transformation processes are already taken care of by global initiatives. Moreover, publicly accessible tools, presentations, and visualizations—as offered by the Institute for Health Metrics and Evaluation, for example—provide insights. These insights follow from *descriptive* research. The last step in the sequence of the definition of analytics, making better decisions, is of a *prescriptive* nature. The decisions prescribe actions, interventions, which intend to result in better health, e.g., higher HALE for individuals and populations. Now, obviously, between a decision and an effect on HALE stands the implementation of the decision. In the prescriptive analytical models developed by operations researchers and management scientists, implementation is often taken for granted. While fully aware that models are simplifications of reality, the model developers assume that the effects of changing the values of the decision variables translate into the effects specified by the mathematical relations of the model. The results of the analysis of the model are acceptable and accepted as the results of the scientific contribution. It is not viewed as necessary to implement the decisions and report the empirical evaluation of the implemented decisions, thus verifying whether the intended effects to improve health have been obtained, and to see which other effects (on health) have empirically resulted from the implementation of the decisions.

Meanwhile, health sciences have taken a different, radically empirical route, referred to as *evidence based*, in their approach to decision making and developing science to improve decision making. As an introduction to this approach, let us consider the influential definition of evidence-based medicine proposed by Sackett et al. [38]: evidence-based medicine is the "conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients" (p. 71). The same authors propose that it "means integrating

Table 2. Does this intervention help?

Level	Evidence
1	Systematic review of randomized trials or n-of-1 trials
2	Randomized trial or observational study with dramatic effect
3	Nonrandomized controlled cohort/follow-up study
4	Case series, case-control studies, or historically controlled studies
5	Mechanism-based reasoning

Source. Oxford Centre for Evidence-Based Medicine.

individual clinical expertise with the best available external clinical evidence from systematic research" (p. 71). Clearly, this definition implies the normative viewpoint that decisions are better as they make more conscientious, explicit, and judicious use of current best evidence and hence as they rely on better evidence as importantly obtained through systematic research.

Over the last two decades, the adoption of *evidence-based practice*, which covers a broader application domain than making decisions about the care of individual patients, has had a considerable impact on health service practice and on science. For practice, it has caused health service provider organizations to systematically introduce and adopt evidence-based practices. For health sciences, it has led to large-scale systematic efforts to generate evidence of sufficient quality to trust that implementation of decisions based on this evidence improve health. Now, following the definitions of health provided above, this asks for evidence that implementation of the decisions cause improvement of HALE for individuals and populations. Following the evidence-based paradigm, we find that a decision (e.g., to provide a new treatment) is considered better than an alternative (e.g., to continue care as usual) if and only if there is evidence that it yields better results. Such evidence on a treatment decision can take on several forms, which we detail out below through a well-known scheme to classify clinical evidence into various levels based on the "strength" of the evidence (Howick et al. [14]); an explanation of terminology can be found in Table 2.

The dominant research design for high-quality evidence is the randomized controlled trial (RCT). A randomized controlled trial is an experimental study, i.e., an empirical study, where the data to evaluate the interventions are collected from an experimental environment designed to exclude external influences on the measured effects. The trial of the intervention is set to be controlled if the same setting is offered to a group of research subjects (e.g., patients) exposed to the intervention and to a group of research subjects not exposed to the intervention. The evidence then consists of significant differences in effects between the thus defined intervention group and control group, or of the absence of significant differences. The experiment is randomized if the subjects are randomly assigned to either of the groups, thus preventing biases that may, for instance, arise from patient choice. The evidence obtained from a single randomized trial is classified as level 2 evidence. Level 1 evidence can be obtained by systematically reviewing (or analysing) all available level 2 evidence.

Level 3 evidence results from less rigorously designed experimental studies, such as a nonrandomized trial. It may also be obtained through nonexperimental yet empirical research. Such research is called observational. (In fact, an observational study with a dramatic effect is even considered level 2 evidence.) A standard design for an observational study is the cohort study, where a fixed population, called a cohort, is followed over time. As time passes, the data may then reveal statistically significant relations between interventions (decisions) and effects (which may occur later). Such studies may reveal that within a cohort of persons with full-time jobs in 1985, the probability of being alive in the year 2000 is significantly lower for those who have continuously worked as truck drivers (until retiring) than for the

cohort in general. Notice that in this case, the cohort study can be viewed as best evidence, because an experimental design where persons are randomly designed to professions is neither ethical nor feasible.

The level 4 classification mainly applies to observational studies in which a population is defined based on being exposed to a certain treatment (case series) or having a certain effect which is absent in the control group (case control). In such studies, the attribution of effects to the intervention is methodologically more constrained, as there are no controls, and confounding factors may be difficult to include in comparison to well-designed cohort studies or experimental studies.

Finally, level 5 evidence is provided by mechanism-based reasoning. Mechanism-based reasoning involves an "inference from mechanisms to claims that an intervention produces a patient-relevant outcome. Such reasoning will involve an inferential chain linking the intervention (such as antiarrhythmic drugs) with a clinical outcome (such as mortality)" (Howick et al. [13], p. 434). Mathematical models that define relationships between variables in constraints and a valuation function can be viewed as inferential chains linking interventions in the form of changes in the values of decision variables to effects in the form of changes in the valuation function, namely, HALE. In other words, results obtained by the nonempirical models commonly used in the operations research and management science communities are qualified as the lowest-quality evidence in a widely adopted classification scheme of evidence-based medicine. This holds true, even if the modeled relationships are based on empirical evidence. Perceived weaknesses of such models, or more generally of mechanism-based reasoning, are as follows: (1) the proposed mechanism follows from a flawed theory, (2) the proposed mechanism lacks theoretical support or evidence, (3) the proposed mechanism is evidence based but ignores some important factors, (4) the probabilistic nature of relationships is ignored, and (5) the model is an oversimplification and ignores relevant related mechanisms (adapted from Howick et al. [13]).

4. Empirical Example: Evidence-Based Management of the Pharmaceuticals Supply Chain in Zambia

Before addressing some of the methodological concerns arising from the differences between clinical sciences and medical decision making and the science of analytics, let us first consider an empirical supply chain management study that delivered evidence of improving HALE (Vledder et al. [46]). The study is an experimental study to test three different designs for managing the pharmaceutical supply chain in Zambia, two different interventions and one care-as-usual design. The designs vary with respect to inventory replenishment policies and structures as follows. The control group adopts the current supply-as-usual practice, where each district has a distribution center that orders from the national warehouse, and the facilities order at the district distribution center. Intervention A consists of training a commodity planner at the district level and having this planner take care of the replenishment. Intervention B additionally adopts a cross-docking approach, where the orders of the facilities are sent to the national warehouse and packed for delivery by the national warehouse. The district center now only serves as a cross-docking facility where the shipments received from the national warehouse are loaded onto local transport vehicles for delivery, without storage. The Vledder et al. [46] study adopts a "cluster randomized evaluation design, with randomization of delivery models conducted at the level of the district" (p. 7). Each district is either an intervention cluster A, an intervention cluster B, or a control cluster, as randomization at, for instance, the level of facilities was likely to experience intervention spillover effects on the control facilities. In the evaluation, the authors considered before–after differences in observed stockouts in each of the clusters. In other words, for each district they measured the stockout frequencies before the start of the experiment and the stockout frequencies after having implemented the aforementioned interventions (or no intervention in the control districts). Moreover, they compared the before–after differences per intervention cluster

with the before–after differences for the control cluster to correct for confounding national developments that may have occurred during the experiment. This difference-in-difference analysis of the randomized controlled experiment showed that cluster A had significantly fewer stockouts than the control cluster for 4 of 15 selected medicines in the set (where significance is set at the 5% level). Likewise, cluster B performed significantly better than cluster A on 4 out of 15 selected medicines and better than the control cluster on 10 out of the same set of 15.

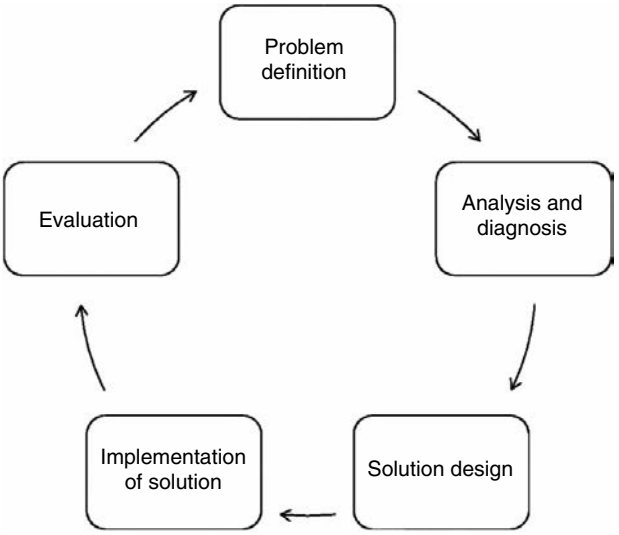
This research clearly demonstrates how studies can be designed to generate high-quality evidence (level 2) and contribute to an evidence base for analytics. Let it be noted, however, that the evidence provided regards stockouts, not health or health-adjusted life expectancy. To assess the effects on health-adjusted life expectancy, the study relies on mechanism-based reasoning. Using a detailed set of evidence-based inferences, Vledder et al. [46] reason that national scale-up of intervention B leads to a reduction in DALYs by averting at least 770,615 YLLs and a cost-effectiveness of USD 22 per YLL averted when implementing intervention B for a five-year period. As much as possible, the mechanism-based reasoning is evidence based. For comparable purposes of mechanism-based reasoning, de Vries et al. [7] systematically review evidence on the relationship between the distance to a health service facility and the effectiveness of treatment for truck drivers in sub-Saharan Africa, as required to optimally locate facilities.

5. Designing and Evaluating Evidence-Based Decisions

While the evidence-based approach has not been without criticism, it has advanced and has also entered the scientific discipline of management. The introduction of evidence-based management in healthcare traces back to the turn of the millennium (Jadad et al. [16], Walshe and Rundall [48]), and it has been introduced to management sciences in general by Rousseau [36] and others (see Center for Evidence-Based Management [3] for extensive additional resources). As argued already by Walshe and Rundall [48], the principles of evidence-based medicine cannot be directly transferred to evidence-based management but need translation. Two of the main limitations that come into play when building an evidence base for management are (1) management outcomes often regard the organization as a unit of analysis instead of the person, thus making it difficult to do large-scale experimental studies as required to contribute strong evidence from statistically significant experimental results, and (2) organizations often form essentially different research subjects, such that the external validity of evidence obtained in a set (population) of organizations may not be valid in others. For an example of the latter, we should not expect that bringing the aforementioned intervention B in the pharmaceutical supply chain from Zambia to the United States will deliver comparable results. For an example of the former, one might also view the Zambia pharmaceutical supply chain study as a single case study, as it regards only the research subject Zambia. Subsequently, the evidence can be classified as weak despite its randomized experimental nature and the large number of persons involved.

While the aforementioned limitations call for caution when appraising evidence regarding managerial decision making, they do not take away the evidence-based orientation of the decision making in healthcare. At management and policy levels, decision makers ask themselves how to put the scarce financial resources to work to best reduce the burden of disease. What are the most cost-effective decisions, given the limited budget? While such decisions problems form an opportunity for analytics to support making better decisions, these problems also form a threat. Analytics can be marginalized if there is no high-quality evidence of its effectiveness. Hence, evidence based practice in health services calls upon analytics not only to transform data into insights for better decisions but also to collect data on the implementation of these decisions, to evaluate the decisions, and thus to contribute to new data, etc. It calls upon analytics to engage in the empirical research cycle

FIGURE 2. The problem-solving cycle.



and show that analytics cost-effectively improves health, whether for individual patients, for populations, or for the organizations/systems involved. This cyclic improvement perspective is captured by the problem-solving cycle model of van Aken et al. [43] and Otte-Trojel et al. [32], depicted in Figure 2.

The cycle first distinguishes the stage of problem definition (e.g., the relative high burden of disease of truck drivers) and next the stage of analysis and diagnosis (e.g., lack of physical exercise, poor access to health services, social behaviors). The next step is a solution design (prevention programs, health service facility location), after which implementation and evaluation follow. Typically, the evaluation reveals next improvement opportunities—for instance, when the implemented solution has not resulted in the expected effects—and hence identifies new problems in turn. Repeated execution of the problem-solving cycle defines a development process, as opposed to a single turn design process. Development processes are especially valuable when involving interventions in complex systems such as health systems and organizations (where a causal inference from intervention to effect is more likely to be incomplete or incorrect) (van Aken et al. [43]).

The problem-solving cycle benefits from an evidence-based approach in several of the stages. First, the problem analysis stage may involve a review of the literature and evidence on the problem defined in the first stage. In many cases the problem at hand has already been addressed and reported in the scientific literature. Ideally, the literature reports solutions for the problem at hand, including the evaluation of the implementation of these solutions. The analysis and diagnosis can then benefit from evidence on hypothesized causes and solutions.

Second, the available evidence may be included in the solution design, i.e., in the decision making. For this stage, it is especially relevant for evidence-based management to appraise relevant evidence, i.e., to assess the strength and validity, and then adapt it to the problem and context at hand. For example, when trying to improve the availability of medicine in health facilities in Malawi, the evidence from Zambia may serve as a basis for an improved design in Malawi. Of all stages, the design stage forms the prime opportunity for the conscientious, explicit, and judicious use of current best evidence in making decisions.

Third, the problem-solving cycle can advance the evidence base by setting up a well-designed implementation of experiments and reporting the findings obtained through scientific evaluation. When analytics thus contributes full circle to improvement of health, it joins other sciences in the health domain to use and produce evidence and to use and produce

Table 3. Diseases with the highest burden of disease, globally and in the United States, plus the number of empirical papers in the OR literature.

Rank	Disease (global)	No. of pubs	Disease (United States)	No. of pubs
1	Ischaemic heart disease	12	*Ischaemic heart disease*	12
2	Lower respiratory infections	1	*COPD*	0
3	Stroke	8	*Low back pain*	1
4	Diarrhea	1	Lung cancer	10
5	HIV/AIDS	48	*Major depressive disorders*	2
6	Malaria	4	(Other) Musculoskeletal diseases	0
7	Lower back pain	1	*Stroke*	8
8	Preterm birth conditions	3	*Diabetes*	13
9	COPD	0	*Road injury*	—
10	Road injury	—	Drug use disorders	1
11	Major depressive disorders	2	Neck pain	0
12	Neonatal encephalopathy	0	Alzheimer disease	0
13	Tuberculosis	7	Anxiety disorders	0
14	Diabetes	13	Self-harm	0
15	Iron deficiency (anemia)	0	Falls	0
Total		98		47

Note. Diseases that occur in both rankings are in italics in the U.S. listing.

data. In so doing, analytics shows the evidence of its contribution to the improvement of health, and provides justification for having scarce financial resources allocated to analytics.

Let it be noted, however, that analytics cannot be fully evidence based. For novel problems or novel contexts, the existing evidence base may be appraised as being relatively small, weak, or of limited validity and hence an insufficient basis for (partial) designs. While this is viewed to be unlikely in relatively mature fields of research (van Aken and Romme [42]), it may more frequently apply in younger fields, as is the case for analytics. Of course, designs can then still find a scientific basis in theory. In such cases, rigorously designed implementation and corresponding evaluation remains valuable as it initiates an evidence base for the benefit of future decision making in a new area.

6. Contributions to Evidence from the Management Science and Operations Research Literature

Let us now proceed by exploring the empirical contributions made in the operations research, management science, and analytics literature to advance the health of populations and individuals. To this end, we consider the empirical contributions made for the 15 diseases causing the highest burden of diseases globally and in the United States. We first list the diseases in Table 3.

We searched for original articles on these diseases with journal title keywords that ensured inclusion of the top 75 journals from the SCImago 2014 list in the category management science and operations research, as well as articles found in other journals that matched the journal keyword criteria (Scimago Lab [39]) and could be classified to fall in the category of management science and operations research. The systematic search included all publications that appeared in 2015 or earlier. On purpose, the search did not include articles using techniques from operations research, management science, or analytics in other journals, to ensure a view of the development of an evidence base in the discipline's own scientific, peer-reviewed literature. The search targeted original peer-reviewed, full-text publications and hence excluded book chapters, conference proceedings, nonoriginal research (e.g., reviews), articles not in English, and articles for which no abstract or full text was available. Moreover, we only included articles reporting effects of research on DALYs, QALYs, YLL, YLD, or other outcome measures on mortality or morbidity. For instance, many papers reported on

mortality in terms of the number of deaths or in terms of prevalence (percentage of the population with the disease) or incidence (percentage of the populations whose health status changes from healthy to having the disease) of a disease. Articles only reporting results on costs, or operational measures such as waiting time and capacity utilization, but not on health outcomes were excluded. We also excluded articles only reporting hospital/ICU readmissions and hospital length of stay, as these are process measures as well if not further specified. Likewise, papers that are not based on empirical data or use empirical data only to test theoretical advancements were excluded. Following the priority for diseases that cause the largest burden of disease, we only considered articles that considered the top 15 diseases (globally and for the United States), which together cause more than half of the total burden of disease (globally and in the United States).

The search led to an initial selection of 2,757 articles, of which in the end 108 met all inclusion criteria and were included as being original empirical papers analysing data and contributing to an evidence base of the top 15 diseases. In case of doubt, publications remained included. Of these publications, 98 focused on diseases occurring in the global burden of disease top 15. Two publications discussed more than one disease, which explains why the total is smaller than the sum of the row entries in Table 3. Likewise, 47 publications addressing conditions in the U.S. top 15 for burden of disease were included. There is considerable overlap between the diseases in the global top 15 and U.S. top 15.

We excluded the disease category road injury for two reasons. First, many of the analyses regarded preventive measures to improve road safety (e.g., design of highway crossings, changing speed limits, road signaling) and hence fell outside of any reasonable health services scope. Second, the health service follow-up to situations leading to road injury often remained unaddressed. Interestingly, however, we note that the transportation literature does provide an extensive evidence base within the domain of the selected journals.

Conversely, let it be noted that publications have been published on diseases (or conditions) that are not in either list of the top 15 conditions considered but do contribute to an evidence base on improving health outcomes in the management science and operations research literature. Added to the fact that the review may have missed some journals within this domain or did not consider journals outside of this domain, the review should not be considered as a systematic review of evidence to improve health outcomes by the sciences of operations research, management science, and analytics and is not intended as such.

Now let us consider the included selection in more depth. First, we notice that HIV/AIDS is the condition that has received the most attention in a number of publications: almost half of the publications on the diseases in the global top 15 regards HIV/AIDS. Somewhat ironically, many of these publications have addressed the incidence and prevalence of AIDS/HIV in high-income countries in the previous century, when the burden of disease was high in high-income countries and the increases in prevalence and incidence were not well understood. Hence these papers do not address the burden of disease in times and countries that carry the burden of disease that places the condition in the global top 15.

A common way of classifying the conditions is to consider communicable (infectious) diseases versus noncommunicable diseases. The communicable diseases occur predominantly in the global top 15 (e.g., lower respiratory infections, diarrhea, HIV/AIDS, malaria, tuberculosis). While their contribution to the global burden of disease is still large, it has been diminishing because of effective prevention and treatment efforts, as well as because of improving economic and public health conditions (e.g., in terms of sanitation). The U.S. top 15 does not include any communicable disease.

For developed countries, most of the present burden of disease is driven by noncommunicable diseases (and conditions). More often than not, these noncommunicable conditions are chronic, thus causing a prolonged burden of disease in terms of YLD, even if not leading to reduced life expectancy (YLL). Many such chronic conditions are in fact related to longer life expectancy, as is the case for Alzheimer disease, falls, diabetes, ischaemic heart disease, and

stroke. Likewise, many of these chronic conditions are associated with (sedentary) lifestyles and changes in the diets of contemporary populations in high-income countries that are, in fact, occurring across the globe (as is the case for ischaemic heart disease, lower back pain, musculoskeletal diseases, stroke, diabetes, and neck pain, for example). Lung cancer is, of course, highly correlated to smoking and not a chronic condition.

Of all publications included, 97 regarded a descriptive analysis of health outcomes. These 97 publications can be classified to fall into the first two stages of the problem-solving cycle. Most commonly, such papers presented models to describe and forecast incidence, prevalence, mortality, quality of life, or burden of disease, for populations or for individuals. Below we will briefly present two such models, compartment models and Markov models. Discrete event simulation has been applied often as well. Forecasting techniques have also been applied often, both econometric techniques and techniques from artificial intelligence.

Contributions that used descriptive models to evaluate the effects of improvement interventions resulting from explicit design efforts or that resulted in specific improved designs (as opposed to, for instance, simulating the effects of already-existing evidence-based interventions) were classified to belong to the third stage of the problem-solving cycle, i.e., solution design. Thirty-six papers fell into this class. This class also contained prescriptive studies, e.g., studies that solved an optimization model to propose an intervention design and studies that resulted in a decision support system for use by decision makers without follow-up scientific involvement of the researchers.

Seven publications report progress beyond designing solutions. Four report having led to implementation, yet without reporting the effects. Three publications report evidence on health effects resulting from implementation, thus initiating an evidence base for practitioners and researchers to build on. The first of these three is Lathrop and Pazzani [20], who consider the design of an optimal HIV patient treatment protocol (drug regimen) with respect to drug resistance for the HIV patient subpopulation that already has experienced at least one treatment failure due to drug resistance, and they report how the results obtained have been evaluated using a small-scale clinical trial. This appears to be the only article reporting experimental evaluation. The small scale of the experimental setup inhibits the thus found evidence to be classified as level 2 evidence, however. The second study, by de Treville et al. [6], considers a case study on inventory replenishment in the tuberculosis drug supply chain, and the authors link the effects obtained in terms of availability (stockouts) with health effects. The evaluation itself, however, does not report effects on health outcomes but presents first results on lead time improvements from a before–after design. Evidence obtained from such designs is considered weak (level 4) because the effects cannot be attributed to the intervention with certainty. The third article is by Monks et al. [26], who discuss health effects obtained from improvement in emergency stroke care that were aided by a simulation model. Like in de Treville et al. [6], the reported evidence is obtained from a before–after design and is not reported directly in terms of health outcomes but in terms of process indicators, such as a threefold increase in the thrombolysis rate for patients under the age of 80. They refer to the medical literature (Emberson et al. [10] for level 1 evidence about the positive net effects of timely thrombolysis treatments on health outcomes.

For now, we establish that the management science and operations management literature has contributed modestly to empirical research on improving health outcomes for the 15 diseases topping the rankings of the global burden of disease. Disregarding road injury, we find that there has been a considerable contribution to empirical work on HIV/AIDS, while for most other diseases, the systematic search and inclusion did not result in more than a handful of publications. With very few exceptions, the empirical embedding takes the form of using (big) data as an input for analysis and improvement design, but it does not contribute to empirical data on (evaluated) implementation of the improvement designs. The studies report little evidence of effectiveness. The studies that report evidence on the evaluation of implementation deliver relatively weak evidence, which relies on before–after

analysis or small sample sizes. Moreover, some of the evidence relies on mechanism-based reasoning in the sense that the studies do not report effects on health outcomes but on other indicators for which there is evidence that they are associated with health outcomes. All in all, our review has delivered very little empirical evidence of valuable contributions from the management science and operations management literature to reduce the burden of the 15 diseases causing the most disability, whether in the United States or globally.

7. Advancing Empirical Evidence on Analytics, and Justification for Alternatives

Before considering enlargement of strong empirical evidence within the management science and operations management literature, let us try to understand the path that has led to the present situation. We have already mentioned some reasons why the empirical work and evidence base on the diseases in the burden of disease top 15 in the management science and operations management literature is modest. For instance, the efforts of this scientific community have focused on other conditions, or the corresponding publications have appeared elsewhere. We will not investigate these hypotheses further here. Alternatively, one might argue that the OR/MS community has focused on contributions of a theoretical nature. For instance, much of the statistical work in evidence-based medicine and health service uses optimization techniques to fit distribution functions or establish maximum likelihood estimations, etc. Likewise, methodological advances on Markov models and discrete event simulations have found their way into related sciences such as health technology assessment. Such contributions have formed valuable contributions to healthcare decision making (Weinstein et al. [50]).

These contributions do not explain, however, why there has been little interest in the management science and operations management literature on the improvement of health outcomes by implementation of applied contributions within the scientific discipline. We conclude this section by posing two hypotheses to explain this limited interest. A first hypothesis is that empirical work on improving health has not been considered a research priority within the community. The four journals that have published the highest number of publications included in the selection (and together have published more than half of the empirical manuscripts included) are *Healthcare Management Science, Socio-Economic Planning Sciences, Interfaces*, and the *Journal of the Operational Research Society*. None of these journals ranks in the top 15 of management science and operations management literature. The three publications identified to report evaluation of implementation have been cited 6 times, 14 times, and 1 time, respectively (according to Scopus).

A second, and related, possible explanation is for researchers to tacitly assume that plausible or earlier reported associations between operations and outcomes remain generally valid. Subsequently, researchers have focused on improving health service operations as a substitute for health outcomes. In fact, this is also the case for two of the three aforementioned evaluated implementations. They report evidence from a before–after case study on improvement on operations, and they additionally present evidence from the literature that the corresponding improvements cause improvement in health outcomes (which is a form of mechanism-based reasoning). Other examples of such assumptions are that improving operating room utilization rates reduces costs or improves health outcomes, and that shorter waiting times cause better health outcomes.

To more formally and rigorously address such mechanism-based reasoning, it is helpful to consider the simple structure-process-outcome model (Donabedian [8]). It posits that health outcomes are obtained through health service processes that take place in health service structures. Hence, to improve outcomes, one cannot directly optimize the outcomes themselves but has to optimize the processes and/or the structures of health service delivery. The next step subsequently entails gaining understanding of the relationships between structures

and processes (together) in relation to health outcomes—and, more specifically, which interventions in the structures and processes improve health outcomes. Once this evidence base is in place, improvement of health outcomes can be ascertained by improving structures and processes, i.e., by improving operations. Mahdavi [21] provides a more extensive treatment of this topic as well as methodological considerations.

For some diseases and conditions, evidence on the relation between operations and outcomes may be lacking. Likewise, existing evidence may not hold validity for the empirical context under consideration. For instance, the effectiveness of a variety of common HIV treatment interventions is shown to depend on prevalence (Mbonigaba [23]). Furthermore, the models to improve structures and processes may have simplified (or disregarded) aspects of reality that are relevant to the relationships with intended outcomes or may cause unintended effects on other outcomes. For instance, the analysis of Glorie et al. [12] suggests that matching donors with recipients in kidney exchange programs with the objective to maximize total health-adjusted life expectancy for the population of potential recipients has no significant effect on the expected number of transplants or the health-adjusted quality-of-life expectancy after transplantation, but it does significantly benefit the health-adjusted life expectancy of the patients left unmatched, in comparison to current kidney exchange practices. Because of such unexpected and unintended effects, evidence-based medicine values a synthesis of multiple empirical studies generating strong evidence on outcomes as best evidence.

Recalling that evidence on effectiveness often plays a decisive role in the allocation of financial and human resources available for improving health outcomes, the state of the art raises the question of whether management science and operations management journals can better promote the contribution of their scientific community by encouraging and publishing more studies that present evidence of health improvement. If the rigorous research designs required to generate strong evidence are difficult to conduct, less rigorous designs may form a first step.

Reasoning along the same lines, the International Society for Pharmacoeconomics and Outcomes Research Task Force on Good Research Practices for modeling studies concludes that model results are not factual statements and as such can not be empirically validated before being implemented (Weinstein et al. [50]). However, they argue that "model assumptions regarding causal structure and parameter estimates should be continually assessed against data, and models should be revised accordingly. Structural assumptions and parameter estimates should be reported clearly and explicitly, and opportunities for users to appreciate the conditional relationship between inputs and outputs should be provided through sensitivity analyses" (p. 9). If these practices are followed, they conclude that "model-based evaluations are a valuable resource for health-care decision makers" (p. 9). Hence the task force justifies their use as an alternative for empirical evidence.

While empirical results on outcomes from multiple well-designed studies may count as best evidence, generating such evidence is indeed not always the preferred mode by which to advance science. Increasingly, diseases that have long been fatal can now be cured, or their progress can be stopped or slowed down. In the latter cases, they are classified as chronic diseases, of which there are many among the leading causes of disability in high-income countries such as the United States—e.g., chronic obstructive pulmonary disease (COPD), lower back pain, depressive disorders, diabetes, neck pain, Alzheimer disease, anxiety disorders—and increasingly also in middle- and low-income countries. The success of slowing the progression of diseases may increase the burden resulting from years lived with disability. As a result, the effects of newly implemented decisions may impact health over many years and cannot be evaluated in the short or medium term from an experiment (or from an observational study). In such cases, the longer-term effects are deduced from models and mechanism-based reasoning. Likewise, situations or interventions may present themselves such that decision making before awaiting the evaluation of rigorously designed

experimental or observational studies is required. Such is the case, for instance, when facing an outbreak of infectious diseases or other urgent conditions, e.g., following from natural disasters or conflict. Experimental designs sometimes are infeasible, impractical, or undesirable for ethical reasons.

8. Contributing to Evidence: Compartmental Models for Communicable Diseases

Traditionally, models to analyze the evolution of health outcomes and the effects of improvement decisions are not based on models of individual health states but are mostly based on analyzing health at the population level. These models offer the convenience of not having to understand which individuals will acquire a disease or how disease progresses depending on individual characteristics, but instead, they allow for understanding incidence and prevalence at the population level. Two common models, which we introduce below as a basis for further discussion, are compartment models and Markov disease models.

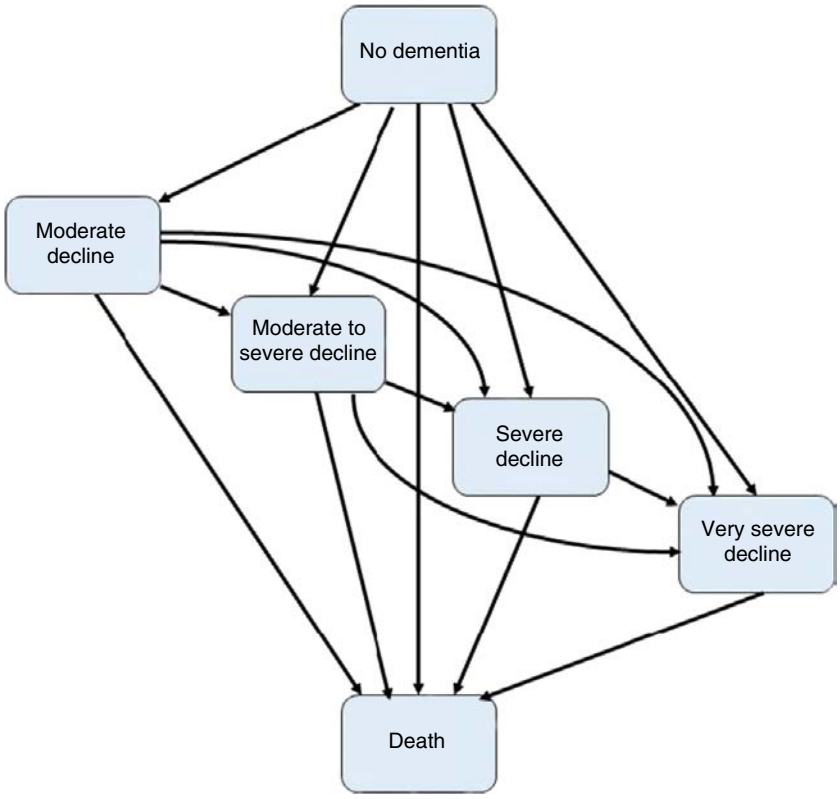
Early models at population levels have focused on the epidemiology of infectious diseases, as they caused the largest burden of disease. The early models divided the population into compartments and are therefore called compartmental models. The SIR model forms a basic model and partitions the population into three compartments: susceptible individuals, infectious individuals, and recovered individuals (see Brauer [1] for an excellent introduction into compartmental models). The SIR model simply models the progress of an infectious disease without intervention and therefore serves as a basic descriptive model that can be used to establish and forecast the health effects of an infectious disease at the population level. The transitions between the compartments are typically described using differential equations (defining a continuous time model) or using difference equations (defining a discrete time model). The causal structures and parameter estimated can be evidence based.

An important extension of the SIR model relevant for decision making is the SIRT model. It contains a fourth compartment—namely, a subpopulation that consists of infectious people who have been treated. Likewise, there has been much attention for models that include vaccination, which results in a compartment of vaccinated persons, originating from the susceptible individuals. SIR models and their extensions have played an important role in the many empirical papers that have appeared in the management science and operations research literature on HIV/AIDS. As already briefly mentioned above, subsequent research on this disease has differentiated different contexts, subpopulations (e.g., drug users), and age categories. While some of such refinements have been analytically included into SIR models and extensions, they complicate the analytical tractability of these original models. Hence subsequent models have, for instance, distinguished more compartments, and the analysis has often relied on simulation studies instead of mathematical analysis. When the refinements progress, e.g., according to gender, location, social status, treatment history, and genetic information, the number of subpopulations under consideration grows, and ultimately, the models may consider a subpopulation of one: the individual. Given the rapid developments in the availability of individual-level data from electronic patient records (with treatment history and health information, social media profiles, genetic profiles, etc.), such more detailed models, which rely on big data, have made their inroads and are bringing insights from the population level to the level of the individual. For instance, Chan and Ghose (Chan and Ghose [5]) study the effects of personal choices in social media behavior on the spread of HIV/AIDS. Naturally, such more refined models require larger data sets to deliver strong evidence at a more refined subpopulation level.

9. Contributing to Evidence: Markov Models for Noncommunicable Diseases

Markov models have become an important basis for the evaluation of decisions on interventions for noncommunicable chronic disease (for which there is no infectious population as

FIGURE 3. Markov disease model for dementia.



is the case in the basis compartmental model, the SIR model). We now provide a simple illustration of such a Markov model for dementia. As illustrated in Table 3, our literature review did not reveal any empirical work on outcomes for dementia (or its most common form, Alzheimer disease) in the management science and operations research literature so far. Alzheimer disease causes the fifth-most years of life lost in developed countries, causing more than 10.5% of all deaths in the United States in 2013 (VanderZanden [45]). Alzheimer disease alone has a current global prevalence of 47.5 million, a number that is expected to triple by 2030 (World Health Organization [52]).

The progression of the chronic condition dementia is often modeled using the staged Global Deterioration Scale (GDS) for assessment of primary degenerative dementia (Reisberg et al. [35]). With the exception of two early stages of (very) mild cognitive decline, the consecutive stages are depicted in Figure 3. The "moderate decline" stage is the first stage, which is diagnosed as dementia. When interpreting this model as a Markov model (Figure 3), it may be incorrect to interpret the model at the level of individuals. Markov models assume that the transition probabilities between stages are memoryless and hence that probabilities do not vary with the length of time a person is already in a certain stage. If there is evidence at a certain stage that these transition probabilities can vary over time, this shortcoming can be remedied by making multiple copies of such a stage, one copy for each length of stay relevant to distinguish and to model the transitions between these copies as well. The model then uses discretized hazard functions to express the transition probabilities. (Continuous models will not be considered in this tutorial.) But even without such time dependencies, the model and transition probabilities can accurately describe the disease progression at the population level.

The state of a patient is often diagnosed using cognitive assessment tests such as the Mini-Mental State Examination (MMSE) questionnaire. Mild dementia refers to an MMSE

score of 20–24. Evidence on the proportion of the population in this state can be collected, for instance, by surveying a population using the MMSE questionnaires, from other original empirical studies, via patient records (diagnosis), or indirectly (e.g., from medication use or insurance claims (e.g., on medication)). Moderate dementia corresponds to a score between 19 and 11, and scores below 10 count as severe dementia. Evidence can be collected as before, where the population suffering from severe dementia in developed countries necessarily lives in a nursing home. Because some of the aforementioned forms of evidence are not exact, and more exact measures may be too costly or impractical to obtain, the transition probabilities of the Markov model may be difficult to establish with certainty, with the exception of the transition to the state of death. Modeling the quality of life—or alternatively, the disability—is difficult as well. It is especially difficult for the later stages of dementia, because of the cognitive decline of the patients involved (see, for instance, Jonsson et al. [17] and Wolfs et al. [51]) for QALY measurements). Even when only available in terms of interval estimates, these probabilities and adjusted life years are necessary to estimate the burden of years of life lost and the burden of years lived with disability (as required to calculate HALE). Moreover, they are needed when trying to establish future budgets and make better decisions—decisions to maximize health within a limited budget.

Because there is no evidence of effective medical interventions to influence the progress of the disease after its onset, the only interventions for which there is evidence that they impact the probabilities in the Markov model regard the incidence, i.e., the transition from the state of no dementia to any of the dementia states. Physical exercise, social contacts, cognitive stimulation, and a Mediterranean diet are all associated with a reduced risk of developing dementia (Mittelman et al. [25], Polidori et al. [33]). For instance, regular physical activity at middle age more than halves the relative risk of developing dementia (Rovio et al. [37]). Likewise, Carlson et al. [2] provide evidence that certain forms of cognitive activity reduce the relative risk by 19%, and moderate daily red wine consumption reduces the relative risk by 47% (Larrieu et al. [19]). Noticing that experimental research designs to establish the cost and effects of decisions to invest in preventive programs to stimulate, for instance, physical exercise will take multiple decennia before delivering evidence, we clearly see that model analysis, mechanism-based reasoning, forms an attractive alternative to deliver evidence within a reasonable decision time frame. Analysis is particularly complex, as the interventions may also affect the likelihood of developing other conditions (e.g., diabetes, stroke) and more generally impact longevity, thus causing a higher probability of developing dementia at a later age. For a Dutch cohort, the combined implementation of these interventions reduces the burden of disease but increases health service cost (van de Klundert et al. [44]).

Present evidence on interventions for improvement in stages 4 and higher primarily regards assigning case managers to coordinate the health service provisioning for cognitively impaired dementia patients and the support of caregivers. The present evidence on case management appears contradictory and inconclusive, perhaps because of differences in interventions, contexts, or patient populations. For the Dutch context, however, there is evidence that caregiver support may reduce the relative risk of nursing home admission by 36% (Voigt-Radloff et al. [47]). As the quality of life of dementia patients is higher when living at home, such interventions are potentially valuable. On the basis of Dutch health service reimbursement data (involving approximately 7 million patients) (van de Klundert et al. [44]), providing caregiver support may indeed result in an improvement in quality of life for the population, without an effect on costs. The latter is partly because dementia patients living at home while being in the advanced stages of the disease are likely to experience costly hospital episodes. Further understanding of the costs and affordability of interventions in relation to the burden of disease therefore is important, as total global costs for dementia already consume 1% of the worldwide gross domestic product, while prevalence is expected to grow threefold until 2050 (World Health Organization [52]).

This brief introduction into Markov models for dementia illustrates that urgent, complex problems await health systems across the globe and that these models are valuable for making the best possible decisions about deploying the scarcely available budgets. More refined models, which include more refined data on the effects of interventions, better distinguish relevant subpopulations, and can describe the combined effects of interventions on dementia as well as on related conditions, are called for. Subsequently, the optimization questions to best deploy scarce (financial) resources can be addressed. Examples of such optimization studies have not yet been conducted for dementia but can be found, for instance, for chronic kidney disease (see Glorie et al. [12] and the references therein). Similar decision-making questions exist for other (chronic) conditions in the U.S. burden of disease top 15 such as diabetes and stroke.

10. Contributing to Evidence: From New Big Data to New Evidence Through Analytics

It is hoped that this tutorial has made clear that very large health data sets are already being collected and exploited, in particular at the population level, and that analytical models can play valuable roles in understanding how management decision making on the structures and processes of health service operations can improve health outcomes. Some of the evidence under consideration already stems from large data. Large data are not, however, synonymous with big data. Definitions of "big data" typically refer to not only the size of the data sets but also the complexity of their structure and behavior, and the technologies by which they are processed (Ward and Barker [49]). Databases with data collected from research, whether empirical studies or observational studies, may be large but are typically well structured and can often be processed using standard tools, such as standardized statistical methods. Existing scientific evidence may therefore not form a likely source of big data. The same holds true for much data from hospital information systems, or insurers databases. These are structured for insight and practical use, and therefore they are kept well structured and readily accessible. The complexity of the data increases, however, when the data start to include additional data from the primary health service processes—in particular, the health behavior of populations and individuals. Such behavioral and process data may regard text data entered into medical records by physicians, nurses, or patients; email messages, video, and texts exchanged between patients and professionals in e-consults; and information provided in health service utilization profiles (e.g., contact frequencies, activity in online communities, responses to automatically generated medication reminder messages). Experiments have been conducted where the movements of dementia patients are tracked through mobile phone applications to detect falls in nursing homes (or "smart homes") or wandering. Robots are being deployed as caregivers and interact with elderly. It will become increasingly possible to enrich data from more classical large databases with more complex types of data to transform them into information for better decisions, which is what analytics is about. This means that, in addition to using data from empirical research, analytics researchers should turn to novel empirical data from practice.

Moreover, we may expect that as more data become available on individuals, more personalized models will gain importance in addition to the population-level models, which have until now been most common. This may be further enforced by advances in medicine that leverage personal characteristics (for instance, genetic information) to tailor custom-made medical interventions. Genetic data can increasingly play a role in healthcare analytics. (Interestingly, one of the three evaluated studies found in the survey considered genetic information back in 2005.) These new types, sources, and structures of information may well form the most challenging opportunities for analytics to contribute to decisions for improved health.

The learning from advances in analytics will remain limited if analytics does not lead to decisions that are evidenced to be better through empirical evaluation. Given the increased

availability of data, and big data in particular, there is great opportunity for analytics to help in designing interventions that optimally reduce the burden of disease. For some types of research questions, models and mechanism-based reasoning may provide the best evidence. This holds particularly true when there is no time to lose with experimental studies or when considering long-term effects (in the case of chronic noncommunicable conditions). In other cases, empirical work is possible and preferable, as investments in analytics will be considered an alternative to clinical research or health services research. Such a comparison will often rely on value for money and cost-effectiveness, and analytics will need to present evidence that it (cost-) effectively contributes to reducing the global burden of disease. Simple before–after case study designs do not produce strong evidence. Other designs are feasible, as witnessed in this tutorial. Healthcare analytics can use big data as input for decisions, but it can also produce big data to provide strong evidence of effectiveness, of decisions that better reduce the global burden of disease.

References

[1] F. Brauer. Compartmental models in epidemiology. F. Brauer, P. van den Driessche, and J. Wu, eds. *Mathematical Epidemiology*, Lecture Notes in Mathetmatics, Vol. 1945. Springer, Berlin, 2008.

[2] M. C. Carlson, J. R. Burke, G. G. Potter, M. J. Helms, D. C. Steffens, and B. L. Plasman. Midlife activity predicts risk of dementia in older male twin pairs. *Alzheimers and Dementia*, 4(5):324–331, 2008.

[3] Center for Evidence Based Management. Resources and tools. Accessed March 2016, http://www.cebma.org/resources-and-tools/.

[4] Centers for Disease Control and Prevention. Bloomberg Data for Health Initiative. Accessed March 2016, http://www.cdc.gov/globalhealth/healthprotection/data.html.

[5] J. Chan and A. Ghose. Internets dirty secret: Assessing the impact of online intermediaries on HIV transmission. *MIS Quarterly* 38(4):955–976, 2013.

[6] S. de Treville, I. Smith, A. Rölli, and V. Arnold. Applying operations management logic and tools to save lives: A case study of the World Health Organization's global drug facility. *Journal of Operations Management* 24(4):397–406, 2006.

[7] H. de Vries, S. Elkhuizen, A. Wagelmans, and J. van de Klundert. A systematic review of the relationship between distance to health service facility and treatment effectiveness for mobile patients. Working paper, Erasmus University Rotterdam, Rotterdam, The Netherlands, 2016.

[8] A. Donabedian. The quality of care: How can it be assessed? *JAMA* 260(12):1743–1748, 1988.

[9] M. F. Drummond, M. J. Sculpher, K. Claxton, G. L. Stoddart, and G. W. Torrance. *Methods for the Economic Evaluation of Health Care Programmes*, 4th ed. Oxford University Press, Oxford, UK, 2015.

[10] J. J. Emberson, K. R. Lees, P. Lyden, L. Blackwell, G. Albers, E. Bluhmki, T. Brott, et al. Effect of treatment delay, age, and stroke severity on the effects of intravenous thrombolysis with alteplase for acute ischaemic stroke: A meta-analysis of individual patient data from randomised trials. *Lancet* 384(9958):1929–1935, 2014.

[11] D. G. Fryback, N. C. Dunham, M. Palta, J. Hanmer, J. Buechner, D. Cherepanov, S. A. Herrington, et al. U.S. norms for six generic health-related quality-of- life indexes from the national health measurement study. *Medical Care* 45(12):1162–1170, 2007.

[12] K. Glorie, G. Xiao, and J. van de Klundert. Health value analysis of allocation policies in kidney exchange. In K. Glorie (Ph.D. Thesis), ed. *Clearing Barter Exchange Markets—Kidney Exchange and Beyond*. Erasmus University Rotterdam, Rotterdam, The Netherlands, 85–106, 2014.

[13] J. Howick, P. Glasziou, and J. K. Aronson. Evidence-based mechanistic reasoning. *Journal of the Royal Society of Medicine* 103(11):433–441, 2010.

[14] J. Howick, I. Chalmers, P. Glasziou, T. Greenhalgh, C. Heneghan, A. Liberati, I. Moschetti, et al. The Oxford levels of evidence 2. [Table.] Oxford Centre for Evidence-Based Medicine, Oxford, UK. Accessed March 2016, http://www.cebm.net/index.aspx?o=5653, 2011.

[15] Institute for Health Metrics and Evaluation. Health data: GBD compare. Accessed March 2016, http://vizhub.healthdata.org/gbd-compare/.

[16] A. R. Jadad, R. B. Haynes, D. Hunt, and G. P. Browman. The Internet and evidence-based decision-making: A needed synergy for efficient knowledge management in health care. *Canadian Medical Association Journal* 162(3):362–365, 2000.

[17] L. Jonsson, N. Andreasen, L. Kilander, K. Soinien, G. Waldemaar, et al. Patient- and proxy-reported utility in Alzheimer disease using the EuroQoL. *Alzheimer Disease and Associated Disorders* 20(1):49–55, 2006.

[18] F. Laden, J. E. Hart, T. J. Smith, M. E. Davis, and E. Garshick. Cause-specific mortality in the unionized U.S. trucking industry. *Environmental Health Perspectives* 115(8):1192–1196, 2007.

[19] S. Larrieu, l. Letenneur, C. Helmer, J. F. Dartigues, P. Barberger-Gateau. Nutritional factors and risk of incident dementia in the PAQUID longitudinal cohort. *Journal of Nutrition, Health, and Aging* 8(3):150–154, 2004.

[20] R. H. Lathrop and M. J. Pazzani. Combinatorial optimization in rapidly mutating drug-resistant viruses. *Journal of Combinatorial Optimization* 3(2):301–320, 1999.

[21] M. Mahdavi. Building the bridge between operations and outcomes. Ph.D. thesis, Institute of Health Policy and Medicine, Erasmus University Rotterdam, Rotterdam, The Netherlands, 2015.

[22] T. Mankowski. Institute for Health Metrics and Evaluation launched. *UW Today* (July 5), http://www.washington.edu/news/2007/07/05/institute-for-health-metrics-and-evaluation-launched/, 2007.

[23] J. Mbonigaba. The cost-effectiveness of intervening in low and high HIV prevalence areas in South Africa. *South African Journal of Economic and Management Sciences* 16(2):183–198, 2013.

[24] Michigan Institute for Data Science. About MIDAS: Mission. Accessed March 2016, http://midas.umich.edu/about/.

[25] S. M. Mittelman, W. E. Haley, O. J. Clay, and D. L. Roth. Improving caregiver well-being delays nursing home placement of patients with Alzheimer disease. *Neurology* 67(9):1592–1599, 2006.

[26] T. Monks, M. Pearson, M. Pitt, K. Stein, and M. A. James. Evaluating the impact of a simulation study in emergency stroke care. *Operations Research for Health Care* 6(September):40–49, 2015.

[27] C. J. Murray. Quantifying the burden of disease: The technical basis for disability-adjusted life years. *Bulletin of the World Health Organizations* 72(3):429–445, 1994.

[28] C. J. L. Murray and A. D. Lopez. Measuring the global burden of disease. *New England Journal of Medicine* 369(5):448–457, 2013.

[29] C. J. L. Murray, M. Ezzati, A. D. Flaxman, S. Lim, R. Lozano, C. Michaud, M. Naghavi, et al. GBD 2010: Design, definitions, and metrics. *Lancet* 380(9859):2063–2066, 2013.

[30] C. J. L. Murray, T. Vos, R. Lozano, M. Naghavi, A. D. Flaxman, C. Michaud, M. Ezzati, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380(9859): 2197–2223, 2013.

[31] P. J. Neumann and M. C. Weinstein. Legislating against use of cost-effectiveness information. *New England Journal of Medicine* 363(16):1495–1497, 2010.

[32] T. Otte-Trojel, A. de Bont, T. G. Rundall, and J. van de Klundert. What do we know about developing patient portals? A systematic literature review. *Journal of the American Medical Informatics Association* 23(e1): e162–e168, 2016.

[33] M. C. Polidori, G. Nelles, and L. Pientka. Prevention of dementia: Focus on lifestyle. *International Journal of Alzheimers Disease* 2010:Article 393579, 2010.

[34] M. Porter and E. O. Teisberg. *Redefining Health Care: Creating Value-based Competition on Results*, 1st ed. Harvard Business Press, Boston, 2006.

[35] B. Reisberg, S. H. Ferris, M. J. de Leon, and T. Crook. The global deterioration scale for assessment of primary degenerative dementia. *American Journal of Psychiatry* 139(9): 1136–1139, 1982.

[36] D. M. Rousseau. Is there such a thing as "evidence-based management?" *Academy of Management Review* 31(2):256–269, 2005.

[37] S. Rovio, I. Kåreholt, E. L. Helkala, M. Viitanen, B. Winblad, J. Tuomilehto, H. Soininen, A. Nissinen, M. Kivipelto. Leisure-time physical activity at midlife and the risk of dementia and Alzheimers disease. *Lancet Neurology* 4(11):705–711, 2005.

[38] D. L. Sackett, W. M. C. Rosenberg, J. A. Muir Gray, R. B. Haynes, and W. S. Richardson. Evidence based medicine: What it is and what it isn't. *British Medical Journal* 312(7023): 71–72, 1996.

[39] Scimago Lab. Scimago journal and country ranking. Accessed March 2016, http://www.scimagojr.com/journalrank.php?category=1803.

[40] J. L. Severens and R. J. Milne. Discounting health outcomes in economic evaluation: The ongoing debate. *Value in Health* 7(4):397–401, 2004.

[41] U.S. Burden of Disease Collaborators. The state of US health, 1990–2010: Burden of diseases, injuries, and risk factors. *JAMA* 310(6):591–606.

[42] J. E. van Aken and G. Romme. Reinventing the future: Adding design science to the repertoire of organization and management studies. *Organization Management Journal* 6(1):5–12, 2009.

[43] J. van Aken, H. Berends, and H. van der Bij. *Problem Solving in Organizations: A Methodological Handbook for Business and Management Students*. 1st ed. Cambridge University Press, Cambridge, UK, 2012.

[44] J. van de Klundert, J. van Wijngaarden, et al. The added value of health insurers in Dutch healthcare. [In Dutch.] Technical report, Institute of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands, 2001.

[45] A. VanderZanden. A closer look at Alzheimers disease worldwide. *IHME Impact* (13), http://www.healthdata.org/newsletters/impact-13/deep-dive, 2014.

[46] M. Vledder, F. M. Sjöblom, J. Friedman, T. Brown, and P. Yadav. Optimal supply chain structure for distributing essential drugs in low income countries: Results from a randomized experiment. Working Paper 1269, Stephen M. Ross School of Business, University of Michigan, Ann Arbor, 2015.

[47] S. Voigt-Radloff, M. Graff, R. Leonhart, K. Schornstein, M. Vernooij-Dassen, M. Olde-Rikkert, and M. Huell. WHEDA study: Effectiveness of occupational therapy at home for older people with dementia and their caregivers—The design of a pragmatic randomised controlled trial evaluating a Dutch programme in seven German centres. *BMC Geriatrics* 9:44, 2009.

[48] K. Walshe and T. G. Rundall. Evidence-based management: From theory to practice in health care. *Milbank Quarterly* 79(3):429–457, 2001.

[49] J. S. Ward and A. Barker. School of computer science. Technical report, University of St Andrews, Fife, UK, 2012.

[50] M. C. Weinstein, B. O'Brien, J. Hornberger, J. Jackson, M. Johannesson, C. McCabe, and B. R. Luce. Principles of good practice for decision analytic modeling in health-care evaluation: Report of the ISPOR Task Force on Good Research Practices—Modeling Studies. *Value in Health* 6(1):9–17, 2003.

[51] C. G. Wolfs, C. D. Dirksen, A. Kessels, J. L. Severens, and F. J. Verhey. Economic evaluation of an integrated diagnostic approach for psychogeriatric patients: Results of a randomized controlled trial. *Archives of General Psychiatry* 66(3):313–323, 2009.

[52] World Health Organization. Dementia. Factsheet 352, World Health Organization, New York, http://www.who.int/mediacentre/factsheets/fs362/en/, 2016.

[53] World Health Organization. Preamble to the Constitution of the World Health organization as adopted by the International Health Conference, New York, 19–22 June, 1946; signed on 22 July 1946 by the representatives of 61 states (Official Records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948. World Health Organization, New York, 1948.