Suggestions for the Ethical Design of a Health Care Treatment Recommender System

Lam, Peterson, Schnall

Harvard University
Professor Barbara Grosz
December 12, 2016

ABSTRACT

Advances in modern medicine have enabled health care to become increasingly precise. This holds promise for patients, however, the increase in the number of possible treatment options can make it difficult for physicians to decide on the best for a particular patient. Developing a treatment recommender system could assist physicians in this task, but due to the high-stakes nature of health care, it is imperative that medical ethics be integrated into its design. After an overview of medical ethics, in this paper, three key design features are investigated – (1) the selection of reinforcement as a learning algorithm, (2) the exploration / exploitation ratio and reward function design, and (3) how recommendations are presented and whether patients should be allowed to opt out of suggestions. Ultimately, ethical design suggests the development of a system that uses batch reinforcement learning with some exploration, a reward function that seeks to maximize a compound quality measure, and allows for patients to opt-out of suggestions. Possible future work could include the development of a treatment recommender system in a lower-risk health care domain (such as in primary care) and optimizing the design before moving to higher-risk domains such as oncology. Due to the incredibly complex nature of health care, treatment recommender systems are likely many years from being completely functional and because of the risks inherent in treating patients, and even if implemented, will require physician oversight for the foreseeable future.

Keywords: Health Care, Recommender System, Reinforcement Learning, Ethics

1. INTRODUCTION

Advances in modern medicine have dramatically improved the health of people worldwide. Life expectancy at birth, a base metric for the health of a society, has increased by approximately 10 years from 1950 and 2007 for both men and women born in the United States (Crimmins, 2011). Outside of life expectancy, people alive today experience greater overall health status as some diseases, such as polio, have been nearly eliminated, while others, like HIV/AIDS, have found better methods of chronic management. These advances in health can be contributed to a combination of complex and interacting factors including social welfare programs; access to better nutrition, vaccination, and public health campaigns; and improvements in medicine not limited to more effective treatment options and increased understanding of disease management.

In the field of oncology, or cancer treatment, there has been a shift towards continually more precise medicine. Beginning with more blunt treatments like radical surgery and general systemic chemotherapy which indiscriminately kill both normal and cancerous cells, cancer therapies have transitioned to complex, targeted treatments designed to only eliminate cancerous cells, reducing the negative impacts of treatment on the patient. One such example is that of PD-L1 inhibitors, which are complex molecules that help the patients own immune system find and eliminate cancerous cells that were previously hidden to the immune system (Brahmer, 2012). Advances in diagnostic capabilities, not limited to biomarker testing and personalized genomics, have helped uncovered new targets that pharmaceutical agents can act on, simultaneously increasing effectiveness and safety. However, as the targets of these treatments become increasingly refined, they also work in fewer and fewer patients, as not every patient will have a cancer with the specific abnormalities that a highly targeted treatment can take advantage of. As the number of treatments available to use is constantly increasing – there were 771 oncology specific drugs in R&D in 2015 – the number of different combinations of treatments and the order in which they can be given is rapidly growing (Buffery, 2015).

While this rapid increase in highly targeted treatment options and diagnostic capabilities has presented potentially dramatic improvements in patient outcomes, it also has made sifting through a patients clinical data and then reviewing medical literature to pick an appropriate, let alone maximally-effective, treatment plan increasingly difficult for physicians. To help compensate for this, doctors are generally supposed to follow clinical guidelines put out by experts in specific disease areas, such as the National Comprehensive Cancer Networks Clinical Practice Guidelines in Oncology. These guidelines, which often are presented in the format of a decision tree, can be useful for determining what care to deliver to patients. However, they are sometimes not available for rare disorders and are not updated as frequently as new treatment options are made available. Additionally, despite the existence of certain clear guidelines, such as performing surgery on patients with operable stage I or II breast cancer, a 2003 study showed that only 55% of patients in the United States receive recommended care (McGlynn, 2003). While this might raise questions of undertreatment, studies show that nearly one third of all clinical Medicare spending, a sum of approximately 650 billion dollars (equivalent to the entire budget for K-12 education), was wasteful and could be avoided without negatively affecting health outcomes (Skinner, 2005). It therefore appears that the provision of treatment guidelines is not sufficient to ensure that patients are receiving optimal care. Given the large degree of wasteful medical practice, lack of provision of standard care, and the inherent difficulties of deciding on an appropriate treatment plan for patients in a timely manner, it is highly important that physicians are provided with better tools for improving the care that they deliver.

One way to do accomplish this would be the development of a treatment recommender system that offers physicians real-time decision support. At a high level, such a system would have an understanding of a given patient – from what disease they have and their age to potentially more granular details like what specific base-pairs are mutated in their DNA – and how other patients with similar characteristics have responded to different therapies. Then it would use these data to help recommend treatments that are likely to be effective for the patient. These systems could build their recommendations on data derived from Electronic Health Records (EHRs), which store data on patient characteristics and treatment administrations as well as other useful data, such as physician notes on how a patient feels and what side-effects they may be experiencing with their treatment. After training on baseline data, such a recommender system could also be capable of learning over time, taking into account how its recommendations affected a patients health, allowing it to ultimately improve what treatments it recommends for similar patients in the future. Akin to physicians performing clinical trials to learn which patients a new drug might be effective in, such a system would need to explore to try to improve. For instance, if a new treatment option became available, the system would need to recommend this option a few times to different patients to learn how well and in what patients it might work. In medicine, unlike in other domains, however, a bad recommendation could result in the avoidable death of a patient. It is important to note that a physician should always have the final decision in what care to give to a patient, being certain to take a patients desires into account. These systems are only meant to assist physicians, however, the mere suggestion of a particular treatment might affect a physician's decision, and so, given these high stakes, it is important that the design of a treatment recommender system be thoroughly considered and guided through concepts found in medical ethics.

2. GUIDING ETHICAL PRINCIPLES

Although no code of medical ethics was documented before the 5th century B.C.E. at the earliest estimate, humans have been practicing some forms of medicine since the stone age (Dorfner, 1999), and a concept of ethics has been inherent in that practice since it began. Medical practice, from prehistoric medicine men to modern pharmacology and surgery, can be understood broadly as attempting to prevent or lessen the harm, suffering or illness individuals experience. The concept that mental and bodily harm should be minimized is in itself an ethical principle, and thus medicine is an inherently ethical discipline.

Historically, Western medical ethics has been understood through sets of codified rules. Contemporary conceptions of medical ethics generally rest on the four prima facie principles of respect for autonomy, beneficence, non-maleficence, and justice (Gillon, 1994). There is no single agreed-upon protocol for handling particular medical situations. Instead, medical ethics prescribes that practitioners and other ethically interested parties ought to understand ethical conflicts in medicine through these principles, and reason among them in the context of the particular situation. Gillon notes that these principles denote prima facie obligations; they are obligations, but in the face of a more pressing duty – perhaps a duty brought by one of the other principles – they may be diminished.

The decision process encouraged by such a code of ethics aligns most closely with rule consequentialism. Where act consequentialism holds that an act is good if its consequences maximize good, rule consequentialism holds that an act is good if it is in accordance with an agreed-upon code of rules; the rules are selected for their capacity to maximize good consequences. Rule consequentialism is not designed necessarily to maximize the good, but to provide a theory to resolve moral disagreements and uncertainties (Hooker, 2000). The four principles of medical ethics should each be met as completely as possible. When the four come into conflict, though, it will sometimes be necessary to leave parts of some principles unmet.

The principle of respect for autonomy concerns the individuals right to deliberated self rule. Individuals should be able to make decisions regarding the course of their own lives. A primary implication in health care of this principle is the obligation to seek and obtain individuals informed consent concerning their treatment. Patients should not be deceived as to what their treatment entails, or coerced into treatment that they do not wish to endure. In the context of medical research, patients should be informed as to the possible risks and benefits of the treatment being tested, and should be informed as to whether the trial is a randomized trial – that is, whether they are actually guaranteed to be receiving the treatment they are expecting to receive, or whether they might instead receive an alternative treatment or placebo.

Gillon argues that the principle of respect for autonomy can also be understood through the Humanity Formulation of Kants Categorical Imperative. The Humanity Formula states that we should never act so as to treat humanity as merely a means to an end; humanity should be treated as an end in itself (Johnson & Cureton, 2008). An implication for clinical research is that individuals ought not to be enrolled in clinical studies that pose no benefit to humanity; this is, importantly, distinct from a proposal that individuals ought only to be enrolled in studies that propose a direct benefit to the individual. There are ethical formulations in clinical research that allow for enrollment in studies that pose to benefit to the individual, but do pose a benefit to society or humanity at a wider level.

Beneficence refers to the promotion of the well-being of others; non-maleficence refers to the avoidance of causing harm. The principles of beneficence and non-maleficence come together to form the medical standard of maximizing benefit to patients while minimizing harms. These are, together, central in any clinical decision-making process, as a physician typically determines which type of treatment to prescribe to a given patient by weighing the associated risks and potential benefits. Several implications for health care result from the consideration of beneficence and non-maleficence (Gillon, 1994). Clinicians need extensive education and training to be able to provide good care and to make decisions about what will bring the greatest net benefit to the patient. Clinicians also must be able to understand the risk and probability of success of various treatments, and the way that these shift depending on the traits of each individual patient. In order to reliably understand the risks and success rates of various treatments, as well as to inform patients completely of the potential risks and benefits of treatments thereby allowing them to give autonomous informed consent, effective medical research is obligatory.

The final core principle is the principle of justice, concerning fair adjudication between competing ethical claims. An Aristotelian view of justice takes into consideration the relevant existing inequalities between individuals when determining just outcomes. Equals should be treated equally, and unequals should be treated unequally, but fairly, in proportion to the inequality that exists between them.

In a competition between moral concerns, there is no guarantee that all concerns can be satisfied nor even a guarantee that all parts of any given concern will be satisfied. Medical ethics holds that a code of principles should be upheld wherever possible, but understands that some deliberation between principles will be necessary in situations of ethical conflict. Thus, medical ethics is not a purely deontological system, as the rules do not entirely dictate conduct, but neither is it an act consequentialist system, as each act is evaluated by its adherence to the code and not by the relative costs and benefits of its consequences alone. Medical ethics, as commonly conceived, most closely resembles a rule consequentialist system.

Medical research is an intrinsic part of the biomedical system, and raises ethical concerns that are somewhat distinct from those handled in medical practice. By Kants Humanity Formula and the principles of beneficence and non-maleficence, no individuals should be enrolled in a study that poses no benefit; additionally, individuals should not be enrolled in a study that poses an undue potential for harm. Yet, all medical research involves some risk to some participants: the objective of medical research is to determine the relative benefit of competing treatment options, and in nearly all cases, one treatment will prove to be preferable to the other. Effective medical research is ethically required in order for clinicians to provide best care and to provide clear information to patients about the relative risks and benefits of treatment options, though, so systems must accommodate some small risk of harm to the individual in order to meet the principles of beneficence and autonomy on a larger scale.

This conflict is often thought to be resolved by the guarantee of informed consent. Patients should be allowed to decide whether or not to participate in research, even if it might pose some risk to them; it is likely that there are at least some individuals who altruistically are inclined to participate in research, and if it can be assumed that they are indeed acting autonomously and freely of coercion, they should be allowed to decide to do so. However, Benjamin Freedman (1987) argues that researchers have an obligation to halt a trial when it is conclusively clear that one treatment is preferable to the other, and to provide the preferred treatment to all participants enrolled in the trial. The individuals autonomy should be respected, but the researcher should not allow a trial to continue once it poses no potential benefit to humanity. For a trial to be ethical, there must be a state of uncertainty in the medical community regarding the benefits of the treatments included in the trial. The trial must be designed so that at the end of the study, this uncertainty will be resolved, and one treatment will be shown conclusively to be better.

The obligation to perform clinical research and the obligation to run studies in an ethical manner are both clarified by the above principles; but for whom and in what circumstances is there an obligation to participate in clinical research? Methodologically, a diverse pool of participants is necessary to produce results that are meaningful for the diverse pool of people that may eventually be prescribed the treatment: pregnant women, for example, are an underrepresented population in most clinical research, as researchers often wish to avoid posing any risk to the fetus (Lyerly et al., 2008). Yet pregnant women require medical treatment, and often respond differently to pharmaceuticals than other adults. Current clinical research fails to produce results that are applicable to underrepresented populations.

Caplan (1984) explains and rejects the social contract theory with respect to the duty to serve as a clinical subject. The theory follows as such: current patients of the biomedical system are benefiting from individuals who served as research participants in past generations. Therefore, current patients have a duty to repay that debt by serving as participants to benefit future generations. Caplan notes a major counter-argument: for past medical patients to have any claim over current patients, they would have had to have undertaken research participation with the belief that the world they were doing needed to be repaid. Most research participants do not understand clinical research in this way. It is pleasant to think that at least some people serve as research participants altruistically, and the social contract theory would undermine their altruistic motivations. Other research participants may have been coerced or tricked into participation. They, too, likely conceived of their participation as something other than a favor that would need to be repaid by future generations. Wendler (2012) agrees with this rejection of contract theory, with a slightly different argument: serving to benefit future generations does nothing to repay a debt to generations past.

A more convincing theory proposed by Caplan takes a Rawlsian approach. Caplan calls this the social cooperative theory. Those who benefit from participation in cooperative social schemes, like a food co-op or neighborhood watch, have obligations to one another to bear the collective risk (Rawls, 1999). People who derive benefit from social schemes without their consent and against their consent do not incur such obligations; by contrast, people who agreeably receive the benefits provided are tacitly consenting to the cooperative schemes.

Caplan uses the example of a teaching hospital as a small biomedical cooperative. Patients who choose to receive their medical care in teaching hospitals gain the benefit of medical care, and generally are understood to incur the responsibility to serve as teaching subjects. Patient consent for individual procedures and protocols is not voided, but teaching activities that do little to infringe upon the autonomy of the patient or pose little to no harm, such as discussion of treatment options with residents, having one qualified medical provide care over another qualified medical professional, or observation of noninvasive treatments are generally allowed by the teaching physicians.

The social cooperative theory extends, Caplan argues, to the biomedical research system, arguing that any competent person who voluntarily seeks out and takes the benefits of care resulting from biomedical research is a legitimately consenting participant in the biomedical research cooperative. The patient retains the right to choose and consent to specific research protocols, but is obligated to participate in activities that are necessary for the upkeep of the cooperative. Faden et al. (2013) agrees with this theory, taking David Humes reciprocal theory of society: All our obligations to do good to society seem to imply something reciprocal. I receive the benefits of society, and therefore ought to promote its interest (Hume, 1987). The health care system is integrated with the health research system by its goals and by the fact that each requires the other to survive. Participation in any healthcare setting that provides benefit should entail obligation to serve in learning activities that benefit the knowledge and effectiveness of the system.

In designing a healthcare recommender system, the above existing framework for evaluation of ethics in health care must be considered. The four principles of medical ethics should be used to analyze the viability of design options, and the ethics of research should be considered. It is now worth considering several key design elements that such a system would have.

3. LEARNING ALGORITHM DESIGN

At its core, a treatment recommender system seeks to answer the question of how best to treat an individual patient based on (1) available patient information such as medical history and panomic information, (2) domain knowledge including past clinical trials and published literature, (3) expertise and experience of the physicians at hand, and (4) prospective goals of the system, such as providing each patient the best possible outcome or optimizing for learning across the field (Tenenbaum et al., 2010). As such, the first major decision that should be made in the design of a treatment recommender system is how exactly the system is going to learn to make its recommendations. Which learning algorithm a system uses can impact its overall function and ultimately what it chooses to recommend. Given the wide variability of patient demographics, the enormous complexity of cancer genomes and the large number of potential treatment plans, a system determining the optimal medical treatment must handle a large state-space with a high-degree of complexity. Selecting an appropriate learning methodology is a primary concern.

Drawing upon a model proposed by Cancer Commons, this process can be broken down into several components (see Figure 1). At a high level, information is first gathered from various sources, including patient electronic health record systems (EHRs), relevant literature, and genomic data, and compiled into an integrated knowledge base. Once this knowledge base is established, various learning methods are used to build predictive models for various problems, such as predicting patients response to treatments or determining clusterings of patients based on similar characteristics. From these predictive models, reinforcement learning can be used to synthesize and optimize treatment plans for individual patients (Tenenbaum et al., 2010).

It is important to note that the process of collecting data and employing suitable learning methods to build predictive models yields its own set of challenges. Data in EHRs is often incomplete, corrupted, and can lack the meaningful data necessary for basing a recommendation, let alone the regulations and restrictions in place that limit data access and use. Necessary efforts for data aggregation and cleaning as well as increased data sharing across medical facilities would help ameliorate this problem and allow a system to learn better. Furthermore, when building a predictive model, a suitable learning method must be able to handle personalized cancer data that is inherently relational and noisy. Tenenbaum et al. (2010) discuss the use of a nonparametric Bayesian Probabilistic model, among others, whose interface is suitable for Monte-Carlo planning for sequential treatment, and its latent representation easily communicable to clinicians. Additional research and experimentation on various learning models used in a clinical setting can further elucidate and help determine a suitable learning method to build these predictive models.

One framework that appears both technically and ethically promising is referred to as batch reinforcement learning. In this method, the recommender system first learns a predictive model to build an initial understanding of the state-space of medical treatment. This model is then improved upon for treatment planning using reinforcement learning (Lange, 2012). The advantages of batch algorithms include efficiency of data as well as stability during the learning process. Traditional reinforcement learning algorithms, such as Q-learning, require

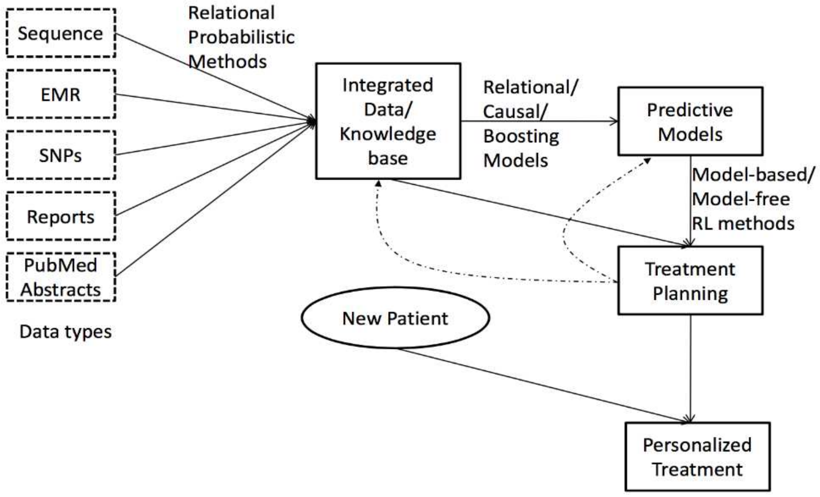
Sequence
EMR
SNPs
Reports
PubMed Abstracts
Data types
Relational Probabilistic Methods
Integrated Data/ Knowledge base
Relational/ Causal/ Boosting Models
Predictive Models
Model-based/ Model-free RL methods
Treatment Planning
New Patient
Personalized Treatment

Figure 1. Model for personalized cancer therapy (Tenenbaum et al., 2010)

a large number of interactions to converge to an optimal solution. Batch reinforcement learning algorithms can converge in a fraction of the time, meaning that the recommender system would be able to provide improved recommendations after a shorter training period. A number of successful real-world systems have been deployed using batch reinforcement learning (Lange, 2012).

A reinforcement learning agent that incorporates past medical data into its initial recommendations reduces the amount of potential harm to patients. A reinforcement-learning system that begins with no understanding of any treatment and their effect on patient health may eventually develop strong and optimal recommendations, but would begin making recommendations without any knowledge. In the field of medicine, the potential risks would be far too high. No practice would employ a human physician with zero training; the possibility that an untrained physician might cause great physical harm is the foundation of the global medical education system.

By employing batch reinforcement learning, the system could avoid recommending treatments that have already been proven to have negative or little value. For example, phase I clinical trials attempt to seek the optimal dosing for a particular drug, oftentimes incrementally raising doses until toxicities appear (Le Tourneau, 2009). Batch reinforcement learning could allow for more accurate dosage suggestions from the start, whereas a reinforcement learning system with no initial seeding might suggest trying a potentially toxic dose that could be avoided. If a physician permitted this dose, perhaps by not noticing the suggestion was in grams instead of milligrams, the system would have contributed in violating aforementioned principle of non-maleficence. To not employ a learning agent that suggests more optimal treatments earlier and faster is less ethically permissible than to do so.

Another reason to use reinforcement learning for the systems learning modality is due to the fact that these algorithms are often more intuitive, especially for clinicians. Reinforcement learning has a more cause-and-effect-esque method of learning. Clinicians learn to provide care by learning from how previous decisions affected outcomes and then doing what had the best outcome, either through personal experience or clinical trials. This similarity to reinforcement learning might help physicians better understand how the system came to a suggestion thereby improving their trust in the system and potentially increase their chances of understanding how a bad suggestion came about.

Reinforcement learning is also a strong method for medical learning because of its flexibility in accommodating for rare case examples. Zhao (2009) demonstrated that reinforcement learning does not have to rely on accurate mathematical models to make decisions. By sequentially selecting treatments and taking into account delayed effects, reinforcement learning can determine the best treatment option from several. As a result, an optimal personalized treatment can be found even if the relationship between treatments and outcomes is not fully known. Additionally, studies have shown that even a single patient who responds exceptionally to a treatment can be identified and learned from, such as in one study where a single patient who completely responded to a cancer treatment had a unique tumor mutation (Shrager, 2014). In the future, patients identified with this same mutation later on will be able to be provided with this effective, targeted treatment and might experience similarly positive outcomes.

As panomic diagnostic techniques, such as transcriptomics, provide increasingly granular patient features, patients with one disease (e.g. lung cancer patients) are able to be broken down into increasingly more detailed subtypes (e.g. lung cancer patients who are EGFR positive). While these additional features might improve how precise a treatment could be, there are also increasingly smaller numbers of patients in each. This means having a system that is capable of learning from a few or a single case(s) is highly important for improving recommendations. The ability to learn through a few instances could also benefit patients with very rare diseases on whom it can be difficult to perform large-scale clinical trials, improving the quality and efficacy of care that they get.

Lastly, reinforcement learnings iterative nature means that every patient of a particular disease should receive better care than the last. In a system that needs to be re-trained to update its recommendations, two patients with similar features should receive similar suggestions if the model has not been updated. Initially this might seem good from a deontological perspective, as each person is being treated similarly. However, the concept of beneficence from medical ethics argues that physicians should continually update training and strive for net benefit (Yelon, 2014). As such, a system designed to continually improve, learning from every decision made by physicians would be in line with these goals. To only periodically retrain a model means that the suggestions the system makes are not necessarily the best possible at every moment in time as the potential learnings from every patient treated since the last re-training of the model are not incorporated into the output of the next recommendation. A system that constantly improves, therefore, is more in accordance with the principles of beneficence and non-maleficence.

In terms of designing a system that is optimized for producing strong recommendations and is ethically sound, the use of reinforcement learning holds promise. By incorporating methods such as batch reinforcement learning, the system could avoid the problems associated with a cold start in a state-space as large as health care. The intuitive mechanism by which reinforcement learning works and its ability to better explain its output mean that a system using this technique might be more well received by doctors. Additionally, reinforcement learnings ability to learn from rarer cases means it will be able to function well for rare diseases and patients with highly detailed disease features. Furthermore, the continually updating aspect of reinforcement learning ultimately means that more patients are able to get better care more quickly than they would if a model reliant on re-training to improve recommendations were used. Overall, a system that incorporates historical medical data and then improved via reinforcement learning would be effective and, in some respects, more ethically sound than other learning algorithms.

4. EXPLORATION-EXPLOITATION RATIO AND REWARD SYSTEM DESIGN

One of the key challenges that arises in reinforcement learning is the tradeoff between exploration and exploitation. A reinforcement learning agent must decide whether to exploit solutions that it has previously found successful or to explore actions that it has not yet taken in an effort to discover better novel solutions. For a stochastic model, a variety of actions must be tried multiple times to gain a reliable estimate of respective rewards while progressively favoring those that appear to be most optimal. This tradeoff can be further described as a tension between short-term and long-term reward (Sutton, 1998; Tenenbaum et al., 2010). Reinforcement learning agents use a discount factor to weight or downweight future reward, encouraging or discouraging short-term exploitation (Tenenbaum et al., 2010). To what extent agents should pursue exploration, favoring future exploitation, is a tenuous topic especially in the medical domain.

In the medical domain, application of a reinforcement learning system implies a model where current patients can be exposed to risk for the purpose of determining the efficacy of a treatment process that might serve more patients in the future. The discount factor essentially determines at what scope the principle of beneficence will be applied. Selecting for a discount factor close to one, an agent would favor minute improvements for many future patients over even drastic outcomes for current patients, such as death (Tenenbaum et al., 2010). Conversely, by selecting for a discount factor close to zero, an agent would be extremely slow to improve, if at all, compromising optimal results for future patients but would suggest the best care it presently knows to a current patient.

The value of the discount factor ultimately affects what care the system delivers over both the long and short-term. Maximizing care for current patients might seem to be the best option with respect to non-maleficence, but the possibility of providing better care to future patients through exposing some of todays patients to some risk might generate more net health in the long-term. This ethical conflict is analogous to the conflicts faced in designing clinical trials today. While clinical research is generally considered a necessary part of the biomedical system – it is very difficult for doctors to claim to provide best care without any way of evaluating different care protocols and determining which are, in fact, the most beneficial (Gillon, 1994) – individual patients enrolled in clinical trials may not face the best outcomes available to them. It is important in designing clinical trials to ensure that the standards laid out by Freedman (1987) are met. A trial should only be conducted when there is genuine uncertainty as to the relative benefits of the treatment options being evaluated.

Thus, from the standpoint of the ethical principles laid out above, a reinforcement learning agent should not include options in its consideration that are considered to be clearly worse for patients. It should explore new treatment options only when there is a genuine possibility that the new options will be as beneficial or more beneficial to the patient that proven treatments. There is, however, an obligation to explore, given that those concerns are met, for without exploration of potential treatments – and without clinical trials of promising new medications and surgeries – treatment quality will stagnate, and the medical system will be failing as a whole to provide best care.

An ideal agent will administer optimal or near optimal (above a certain threshold) treatments for patients, but also take calculated opportunities to explore the treatment space to determine more effective treatment pathways. Thus, a reinforcement learning agent must be able to induce an effective mapping from states to actions to be taken (Sutton, 1998): for example, the system must correctly provide erlotinib to patients with EGFR+ non-small-cell lung cancer). As in batch reinforcement learning, this policy must be based on and incorporate past data and literature and draw on statistical models. While there have been a number of successful real-world cases using batch reinforcement learning (Lange, 2012), Tenenbaum et al. (2010) note that there has been limited work on the use of batch reinforcement learning on clinical data. As such, there remain many questions and challenges that will need to be addressed.

There are a variety of model-based and model-free approaches for a batch reinforcement learning agent. A model-based approach faces the challenge of having to learn a model from the enormous state-space of cancer treatment and derive a policy, via planning, from that model (Lizotte, 2012). However, advances in planning algorithms lag behind those of model learning algorithms (Tenenbaum et al., 2010). Alternative model-free learning methods have been explored as well, and while model-free models avoid both planning and model learning, these approaches induce a single policy for all patients across different groups, lacking customization for individual patients (Ernst, 2005; Lagoudakis, 2003; Tenenbaum et al., 2010). Tailoring care decisions to each individual patient is an important standard of care as each may experience various responses to treatments, including negative side effects; may have different contraindications against certain treatment options; or may have different personal preferences. One patient may wish to avoid surgery whereever possible, and another might value reduction of pain over a marginal extension of life.

Tenenbaum et al. (2010) discusses the potential of promising new model-based approaches given recent advancements in planning algorithms. One such algorithm, UCT, extends Monte-Carlo planning and was shown to be consistent and significantly more efficient in finding near-optimal solutions in a large state-space (Kocsis, 2006). In traditional Monte-Carlo planning, a domain specific heuristic is used to determine which states are most worth exploring. As a result, an extended Monte-Carlo planner will spend an additional, albeit expensive, amount of computation at each time step to derive the most optimal decision from the current state. However, these decisions can be tailored and personalized to match the individual information of a patient candidate at the expense of computational costs (Kocsis, 2006; Tenenbaum et al, 2010).

When deploying such learning and planning algorithms in practice, it is imperative that a planning algorithm factor in customization for individual patients while respecting their autonomy, maintain optimal or near optimal treatments recommendations for patients, and satisfy the need for exploration so as to ensure that future benefit and progress is made. To satisfy these constraints, a tradeoff between computational costs and the quality of the suggestions must be made. In the framework of medical ethics, spending more on computation in order to ensure that an optimal solution is obtained that promotes maximal health is well worth it.

4.1 Reward Function

Another key feature of a reinforcement learning agent is its reward function, which establishes the goal of the agent. More formally, a reward function defines a mapping from a state in an environment to a numerical value, its reward, indicating the intrinsic desirability of a potential state (Sutton, 1998). When defining a reward function in the context of the medical domain, various factors must be considered, from symptom measurements for individual patients to the side effects caused by various treatments (Lizotte, 2010).

A majority of existing reinforcement agents for clinical data use a weighted combination of these various factors. Reward functions in the proposed system would be calculated based on three core measures, with some flexibility for side constraints: (1) maximizing Quality Adjusted Life Years (QALYs), (2) minimizing negative side effects, and (3) patient preferences. This compound measure would account for three of the core ethical principles explained above. A measure of maximum QALYs would be a proxy for measuring the good done by following various treatment paths. A measure of negative side effects would allow for the reward function to minimize harm. Accounting for patient preferences is necessary to uphold the principle of respect for patient autonomy.

Preferences are likely to vary widely among patients. It is likely that patients and practitioners may have difficulty specifying precisely how they wish these preferences to be weighted. Tenenbaum et al. (2010) suggest that one possible method to quantify patient preferences would entail having patients make judgments of potential treatment sequences. A patient could rank pairs of treatment sequences based on her own preferences, and the rankings could then be used either directly in the rankings of various treatment options or could be used to construct some reward function that respects those preferences.

The weighting of QALYs against side-effect avoidance is complicated, and there is no ethical principle that provides a clear guide as to which should be weighted more heavily. Aggressive treatments that reduce symptoms severely might introduce debilitating negative side effects, while less aggressive treatments that introduce fewer side effects might also be far less effective at eliminating symptoms (Lizotte, 2008). Again, patient preferences might provide a guiding measure as to how rewards should be weighted.

Both the discount factor and the reward values can be optimized for patient preferences. If a patient is more concerned with short-term relief from suffering than with extension of life, the discount factor may be oriented to optimize for the short term. If a patient finds the prospect of negative side effects entirely intolerable, rewards for avoiding side effects might be weighted higher. It is imperative that the discount and reward factors are dynamic. Providing a fixed discount factor or reward function would be tantamount to assuming that all patients have the same wishes and concerns about their treatment, and that every patient would make the same decision if offered a given set of options. This would be a concerning position to take, given the need to respect patient autonomy.

4.2 Resource Allocation

As a system that functions as an aid to physicians in making decisions about individual patients, considerations such as allocation of funds and resources are beyond the scope of the recommendations made by this system. The system should not take factors such as financial cost into account when making decisions, and should not make systems-level comparisons. The system should be optimized to recommend the best care for the present patient. In a scenario where there is only one kidney available, and a physician has two patients in need of a kidney, the system should make recommendations to each patient as in a tabula rasa state. The recommender system should never be in a position of deliberating between patients or making value judgments of patients to determine which should receive better care in the instance of resource scarcity. These judgments should be reserved for the human physician.

5. RECOMMENDATION PRESENTATION AND ADHERENCE DESIGN

While other design decisions exist and will eventually need to be investigated both ethically and technically, a salient consideration is the way with which recommendations are ultimately made to the physician.

Recommendations should be presented in such a way that the amount of information presented to the physician is maximized. The recommender systems algorithm can be thought of in itself as a kind of low-risk research, for it results in statistical information about the viability and effectiveness of treatment options for a given patient. In order for the principles of beneficence and nonmaleficence to be upheld, the physician should be provided with enough information to make an effective decision as to how to treat her patient.

Thus, the way that recommendations are ultimately presented to the physician has great significance for the ethical status of the system. How many options are ultimately presented to the physician – is a single treatment recommended, with little information about how or why the system arrived at that decision, or is a ranked list of treatment options presented? Does the system select one option as the best option, taking all variables into account, or does it present one option that has the lowest risk, one that has the greatest chance of success, and one that will be the least expensive financially and emotionally to the institution and to the patient?

5.1 Lying Recommender Systems

In some recommender systems, lying is a key design feature. Like a recommender that would guide the user through a list of stores at which to shop for a computer by presenting information about price ranges of computers at each store, but lie about the probability of a computer being a very low price in order to influence the user to shop at that particular store, a treatment recommender could lie about or conceal information about a given treatment in order to influence the physician to take the course of action that the system found statistically to be most beneficial.

Though a lying recommender system might indeed be able to convince users to follow the most statistically beneficial course of treatment, medical ethics would not allow for any system that provided false information to the physician. This would be an infringement of both the physicians ability to uphold the principle of beneficence – perhaps the suggested treatment would have side effects that the physician is aware would have a particularly negative effect on the patient – and the physicians ability to respect the patients autonomy. The patients autonomy in choosing and consenting to his treatment is dependent upon his understanding of the treatment options available. If he is not given complete information or is given false information, the physician is in effect acting coercively, and the patients decisions can no longer be autonomous.

This would be a serious violation of medical ethics. There is certainly a broad precedent in medicine of deceiving patients in order to meet the principle of beneficence. However, this deception is only allowed when the patient has failed to meet a standard of competence. If a patient is not able to competently make decisions, the argument holds, his autonomy is impossible in any regard, and so the principle of beneficence should be upheld over all (Gillon, 1994). In any context in which the patient is indeed competent and would be granted the respect to consent to his treatment, he should not be deceived.

5.2 Presentation of Multiple Options

A system would do a far better job of meeting the principles of medical ethics as laid out by Gillon (1994) if it were to present multiple options for treatment. The physician should be able to deliberate between options for treatment in order to determine the course of action that would most maximize benefit and minimize harm for the patient in front of her. When the recommender system performs this deliberation in entirety, from selection of the best few options to determination of a single care plan, the physicians ability to guarantee the principle of beneficence is met is severely hindered.

An ideal system would present multiple options, ranking more highly the options that it believes to most maximize good and minimize harm, but would present them alongside information about their likelihood of success, possible harms, and a clear explanation of the way it evaluated those harms and benefits in making its decision.

A particular treatment option might have a higher rate of success, for example, but might be far more expensive. The learning algorithm might learn how to make a decision in this situation following rules that resemble the ethical principles held by physicians, but ultimately, the principles of beneficence and non-maleficence would hold that the final decision should be made by the physician, with the consent of the patient.

5.3 Opting Out

An important consideration is whether or not the physician should be allowed to opt out of the treatment(s) recommended by the system and instead follow an alternative course of treatment. The scope of interpretation of the principles of beneficence is at question here, as it is the the standard course of prescription of treatment by human physicians.

If applied at the level of the patient, the physician should disregard the recommendation made by the system – or choose among the recommendations made – if the physician knows with certainty that one treatment will leave the patient worse off. This might mean opting out of a slightly less effective treatment that would teach the system far more, and allow it to make better recommendations for future patients. At the level of the community of physicians using this recommender system, then, the physician should heed the recommendation made by the system: this would provide the system with a learning opportunity and the chance to modify and improve its diagnostic ability for the benefit of future patients (and, potentially, for future treatment of the same patient).

A clear analogy can be drawn to the tension between a physician recommending that a patient enroll in a clinical trial when she has a compelling belief that one treatment option is better than the other. If no one were to enroll in clinical trials, there would be no way to advance knowledge of better treatment and quality of care would be static indefinitely. If some physicians and patients are willing to assume minimal risk, the benefits to the system are large.

In the same vein, if many physicians were to opt out of the treatment recommended by the system, it would be impossible for it to learn from the outcomes of its recommendations, and the quality of the recommendations made would never improve.

Shrager and Tenenbaum (2014) propose that several available options for treatment be presented, ranked on the basis of the treating physicians knowledge and the standard of care (see Figure 2 for a model). When there is no clear superior option, they suggest the available options be presented in order of the information gathering value of each. As they note, clinical trials and non-standard treatments would only be proposed in the event that no standard treatment options are available, even though those options might present a large learning opportunity to the system. First and foremost, options are ranked to maximize the good for the present patient; the learning of the system and the good for future patients are only considered when all options present the same benefit to the present patient.

A system that offers several ranked options to the patient and makes clear both which ones are statistically likely to be superior and the decision process by which it arrived at those conclusions would allow the physician to evaluate recommendations in the context of the individual patient and the practice. This would be an ideal system to meet the ethical demands of medical practice.

For the present moment, a system that does not offer a ranked list of options but instead provides only one recommendation will still be a highly useful decision-making tool. The recommendation may provide valuable insight to the prescribing physician, especially if she is trained as to the decision-making calculus used by the system and is able to draw her own conclusions from its recommendation. However, for the system to be ethically permissible, the physician must be allowed to opt out of the recommendation made and, indeed, encouraged to do so when prevailing factors indicate that an alternative course of treatment would maximize the good in a particular instance. This will not give the system as great an opportunity to learn, but will ensure that the right to autonomy and the standard of beneficence are upheld.

6. CONCLUSION

Modern medicine has allowed for an unprecedented number of treatment options for patients. As these become increasingly targeted and numerous while more and richer data become available for physicians on their patients, it has and will become more difficult to pick maximally effective treatments for the reduction of disease. Development of a treatment recommender system that uses EHR data, past literature, and patient preferences

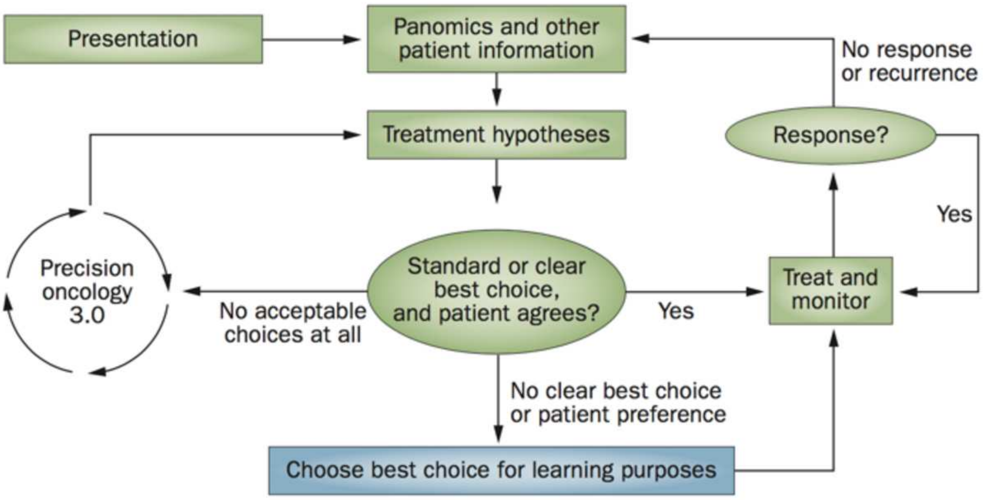
Presentation
Panomics and other patient information
No response or recurrence
Treatment hypotheses
Response?
Yes
Precision oncology 3.0
Standard or clear best choice, and patient agrees?
No acceptable choices at all
Yes
Treat and monitor
No clear best choice or patient preference
Choose best choice for learning purposes

Figure 2. Global Cumulative Treatment Analysis (Shrager & Tenenbaum, 2014).

could assist physicians in this task. Medicine is a high-stakes field, and poor recommendations could negatively impact humans. Implementing such a system will require intense scrutiny of certain design choices and their effect upon the recommendations made by the system. The field of medical ethics, founded upon its four key tenets of respect for patient autonomy, beneficence, non-maleficence, and justice will be critical in ensuring these design choices are ethically sound.

Key decisions needed to be made with respect to design include what learning algorithm the recommender system uses, how it navigates the trade-off between exploration and exploitation and is rewarded, and how recommendations are presented and whether patients should be able to opt out of being treated with these recommendations. With respect to the learning algorithm, a system that uses batch reinforcement learning seems ethically and technically promising, as it could use past medical data to initially learn from, getting to better recommendations more quickly, and it has an intuitive design analogous to the decision process of physicians, potentially increasing its adoption. Additionally, this algorithm works well for rarer cases and has a constantly improving model. All of these features mean such a system would develop better recommendations more quickly, thereby functioning in accordance with the concept of beneficence.

The decisions made surrounding how a system should trade-off between exploration and exploitation and is rewarded is more complex. Ethics surrounding medical research can help guide this decision, indicating that some exploration should be included in a systems recommendations to work towards better long-term recommendations, however, a system should not risk patients lives when strong alternatives exist. A system that is rewarded via a compound measure that takes into account maximizing quality adjusted life years, minimizing negative side-effects, and adherence to patient preferences would be designed around the concepts of beneficence, non-maleficence, and respect for patient autonomy respectively.

While many other design considerations are present and will require future investigation, determining how the system presents its options and whether patients are able to opt out of these recommendations are salient choices that need to be made early on in the design process. For these, an ethically sound system would not lie to a physician or patient, would present multiple treatment options with valuable information about the benefits and risks of each, and would allow physicians and patients to opt out of any recommendation made. Doing so would respect the physicians decision making capabilities and patient autonomy. In the end, any treatment made by one of these systems is a recommendation and a physician should have the final say in a patients care, taking care to respect the patients desires and autonomy.

Developing a treatment recommender system that provides strong treatment recommendations founded in clinical data and ethical design will require considerable future work. The data on which such a system operates will need to be improved. Currently EHR data is often left incomplete and does not capture all meaningful data for basing a recommendation. Efforts in data cleaning and aggregation, as well as increased data sharing across care providing facilities would improve this problem, allowing a system to learn better. Additional research will need to be done on how such a system would work in a clinical setting in real-time, and extensive effort would need to be put into testing and optimizing the reward functions and exploration-exploitation ratios that systems use. Moving into new diseases spaces will require different models, and dealing with co-morbidities (e.g. having cancer and diabetes) will present unique challenges. As more is learned about various diseases, models may need to be radically revised to ensure that the best care possible is provided to patients. Lastly, significant work needs to be done to make sure a recommender system is able to explain to doctors how it came to its result. Without this, such a system does not provide much value to a physician as it would either only confirm an already thought of course of treatment or confuse a physician with its output. Overall, the design of a system that functions well in health care treatment recommendations should be pursued due to the potentially immense benefits it could have on individuals and society.

REFERENCES

Brahmer, J. R., Tykodi, S. S., Chow, L. Q., Hwu, W. J., Topalian, S. L., Hwu, P., ... Pitot, H. C. (2012). Safety and activity of anti PD-L1 antibody in patients with advanced cancer. New England Journal of Medicine, 366(26), 2455-2465.

Buffery, D. (2015). The 2015 oncology drug pipeline: innovation drives the race to cure cancer. American health drug benefits, 8(4), 216.

Caplan, A. L. (1984). Is there a duty to serve as a subject in biomedical research?. IRB: Ethics Human Research, 6(5), 1-5.

Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., Elhadad, N. (2015, August). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1721-1730). ACM.

Crimmins, E. M., Preston, S. H., Cohen, B. (Eds.). (2011). Explaining divergent levels of longevity in high-income countries. National Academies Press.

Dorfer, L., Moser, M., Bahr, F., Spindler, K., Egarter-Vigl, E., Giullen, S., Kenner, T. (1999). A medical report from the stone age?. The Lancet, 354(9183), 1023-1025.

Ernst, D., Geurts, P., Wehenkel, L. (2005). Tree-based batch mode reinforcement learning. Journal of Machine Learning Research, 6(Apr), 503-556.

Faden, R. R., Kass, N. E., Goodman, S. N., Pronovost, P., Tunis, S., Beauchamp, T. L. (2013). An ethics framework for a learning health care system: a departure from traditional research ethics and clinical ethics. Hastings Center Report, 43(s1), S16-S27.

Faust, A. (2012). Reinforcement Learning as a Motion Planner-A Survey. Technical report, University of New Mexico, Department of Computer Science, 2012. [Online: http://www.cs.unm.edu/ pdevineni/papers/Faust.pdf].

Freedman, B. (1987). Equipoise and the ethics of clinical research. N Engl J Med, 317(3), 141-145.

Gillon, R. (1994). Medical ethics: four principles plus attention to scope. Bmj, 309(6948), 184.

Hooker, B. (2000). Ideal code, real world: A rule-consequentialist theory of morality. Oxford University Press.

Hume, D. (2010). Of suicide. Life, Death, and Meaning: Key Philosophical Readings on the Big Questions, 291.

Johnson, R. (2008). Kant's moral philosophy. Stanford encyclopedia of philosophy.

Kocsis, L., Szepesvri, C. (2006, September). Bandit based monte-carlo planning. In European conference on machine learning (pp. 282-293). Springer Berlin Heidelberg.

Lagoudakis, M. G., Parr, R. (2003). Least-squares policy iteration. Journal of Machine Learning Research, 4(Dec), 1107-1149.

Lange, S., Gabel, T., Riedmiller, M. (2012). Batch reinforcement learning. InReinforcement learning (pp. 45-73). Springer Berlin Heidelberg.

Le Tourneau, C., Lee, J. J., Siu, L. L. (2009). Dose escalation methods in phase I cancer clinical trials. Journal of the National Cancer Institute.

Lizotte, D. J., Gunter, L., Laber, E., Murphy, S. A. (2008). Missing data and uncertainty in batch reinforcement learning. In Neural Information Processing Systems (NIPS).

Lizotte, D. J., Bowling, M. H., Murphy, S. A. (2010). Efficient reinforcement learning with multiple reward functions for randomized controlled trial analysis. In Proceedings of the 27th International Conference on Machine Learning (ICML-10) (pp. 695-702).

Lyerly, A. D., Little, M. O., Faden, R. (2008). The second wave: Toward responsible inclusion of pregnant women in research. IJFAB: International Journal of Feminist Approaches to Bioethics, 1(2), 5-22.

McGlynn, Elizabeth A., et al. "The quality of health care delivered to adults in the United States." New England Journal of Medicine 348.26 (2003): 2635-2645.

Rawls, J., 1999. A Theory of Justice. Cambridge, Mass: Belknap Press of Harvard University Press.

Shrager, J., Tenenbaum, J. M. (2014). Rapid learning for precision oncology. Nature reviews Clinical oncology, 11(2), 109-118.

Skinner, J., Fisher, E. (2010). Reflections on geographic variations in US health care. Dartmouth Institute for Health Policy and Clinical Practice, May, 12, 2013.

Sutton, R. S., Barto, A. G. (1998). Reinforcement learning: An introduction (Vol. 1, No. 1). Cambridge: MIT press.

Tenenbaum, M., Fern, A., Getoor, L., Littman, M., Manasinghka, V., Natarajan, S., Page, D., Shrager, J., Singer, Y. Tadepalli, P. (2010, Dec.). Personalized Cancer Therapy via Machine Learning. In Neural Information Processing Systems Foundation Workshop. Lecture conducted from Hilton Whistler, Whistler, BC.

Wendler, D. (2009). The Ethics of Clinical Research. The Stanford Encyclopedia of Philosophy. Ed. Edward N. Zalta. N.p., 2012. Web. 10 Dec. 2016

Yelon, J. A. (2014). Geriatric trauma and critical care. F. A. Luchette (Ed.). Springer New York.

Zhao, Y., Kosorok, M. R., Zeng, D. (2009). Reinforcement learning design for cancer clinical trials. Statistics in medicine, 28(26), 3294-3315.