

## **Title Page**

### **Article Title:**

Quickest Way to Less Headache Days: an operational research model and its implementation for Chronic Migraine

### **Authors:**

Irene Lo, Ph.D.

Pengfei Zhang, MD

### **Author affiliations:**

From the Department of Neurology, Rutgers Robert Wood Johnson Medical School, New Brunswick, New Jersey, USA (Pengfei Zhang)

From the Department of Management Science & Engineering, Stanford University, Stanford, California, USA (Irene Lo)

### **Conflict of Interest Statement:**

Irene Lo: No conflict

Pengfei Zhang: He has received honorarium from Alder Biopharmaceuticals, Board Vitals, Fieve Clinical Research. He collaborates with Headache Science Incorporated without receiving financial support. He has ownership interest in Cymbeline LLC.

**Key Words:** Operations Research, Chronic Migraine, Headache, Medication Selection, Medical Decision Research, Markov Decision Process, Multi-armed bandit process

**Financial Support:** None

**Trial Registration:** Not applicable.

**Abbreviations:** International Headache Congress (IHC), Operations Research (OR), Human Immunodeficiency Virus (HIV), Markov Decision Process (MDP), calcitonin gene-related peptid (CGRP)

### **Acknowledgements:**

Laura Hester Ph.D for feedback on our project.

## Structured Abstract

### Background:

Choosing migraine prevention medications often involves trial and error. Operations research methodologies, however, allow us to derive a mathematically optimum way to conduct such trial and error processes.

### Objective:

Given probability of success (defined as 50% reduction in headache days) and adverse events as a function of time, we seek to develop and solve an operations research model, applicable to any arbitrary patient, minimizing time until discovery of an effective migraine prevention medication. We then seek to apply our model to real life data for chronic migraine prevention.

### Methods:

We develop a model as follows: Given a set of  $D$  many preventive medications, for drug  $i$  in  $D$ , we describe the likelihood of reaching 50% headache day reduction over the course of time,  $(t_{i,1} \leq t_{i,2} \leq \dots)$  by probability  $(p_{i,1} \leq p_{i,2} \leq \dots)$ . We additionally assume a probability of adverse event  $(q_{i,1} \leq q_{i,2} \leq \dots)$ . We then solve for a sequence of prescription trials that minimizes the expected time until an effective drug is identified.

Once we identify the optimal sequence for our model, we estimate  $p$ ,  $t$  and  $q$  for topiramate and OnabotulinumtoxinA based on the FORWARD study by Rothrock et al as well as erenumab data published by Barbanti et al. at IHC 2019.

### **Results:**

The solution for our model is to order the drugs by probability of efficacy per unit time.

When the efficacy of each drug  $i$  is known only for one period  $t_i$  and there are no adverse effects, then the optimum sequence is to administer drug  $i$  for  $t_i$  periods in decreasing order of  $p_i/t_i$ . In general, the optimum sequence is to administer drug  $i$  for  $t_{i,k^*}$  periods in decreasing order of the Gittins index  $\sigma_{ik^*}$ .

Based on the above data, the optimum sequence of chronic migraine prevention medication is a trial of erenumab for 12 weeks, followed by a trial of OnabotulinumtoxinA for 32 weeks, followed by a trial of topiramate for 32 weeks.

### **Conclusion:**

We propose an optimal sequence for preventive medication trial for patients with chronic migraine. Since our model makes limited assumptions on the characteristics of disease, it can be readily applied also to episodic migraine, given the appropriate data as input.

Indeed, our model can be applied to other scenarios so long as probability of

success/adverse event as a function of time can be estimated. As such, we believe our model may have implications beyond our sub-specialty.

## Introduction

Migraine has an approximate global prevalence of 14.7%.<sup>1</sup> Unfortunately, the process of finding an effective prevention therapy for patients is often largely trial and error process.<sup>2</sup> Medication trial failures frequently cause frustrations and a sense of helplessness among migraine sufferers.<sup>3,4</sup> Therefore an expedient way of determining effective prevention therapy can significantly benefit our patient population.

Operations research (OR) is the study of optimal decision-making using mathematical models. In healthcare, OR has been used to inform clinical trial designs and medical decision making in chronic diseases.<sup>5-7</sup> For example, OR models have been used to propose selections of chemotherapy combinations for clinical trials in oncology, to identify optimal timing and criteria for cancer screening, and to inform the optimal criteria for initiating medications such as statin in type 2 diabetics or highly active antiretroviral therapy in HIV patients.<sup>8-10</sup> OR models can also be used to support surgical decision-making, determining when to surgically correct abdominal aortic aneurysm or to receive kidney transplantation.<sup>11,12</sup> To the best of our knowledge, however, operations research methods have not been applied to medication selection in headache.

The aim of this paper is to propose an operations research model for migraine medication selection. We assume that given any randomly selected patient one can estimate the

probability of success for individual prevention medications as a function of time. We then solve our model for the optimal sequence of medication trial, minimizing time for the discovery of an effective treatment. As a proof of concept, we utilize published data to demonstrate that such a model can be implemented for real world decision-making purposes in chronic migraine prevention.

(Of note, our project was presented as a poster presentation during the American Headache Society's June 2020 Virtual Meeting.)

## **Methodology**

Our methodology and results outline our operations research model for the clinician and layperson. We have included an in depth technical methodological portion for the technical reader in appendix A. Key proofs of our result are also included in appendix A.

### **Setting up the model:**

For modeling purposes, we define successful/effective treatment response as 50% improvement in headache days. This value is not intrinsic to our model but is intrinsic to the input data. We assume that given any two medications, their respective probabilities of success are statistical independence. We assume that any intolerable adverse event would

be discovered before observing 50% improvement in headache days. Finally, only one medication is allowed to be tried at a time.

Our model is as follows:

Given a set of  $D$  many preventive medications, for drug  $i$  in  $D$ , we describe the likelihood of reaching 50% headache day reduction over the course of time,  $(t_{i,1} \leq t_{i,2} \leq \dots)$  by probability  $(p_{i,1} \leq p_{i,2} \leq \dots)$ . We additionally assume a probability of adverse event  $(q_{i,1} \leq q_{i,2} \leq \dots)$ . We then solve for a sequence of prescription trials that minimizes the expected time until an effective drug is identified.

Topiramate, Botulinum Toxin (OnabotulinumtoxinA), and CGRP antagonists are the only medications widely accepted to have evidenced based efficacy in chronic migraine prevention.<sup>13,14</sup> We use published 50% responder rates for week 12 and week 32 as well as adverse event rates from the FORWARD study as estimates for  $(p_{i,1} \leq p_{i,2} \leq \dots)$  and  $(q_{i,1} \leq q_{i,2} \leq \dots)$  for topiramate and OnabotulinumtoxinA.<sup>15</sup> We use 50% responder rates from Barbanti and colleagues' retrospective study to estimate probability of real world efficacy as a function of time for erenumab.<sup>16</sup> We have decided against using clinical trial data due to methodological concerns; galcanezumab and fremanezumab are therefore not

included in this study. (See Limitation Section.) We estimate adverse event for erenumab based on package insert.<sup>17</sup> Our input data for the model is summarized in Table 1.

(A quick note on notation: The set  $D$  in our case represents the set of three medications: OnabotulinumtoxinA, topiramate, erenumab. The subscript  $i$  therefore represents one of these medications. The second subscript  $k$ , represents the epoch of time under consideration. The notation  $P(x)$  represents probability of  $x$  and  $E(x)$  represents expected value of  $x$ , in accordance with traditional mathematics parlour. The asteroid (\*) next to a variable is used to denote optimum solution.)

## Theoretical Results

Our model a case of the multi-armed bandit problem in operations research. Solution to this class of problem can be solved through calculation and identification of a ratio called the Gittins Index at every decision point.<sup>23</sup> In other words, index ratios are calculated for all possible choices at every decision point; one then identifies the next optimum decision by picking the decision with the highest index. We denote the calculation of these indices as  $\sigma_{i,k}$ . We denote the Gittins index as  $\sigma_{i,k}^*$ .

The solution for our model is to order the drugs by probability of efficacy per unit time. Let us first consider an unrealistic but simplified scenario to our problem: When the efficacy of



each drug  $i$  is known only for one period  $t_i$ , then the optimum sequence is to administer drug  $i$  for  $t_i$  periods in decreasing order of  $p_i/t_i$ . For example, under this simplified scenario, suppose a hypothetical antiepileptic has a 90% probability of being effective (let's call this  $p_{AED}$ ) only by the end of 3 months ( $t_{AED}$ ) whereas a hypothetical CGRP mab has a probability of 50% ( $p_{CGRP}$ ) only by the end of 1 month ( $t_{CGRP}$ ). Then it is wise to try the hypothetical CGRP first for 1 month followed by trying the antiepileptic for 3 months if the goal is to discover an effective medication as quickly as possible. Note that in this simplified scenario we assume that one does not see any intermediate benefit between month 0 and month 3 for the antiepileptic nor any intermediate benefit between day 0 and day 1 for the CGRP. This, of course, is not realistic. (The proof of this result is under the Special Case section in Appendix A.)

Our problem is of course more complicated than the simplified version above and a generalized, and therefore more realistic, solution to the above can be derived: the optimum sequence is to administer drug  $i$  for  $t_{i,k^*}$  periods in decreasing order of the index  $\sigma_{i,k^*}$ , where

$$\begin{aligned}\sigma_{i,k} &= \frac{P(\text{efficacy with no adverse effects if administered until time } t_{i,k})}{E(\text{time administered if administered at most until time } t_{i,k})} \\ &= \frac{\sum_{j \leq k} (p_{i,j} - p_{i,j-1})(1 - q_{i,j})}{\sum_{j \leq k} (t_{i,j} - t_{i,j-1})(1 - p_{i,j-1})(1 - q_{i,j-1})}\end{aligned}$$

is the probability of efficacy per unit time if the drug were administered until time  $t_{i,k}$ , and

$$k^* = \operatorname{argmax}_k \sigma_{i,k}$$

is the number of time periods that maximizes the efficacy per unit time of the drug. Note that in order to determine which drug to try next, the indices for the trialed drug  $i$  need to be recalculated after trying it for time  $t_{i,k^*}$ . (Proof of the above results are included in Appendix A.)

In other words,  $\sigma_{i,k}$  represents a generalized version of  $p_i/t_i$ . The numerator of  $\sigma_{i,k}$  represents the probability of the drug efficacy if administered until time  $t_{i,k}$ . The denominator of  $\sigma_{i,k}$  represents the time the drug is administered to the patient before time  $t_{i,k}$ , taking into account that this time may be shorter than  $t_{i,k}$  if the drug is successful earlier in the process.

To calculate the index for individual drugs, we must first find the optimal medical trial time frame  $t_{i,k^*}$ , represented in  $k$ , by picking  $k^*$  such that it maximizes  $\sigma_{i,k}$ . This is the

calculation of the Gitten's index. Once these selections are obtained, then we simply rank them in decreasing orders of  $\sigma_{i,k^*}$ , trying each drug for  $t_{i,k^*}$  periods.

### **Application of model to chronic migraine data:**

Based on FORWARD as well as Barbanti study data, we calculated the above indices for the first iteration and present the results in Table 2. (Here,  $\sigma_{ik}$  represents the index for each medication at specific time k for drug i.) Input probabilities are adjusted to ensure probabilities are increasing over time, this is done by simply taking the maximum probability of efficacy up to each specific time frame.

Note that regardless of potential erenumab trial duration, the resulting index is larger than the largest indices for both of the other drugs. Therefore it is optimal to try erenumab first for the full 12 weeks. As such, one optimal sequence of chronic migraine prevention medication is:

- 1) A trial of erenumab for 12 weeks
- 2) If the above fails, then proceed with a trial of OnabotulinumtoxinA for 32 weeks
- 3) If the above fails, then proceed with a trial topiramate for 32 weeks.

Since estimates for OnabotulinumtoxinA and topiramate are based on a lower and upper bound based on our input data (the FORWARD study contains these values), it is possible that the following alternative sequence is also optimal:

- 1) A trial of erenumab for 12 weeks
- 2) If the above fails, then proceed with a trial of OnabotulinumtoxinA for 12 weeks
- 3) If the above fails, then proceed with a trial of topiramate for 12 weeks
- 4) If the above fails, switch back to a trial of OnabotulinumtoxinA for 20 further weeks
- 5) If the above fails, switch back to a trial of topiramate for 20 further weeks.

This alternative optimal solution is, of course, only a theoretical one given that switching between OnabotulinumtoxinA and topiramate would be logistically difficult and clinically peculiar.

### **Instrumentation:**

Our model is not a simulation and therefore the results can be calculated by hand. We, have, however, conducted our calculations through the use of Microsoft Excel.

### **Discussion:**

In this proof of concept study, we frame the process of migraine preventive medication selection as an operations research problem. We then outline a methodology for the construction of an operations research model, its mathematical solution, and its real life implementation. We believe such a perspective will be invaluable to clinical practice. Applications of operations research to medical decision making/clinical trial design have relied mostly on Markov Decision Process (MDP).<sup>18, 19</sup> Traditionally, medical decision-making MDP models are personalized, having the benefit of producing optimizing policy through individual-specific information as well as treatment responses.<sup>9-12</sup> However, solutions to personalized MDP models can be computationally demanding, requiring large amount of data as input and are often difficult to interpret for practitioners. These limitations have prevented widespread use of personalized MDP models clinically.<sup>10, 20</sup> As an alternative to personalized MDP, our paper utilizes a multi-armed bandit process for medication selection. A multi-armed bandit process is a special case of MDP well studied in OR and utilizes an “index” method. Our method is based on an indexing method proposed previously for evaluation of sequential clinical trial design.<sup>5, 6</sup> We are unable to find a prior instance of utilization of this method in clinical decision-making for medication selection.

We want to be very clear, however, that our model is not intended to be utilized to the exclusion of all other patient centered concerns such as comorbidities, access, or drug-drug interactions. Indeed, mathematical modeling is not a substitute for good doctoring. For example, obviously we do not recommend practitioners initiate topiramate if the patient has a history of nephrolithiasis, nor do we think that one should start off with erenumab as a treatment should the patient has a history of severe hypertension or severe constipation. Furthermore, mathematical modeling is not a panacea for addressing medication overuse headaches, obtaining insurance approved to prevention medication, avoiding medication interactions, assessing patient preference, and all other clinical decision making processes. Addressing these real-life concerns are beyond the scope of this paper. In other words, mathematical model, just like animal models in biological sciences, is meant to inform clinicians/scientists.

By using a multi-arm bandit process, our model allows for an important intuitive result: that we can estimate optimal sequence for prevention medication selection with the index  $p_i/t_i$ . This provides a clinically intuitive way of understanding medication optimization – prioritizing and select medications that maximizes probability of efficacy per unit of time. In accordance with our model, this index should be re-evaluated in real time and on each follow up visits to offer continual guidance for the next medication choice.

Although the focus of this paper is chronic migraine, our model and its mathematical solution can be readily translated to episodic migraine as long as probabilities for success/adverse events as a function of time can be estimated. The insight for optimizing the index  $p_i/t_i$  can be directly translate to episodic migraine selection as well.

Finally, we believe that operations research models for medication selection should not be limited to our subspecialty. For example, one can imagine applying a similar approach to managing epilepsy and multiple sclerosis. Given that we make no assumption on the underlying disease process, our model and its solution may be directly applicable to a wide range of clinical medicine scenarios outside our discipline.

### **Strength and Limitations:**

Although we defined success in our model as 50% improvement, this number is not intrinsic to our model. In other words, the mathematics of our model would work equally well for any other parameter so long as the input data contains efficacy data as a function of time for the same parameters. That is, the logic for our model can be carried over to efficacies defined as, for example, 25% improvement, or 75% improvement, or even 50% decrease in headache severity with the caveat that we have data on how these parameters change as a function of time given specific medications. Indeed our model is limited by the availability of real life data.

Our model's assumptions are also its limitations: First, we assume mono-therapy in medication selection. While we believe this to be a reasonable assumption, we acknowledge that this is not often true in the real world. For example, it is not uncommon for patient to be prescribed topiramate while waiting for OnabotulinumtoxinA, resulting in eventual continuation of both OnabotulinumtoxinA and topiramate for prevention.

Challenging our paradigm of monotherapy is also the open question whether synergism exists between OnabotulinumtoxinA and CGRP antagonists. Of course, one can argue that the quickest way for treating chronic migraine would be to simply provide OnabotulinumtoxinA, CGRP, and topiramate concurrently. This solution is possible but often not feasible in real life due to economic limitations and clinical considerations.

Our model also does not take into account how patient-specific factors affects efficacies of individual medications. This is, in part, due to the lack of data for patient-specific factors for various medications as a function of time. For example, an extension of our model may be able to account for efficacies of topiramate as a function of time when body mass index (BMI) is taken into consideration. However, we would then require multivariable data on topiramate efficacy as a function of both time *and* BMI.

Our most significant limitation is therefore the scarcity of available data for our model.

Firstly, as a proof of concept paper we are not wedded to the idea of using only



FORWARD nor Barbanti data for our model; the primary purpose of this paper is a methodological one after all. Secondly, we have decided to only include real-life data based on the critique that clinical trial responder rates may not translate directly to real world settings. This methodological concern is especially regrettable in regards to CGRP data: While Forderreuther et al as well as Tepper et al estimated 50% response rates of erenumab and galcanezumab as a function of time, these study were extrapolated from clinical trial, therefore limiting their application to this paper.<sup>21, 22</sup> Finally, although multiple real-life data exist for OnabotulinumtoxinA efficacy, apart from the FORWARD study we are unable to locate a study where 50% responder rate as a function of time is documented.

#### **Future directions:**

Our model's limitations are sources for future directions: In order to apply our solution to real life decision-making, we require real life data in the form of responder rate as a function of time. Our model shows that these real-world data can significantly affect clinical decision-making. We hope that this paper will alert future researchers to the importance of real-world data, allowing for improvement in implementation of our method.

#### **Conclusions**

An operation research model can be constructed for optimal medication decision-making in headache prophylaxis. We outline the methodology for such an operations research model

and its empirical implementation. Although our model supports the sequence of erenumab, followed by OnabotulinumtoxinA, followed by a trial of topiramate, more data is needed to support this sequence as being optimal. Our model can be applied to episodic migraine should more real life data be made available. We believe a similar approach can be applied to other specialties.

### **Article Highlights:**

- We outlined a methodology for the construction of an operations research model, its mathematical solution, and its real life implementation for selection of migraine prevention medications minimizing time until efficacy.
- One can estimate optimal sequence for prevention medication selection with the index  $p_i/t_i$ . (A generalized and more realistic version of the above index is the Gittins Index.) This provides a clinically intuitive way of understanding medication optimization – prioritizing and select medications that maximizes probability of efficacy per unit of time.

## References:

1. Steiner TJ, Stovner LJ, Birbeck GL. Migraine: the seventh disabler. *The Journal of Headache and Pain*. 2013;14:1
2. Silberstein SD, Holland S, Freitag F, Dodick DW, Argoff C, Ashman E. Evidence-based guideline update: pharmacologic treatment for episodic migraine prevention in adults: report of the Quality Standards Subcommittee of the American Academy of Neurology and the American Headache Society. *Neurology*. 2012;78:1337–1345.
3. Sprenger T, Goadsby PJ. Migraine pathogenesis and state of pharmacological treatment options. *BMC Medicine*. 2009;7:71.
4. Palacios-Ceña D, Neira-Martín B, Silva-Hernández L, Mayo-Canalejo D, Florencio LL, Fernandez-de-las-Penas C, Gracia-Moreno H, Garcia-Azorin D, Cuadrado ML. Living with chronic migraine: a qualitative study on female patients' perspectives from a specialised headache clinic in Spain. *BMJ Open* 2017;7:e017851.
5. Berry DA. Modified Two-Armed Bandit Strategies for Certain Clinical Trials. *Journal of the American Statistical Association*. 1978;73:339-345
6. Berry DA, Pearson LM. Optimal designs for clinical trials with dichotomous responses. *Statistics in Medicine*. 1985;4:497-508

7. Ahuja V, Birge JR. Response-adaptive designs for clinical trials: Simultaneous learning from multiple patients. *European Journal of Operational Research*. 2016;248:619-633
8. Bertsimas D, O'Hair A, Relyea S, Silberholz J. An Analytics Approach to Designing Combination Chemotherapy Regimens for Cancer. *Management Science*. 2016;62:1225-1531
9. Steimle LN, Denton BT. Markov Decision Processes for Screening and Treatment of Chronic Diseases. In: Boucherie R, van Dijk N, eds. *Markov Decision Processes in Practice*. International Series in Operations Research & Management Science. Boston, MA: Springer International Publishing; 2017;248:189-222.
10. Schaefer AJ, Bailey MD, Shechter SM, Roberts MS. Modeling Medical Treatment Using Markov Decision Processes. In: Brandeau ML, Sainfort F, Pierskalla WP, eds. *Operations Research and Health Care*. International Series in Operations Research & Management Science. Boston, MA: Springer International Publishing; 2005;70:593-612
11. Mattila R, Siika A, Roy J, Wahlberg B. A Markov decision process model to guide treatment of abdominal aortic aneurysms. 2016 IEEE Conference on Control Applications (CCA), Buenos Aires, 2016;436-441

12. Alagoz O, Hsu H, Schaefer AJ, Roberts MS. Markov decision processes: a tool for sequential decision making under uncertainty. *Medical Decision Making*. 2009;30:474-483.
13. Diener HC, Solbach K, Holle D, Gaul C. Integrated care for chronic migraine patients: epidemiology, burden, diagnosis and treatment options. *Clinical Medicine*. 2015;15:344–350.
14. Loder E, Rizzoli P. Pharmacologic Prevention of Migraine: A Narrative Review of the State of the Art in 2018. *Headache*. 2018;58:218-229
15. Rothrock JF, Adams AM, Lipton RB, Silberstein SD, Jo E, Zhao X, Blumenfeld AM. FORWARD Study: Evaluating the Comparative Effectiveness of OnabotulinumtoxinA and Topiramate for Headache Prevention in Adults With Chronic Migraine. *Headache*. 2019;59:1700-1713
16. Barbanti P, Aurilia C, Egeo G, Fofi L, Vernieri F. Erenumab in migraine: the first Italian real-life data. *Cephalalgia*. 2019;39(1S):262-263
17. Erenumab [Package Insert] *Amgen Inc*; 2019.
18. Denton BT, Alagoz O, Holder A, Lee EV. Medical Decision Making:open research challenges. *IIE Transactions on Healthcare system engineering*. 2011;1:161-167.

19. Tunc S, Alagoz O, Burnside E. Opportunities for Operations Research in Medical Decision Making. *IEEE Intelligence System*. 2014;29:59-62
20. Schell GJ, Marrero WJ, Lavieri MS, Sussman JB, Hayward RA. Data-Driven Markov Decision Process Approximations for Personalized Hypertension Treatment Planning. *Medical Decision Making Policy & Practice*. 2016: doi: 10.1177/2381468316674214.
21. Tepper SJ, Lucas S, Ashina M, Schwedt TJ, Ailani J, Scanlon JV, Klatt J, Chou DE, Richards S, Lima GPS. Temporal Response Patterns to Erenumab in Patients with Chronic Migraine. *Cephalalgia* 2019;39(1S):211
22. Förderreuther S, Zhang Q, Stauffer VL, Aurora SK, Lainez MJA. Preventive effects of galcanezumab in adult patients with episodic or chronic migraine are persistent: data from the phase 3, randomized, double-blind, placebocontrolled EVOLVE-1, EVOLVE-2, and REGAIN studies. *The Journal of Headache and Pain*. 2018;19:121
23. Gittins JC, Jones DM. A dynamic allocation index for the discounted multiarmed bandit problem. *Biometrika*. 1979; 66:561-565.

## Appendix A: Proof of the key results

Let

$$\begin{aligned}\sigma_{i,k} &= \frac{\text{P(efficacy with no adverse effects if administered until time } t_{i,k})}{\text{E(time administered if administered at most until time } t_{i,k})} \\ &= \frac{\sum_{j \leq k} (p_{i,j} - p_{i,j-1})(1 - q_{i,j})}{\sum_{j \leq k} (t_{i,j} - t_{i,j-1})(1 - p_{i,j-1})(1 - q_{i,j-1})}.\end{aligned}$$

We show that the expected time to effective treatment is minimized using the following procedure:

1. Calculate  $\sigma_{i,k}$  for all arms  $i$  and periods  $t_{i,k}$ .
2. Find the largest  $\sigma_{i,k}$ , call this  $\sigma_{i,k^*}$
3. Administer drug  $i$  for  $t_{i,k^*}$  periods
4. If drug  $i$  is not effective after  $t_{i,k^*}$  periods, recalculate  $\sigma_{i,k}$  for all subsequent time periods for arm  $i$  and repeat steps 2-4.

### Special Case:

To illustrate why such a procedure should work, we first show that it minimizes the time to effective treatment when the efficacy of each drug  $i$  is known only for one period  $t_i$  and there are no adverse effects each drug. In this case

$$\sigma_{i,k} = \frac{p_i}{t_i}$$



and we show that the optimal prescription sequence is to prescribe drugs in decreasing order of  $p_i/t_i$ .

Suppose not, so there is some drug prescription sequence  $S$  that does better. Let us assume that the drugs are labeled so that the sequence  $S$  is in the order  $1, 2, \dots, n$ . Since the drugs are not in decreasing order of  $p_i/t_i$ , there are two drugs  $i, j$  such that  $p_i/t_i > p_j/t_j$  and drug  $i$  is prescribed after drug  $j$ . (In fact we can assume that drug  $i$  is prescribed right after drug  $j$ , so  $j = i - 1$ ). Then

$$\begin{aligned} & E[\text{time to effective treatment with sequence } S] \\ &= \sum_k P(\text{drug } k \text{ is first effective drug in sequence } S) E[\text{time to effective treatment} \mid \text{drug } k \text{ is first effective}] \\ &= \sum_k P(\text{drug } k \text{ is first effective drug in sequence } S) (t_1 + \dots + t_k). \end{aligned}$$

Let  $S'$  be the sequence where the order of drugs  $i$  and  $j$  are swapped. We show that the expected time to effective treatment with  $S'$  is strictly less than with  $S$ :

$E[\text{time to effective treatment with sequence } S']$

$$\begin{aligned}
 & - E[\text{time to effective treatment with sequence } S] \\
 & = P(\text{drug } i \text{ is first in } S')(t_1 + \dots + t_{i-2} + t_i) \\
 & + P(\text{drug } i - 1 \text{ is first in } S')(t_1 + \dots + t_i) \\
 & - P(\text{drug } i \text{ is first in } S)(t_1 + \dots + t_i) \\
 & + P(\text{drug } i - 1 \text{ is first in } S)(t_1 + \dots + t_{i-1}) \\
 & = P(\text{drug } i \text{ is first in } S')(t_i) + P(\text{drug } i - 1 \text{ is first in } S')(t_{i-1} + t_i) \\
 & - P(\text{drug } i \text{ is first in } S)(t_{i-1} + t_i) - P(\text{drug } i - 1 \text{ is first in } S)(t_{i-1}) \\
 & = P(\text{drugs } 1, \dots, i - 2 \text{ not effective})[p_i \cdot t_i + (1 - p_i)p_{i-1}(t_{i-1} + t_i) \\
 & - (1 - p_{i-1})p_i(t_{i-1} + t_i) - p_{i-1}(t_{i-1})] \\
 & = P(\text{drugs } 1, \dots, i - 2 \text{ not effective})[p_{i-1} \cdot t_i - p_i \cdot t_{i-1}] < 0
 \end{aligned}$$

where the final inequality is because  $p_i/t_i > p_{i-1}/t_{i-1}$  and so  $p_{i-1} \cdot t_i < p_i \cdot t_{i-1}$ .

### General Case:

We now show that when drug efficacy and probability of adverse effects are known for multiple periods the procedure outlined above minimizes expected time to effective treatment.

For each drug, we can think of taking one of two actions: either trying the drug for another week, or not trying the drug for another week. For every week we try the drug we pay a

cost of 1. If we try the drug for another week and it is effective with no side effects, we receive a reward of  $R$ , otherwise we receive a reward of 0. Once the drug is effective or the patient exhibits adverse effects we receive no further rewards for trying the drug. The decision-maker must sequentially choose every week at most one drug to prescribe for that week, with the goal of minimizing the expected time until efficacy. For sufficiently large  $R$  this goal is the same as maximizing the total reward minus cost.

This kind of problem is precisely a **multi-armed bandit process**, a special case of a Markov decision process that is well-understood by the operations research literature. A general multi-armed bandit process is defined as follows. There are  $n$  arms, each arm  $i$  starting in a given state  $x_i$ . For each arm  $i$ , we can take one of two actions: either pulling the arm this time period, or not pulling the arm this time period. If we pull the arm we receive some random reward  $r(x_i)$  that is a function of the state of the arm, and the arm randomly transitions to a new state that is a function of the current state of the arm. The decision-maker must sequentially choose in every period which arm to pull, with the goal of maximizing the total reward. There is a well-known solution to the multi-armed bandit process, which was first formulated by J. C. Gittins and D. M. Jones and is given by the following:

**Theorem.** The expected discounted reward obtained from a multi-armed bandit problem is maximized by pulling the arm with the greatest Gittins index:

$$G_i(x_i) = \sup_{\tau \geq 1} \frac{\mathbb{E}[\sum_{t=0}^{\tau} r_i(x_i(t))\beta^t | x_i(0) = x_i]}{\mathbb{E}[\sum_{t=0}^{\tau} \beta^t | x_i(0) = x_i]}$$

where  $\tau$  is a past-measurable stopping time.<sup>23</sup>

Let us rewrite the drug prescription problem as a multi-armed bandit problem. Let  $T_i = \max_k t_{i,k}$  be the longest amount of time we can prescribe drug  $i$ . Each arm  $i$  (representing drug  $i$ ) can be in one of the states  $\{1, \dots, T_i, s, -1\}$ , where 1 represents the initial state,  $s$  represents the state when the drug is first successful,  $-1$  represents the state where no more rewards are received, and  $k \in \{1, 2, \dots, m_i\}$  represents the number of times drug  $i$  has been tried. Arm  $i$  starts in the state ‘1’. From each state  $k \in \{1, \dots, T_i\}$ , if arm  $i$  is pulled the ‘reward’ is  $-1$  (i.e. paying a cost of 1 for each week the drug is tried). From each state  $k \in \{t_{i,1}, t_{i,2}, \dots\}$  and the state changes randomly according to the following probabilities: with probability  $p_{i,k}(1 - q_{i,k})$  the drug is successful with no adverse effects and the state changes to  $s$ , with probability  $q_{i,k}$  there are adverse effects and the state changes to  $-1$ , and with probability  $(1 - p_{i,k})(1 - q_{i,k})$  the drug is not successful with no adverse effects and the state changes to  $k + 1$  (or to state  $-1$  if  $k = m_i$ ). From state  $s$  if arm  $i$  is pulled the reward is  $R$  and the state changes to  $-1$ , and from state  $-1$  if arm  $i$  is pulled the reward is 0

and the state stays at  $-1$ . Finally, from all other states  $k \in \{1, \dots, T_i\} \setminus \{t_{i,1}, t_{i,2}, \dots\}$  the reward is  $-1$  and the state changes to  $k + 1$ . The goal is to maximize the expected total reward, which is given by

$$R(\text{probability success}) - (\text{time until find successful drug with no adverse effects}).$$

For sufficiently large  $R$  this is the same as minimizing the expected time until success.

Our rewritten drug prescription problem is a special case of a multi-armed bandit problem where the reward is deterministically 0 after some finite time, and so the expected total reward is maximized by pulling the arm with the greatest Gittins index:

$$G_i(x_i) = \sup_{\tau \geq 1} \frac{E[\sum_{t=1}^{\tau} r_i(x_i(t)) | x_i(0) = x_i]}{E[\tau | x_i(0) = x_i]}$$

where  $\tau$  is a past-measurable stopping time. In this case the only values of  $\tau$  where we can receive positive reward (and hence the only values of  $\tau$  we have to care about) are

$\{\tau_{i,1}, \tau_{i,2}, \dots\}$  where is the random stopping time  $\tau_{i,k}$  corresponding to pulling the arm until the first time before  $t_{i,k}$  there is a success, and then pulling once more. Hence

$$\begin{aligned} G_i(1) &= \sup_k \frac{R \cdot E[\text{success before time } t_{i,k}] - E[\tau_k]}{E[\tau_k]} = \sup_k \frac{P[\text{success before time } t_{i,k}]}{E[\tau_k]} \\ &= \sup_k \sigma_{i,k}. \end{aligned}$$