MIT MANAGEMENT SLOAN SCHOOL

Master of Business Analytics

15.095: Machine Learning under a Modern Optimization Lens

Reevaluating Former Standard Therapy on Interpretable Subgroups of Primary Biliary Cirrhosis Patients

Guillaume BONHEURE

Max PETRUZZI

December 9, 2022

Contents

1 Introduction **1**

2 Problem Statement **1**

3 Methodology **1**
3.1 Data Set 1
3.2 Treatment Effect 2
3.3 Formulation 3
3.3.1 Parameters 3
3.3.2 Model 3

4 Finding optimal D-penicillamine subgroups **4**
4.1 Establishing a baseline 4
4.2 Optimal subgroup versus control 4
4.3 Optimal subgroup bar one versus control 5
4.4 Who not to give D-penicillamine? 6

5 GLOBE Scores **7**

6 Can we beat current standard practice? **7**

7 Conclusion **8**

8 Team Contributions **9**

A Model Formulation **11**

1 Introduction

The first published randomized controlled trial in medicine occurred in 1946 and has been since regarded as the gold standard for evaluating the safety and efficacy of new therapeutics.[1] In the United States, a multidisciplinary team at the Food and Drug Administration (FDA) reviews the data from each stage of the trials and decides whether to approve the new drug for sale.[2] Once the drug enters the market, the FDA continues to monitor its safety within the population. Over time, the medical community decides whether to adopt this drug as the *standard therapy*, meaning the widely accepted treatment for a certain disease. This distinction is fluid, meaning that as the risks and side effects of the drug become better understood over time, and as new therapeutics become available, the standard therapy can change.

Primary biliary cirrhosis (PBC) is a rare autoimmune disease that causes progressive damage to the ducts that transport bile.[3] If left untreated by medication or liver transplant, this may lead to liver failure and death. A common feature of PBC is the accumulation of copper in the liver, and research in the 1970s suggested that there was a causal relationship between copper toxicity and the progression of PBC.[4] Consequently, the corticosteroid D-penicillamine became a promising treatment candidate due to its ability to decrease copper in the liver as well as suppress the autoimmune attack on the bile ducts.

2 Problem Statement

D-penicillamine became the standard therapy for several years until 1985 when it was replaced by a different drug due to its lesser propensity for side effects. However, one inherent limitation of the aforementioned process for drug approval and standard therapy determination is that it is generally conducted at a macro level. In other words, the risk of side effects for the general population is weighed against the average benefit of treatment among all patients taking the drug. Using a more personalized lens, certain subpopulations may exist where the treatment effect is substantially greater, outweighing the differences in risk. In fact, these subpopulations may be better candidates for D-penicillamine than the current standard therapy, despite the fact that D-penicillamine is only currently prescribed to about 2.8% of patients with PBC.[4]

The goal of this project is to identify these subpopulations of viable candidates for D-penicillamine, if they exist, as well as to provide a framework for analyzing other drugs which were replaced as the standard therapy for a particular disease.

3 Methodology

3.1 Data Set

The data set used contains the results of a randomized controlled trial for the drug D-penicillamine, conducted by the Mayo Clinic between 1974 and 1984. It consists of 19 variables across the 312 randomized PBC patients that met the eligibility criteria. Among these patients, 158 patients received D-penicillamine and 154 received a placebo, none of whom had knowledge of which treatment they received. The features for each patient include demographics, e.g., age and sex, as well as metrics of the patient's condition before treatment, e.g., serum bilirubin (mg/dl), urine copper (μg/day), and presence of hepatomegaly (binary). Because the laboratory tests and other metrics were only taken before treatment, they are not appropriate for evaluating the treatment effect of the drug.

3.2 Treatment Effect

In order to quantify the treatment effects, we examined the reported time that one of three possible outcomes occurred: death, liver transplant, or completion of observation. Because advanced PBC can lead to liver failure and death, the goal of treatment is to prolong the life of the individual and the time before requiring a liver transplant (if necessary). The best possible outcome is that, at the completion of observation, the patient is living and did not require a liver transplant. However, not all patients are observed for the same duration. So, the treatment effects of patients with this outcome, designated as `Alive`, are rewarded (given increased weighting) if they are observed for a longer duration because that means that they definitively survived for longer. Certainly, a patient who was observed as `Alive` for 100 days is more significant than one who was observed as being `Alive` for 50. Because that latter patient was not observed after day 50, we are blind to the status of that patient between days 50 and 100. Therefore, we simply weigh its treatment effect less heavily than the former patient.

An additional complication is how to quantify treatment effect with the same metric among each of the three outcomes since the reported times between different outcomes often overlap. For example, if one patient was `Alive` after only 50 days of observation, it is difficult to compare how positive that outcome is with another patient who was `Deceased` after a reported 100 days. As mentioned previously, we are blind to whether that former patient died or had a transplant between days 50 and 100.

In order to resolve this issue, a two-step approach was taken to develop the treatment effect score for each patient. First, for each outcome type, reported times outside the range of ± 1.282 standard deviations from the mean were truncated to remove outliers and improved the robustness of our results. This process was done such that the top and bottom 10% of the reported times were removed based on the area under the curve. This is in accordance with the recommendation by Bertsimas et al. (2019), which was to truncate the top and bottom 10% ATEs for robustness.[5]

The second step is to apply min-max scaling to all times in the truncated data set, using the min and max times for that particular outcome. This will produce times for each particular outcome such that they range from 0 to 1, where 0 represents the minimal time reported for that outcome and 1 represents the maximum. Now, we want to ensure that patients with the outcome `Transplant` have a smaller treatment effect score than all patients with outcome `Deceased` since a transplant is a better outcome than death and should always be associated with a higher score. In this way, we will add 1 to each time in the `Transplant` outcome so that it ranges from 1 to 2 and is always greater than a score from the `Deceased` outcome. Further, we will add 2 to each time in the `Alive` outcome so that it ranges from 2 to 3 and is always greater than a score from the `Transplant` or `Deceased` outcomes.

3.3 Formulation

The formulation utilized was developed based on the course 15.095 presentation for *Lecture 16: Identifying Exceptional Responders in Randomized Trials* which was adapted from the journal article, Bertsimas et al. (2019). An abbreviated explanation of this formulation will be highlighted in this paper for context.

3.3.1 Parameters

$\mathcal{T}_1$	Set of patients i in experimental group ($t = 1$)
$\mathcal{T}_0$	Set of patients i in control group ($t = 0$)
v_i	Treatment effect for patient i
z_i	Indicator variable, patient i in subgroup
$\bar{N}$	Maximum number of patients allowed to be in the subgroup
$\underline{N}$	Minimum number of patients required to be in the subgroup
ζ_{ij}	Indicator variable, $1 \iff z_i = 1$ and $\theta_j = 1$
θ_j	Indicator variable, if j between $\underline{N}$ and $\bar{N}$ is equal to the number of patients from treatment group t within the subgroup
γ_{sk}	Cut-point k along dimension s that designates where a potential subgroup might be delineated
K_s	Largest cut-point index for a particular dimension
S	Largest index for the dimensions in the data set
$L_{s,k}$	Lower bound indicator variable, $1 \iff$ cut k is the lower bound for feature s in the interpretable subgroup
$U_{s,k}$	Upper bound indicator variable, $1 \iff$ cut k is the upper bound for feature s in the interpretable subgroup
q_s	Indicator variable, $1 \iff$ feature s has a non-extremal lower and/or upper bound in the interpretable subgroup (i.e., feature s is part of the definition of the subgroup)
S_0	Maximal amount of features that define a subgroup

3.3.2 Model

For reference, the full model formulation can be found in Appendix A. What follows is a high-level description of how the model works.

First and foremost, the objective of the model is to find a subgroup of patients within a clinical trial dataset for which the difference in average treatment effect (ATE) between patients receiving the treatment and the control group is maximal. In other words, the model aims to find a subgroup of patients for which the drug has the greatest average benefit. To this end, we define a set of possible cuts for each feature under consideration, and a subgroup is then defined in terms of which cuts are selected per feature. Simply put, the model characterizes the optimal subgroup by returning a range of possible values for each feature (i.e., a unique lower and upper bound per feature). Any patient that lies within these ranges is part of the optimal subgroup, all other patients are not. In an effort to keep subgroups interpretable, the amount of features that define such a group can be limited by the parameter S_0, which we set to 4 for our project.

When performing our experiments, we also wanted to see if there was another, slightly less optimal subgroup present in the clinical trial data. This secondary exceptional responders group should contain minimal patients from the first subgroup in an effort to capture a population that is perhaps defined by other characteristics. To this end, we added one additional constraint, which ensures an upper bound on the number of patients that can overlap between both optimal responder groups. With Z_1 being the patients in the first subgroup and $\rho = 0.2$ as per Bertsimas et al. (2019), the extra constraint is:[5]

	Control	Treatment	Difference	P-Value
Overall mean ATE	1.283	1.351	5.3%	0.513

Table 1: Average treatment effect (ATE) of all control versus treatment group patients

$$\sum_{i \in Z_1} z_i \leq \rho \cdot \left(|Z_1| + \sum_{i \notin Z_1} z_i \right)$$

4 Finding optimal D-penicillamine subgroups

4.1 Establishing a baseline

Before we are able to say anything about a potentially better treatment effect compared to the control, we want to establish the overall effectiveness of D-penicillamine. To this end, we calculate the average treatment effect (ATE) across our entire clinical trial data set, for both the control group and the treatment group. Next, we determine whether there is a statistically significant difference between both ATE values according to an independent samples T-test.

The mean treatment effects of both overall control and treatment groups can be seen in Table 1. At first glance it might seem obvious that ATE for the treatment group would be higher than for the control; however, the difference is not large and is likely due to chance. The p-value for the comparison between treatment and control groups is 0.513, which is not enough to claim that this difference is significant (95% statistical significance would require a p-value of 0.05 or lower). In other words, there is no significant evidence that the treatment group is better off according to our definition of ATE compared to the control group.

This is less surprising in the context that this drug is no longer being used. As mentioned previously, part of the reason for this is the high propensity of D-penicillamine for side effects. An additional reason is that, despite being at one point the standard therapy, there were inconsistent results for this drug in clinical trials.[4] We propose that there are particular subgroups of patients for which the administration of D-penicillamine may produce more definitive treatment effects. In fact, these accentuated treatment effects may overshadow the risk of side effects, and make D-penicillamine a viable drug candidate for patients in those groups.

4.2 Optimal subgroup versus control

We now attempt to find a subgroup within the overall patients dataset for which there is a significant improvement in ATE for the treated population compared to the control. To this end, we use the following parameters in the algorithm:

- $K = 15$: We consider 15 cuts for each feature, which proved a reasonable balance between computation time and how fine-grained we can split the features. [1]
- $S_0 = 4$: As we want our subgroups to remain interpretable, we decided to limit the number of features that define a subgroup to at most 4. This proved a good trade-off between achieving optimality in the subgroup and still having easily understandable subgroups.

The optimal subgroup we identified is defined by the features and ranges shown in Table 2. The difference in ATE between control and treatment groups inside this optimal subgroup is

[1] Binary features always have only 3 cuts. This allows the algorithm to pick either $[0 - 0.5]$, $[0.5 - 1]$ or $[0 - 1]$, capturing only 0, 1 or both classes respectively.

Feature	Lower Bound	Upper Bound
`Age`	37.5	56.1
`Hepato`	1	1
`Copper`	4.0	296.0
`Protime`	9.0	11.3

Table 2: Features and ranges that define the subgroup with the largest difference in ATE between the control and treatment populations. `Hepato` is a binary indicator variable for presence of hepatomegaly, `Copper` is urine copper (μg/day), and `Protime` is prothrombin time in seconds

.

	Control	Treatment	Difference	P-Value
Overall mean ATE	1.283	1.351	5.3%	0.513
Opt. Subgroup 1 mean ATE	1.038	1.897	82.8%	0.0009

Table 3: Average treatment effect (ATE) of control versus treatment group patients in optimal subgroup compared to overall

shown in Table 3. The difference in ATE between treatment and control appears to be massive in our optimal subgroup, but once again we cannot say for certain whether this is a significant effect without first performing a t-test. The difference in ATE in our optimal subgroup has an associated p-value of 0.0009, which makes this result significant at 99.9%. Hence, we truly have found a subpopulation of patients for which D-penicillamine works really well.

In fact, the subgroup members receiving the placebo have an average treatment effect that is indistinguishable from the overall control group (p=0.186). However, the subgroup members receiving treatment have significantly better outcomes on average than the overall treatment group (p=0.01, so significant at 99%). This means that for these people the drug truly does work wonders as they are benefitting more from the drug than the general population.

4.3 Optimal subgroup bar one versus control

Finding an optimal subgroup in this data set raises the question if patients with the characteristics in Table 2 are the only ones to exhibit a significant difference in ATE between treatment and control. To this end, we added the additional constraint outlined at the end of Section 3. Re-running the model did provide us with a second, slightly less optimal but still significant subgroup.

The features and ranges that make up this second subgroup are shown in Table 4 and differ in all features but `protime`. The ATE for treatment and control groups in this second subgroup is shown in Table 5, with the results for subgroup 1 repeated for comparison. Interestingly, while the group appears to be defined by different characteristics, the difference in ATE appears to be comparable between optimal subgroup 1 and this new optimal subgroup 2. The difference in ATE between treatment and control for this subgroup has been determined to be significant as well, with a p-value of 0.001, which is only slightly less than in optimal subgroup 1.

Similarly as before, the ATE of the subgroup's treated patients is significantly higher than the overall treated ATE (p=0.0187), but the ATE of the subgroup's control patients is not significantly different from the overall control ATE (p=0.158).

Feature	Lower Bound	Upper Bound
`Alk.phos`	1258.5	11923.3
`Ast`	57.1	149.5
`Platelet`	133.6	491.4
`Protime`	9.6	12.5

Table 4: Features and ranges that define optimal subgroup 2. `Alk.phos` and `Ast` are measures of enzyme activity in units of U/liter, and `Platelet` is number of platelets per cubic mL / 1000

	Control	Treatment	Difference	P-Value
Overall mean ATE	1.283	1.351	5.3%	0.513
Opt. Subgroup 1 mean ATE	1.038	1.897	82.8 %	0.0009
Opt. Subgroup 2 mean ATE	1.004	1.802	79.5%	0.001

Table 5: Average treatment effect (ATE) of control versus treatment group patients in optimal subgroup and secondary optimal subgroup compared to overall

4.4 Who not to give D-penicillamine?

Finally, we wanted to observe if this formulation could be used to identify subpopulations for which D-penicillamine would have detrimental effects compared to the control. Essentially we want to consider the inverse case: can we find the subgroup of patients that have a significantly worst treatment response to this drug? It suffices to edit the formulation of Chapter 3 into a minimization problem instead of a maximization problem, as this will ensure that both ATE for the treatment group is as low as possible and ATE for control is as high as possible within the calculated subgroup.

We identified an "inversely optimal" subgroup defined by the features and ranges in Table 6 and with treatment effects detailed in Table 7. This effect is significant at the highest level we have observed so far, with a p-value of 0.0001. We now observe the opposite of before, as the control group of the subgroup now has an indistinguishable ATE compared to both the overall control and treatment groups ($p=0.271$ against overall control and $p=0.442$ against overall treatment), but the subgroup treatment group has lower ATE than both the overall treatment and control groups (significant, $p<0.0001$ against overall treatment and $p=0.0001$ against overall control).

Identifying such a subgroup can help providers determine which types of patients to absolutely not give a certain drug to, and provides a clear research objective to further understand why these characteristics effectively make the drug harmful to certain populations.

Feature	Lower Bound	Upper Bound
`Edema`	0.0	0.33
`Chol`	238.2	356.4
`Trig`	73.4	154.1
`Platelet`	62.0	348.3

Table 6: Features and ranges that define inverse subgroup. `Edema` is a categorical variable with three levels, `Chol` is serum cholesterol in units of mg/dL, and `Trig` are triglycerides in units of mg/dL.

	Control	Treatment	Difference	P-Value
Opt. Subgroup 2 mean ATE	1.510	0.584	-61.3%	0.0001

Table 7: Average treatment effect (ATE) of control versus treatment group patients in inverse subgroup

5 GLOBE Scores

One limitation of generating groups based primarily on clinical metrics and laboratory tests is that it is difficult for someone outside the medical profession to interpret the commonalities that define our subgroups. For example, it would be useful to understand whether our subgroups represent individuals that typically presented with more severe PBC at the beginning of the study.

For `Optimal Subgroup 1`, one of the features that bound this subgroup is a prothrombin time between 9.0 and 11.3 seconds. Prothrombin is a protein produced by the liver that is involved in blood clotting.[6] So, if blood clotting is observed to take a longer time, that means that the liver is not producing an adequate amount of prothrombin, and thus there is likely a higher degree of liver damage or cirrhosis. However, it would require a medical professional to understand how this metric, in combination with all of the other features that define our subgroup, can elucidate the average health of the patients in the group.

For this, we utilized a metric called a GLOBE score. A GLOBE score is produced by a statistical model that predicts the short-term survival probability of a patient with PBC.[7] Fortunately, the determination of the GLOBE score requires the exact metrics that were found for each patient in our data set. This allowed us to calculate the average GLOBE score for our overall data set and for each of the subgroups we defined. It should be reiterated that, because all of these metrics were taken at the beginning of the study, GLOBE scores represent the pre-treatment status of the patient and do not reflect the treatment effect of the drug.

As shown in Table 8, a higher GLOBE score indicates an increased risk of death and thus results in lower survivability at each time point. **Optimal Subgroup 2** is the subgroup that is most similar to the overall patient population in terms of GLOBE score. On the other hand, the features that define **Optimal Subgroup 1** mean that those patients were healthier on average at the beginning of the study. The **Inverse Subgroup** contains patients that were much less healthy than average at the beginning of the study.

This interpretation of the subgroups is useful because it allows people outside the medical profession to understand how our groups are defined, without needing to know what the laboratory tests indicate. However, it should be noted that we do not expect that one's status of being more or less healthy pre-treatment will impact the selection of our subgroups. This is because our subgroups were selected based solely on the difference in treatment effect between subgroup members from the experimental group versus those in the control group.

6 Can we beat current standard practice?

So far, it has been shown that D-penicillamine utilized by patients within our optimal subgroups results in a substantially greater treatment effect than that of the general population. We have also found a subgroup for which administering D-penicillamine actually has the greatest negative effect, where the experimental group is actually worse off than the control. However, in order to make these findings practical, they should be compared with the current standard therapy, Ursodeoxycholic Acid (UDCA), which is most commonly known by the brand name Actigall.

	All Patients	Opt. Subgroup 1	Opt. Subgroup 2	Inv. Subgroup
GLOBE score	8.8336	8.2451	9.1168	9.8958
% survival at:				
3 months	94	97	92	84
6 months	89	93	85	70
9 months	87	93	83	67
12 months	85	91	80	62
15 months	81	89	75	54
18 months	74	84	67	41
24 months	71	83	64	38

Table 8: GLOBE scores and predicted percent survival at different time points for all patients and each subgroup. Higher GLOBE scores indicate an increased risk of death and thus lower survivability.

	D-penicillamine			UDCA
	Overall	Opt. Subgroup 1	Opt. Subgroup 2	Overall
P-Value	0.513	0.0009	0.001	0.0490

Table 9: Features and ranges that define the subgroup with the largest difference in ATE between the control and treatment populations.

To make this comparison, a data set from a different clinical trial, one testing the effectiveness of UDCA, was utilized. This dataset contains data on 170 patients with 86 in the treatment group and 84 in control. Because the data set included the dates for each of the three outcomes we considered, we were able to calculate the treatment effect for this trial exactly as we did for that of D-penicillamine.

In order to quantify the effectiveness of UDCA, we conducted another independent samples T-test between the control and treatment groups in that study, which yielded a p-value of 0.0490. Because this value is less than the 0.05 cutoff, this means that there is a statistically significant difference between the experimental and control groups for the UDCA trial. This is in line with what we would expect for this drug because UDCA was shown as more effective through more consistent results in clinical trials.

However, in comparison with UDCA, D-penicillamine administered to our optimal subgroups results in a much greater difference in treatment effect. **Optimal Subgroup 1** had a p-value of 0.0009 and **Optimal Subgroup 2** had a p-value of 0.001, meaning that both subgroups had an even lower probability that their treatment effect differences for D-penicillamine were due to chance. In other words, there is evidence to suggest that administering D-penicillamine to a patient within one of our subgroups would result in a better improvement of outcomes than a patient from the general population would receive from UDCA.

7 Conclusion

We identified two subgroups within the patient population of the Mayo Clinic 1974-1984 D-penicillamine clinical trial for PBC in which the difference in ATE between the treated patients and control patients was maximal. This procedure could likely be repeated to find even more subgroups, but for the purposes of this project, this would not have improved our findings. We also identified a subgroup of patients for which the effects of the drug are particularly detrimental, informing caregivers which patients not to use the drug on and opening avenues for further

research. Identifying optimal subgroups for a drug that is no longer in widespread use can be impactful, as demonstrated by our comparison against UDCA, the current standard practice for PBC, where we showed that for these optimal subgroups of the D-penicillamine clinical trial the treatment effect difference was greater than that observed in the UDCA trial.

A major point of contention remains that these are not comparable populations, as we only compare a highly specific group of patients against the entire set of patients in the UDCA trial. This was as far as we could go in our project due to a lack of further data on UDCA, but it is reasonable to state that our results would be even more significant if we considered differences in treatment effectiveness between D-penicillamine and UDCA for similar patient subgroups. We would be able to do this given UDCA trial data that contains the same measurements as with the D-penicillamine trial, but unfortunately this data was not publicly available to the best of our knowledge.

8 Team Contributions

The team that conducted this report consisted of complementary backgrounds that allowed for great synergy. Guillaume has a strong background in computer science and mathematics. So, his contributions included writing the majority of Julia code and performing the validation with the different parameters. Guillaume was also responsible for reviewing literature in the field of optimization, primarily Bertisimas et al., in order to follow established guidelines that would maximize the performance of our model. Guillaume made innovative contributions to the project, including defining a "2nd best subgroup" and an "inverse subgroup." Max has a background in biomedical sciences and has conducted biomedical research. Max was able to contribute domain knowledge by helping to establish a metric for the treatment effect and incorporating statistical analyses. Max also conducted a literature review in the biomedical field to learn about the disease and the nature of both drugs we considered, D-penicillamine and UDCA, and incorporated an interpretation of GLOBE scores. Both team members worked collaboratively in defining the formulation and building the deliverables. Despite each team member having specialized knowledge, both team members had a strong understanding of the entire scope of the project across both disciplines, which was a product of effective communication.

References

[1] Bhatt A. "Evolution of Clinical Research: A History Before and Beyond James Lind". In: *Perspect Clin Res.* 1.1 (2010), pp. 6–10.

[2] *Development & Approval Process | Drugs.* 2022. URL: https://www.fda.gov/drugs/development-approval-process-drugs#:~:text=A%5C%20team%5C%20of%5C%20CDER%5C%20physicians,drug%5C%20is%5C%20approved%5C%20for%5C%20sale..

[3] *Primary biliary cholangitis.* 2022. URL: https://www.mayoclinic.org/diseases-conditions/primary-biliary-cholangitis/symptoms-causes/syc-20376874.

[4] Gluud C. Gong Y Frederiksen SL. "D-penicillamine for Primary Biliary Cirrhosis". In: *Cochrane Database Syst Rev.* 4 (2004).

[5] Alexander M. Weinstein Dimitris Bertsimas Nikita Korolko. "Identifying Exceptional Responders in Randomized Trials: An Optimization Approach". In: *INFORMS Journal on Optimization* 1.3 (2019), pp. 187–199.

[6] *Prothrombin time.* 2022. URL: https://www.hepatitis.va.gov/hcv/patient/diagnosis/labtests-prothrombin-time.asp#:~:text=When%5C%20the%5C%20PT%5C%20is%5C%20high,serious%5C%20liver%5C%20damage%5C%20or%5C%20cirrhosis..

[7] Wiesner RH Kim WR Therneau TM. "A revised natural history model for primary sclerosing cholangitis". In: *Mayo Clin Proc* 75.7 (2000), pp. 688–694.

A Model Formulation

$$
\begin{aligned}
\max_{\mathbf{z},\mathbf{q},\mathbf{L},\mathbf{U},\zeta,,} \; & \sum_{i\in\mathscr{T}_1}\sum_{j=\underline{N}}^{\bar{N}}\frac{1}{j}v_i\zeta_{ij} - \sum_{i\in\mathscr{T}_0}\sum_{j=\underline{N}}^{\bar{N}}\frac{1}{j}v_i\zeta_{ij} \\
\text{s.t. } & z_i + \sum_{s=1}^{S}\left[\sum_{k:\gamma_{sk}>x_{is}} L_{sk} + \sum_{k:\gamma_{sk}<x_{is}} U_{sk}\right] \geq 1, && \forall i = 1,\ldots,n, \\
& z_i + L_{sk} \leq 1, && \forall s = 1,\ldots,S, k = 1,\ldots,K_s,\, i : x_{is} < \gamma_{sk}, \\
& z_i + U_{sk} \leq 1, && \forall s = 1,\ldots,S, k = 1,\ldots,K_s,\, i : x_{is} > \gamma_{sk}, \\
& \sum_{k=1}^{K_s} L_{sk} = 1, && \forall s = 1,\ldots,s, \\
& \sum_{k=1}^{K_s} U_{sk} = 1, && \forall s = 1,\ldots,S, \\
& q_s + L_{s1} \geq 1, && \forall s = 1,\ldots,S, \\
& q_s + U_{sK_s} \geq 1, && \forall s = 1,\ldots,S, \\
& q_s + L_{s1} + U_{sK_s} \leq 2, && \forall s = 1,\ldots,S, \\
& \sum_{s=1}^{S} q_s \leq S_0, \\
& \underline{N} \leq \sum_{i\in\mathscr{T}_t} z_i \leq \bar{N}, && \forall t = 0,1, \\
& \zeta_{ij} \leq \theta_j^{(T_i)}, && \forall i = 1,\ldots,n, j = \underline{N},\ldots,\bar{N}, \\
& \zeta_{ij} \leq z_i, && \forall i = 1,\ldots,n, j = \underline{N},\ldots,\bar{N}, \\
& \zeta_{ij} \geq \theta_j^{(T_i)} + z_i - 1, && \forall i = 1,\ldots,n, j = \underline{N},\ldots,\bar{N}, \\
& \sum_{i\in\mathscr{T}_t}\sum_{j=\underline{N}}^{\bar{N}}\frac{1}{j}\zeta_{ij} = 1, && \forall t = 0,1, \\
& \sum_{j=\underline{N}}^{\bar{N}} \theta_j^{(t)} = 1, && \forall t = 0,1, \\
& 0 \leq \zeta_{ij} \leq 1, && \forall i = 1,\ldots,n, j = \underline{N},\ldots,\bar{N}, \\
& \mathbf{z},\mathbf{q},\mathbf{L},\mathbf{U},\boldsymbol{\theta} \in \{0,1\}.
\end{aligned}
$$