Efficient Discovery of Cost-effective Policies in Sequential, Medical Decision-Making Problems

Narges Mohammadi
Imperial College Business School, Imperial College London, London
n.mohammadi19@imperial.ac.uk

M. Reza Skandari
Imperial College Business School, Imperial College London, London
r.skandari@imperial.ac.uk

Anand Shah
The Royal Brompton Hospital, London
Imperial College London, London
s.anand@imperial.ac.uk

Cost-effectiveness analysis (CEA) is extensively employed by healthcare policymakers to guide funding decisions and inform optimal design of medical interventions. In the CEA literature, willingness to pay (WTP) serves as a common metric for converting health benefits into monetary value and defining the net monetary benefit of an intervention. However, there is no universally accepted value for WTP. To address this issue, we propose presenting policymakers with a comprehensive menu of strategies that are proven cost-effective across a reasonable range of WTP values. In our approach, we consider a setting where the medical decision-making process can be formulated as a parametric linear programming model. We have developed a novel algorithm aimed at efficiently constructing the menu of cost-effective policies. Our algorithm is particularly suited for Constrained Markov Decision Process (CMDP) and Constrained Partially Observable Markov Decision Process (CPOMDP) models, which are commonly utilized modeling frameworks for addressing sequential medical decision-making problems. We have applied our modeling framework to design hearing loss screening strategies for cystic fibrosis patients. Informed by a validated, data-driven model, we have developed several heuristic and approximate policies, allowing policymakers to balance between performance and ease of implementation.

Key words: Constrained, non-linear partially observable Markov decision process; cost-effectiveness analysis; grid-based approximation; sequential medical decision making

1. Introduction

A primary responsibility of public policymakers is to enhance public life by efficiently allocating a limited budget to programs (i.e., sets of interventions) that improve people's lives. A systematic, data-driven approach to budget allocation involves developing mathematical models to optimize public benefits, such as public health, within budget constraints. These models facilitate the identification and prioritization of cost-effective interventions for funding. However, as discussed in Glassman et al. (2017), such approach involves solving a large-scale, multi-period model and requires

detailed data on intervention costs, benefits, and resource needs. To address model complexity, it may be necessary to simplify intervention details. However, this approach could potentially introduce notable inaccuracies concerning intervention design and resource needs. Another challenge is budget uncertainty, as funds are allocated and reallocated dynamically among interventions and may be borrowed from future allocations. These complexities are not easily captured by a centralized planning mathematical model (Drummond et al. 2015).

An alternative and more common approach is a decentralized method, where the inclusion of each intervention into a program is determined by an effectiveness threshold. This threshold serves as a monetary exchange rate for an intervention's benefit, determining its net monetary benefit (NMB) by subtracting its cost from its monetary value. Within this context, interventions with a positive NMB are considered cost-effective and included in the program. The composition of the program depends on the chosen effectiveness threshold value; as the threshold increases, a broader set of interventions is included, leading to an increase in the program's overall cost. Thus, this threshold serves as a straightforward mechanism for efficiently managing the constrained budget (Glassman et al. 2017).

Another closely related topic in cost-effectiveness analysis involves determining the optimal configuration of a single intervention, such as the screening schedule for a specific chronic disease. In this context, the cost-effective policy among all variations can be identified by selecting the intervention configuration that yields the highest NMB. Alternatively, Incremental Cost-Effectiveness Ratio (ICER), another metric for assessing cost-effectiveness, can be employed. ICER quantifies the additional cost required to achieve a one-unit improvement in effectiveness compared to a comparator policy, usually the standard of care. The best configuration among all intervention variations is the one with the largest ICER below the target effectiveness threshold (Drummond et al. 2015).

In the healthcare domain, the Quality-Adjusted Life Expectancy (QALE) is a widely adopted metric for quantifying the health benefits of interventions. It offers a comprehensive valuation by considering both the quantity and quality of life. The willingness to pay (WTP), which represents the maximum cost society is willing to pay for one additional unit of QALE gain, is commonly used as the effectiveness threshold in the healthcare Cost-Effectiveness Analysis (CEA) (Macones et al. 1999).

Determining WTP poses significant ethical challenges, particularly when the primary aim is to enhance a patient's well-being (Neumann et al. 2010). Policymakers, recognizing these complexities, often consider a range of values for WTP rather than relying on a single estimate. For instance, the United Kingdom's National Institute for Clinical Excellence (NICE), tasked with offering national guidance to improve health and social care, takes a nuanced approach by incorporating a spectrum

of WTP values and considering the unique circumstances of individual cases (Rawlins and Culyer 2004). Additionally, recognizing the evolving nature of efficiency, WTP values require periodic reassessment to adapt to changes over time (McCabe et al. 2008).

Given the ambiguity of WTP, scenario analysis is often conducted to assess its impact on the optimal policy. As discussed later, conducting scenario analysis without a careful selection of candidate WTPs is not computationally efficient and may overlook cost-effective policies. We propose an alternative approach: presenting policymakers with a comprehensive range of cost-effective policies. Providing multiple options enhances the flexibility and adaptability of healthcare decision-making, accommodating individual circumstances, budget constraints, and resource limitations.

To support this approach, we introduce an efficient algorithm that generate cost-effective policies, policies optimizing the NMB for an arbitrary WTP, along with their cost and QALE performances. The latter is referred to as the cost-effectiveness frontier in the CEA literature (Drummond et al. 2015, Glassman et al. 2017). The developed algorithm achieves this objective through iterative solutions to the ICER minimization problem, updating the comparator policy in each iteration. As the process unfolds, the comprehensive cost-effectiveness frontier and the corresponding policies emerges.

The outlined framework has broad applications in medical decision-making, covering chronic disease management across screening, treatment planning, monitoring, and surveillance pillars. In this paper, we demonstrate the practicality of our approach by focusing on developing cost-effective policies for screening hearing loss in a population affected by cystic fibrosis (CF) disease. To formulate the problem, we employ a Constrained Partially Observable Markov Decision Process (CPOMDP) approach. This choice is motivated the framework's capability to accommodate screening assessments with limited accuracy.

2. Literature Review

The NMB maximization problem can be framed as a bi-objective problem, aiming to optimize a weighted sum of conflicting objectives, namely QALE and cost. The assigned weights represent the trade-off between costs and QALE outcomes, enabling decision-makers to balance these objectives based on their priorities. This paper focuses on constructing a Pareto frontier using the weighted sum approach for bi-objective MDP and POMDP problems. This approach filters out policies exhibiting strong, weak, or extended dominance, aligning with recommendations in the cost-effectiveness analysis literature (refer to Section 8 for further discussions).

Roijers et al. (2013) presents a comprehensive survey of algorithms for solving multi-objective Markov decision processes, considering various criteria, including the weighted sum. While the conversion of POMDPs to MDPs, a common approach for solving POMDPs, enables the application

of suggested algorithms, it should be noted that this conversion leads to an uncountable state space. Consequently, none of the surveyed methods is readily adaptable to yield exact solutions for multi-objective POMDPs with extensive or infinite horizon lengths. Some of the examined algorithms handle the uncertainty of objective weights by converting an MDP to a POMDP, which further restricts their applicability to problems originally formulated as a POMDP.

Parametric optimization provides an effective framework to model the uncertainty associated with objective weights. As MDPs can be solved using linear programming, the utilization of parametric linear programming offers an alternative solution strategy. Existing literature on parametric optimization predominantly concentrates on identifying the parameter range within which a specific policy remains optimal. However, these studies often fall short of developing algorithms capable of navigating the entire parameter space. In contrast, some works in this field, such as Holder (2010), primarily aim at determining the optimal objective function, such as the NMB in our context, for different weight values. This focus tends to exclude the generation of the Pareto frontier or the corresponding policies. Moreover, the suggested algorithm faces similar computational challenges arising from the conversion of POMDPs to MDPs.

Suen and Goldhaber-Fiebert (2016) present an algorithm for constructing the cost-effectiveness frontier, but it comes with several significant limitations. Firstly, the algorithm is restricted to scenarios with a finite, predetermined set of policies. Secondly, it assumes that these policies have already been evaluated, and their performance metrics (cost and QALE) are available as model inputs. Thirdly, while they explore the relationship between ICER and NMB objectives, their discussion is confined to cases with pre-defined set of policies and does not readily generalize to problems formulated as mathematical programming models. Finally, they do not provide a proof of computational efficiency for their algorithm.

Scenario analysis, which involves considering a pre-determined set of WTP values, is a prevalent approach in the literature for investigating the impact of WTP variations on the optimal policy. For instance, Mason et al. (2014), Chen et al. (2018), and Helmeczi et al. (2023b) have utilized this approach to ascertain cost-effective strategies for various clinical problems modeled as POMDP or CPOMDP. Although scenario analysis is straightforward and numerically attractive, it has several limitations. Firstly, since policies may remain optimal over a range of WTP values, some of the scenario analyses become redundant, leading to computational inefficiency. Secondly, time constraints may limit the number of WTP values considered for evaluation, potentially resulting in a failure to capture a broader range of cost-effective policies. Lastly, this approach fails to provide policy recommendation for WTPs outside the considered set.

Given the inaccuracies in observing patient state within our decision model, the NMB optimization problem can be formulated as a POMDP. Similarly, the ICER optimization model takes the

form of a CPOMDP, as the ICER problem requires policies to surpass the comparator policy in the QALE metric. The study by Helmeczi et al. (2023a) offers a thorough review of solution methods for POMDPs and CPOMDPs. This work underscores that while exact solutions do exist for POMDP problems, they are typically only practicable for small-scale scenarios. Lusena et al. (2001) discuss the inherent complexity of POMDPs and caution that one must 'choose between performance guarantees and efficient computation'. In practice, approximation methods are commonly employed to tackle POMDP and CPOMDP problems.

Grid-based approximation, a prevalent method used to solve C/POMDPs, involves approximating the belief space using a finite grid. Kavaklioglu and Cevik (2022) substantiate the efficacy of grid-based approximation through numerical experiments, highlighting its high-quality results. The grid-based approximation not only serves as a solution method for solving POMDPs but, as established by Lovejoy (1991), it yields a lower bound for the optimal value in minimization problems. In Poupart et al. (2015), the authors extend this result to a subset of CPOMDPs where both the objective function and constraints are linear in the vector of expected total rewards. Considering the potential value of this result in assessing the optimality gap of the approximation method, we broaden its scope to encompass non-linear CPOMDPs.

POMDP models have been extensively used in various medical decision-making problems, including the optimization of prostate cancer screening strategies (Zhang et al. 2012), breast cancer screening approaches (Ayer et al. 2016), sepsis detection systems (Liu et al. 2022), and the individualization of patient monitoring in ICUs (Piri et al. 2022). For an in-depth exploration of additional applications of POMDPs in medical decision-making, we recommend that readers refer to Li et al. (2023).

CPOMDPs have also found applications in healthcare, particularly in scenarios where available resources, such as budget, are constrained. For example, Gan et al. (2019) explored the optimization of interventions for addressing opioid use disorder within the constraints of a limited budget. Similarly, Cevik et al. (2018) optimized the breast cancer screening problem, imposing an upper limit on the number of screenings. To generate deterministic solutions for the CPOMDP, mixed-integer linear programming is commonly employed. The majority of the reviewed papers utilized approximate approaches to solve both POMDP and CPOMDP models, except in instances where the models were of a sufficiently small scale, enabling the derivation of exact solutions.

3. Contributions

We contribute to the literature on cost-effectiveness analysis by developing an efficient algorithm that discovers the cost-effectiveness frontier for sequential, medical decision-making problems. This

is a step forward from cost-effectiveness analysis using a single point-estimate for WTP or performing scenario analysis. The algorithm, tailored specifically for bi-objective linear programming models, presents a novel method to generate exact Pareto frontiers for finite and infinite horizon C/MDPs, as well as C/POMDPs with relatively short horizons. However, the algorithm yields an approximate frontier when the horizon is large. We also establish the relationship between the ICER and NMB optimization problems and characterize the comprehensive sets of optimal policies for all WTP values. To the best of our knowledge, our work is the first to incorporate ICER as the objective function in a mathematical programming model for the identification of cost-effective policies.

The existing literature on generating the cost-effectiveness frontier is limited to finding the frontier for a given finite set of evaluated policies, e.g., see Suen and Goldhaber-Fiebert (2016). In contrast, our proposed framework defines the domain of feasible policies using a mathematical programming model. In this model, the policymaker does not need to generate and evaluate policies in advance, nor is restricted to a finite set of policies. Instead, they can determine a set of operational, budgetary, or other types of constraints that define feasibility of policies.

Moreover, we contribute to the literature of CPOMDPs and grid-based approximation methods by generalizing a well-known bounding result (Lovejoy 1991, Poupart et al. 2015) to non-linear CPOMDPs and provide an alternative proof strategy. This result asserts that grid-based approximations with linear interpolation produce lower bounds for the objective value in CPOMDP minimization problems. Moreover, the generated policies can be utilized to derive upper bounds on the optimal value. We can utilize these bounds to evaluate the optimality gap of the approximation method. Moreover, we extend the findings in Wagner (1960) by proving that the extreme points of the set of policy occupancy measures in general MDPs/POMDPs with finite or infinite horizons correspond to pure (deterministic) policies. Building upon this result, we establish that certain non-linear CMDP/CPOMDPs, including ICER minimization, admit deterministic policies. From a computational standpoint, this result obviate the need for employing integer programming techniques to enforce deterministic policies, as demonstrated in Cevik et al. (2018). Furthermore, this result guarantees the efficiency of deterministic policies, which are particularly well-suited for implementation in medical practice. Furthermore, our contribution extends to the CF clinical literature by developing cost-effective policies for hearing loss screening. We also devise several easy-to-implement, approximate policies as well as other heuristics and compare their performance against that of the optimal policy. A policymaker can use the policy evaluation results to decide the trade-off between performance and ease of implementation.

Finally, we contribute to the literature on screening problems within the context of medical decision-making. Our particular emphasis is on determining whether screenings should be conducted only when a specific adverse event occurs for the patient, guiding screening decisions during

a (random) subset of cycles. This targeted and event-driven approach sets our model apart from the more generic screening scenarios examined in prior studies.

4. Solution to WTP Ambiguity

In this section, we propose an algorithm to generate the menu of cost-effective policies and the cost-effectiveness frontier. We start by defining and formalizing these concepts using mathematical notation. Let a policy be denoted by a real-valued vector $x \in X$, where X, the set of all acceptable policies, is a non-empty polytope. We assume that the cost and QALE associated with policy x is linear in x. Let $\phi : X \to \Re^2$ be the linear policy-to-performance map defined as follows: $(q, c) = \phi(x)$. In this equation, $q = ax$ and $c = bx$, where c and q represent the cost and QALE of the policy denoted by x, and a and b are coefficients for QALE and cost, respectively. This setup encompasses many formulations for which there is an exact or approximate linear programming model. When policies are represented by their state/action occupancy measures, C/MDPs and C/POMDPs, the standard frameworks for solving sequential decision-making problems, conform to this framework (Ross 2014).

A cost-effective policy is defined as a policy that maximizes NMB for a specified WTP. In other words, all solutions to the following problem represent cost-effective policies.

$$\begin{aligned} v(\lambda) = \max_{x} \quad & \text{nmb}(x;\lambda) \\ \text{s.t.} \quad & x \in X. \end{aligned} \tag{1}$$

In this model, $\text{nmb}(x;\lambda)$ represents the NMB function, which satisfies $\text{nmb}(x;\lambda) = (\lambda a - b)x$, where $\lambda \geq 0$ is the WTP, and $v(\lambda)$ is the optimal NMB value for WTP λ.

With variations in the value of WTP, distinct cost-effective interventions are generated. Consequently, our objective is to obtain the set of all cost-effective interventions by solving Problem 1 across a reasonable range of WTP values. In essence, our aim is to identify the following set.

$$V = \{x \in X : \exists \lambda \geq 0 \text{ s.t. } \text{nmb}(x;\lambda) = v(\lambda)\}.$$

Set V encompasses all cost-effective policies, which are maximizers to Problem 1 for all values of $\lambda \geq 0$.

The cost-effectiveness frontier is defined as the set of QALE and cost outcomes associated with cost-effective policies (Glassman et al. 2017). Mathematically, the cost-effectiveness frontier, denoted as L, satisfies $L = \phi(V)$, meaning it contains the QALE and cost of policies within the set V.

Alternatively, we can construct the cost-effectiveness frontier by solving the following problem.

$$\begin{aligned} c^*(z) = \min_{c,q} \quad & c \\ \text{s.t.} \quad & q \geq z, \\ & (q, c) \in \Psi, \end{aligned} \tag{2}$$

where $\Psi = \phi(X)$ represents the set of performance metrics of all feasible policies.

In this problem, we leverage the concept that a cost-effective policy is a feasible policy minimizing the cost while meeting a minimum requirement on QALE (parameter z in this problem). To ensure feasibility, z should belong to interval $[q_{\min}, q_{\max}]$, where $q_{\max}$ is the highest QALE achievable by a feasible policy, and $q_{\min}$ is the highest QALE of all policies that have the lowest achievable cost. To generate the frontier, one needs to solve this problem over the range $[q_{\min}, q_{\max}]$.

Through the following results, we demonstrate the relationship between the two concepts of cost-effectiveness frontier. Specifically, we establish that L, the cost-effectiveness frontier, is equal to the graph of function $c^*(z)$. Since $c^*(z)$ is continuous, convex, piece-wise affine, and increasing in z, the set L can be constructed by identifying the breakpoints of the function c^*.

To accomplish this, let vectors (q_i, c_i), with $i = 1, \ldots, m$, correspond to the breakpoints of the piece-wise affine function $c^*(z)$. Additionally, let $\lambda_i := (c_{i+1} - c_i)/(q_{i+1} - q_i)$ represent the slopes of its individual segments. We show that the set of solutions to Problem 1 for a specific λ, denoted by $V(\lambda)$, can be characterized by comparing λ with the slopes λ_i's as follows.

LEMMA 1. We have the following.

(a) If $\lambda_i < \lambda < \lambda_{i+1}$ for some i, we have $V(\lambda) = \phi^{-1}(q_{i+1}, c_{i+1})$,

(b) If $\lambda < \lambda_1$, we have $V(\lambda) = \phi^{-1}(q_1, c_1)$,

(c) If $\lambda > \lambda_{m-1}$, we have $V(\lambda) = \phi^{-1}(q_m, c_m)$,

(d) If $\lambda = \lambda_i$ for some i, we have $V(\lambda) = \phi^{-1}(L_i)$,

where L_i is the convex hull of points (q_i, c_i) and (q_{i+1}, c_{i+1}), and $\phi^{-1}(Z)$ denotes the set of all feasible policies with QALE q and cost c in any arbitrary set Z, i.e., $(q, c) \in Z \subset \Psi$.

The lemma asserts that we can use the slopes of the pieces of c^* to partition the range for WTP. Depending on which set in the partition contains the target WTP, policies with performance lying on one end point of the piece (cases (a)-(c)) or the whole piece (case (d)) optimize the NMB. We leverage this result in Algorithm 1 to iteratively generate the breakpoints of the cost-effectiveness frontier and the desired menu of policies. Since there might be multiple optimal solutions to Problem 1 for a specific value of λ, the primary objective of this algorithm is to identify a subset of V that contains at least one solution for any desired $\lambda \geq 0$.

The algorithm's logic relies on a specific relationship between the ICER and NMB optimization problems, as outlined in the following proposition.

PROPOSITION 1. Assume (u, v) lies on the frontier, i.e., it satisfies $q_{i-1} \leq u < q_i$ for some i, and $v = c^*(u)$. Consider the following ICER minimization problem.

$$\begin{aligned} r^*(u, v) = \inf_x \quad & r(x; u, v) \\ \text{s.t.} \quad & ax > u, \\ & x \in X, \end{aligned} \tag{3}$$

where $r(x;u,v)=\frac{bx-v}{ax-u}$. Then, $\phi^{-1}(q_i,c_i)$ optimizes Problem 3 and $r^*(u,v)=\lambda_i$.

The proposition asserts that given the performance (cost and QALE) of a non-dominated policy, (a point on the cost-effectiveness frontier), the minimum ICER is attained by the policies associated with the next breakpoint on the cost-effectiveness frontier. Furthermore, the optimal ICER is equal to the the slope of the segment containing the point.

Data:

Vectors $a,b\in\Re^n$ and polytope $X\subset\Re^n$.

Result:

Finite set of m points $x_i\in X$, $d_i\in\Re^2$, and $m-1$ slopes $\lambda_j\in\Re$.

Initialize:

Let $i\leftarrow 1$ and find u,v and $q_{\max}$ as follows:

$$v=\min bx \quad \text{subject to} \quad x\in X; \tag{4}$$

$$u=\max ax \quad \text{subject to} \quad x\in X;\ \ bx\le v. \tag{5}$$

$$q_{\max}=\max ax \quad \text{subject to} \quad x\in X;$$

Let $d_1\leftarrow(u,v)$, and x_1 be a solution to Problem 5.

while $u<q_{\max}$ **do**

Solve

$$x^*\in\arg\min_x \quad r(x;u,v)$$
$$\text{s.t.,} \quad ax>u$$
$$x\in X.$$

Let $\lambda_i^*=r(x^*;u,v)$ and solve:

$$x_{i+1}\in\arg\max_x \quad ax$$
$$\text{s.t.,} \quad (\lambda_i^* a-b)(x-x^*)=0 \tag{6}$$
$$x\in X.$$

Let $u\leftarrow ax_{i+1}$, $v\leftarrow bx_{i+1}$, and $d_{i+1}\leftarrow(u,v)$.

$i\leftarrow i+1$

end

return *x_i and d_i for $j\le i$, and λ_j for $j<i$.*

Algorithm 1: Iterative ICER Algorithm

In each iteration of Algorithm 1, we first solve the ICER problem and determine λ_i^*, the slope of the next piece of the frontier, and identify a new policy that lies on that piece, x^*. Note that

x^* does not necessarily correspond to a breakpoint on the frontier; it may strictly lie on the line connecting the two adjacent breakpoints. This point is related to Lemma 1 (d). As a result, we solve Problem 6 to find the next breakpoint of the cost-effectiveness frontier, which is the right-most point on that segment of the frontier. The latter problem finds a policy with the highest QALE among all optimal policies that optimize the NMB for $\lambda = \lambda_i^*$, i.e., the policies that have the same NMB as x^*'s. We subsequently use the newly discovered breakpoint as the new comparator policy in the next iteration of the algorithm. The algorithm terminates once we have reached the highest possible QALE. In the following proposition, we establish the efficiency and efficacy of Algorithm 1.

PROPOSITION 2. Algorithm 1 terminates in exactly $m-1$ iterations. Furthermore,

- d_i's are breakpoints of c^*.
- $m < \infty$, meaning that the algorithm terminates in finite number of iterations.
- Let W be a collection of x_i's. Then, W is the smallest set containing at least one solution to Problem 1 for any given $\lambda \geq 0$.

This result implies that in each iteration of the algorithm, a non-redundant policy is discovered, proving the algorithm's computational efficiency. It also demonstrates that the algorithm generates a set with smallest cardinality among all sets containing at least one solution to Problem 1 for any given $\lambda \geq 0$, proving its efficiency and efficacy.

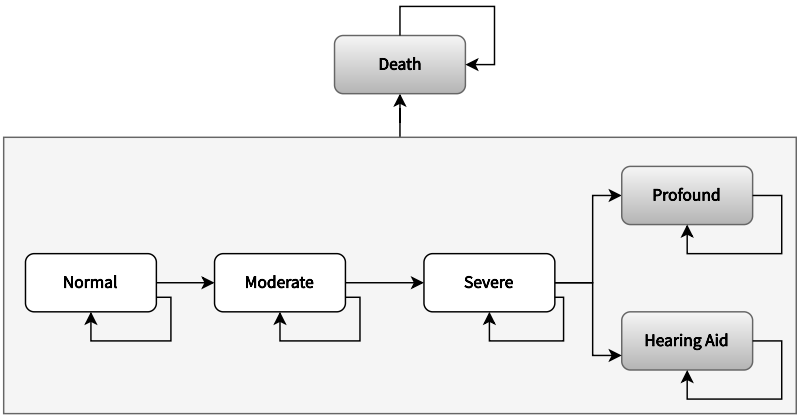
5. Clinical Case

5.1. Case Introduction

In 2018, there were over 100,000 patients worldwide diagnosed with cystic fibrosis (CF) disease (Choi and Pietrangelo 2019). The average life expectancy of patients with CF in developed countries is between 42 and 50 years (Nazareth and Walshaw 2013, Ong and Ramsey 2015). Pulmonary disease is the main cause of morbidity and mortality in CF patients (Ratjen et al. 2015). The disease trajectory of CF patients is punctuated by acute episodes of pulmonary function deterioration, referred to as pulmonary exacerbations (PEx) (Sanders et al. 2011). PEx's are predominantly caused by infection and colonization of bacteria, viruses, fungi, and yeasts. Lung infections and PEx's, regardless of the infection source, are treated by antibiotics (Westerman et al. 2004).

Aminoglycosides (AGs) are commonly used as first-line agents for treating severe infections. However, their intravenous administration, in particular, is associated with ototoxicity, leading to potential hearing loss in patients (Prayle and Smyth 2010). Due to their effectiveness, intravenous (IV) AGs are routinely used in CF to treat severe lung infections, despite their ototoxicity (Gleser and Zettner 2018). Repeated exposure to IV AG can lead to significant progression of hearing

Figure 1: State transition diagram of the Markov process.



Note. Gray/white boxes show observable/unobservable states. Hearing status may only progress with PEx incidence.

impairment, necessitating the use of hearing assistance devices. Several studies indicated a 40%–56% prevalence of hearing loss in adult CF patients (Vijayasingam et al. 2020) and $< 29\%$ in pediatric CF patients (Farzal et al. 2016). Ototoxic treatments such as AGs are suspected to be the main cause. There is a general agreement in the clinical community that early detection of hearing loss can potentially improve a patient's health outcomes (Garinis et al. 2017). For example, hearing loss impedes the natural speech and language development of children. We can alleviate the negative consequences of hearing loss by early detection through frequent screening.

There are various methods to assess a patient's hearing status, varying in precision, cost, capacity, ease of use, patient adherence and safety. Although the American Speech-Language-Hearing Association and the American Academy of Audiology recommended hearing screening for ototoxic medications, screening practices vary between clinics (American Speech-Language-Hearing Association 1994, Durrant et al. 2009, Huang et al. 2021). To the best of our knowledge, there is no established local or national guideline with evidence-based recommendations on the modality and/or frequency of hearing loss screening in CF populations.

5.2. Model Description

We categorize hearing status into several distinct states: normal hearing, as well as moderate, severe, and profound hearing loss. A hearing aid is prescribed when the hearing status is identified as severe hearing loss and subsequently confirmed. However, if the progression leads to profound hearing loss, a cochlear implant, which is a lifelong and permanent solution, becomes necessary. The underlying Markov process for the state transitions is depicted in Figure 1.

In the effort to detect hearing loss and provide timely intervention through hearing aids, two primary assessment methods are available:

- **Formal Audiometry**: This method serves as the gold standard for detecting hearing loss. It is typically performed in an audiometric booth by a trained audiologist. These booths, commonly located within hospital audiology departments, are designed to minimize background noise interference, ensuring precise measurements. While formal audiometry is highly accurate and reliable, it is also associated with substantial costs and limited availability due to capacity constraints.
- **Mobile Audiometry:** In contrast, mobile audiometry offers a point-of-care testing approach that can be administered by non-specialists. This method only requires basic equipment, including a tablet or mobile phone, specialized headphones, and appropriate software licensing. As a result, it is highly accessible and incurs lower costs compared to formal audiometry. However, it is essential to acknowledge that mobile audiometry may be less precise and occasionally yield inaccurate hearing status measurements.

A cost-effective screening strategy aims to strike a balance between the costs and benefits of screening. While formal audiometry excels in accuracy, the affordability and accessibility of mobile audiometry can play a valuable role in expanding access to hearing assessments, particularly in resource-constrained settings.

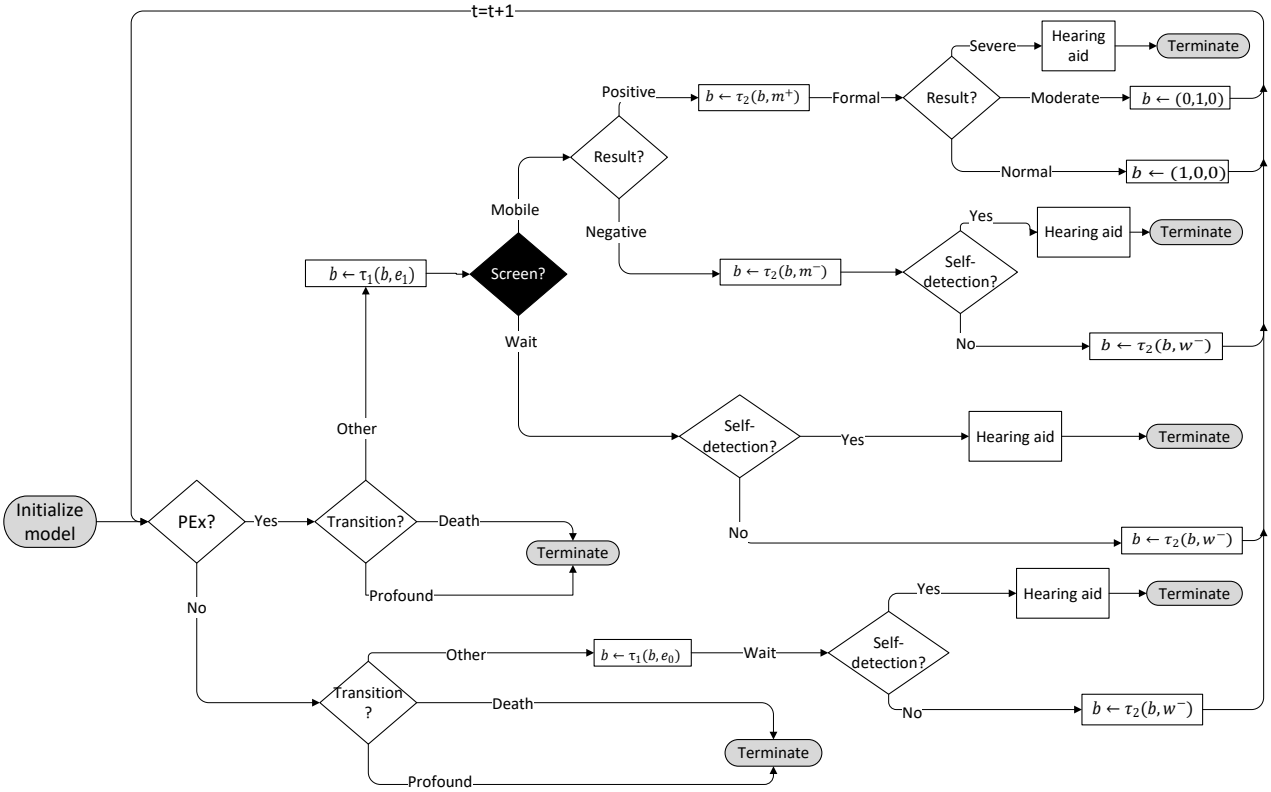
The outcome of mobile audiometry can be either positive, suggesting potential hearing loss, or negative, indicating no hearing impairment. When mobile audiometry yields a positive result, the next step involves confirming this finding through formal audiometry; the prescription of a hearing aid is contingent upon the confirmation of severe hearing loss through formal audiometry. To minimize the patient's visits to the hospital and reduce her exposure to infections, formal audiometry is conducted only after a positive mobile audiometry result.

Since profound hearing loss is detectable by the patient, screening is no longer necessary once the patient is prescribed a hearing aid or a cochlear implant. Screening may only be considered after a PEx incidence since hearing deterioration does not occur in the absence of PEx treatments. In each cycle with a PEx, we may conduct mobile audiometry, followed by formal audiometry if the mobile audiometry result is positive. Alternatively, we can choose to postpone screening until the next PEx. If no screening is conducted or if the result of mobile audiometry is negative, the patient may still become aware of her hearing impairment through self-detection.

5.3. Model Formulation

In this section, we construct a finite horizon POMDP model to develop cost-effective hearing loss screening strategies. The decision-making process and the sequence of events in the POMDP model are illustrated in Figure 2. The following outlines the key components of our model.

Figure 2: Decision process for hearing loss screening.



Note. The black box denotes the optimization layer. The beliefs (1, 0, 0) and (0, 1, 0) correspond to noise-free observations of the normal and moderate hearing loss states, respectively.

- **Decision Epochs:** $t = 1, 2, ..., T$. Given the frequency of severe PEx's and following our clinician's recommendation, we assume that screening decisions are made every three months. We let t denote the number of quarters since age three, the earliest age mobile audiometry can be used (Yeung et al. 2013). Our decision horizon is set at age 80, which is the maximum age reported in the CF literature (O'Brien et al. 2014), resulting in $T = 309$ epochs.
- **Core States:** The state is comprised of the patient's hearing status and whether the patient is alive or not. We consider five hearing states for the patient, normal hearing status (h_n), moderate, severe, and profound hearing loss states (h_m, h_s, h_p, respectively), and hearing aid (h_a). Reaching death (D), profound hearing loss, and hearing aid states terminates our decision process. A patient with severe hearing loss stays in this state until the hearing loss is detected and confirmed, in which case the patient receives a hearing aid or transitions to profound hearing loss or death. A patient with profound hearing loss receives a cochlear implant immediately. We represent the decision making state space with $\bar{S} = \{h_n, h_m, h_s\}$, and terminal state space with $\hat{S} = \{h_a, h_p, D\}$.

- **Action Space:** At each decision epoch, we have the option to either conduct mobile audiometry, which may be subsequently followed by formal audiometry, or wait until the next decision epoch. Action is represented by $a \in A = \{w, m\}$, where w and m are the wait and screen actions, respectively.
- **Belief Space:** We define the belief state, denoted by b, as a probability distribution over the hidden states $\{h_n, h_m, h_s\}$. Our belief space, denoted as B and representing all possible beliefs, is defined as follows:

$$B = \{b \in [0,1]^3 : \sum_{s \in \bar{S}} b(s) = 1\}. \tag{7}$$

  In this equation, $b(s)$ represents the probability of being in state $s \in \{h_n, h_m, h_s\}$. Since the belief state serves as a sufficient statistic for the available observations, a policy in this context maps the belief state at time t to an action.
- **Observation Probabilities**: We represent the conditional probability of observing o when the true state of the patient is $s \in \{h_n, h_m, h_s\}$ as $k(o|s)$.
  - Formal audiometry accurately reveals the true hearing state of the patient. Therefore, we have $k(o|s) = 1$ for any $o = s$, with $o, s \in \{h_n, h_m, h_s\}$ and $k(o|s) = 0$ otherwise.
  - With mobile audiometry, we may obtain either a positive result indicating some hearing loss, denoted as m^+, or a negative result indicating normal hearing, denoted as m^-. Let $1-\alpha$ represent the specificity of mobile audiometry, which is the probability of obtaining a negative result when the true state is normal hearing. Additionally, let $1-\beta(s)$ denote the sensitivity of mobile audiometry, which is the probability of obtaining a positive result when the patient is in one of the two hearing loss states. We then have the following.

$$\begin{cases} k(m^-|h_n) & = 1-\alpha, \\ k(m^+|s) & = 1-\beta(s), \quad \forall s \in \{h_m, h_s\}. \end{cases}$$

  - Hearing loss may be detected with probability p^{sd} by the patient (self-detection) only when the patient is in the state of severe hearing loss. We define the observation set for the 'wait' action as $O_w = \{w^+, w^-\}$, where w^+ represents self-detection of hearing loss, and w^- represents the lack of self-detection. Therefore, we have the following.

$$\begin{cases} k(w^+|h_s) & = p^{sd}, \\ k(w^+|s) & = 0, \quad \forall s \in \{h_n, h_m\}. \end{cases}$$

  - Since a positive mobile audiometry result is followed by formal audiometry, and a negative result may be followed by self-detection, the set of observations for mobile audiometry satisfies $O_m = \{m^+h_n, m^+h_m, m^+h_s, m^-w^+, m^-w^-\}$. Under the assumption of observation independence, we have the following.

$$k\big([o_1o_2]|s\big) = k(o_1|s).k(o_2|s), \quad [o_1o_2] \in O_m.$$

- **Transition Probabilities:** In period t, the patient may have a pulmonary exacerbation denoted by e_1 or not denoted by e_0. The transition probabilities between the core states are denoted by $\mathbf{P}^e(s'|s)$, and the probabilities of death and survival in period t are represented by d_t^e and $\bar{d}_t^e$, respectively. In both notations, $e=0$ represents the absence of a PEx, while $e=1$ represents the presence of a PEx.
- **Belief Update:** The patient's hearing status may progress when experiencing a PEx. Consequently, the presence of a PEx provides valuable information, allowing for the calculation of a posterior belief. The model's decisions are based on this updated belief. Following the screening action (screen or wait), the posterior is further updated based on the acquired observation. The initial posterior update is as follows.

$$b \leftarrow \tau_1[b,e](s') = \sum_{s\in\bar{S}} b(s)\mathbf{P}^e(s'|s), \quad \forall s' \in \{h_n, h_m, h_s\}, e \in \{0,1\}. \tag{8}$$

We proceed to adjust our belief in response to the observed screening action as follows.

$$b \leftarrow \tau_2[b,o](s') = \frac{b(s')k(o|s')}{\sum_{s\in\bar{S}} b(s)k(o|s)} \quad \forall s' \in \{h_n, h_m, h_s\}. \tag{9}$$

In these equations, $\tau_i[b,o](s'), i=1,2$ is the posterior probability of being in state s' after observing o. Depending on the action taken, the set of possible observations would differ. If we do not screen, a self-detection may occur, and thus $o \in O_w$. If we choose to screen, indicating the presence of a PEx, we observe $o \in O_m$.

- **Rewards:** Our objective functions, defined in the next section, optimize functions of the expected total cost and total health utilities. The expected total health utilities are also referred to as QALE. The per period health utilities and expected costs are denoted by $u_t^e(s)$ and $c_t(s,a)$, respectively, for state $s \in \{h_n, h_m, h_s\}$, action $a \in A$, $t < T$, and $e \in \{0,1\}$. Mobile audiometry and wait actions are followed by a confirmatory formal audiometry test in the case of a positive result or self-detection. We denote the cost of formal audiometry as c_f and the cost of mobile audiometry as c_m. Therefore, we have $c_t(s,w) = c_f k(w^+|s)$ and $c_t(s,m) = c_m + c_f[k(m^+|s) + k(m^-|s)k(w^+|s)]$. The screening actions do not have an immediate effect on health utilities, and the potential benefit, such as receiving a hearing aid, is modeled as a terminal reward.

  Upon reaching absorbing states, which render further screening unnecessary, patients are granted terminal lump-sum rewards in terms of health utilities (QALE) and cost. The terminal cost attributed to profound hearing loss encompasses the fixed and expected lifetime maintenance cost associated with the cochlear implant. In contrast, the terminal cost associated with the hearing aid state includes the fixed and expected lifetime maintenance expenses

of the hearing aid, and it may also encompass the costs of the cochlear implant if the patient transitions to profound hearing loss before her death. We represent the terminal health utility and cost for the absorbing states $s \in \{h_a, h_p\}$ as $R_t^u(s)$ and $R_t^c(s)$, respectively.

The health utility for a belief state is defined as the patient's expected health utility with respect to the unknown core state. The expected health utility, denoted as $q_t^e(b)$, satisfies the following.

$$q_t^e(b) = \sum_{s \in \bar{S}} \tau_1(b,e)(s).u_t^e(s), \quad e \in \{0,1\}.$$

In this equation, $\tau_1(b,e)$ represents the posterior belief after observing the PEx status. Similarly, the expected cost is defined as follows.

$$c_t^e(b,a) = \sum_{s \in \bar{S}} \tau_1(b,e)(s) c_t(s,a), \quad e \in \{0,1\}, a \in \{w,m\}.$$

6. Solution Methodology

In this section, we provide a brief overview of the grid-based approximation technique, a commonly used approach for solving POMDP and CPOMDP problems. Subsequently, we adapt this approach to solve our NMB and ICER optimization models.

6.1. Grid-based Approximations

One approach to solve POMDPs involves solving the equivalent MDP model on the set of reachable beliefs (Altman 1999). This method necessitates the enumeration of reachable beliefs in advance and subsequently solving the MDP using established MDP solution techniques, such as value iteration or linear programming. While in finite-horizon POMDPs with finite sets of actions, observations, and core-states, the set of reachable beliefs is indeed finite, it is important to note that this set grows exponentially as the horizon length increases. This exponential growth poses a significant computational challenge, rendering the problem computationally intractable for longer horizons. Approximation approaches, including the grid-based approximation method, have been widely employed in the literature. Numerical and theoretical studies have demonstrated the efficacy of grid-based approximation in generating high-quality policies (Lovejoy 1991, Kavaklioglu and Cevik 2022). Grid-based approximation can solve both unconstrained and constrained POMDPs by transforming them into unconstrained or constrained MDP approximations.

In a grid-based approximation approach, the entire belief simplex is approximated by a limited set of beliefs, known as grid points. To account for beliefs that are not explicitly part of the grid, various interpolation strategies can be employed. Linear interpolation stands out as a significant interpolation technique due to its favorable theoretical characteristics. In linear interpolation, a

belief b that is not among the grid points is approximated as a convex combination of the beliefs within the grid set. This convex combination must accurately reproduce the target belief, which can be expressed mathematically by ensuring that the convex weights satisfy the following condition.

$$b = \sum_{j=1}^{|G|} \theta_j^b b_j^G, \tag{10}$$

where b_j^G are the grid points in the grid set G, and θ's are the weights. It is worth noting that, in general, the weights θ are not uniquely defined, and their selection process can significantly impact the quality of the approximation. Consequently, the literature has proposed various weighting methods, which vary in terms of their performance and computational efficiency (Lovejoy 1991, Poupart et al. 2015, Hauskrecht 2000). After the weights are determined, regardless of the weighting mechanism used, we can calculate the transition probabilities between the beliefs represented by the grid points. Through this process, a POMDP can be transformed into an MDP, with the grid set serving as the state space. The following steps are required to compute the transition probability between grid points.

- Use Eq. 8 and Eq. 9 to compute the posterior belief for all combinations of beliefs in the grid set, observations, and actions. We use b_j^{ao} to denote the posterior belief corresponding to the initial belief b_j^G, action a, and observation o.
- Utilize the weighting mechanism of choice to calculate weights for the posterior b_j^{ao}. We denote weights of grid point b_i^G corresponding to posterior b_j^{ao} as θ_{ji}^{ao}. These weights should satisfy the following.

$$b_j^{ao} = \sum_i \theta_{ji}^{ao}.b_i^G, \quad \forall b_i^G \in G.$$

- Use the following formula to compute the transition probabilities between grid points b_i^G and b_j^G in the grid. In this formula, we calculate the total weight for grid point b_j^G by summing all weights over all possible posteriors of b_i^G.

$$\mathbf{P}^e(b_j^G|b_i^G, a) = \sum_{o\in O_a} k(o|\tau_1(b_i^G, e))\theta_{ji}^{ao}, \qquad \forall b_j^G, b_i^G \in G, e = 0, 1, a \in \{w, m\}.$$

In this equation, $k(o|b)$ represent the total probability of observing o when the belief is b, satisfying the following equation.

$$k(o|b) = \sum_{s\in\bar{S}} b(s)k(o|s).$$

Note that transitions occur after observing the PEx status and making a decision. Therefore, the transition probabilities are calculated for a values of PEx status e and action a.

It is a well-established fact that grid-based approximation with linear interpolation yields a lower bound for a POMDP with a minimization objective, as demonstrated in Lovejoy (1991). As a result, employing a grid-based approach to solve the NMB maximization problem provides an upper bound on the optimal NMB value. However, the ICER minimization problem presents a different challenge. This problem constitutes a non-linear CPOMDP, where the single constraint and the non-linear objective are defined over two expected total rewards (costs and utilities). Thus, the earlier mentioned result does not directly apply to this context.

Poupart et al. (2015) proposed an extension of the bounding result for CPOMDPs when both constraints and the objective function are linear in the expected total rewards. Due to the non-linearity of the ICER objective, this extension also does not directly apply to our problem. In the subsequent sections, we will extend this result to a general non-linear CPOMDP and provide an alternative proof to the result in Poupart et al. (2015).

Let us start by formally defining the model setting. We are addressing a sequential decision-making problem where the system's state evolves according to a hidden Markov model. During each time period, multiple reward streams are accumulated. These rewards can be contingent on both the chosen action and the underlying state, and they may have different natures, including cost (negative reward) and utility. We can mathematically represent a CPOMDP as follows.

$$\begin{aligned} \inf_{\pi} \quad & \varphi(\bar{r}_1(\pi),\dots,\bar{r}_m(\pi)) \\ \text{s.t.} \quad & (\bar{r}_1(\pi),\dots,\bar{r}_m(\pi)) \in \chi, \\ & \bar{r}_i(\pi) = \mathbb{E}[\sum_{t<T} \gamma^t r_i(s_t,a_t) + \gamma^T R_i(s_T)|\pi], \end{aligned} \tag{11}$$

In this model, $\bar{r}_i(\pi)$ represents the expected total reward for reward stream i under policy π. The tuple $(\bar{r}_1(\pi),\dots,\bar{r}_m(\pi))$ forms the rewards vector, and χ represents the feasibility set for the rewards vector. In other words, it represents the constraints on the cumulative rewards. Lastly, $\varphi : \Re^m \to \Re$ represents the non-linear mapping for the objective function.

We can transform Problem 11, which is a CPOMDP, into an equivalent constrained belief state MDP using the grid-based approximation and linear interpolation techniques described earlier. Let $\mathcal{R}(\pi) \in \Re^m$ and $\hat{\mathcal{R}}(\pi) \in \Re^m$ denote the vectors of expected total rewards in the original CPOMDP and the approximation CMDP, respectively. With these definitions, we can reformulate both the CPOMDP and its corresponding approximation CMDP as follows.

$$\begin{aligned} v = \inf_{\pi} \quad & \varphi(\mathcal{R}(\pi)) \\ \text{s.t.} \quad & \mathcal{R}(\pi) \in \chi. \end{aligned} \tag{12}$$

$$\begin{aligned} \hat{v} = \inf_{\pi} \quad & \varphi(\hat{\mathcal{R}}(\pi)) \\ \text{s.t.} \quad & \hat{\mathcal{R}}(\pi) \in \chi. \end{aligned} \tag{13}$$

Let $\mathcal{A}$ and $\mathcal{B}$ be sets of reward vectors $\mathcal{R}(\pi)$ and $\hat{\mathcal{R}}(\pi')$ for all admissible policies π and π' in the original CPOMDP and the approximation CMDP problems, respectively. To clarify, we can define

$\mathcal{A}$ as $\mathcal{R}(\Pi^{HR})$ and $\mathcal{B}$ as $\hat{\mathcal{R}}(\hat{\Pi}^{HR})$, where Π^{HR} and $\hat{\Pi}^{HR}$ represent the sets of all history-dependent, random policies for each respective problem. In Theorem 1 below, we establish that $\mathcal{A} \subset \mathcal{B}$, implying $v \geq \hat{v}$. Therefore, we can conclude that the grid-based approximation of minimization CPOMDPs produces a lower bound for the true objective value.

THEOREM 1. We have $\mathcal{A} \subset \mathcal{B}$, and hence, $v \geq \hat{v}$.

Hence, we can utilize Theorem 1 to infer that the optimal ICER obtained through grid-based approximation serves as a lower bound for the true optimal ICER value. Similarly, we can utilize this result to prove that the frontier generated using the grid-based approximation is positioned either below or on the true frontier. This is because the approximation generates lower bounds on the optimal cost for any given QALE. We leverage this result in Section 7.4 to evaluate the quality of the approximation.

6.2. NMB Objective

Optimality Equations: Consider $v_t(b)$ as the value function at belief state b during period t. For the absorbing states $s \in \{h_a, h_p, D\}$, we define $v_t(s)$ as the terminal NMB reward, which is governed by the equation $v_t(s) = \lambda \cdot R_t^u(s) - R_t^c(s)$. Here, we explicitly set $v_t(D) = 0$. For the non-absorbing states, which are the belief states, we define the immediate reward as follows: $r_t^e(b,a) = \lambda q_t^e(b) - c_t^e(b,a)$.

Given that patient outcomes are contingent on the PEx status and whether transitions to absorbing states occur, we introduce two additional value functions. Let $v_t^e(b)$ denote the value function conditioned on PEx status, where $e = 0, 1$. We then have the following.

$$v_t(b) = p(e_1).v_t^1(b) + p(e_0).v_t^0(b), \quad \forall b \in G. \tag{14}$$

In this equation, $p(e_i)$ with $i = 0, 1$, represents the probability of the absence and presence of a PEx, respectively. Define $z_t^e(b)$ as the value function at belief state b during period t, conditioned on remaining in the hidden states without transitioning to either death or the profound hearing state. For all $b \in G$ and $e \in \{0, 1\}$, the following equation holds.

$$v_t^e(b) = \bar{d}_t^e \left\{ \bar{\mathbf{P}}^e(h_p|b).z_t^e(\tau_1(b,e)) + \mathbf{P}^e(h_p|b).v_t(h_p) \right\} + d_t^e.v_t(D), \tag{15}$$

where $\mathbf{P}^e(h_p|b)$ represents the probability of transitioning to profound hearing loss from belief state b and adheres to the following equation:

$$\mathbf{P}^e(h_p|b) = \sum_{s \in \bar{S}} b(s).\mathbf{P}^e(h_p|s), \quad e = 0, 1.$$

We employ $\bar{p}$ to represent the complement of the probability p, defined as $\bar{p} = 1 - p$. In the presence of a PEx, we determine our screening decision as follows.

$$z_t^1(b) = \max_{a \in A} \{ r_t^1(b,a) + \gamma \sum_{b' \in G} \mathbf{P}^1(b'|b,a) v_{t+1}(b') \}, \quad \forall b \in G.$$

In the absence of the PEx event, the following equation applies.

$$z_t^0(b) = r_t^0(b,w) + \gamma \sum_{b' \in G} \mathbf{P}^0(b'|b,w) v_{t+1}(b'), \quad \forall b \in G.$$

Linear Programming: As an alternative to value-function-based methods, we can approach the problem using a linear programming model, which is detailed in the following equations.

$$\max \ \lambda q - c \tag{16a}$$

$$\text{s.t.} \quad \sum_{a \in A} y_1(b,a) = \delta(b), \quad \forall b \in G, \tag{16b}$$

$$y_1(s) = 0, \quad \forall s \in \{h_a, h_p\}, \tag{16c}$$

$$\sum_{a \in A} y_t(b',a) = \gamma \sum_{i=0,1} p(e_i) \bar{d}_{t-1}^i \sum_{b \in G, a \in A} \bar{\mathbf{P}}^i(h_p|b) \mathbf{P}^i(b'|b,a) y_{t-1}(b,a), \quad \forall b' \in G, 1 < t < T, \tag{16d}$$

$$y_t(h_a) = \gamma \sum_{i=0,1} p(e_i) \bar{d}_{t-1}^i \sum_{b \in G, a \in A} \bar{\mathbf{P}}^i(h_p|b) \mathbf{P}^i(h_a|b,a) y_{t-1}(b,a), \quad 1 < t, \tag{16e}$$

$$y_t(h_p) = \gamma \sum_{i=0,1} p(e_i) \bar{d}_{t-1}^i \sum_{b \in G, a \in A} \mathbf{P}^i(h_p|b) y_{t-1}(b,a), \quad 1 < t, \tag{16f}$$

$$q = \sum_{a \in A, b \in G, t} q_t^1(b) y_t(b,a) + \sum_{s \in \hat{S}, t} R_t^u(s) y_t(s), \tag{16g}$$

$$c = \sum_{a \in A, b \in G, t} c_t^1(b,a) y_t(b,a) + \sum_{s \in \hat{S}, t} R_t^c(s) y_t(s), \tag{16h}$$

$$y_t(b,a), y_t(s) \geq 0, \qquad \forall b \in G, a \in A, t < T, s \in \{h_a, h_p\}. \tag{16i}$$

In the linear programming model outlined above, $\delta(b)$ represents the initial distribution over the belief states. The continuous decision variable $y_t(b,a)$ signifies the discounted state-action occupancy measure at time $t < T$, and similarly, $y_t(s)$, where $s \in \{h_a, h_p\}$, denotes the discounted probability of occupying the absorbing states. Eq. 16i imposes logical constraints on the decision variables. The solution to the set of equations Eq. 16b−16f corresponds to the occupancy measures of the policy described in equation Eq. 17 below (Ross 2014).

$$\pi_t(a|b) = \frac{y_t(b,a)}{\sum_{a' \in A} y_t(b,a')}. \tag{17}$$

In this formula, the policy $\pi_t(a|b)$ represents the probability of taking action a in belief b at time t. Therefore, once the model is solved, one can extract an (possibly random) optimal policy π_t^* by normalizing the optimal occupancy measures to 1 as shown in Eq. 17.

6.3. ICER Formulation

Since the ICER minimization problem is a constrained non-linear POMDP, traditional value-function-based algorithms such as the value iteration method are no longer applicable. As an alternative approach, we extend the linear programming model designed for the NMB maximization and formulate the following constrained, non-linear programming model.

$$\begin{aligned} \min \quad & \frac{c-c_0}{q-q_0} \\ \text{s.t.} \quad & \text{Eq. } 16b-16i, \\ & q > q_0. \end{aligned}$$

We will demonstrate later that the strict inequality $q > q_0$ does not compromise the validity of the mathematical model presented above. To handle the non-linearity of the objective function in the model described above and make it amenable to linear programming, we can employ the Charnes-Cooper transformation method and develop the following linear programming model (Charnes and Cooper 1973).

$$\min \sum_{a\in A, b\in G, t} c_t^1(b,a)\zeta_t(b,a) + \sum_{s\in \hat{S},t} R_t^c(s)\zeta_t(s) - c_0 m \tag{18a}$$

$$\text{s.t.} \quad \sum_{a\in A, b\in G, t} q_t^1(b)\zeta_t(b,a) + \sum_{s\in \hat{S},t} R_t^u(s)\zeta_t(s) - q_0 m = 1, \tag{18b}$$

$$\sum_{a\in A} \zeta_1(b,a) = \delta(b)m, \quad \forall b \in G, \tag{18c}$$

$$\zeta_1(s) = 0, \quad \forall s \in \{h_a, h_p\}, \tag{18d}$$

$$\sum_{a\in A} \zeta_t(b',a) = \gamma \sum_{i=0,1} p(e_i)\bar{d}_{t-1}^i \sum_{b\in G, a\in A} \bar{\mathbf{P}}^i(h_p|b)\mathbf{P}^i(b'|b,a)\zeta_{t-1}(b,a), \quad \forall b' \in G, 1 < t < T, \tag{18e}$$

$$\zeta_t(h_a) = \gamma \sum_{i=0,1} p(e_i)\bar{d}_{t-1}^i \sum_{b\in G, a\in A} \bar{\mathbf{P}}^i(h_p|b)\mathbf{P}^i(h_a|b,a)\zeta_{t-1}(b,a), \quad 1 < t, \tag{18f}$$

$$\zeta_t(h_p) = \gamma \sum_{i=0,1} p(e_i)\bar{d}_{t-1}^i \sum_{b\in G, a\in A} \mathbf{P}^i(h_p|b)\zeta_{t-1}(b,a), \quad 1 < t, \tag{18g}$$

$$\zeta_t(b,a), \zeta_t(s), m \geq 0, \qquad \forall b \in G, a \in A, t < T, s \in \{h_a, h_p\}. \tag{18h}$$

We can obtain a solution to the original problem by applying the following transformations.

$$\begin{aligned} y_t(b,a) &= \frac{1}{m}\zeta_t(b,a), \forall b \in G, a \in A, t < T, \\ y_t(s) &= \frac{1}{m}\zeta_t(s), s \in \{h_a, h_p\}. \end{aligned} \tag{19}$$

We now discuss a few important points regarding the ICER problem and its reformulation.

- The system of equations above is feasible when $q_0 < q_{\max}$. We prove this result by constructing a solution to Eq. 18 based on a policy achieving the maximum QALE. Let y represent the

occupancy measures associated with one such policy satisfying the constraints in Eq. 16. We can use Eq. 16g and the transformation in Eq. 19 to reformulate Eq. 18b as $qm - q_0 m = 1$. Let $m = (q_{\max} - q_0)^{-1}$ and calculate ζ from y using Eq. 19. It can easily be verified from Eq. 16 and Eq. 18 that the constructed variables m and ζ collectively form a feasible solution to the transformed model.

- The transformations in Eq. 19 is well-defined since $m > 0$ for any feasible solution to Eq. 18. To show this, note that $qm - q_0 m = 1$ implies $m \neq 0$. Since $m \geq 0$, we have $m > 0$
- The strict inequality $q > q_0$ holds in any feasible solution to Eq. 18, thereby establishing the validity of the ICER model despite its inclusion of a strict inequality. We can prove this by noting that $qm - q_0 m = 1$ implies $q - q_0 = \frac{1}{m} > 0$.

In the forthcoming Corollary 1, we establish that a deterministic, optimal policy exists for any well-defined ICER optimization problem. An ICER optimization problem is considered well-defined when the comparator policy is non-dominated, and its QALE can be strictly enhanced by a feasible policy, which holds when $q_0 < q_{\max}$. Let Y be the set of all occupancy measures for an MDP or a POMDP, with either finite or infinite horizon. We now present the following intermediate result, which extends the findings of Wagner (1960) to encompass arbitrary MDPs and also provides an alternative proof strategy.

Lemma 2. The extreme points of Y are associated with deterministic policies.

This result suggests that when an optimal solution for a (potentially constrained) MDP/POMDP resides at an extreme point of Y, the problem naturally lends itself to deterministic, optimal policies. One immediate application of this result is to establish that unconstrained POMDPs minimizing a quasi-concave objective admit deterministic policies. This conclusion builds on Theorem 3.2 in Bereanu (1974), which demonstrates the existence of a minimizer at the extreme points of a compact, convex feasible region in a minimization problem with a quasi-concave objective function.

Corollary 1. ICER minimization problems admit a deterministic, optimal policy.

This result ensures the cost-effectiveness of deterministic policies, which are desirable for their straightforward implementation in medical practice. Moreover, it eliminates the need for using integer programming techniques to enforce deterministic policies.

7. Numerical Results

In this section, we conduct a numerical study to explore the optimal design of cost-effective strategies for screening hearing loss in individuals with cystic fibrosis. We solve a data-driven model using the methodology described in Sections 4–6. We use a regular grid with resolution $m = 40$,

resulting in 861 grid points. We utilize the Freudenthal triangulation method for linear interpolation (Lovejoy 1991). Increasing the resolution did not yield a notably discernible effect on the model outcomes. We implemented the value iteration method in R to solve the NMB problem (R Core Team 2022). The linear programming model for ICER minimization was solved using GAMS 36 (GAMS Development Corporation 2021). All computations were carried out on a quad-core Intel 3.20 GHz processor with 16 GB of RAM. The computation times for solving the optimization problems are discussed in the following sections.

7.1. Model Parameters

The model parameters needed for the numerical experimentation were extracted from the literature (refer to Table 1 for details). We briefly describe them below.

- **State transitions:** At each cycle, the patient may experience a PEx with a quarterly probability of 0.106 (Waters et al. 2015). The presence of PEx during a period has a detrimental effect on both survival and hearing. We used the survival model in Keogh et al. (2018) along with the mortality hazard ratio linked to PEx versus non-PEx conditions (6.17, as reported in Stephenson et al. (2015)) to calculate the survival probabilities. The patient's hearing deteriorates exclusively during periods marked by the incidence of a PEx. The chances of transitioning from normal hearing and moderate and severe hearing loss states to a one-degree worse hearing state after experiencing a PEx are 0.132, 0.097, and 0.193, respectively (Garinis et al. 2021). Considering the natural, gradual progression of hearing status, it is improbable for there to be a deterioration of more than one degree or any significant improvement.
- **Observation probabilities:** For the mobile audiometry, we consider a specificity of 88% and sensitivities of 87.5% and 93.3% for the moderate and severe hearing loss states, respectively (Yeung et al. 2013, Saliba et al. 2017, Vijayasingam et al. 2020). We calibrated the self-detection probability using the average hearing loss detection delay from McMahon et al. (2013) (10 years).
- **Rewards (costs and health utilities):** The costs of formal and mobile audiometry visits are \$241 and \$31.31, respectively (American Speech Language Hearing Association 2019, United States Bureau of Labor Statistics 2019). The one-time cost of the hearing aid and annual maintenance cost are $\$2,325$ and \$574, respectively (Gillard and Harris 2020). The initial cost of the cochlear implant is \$51,084, and the annual maintenance cost is \$1,253 (Laske et al. 2019). All costs have been adjusted to costs in 2022 based on published inflation rates (United States Government 2022a).

  The patient's health utility is influenced by the incidence of a PEx and her hearing status, calculated using the multiplicative approach. The baseline utilities of a CF patient with and

Table 1: Model input parameters.

Inputs	Value	Source
Survival model	-	Keogh et al. (2018)
Mortality hazard ratio PEx vs. No PEx	6.17	Stephenson et al. (2015)
Hearing loss detection delay	10 years	McMahon et al. (2013)
State transition probabilities		
Quarterly chance of PEx	0.106	Waters et al. (2015)
Normal to moderate	0.132	Garinis et al. (2021)
Moderate to severe	0.097	Garinis et al. (2021)
Severe to profound	0.193	Garinis et al. (2021)
Utilities		
Baseline without PEx	0.83	Michael et al. (2003)
Baseline with PEx	0.76	Solem et al. (2016)
Moderate hearing loss	0.89	Abrams et al. (2005)
Severe hearing loss	0.77	Abrams et al. (2005)
With hearing aid	0.83	Abrams et al. (2005), Barton et al. (2004)
With cochlear implant	0.8	Abrams et al. (2005)
Costs ($)		
Formal audiometry	241	American Speech Language Hearing Association (2019)
Mobile audiometry	31.31	United States Bureau of Labor Statistics (2019)
Hearing aid new user	2,325	Gillard and Harris (2020)
Hearing aid maintenance	574	Gillard and Harris (2020)
Cochlear implant new user	51,084	Laske et al. (2019)
Cochlear implant maintenance	1,253	Laske et al. (2019)
Mobile audiometry accuracy (%)		
Specificity	88	Vijayasingam et al. (2020)
Sensitivity at state h_m	87.5	Yeung et al. (2013)
Sensitivity at state h_s	93.3	Yeung et al. (2015)

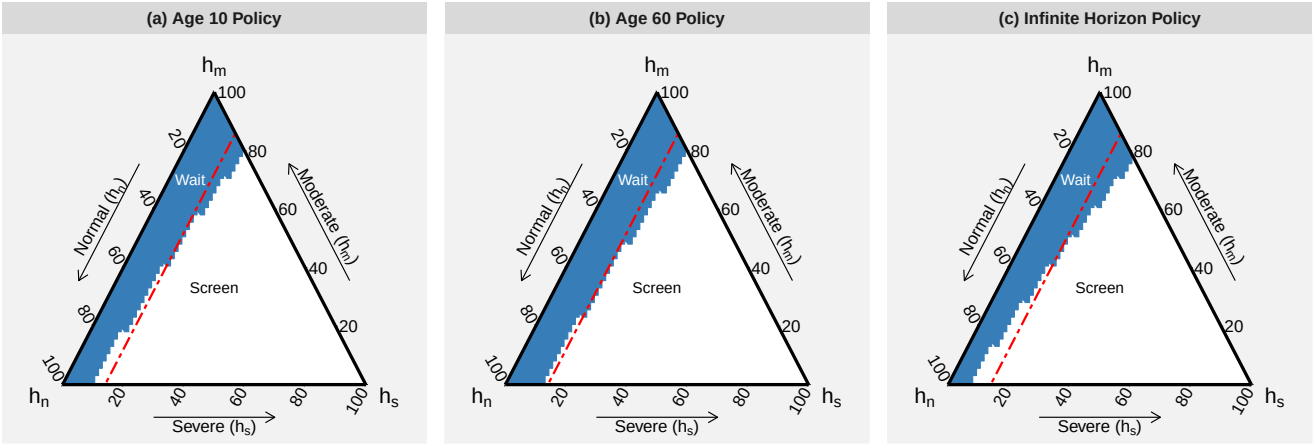
Costs are converted to 2022 USD when used in the model (United States Government 2022a).

without PEx are 0.83 and 0.76, respectively (Michael et al. 2003, Solem et al. 2016). The health utilities of moderate and severe hearing loss are 0.89 and 0.77, respectively (Abrams et al. 2005). The hearing aid increases the utility by 0.06 (Barton et al. 2004). A patient with profound hearing loss depends on a cochlear implant, which has a utility of 0.8 (Abrams et al. 2005). The lump-sum total rewards for the terminal states of profound and hearing aid are determined by calculating the expected (discounted) rewards of a Markov reward model. All rewards are discounted at an annual rate of 5%.

7.2. Optimal Policy

We utilize a base-case WTP of £20,000, a value commonly employed by NICE in the UK (McCabe et al. 2008). This amount is approximately equivalent to $27,000 (United States Government

Figure 3: Optimal policy for the finite and infinite horizon model with WTP=£20,000



Note. Ternary plots for the optimal policies for age 10 (a) and 60 (b) and the infinite horizon model (c). The red line shows a threshold of 17% for the severe hearing loss state.

2022b). To visually represent the optimal policy across all beliefs within the belief simplex, we employ a ternary diagram (Weisstein 2021). In this diagram, axes corresponding to the probabilities of being in each hidden state $s \in \{h_n, h_m, h_s\}$ are positioned along the sides of a triangle. We illustrate the optimal policy for CF patients in the age groups of 10 and 60 in Figure 3. A more detailed plot is provided in Appendix C.

The consistency in the optimal policy observed across different age groups underscores the notion that age may not significantly influence screening decisions within the typical life expectancy range for CF patients (between 42 to 50 years). This suggests that a transition from a finite horizon to an infinite horizon model may be appropriate. Such a transition streamlines the exploration and communication of optimal policy dynamics regarding changes in model inputs, such as WTP. Policies that are not influenced by age are more straightforward to comprehend and implement within a clinical setting. Additionally, this approach leads to decreased computational time. Solving instances of NMB and ICER problems in the finite horizon model took approximately one minute and 20 minutes, respectively. However, after the transition, these times were reduced to less than a second. Furthermore, there is theoretical support for transitioning to an infinite horizon model when a finite horizon Markov model contains a significant number of cycles, as discussed in Section 6.8 of Puterman (2014). Numerically, it is evident that infinite horizon policies exhibit similarity in shape and perform well, as illustrated in Figure 3 and Figure 5.

7.3. Easy-to-Implement Policies

In this section, our focus is on deriving practical and effective policies for real-world healthcare settings. We aim to bridge the gap between advanced mathematical models and practical healthcare

decisions, translating our research findings into actionable strategies that benefit patients and healthcare systems.

7.3.1. Threshold-based Policy The pursuit of effective threshold-based policies is motivated by their ease of implementation. Ternary plots in Figure 3, depicting optimal policies for finite and infinite horizon problems, suggest that regions where screening and waiting are optimal can be roughly differentiated by a discriminant line. The choice of this line can be guided by various metrics. One approach may seek to minimize classification errors, while another could focus on identifying the optimal discriminant line associated with more effective approximate policies. Both approaches are likely to generate threshold policies that rely on the entire belief state. Another method involves finding a line that delineates belief states based on just one of the three belief probabilities. If this approximate policy is demonstrated to be effective, it offers even greater simplicity in implementation. We adopt the latter approach and demonstrate its efficacy through simulation results.

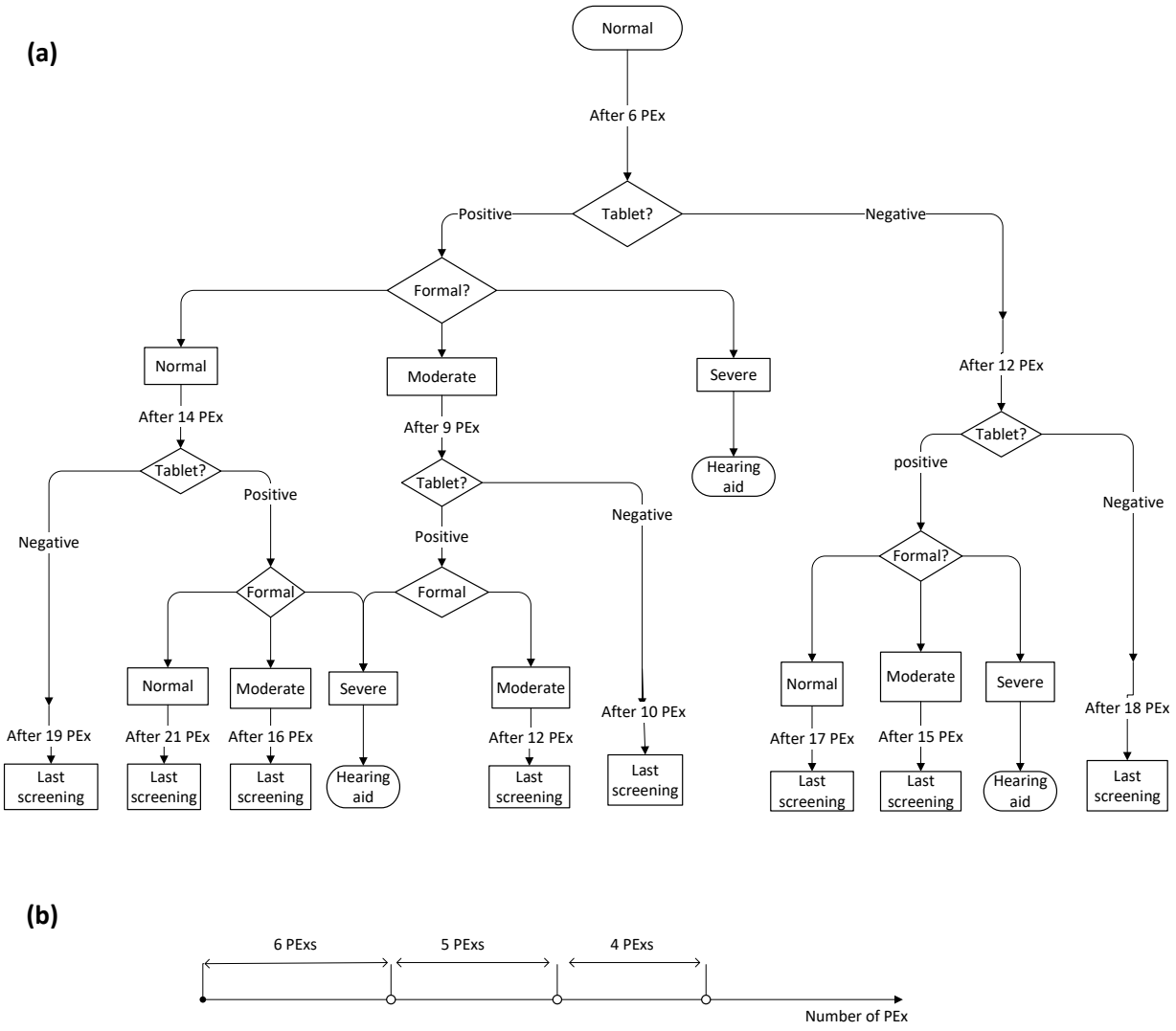
As demonstrated in Figure 3, we can approximately differentiate the two optimal policy regions based on the likelihood of severe hearing loss. Thus, we determine a line that minimizes total deviations (classification errors) from the original policy. Numerically, a threshold of 17% yields the lowest classification error for a WTP of £20,000.

7.3.2. Simulation-based Approximate Policies Our pursuit of practical policies leads us to leverage insights gained from investigating the simulated patient trajectories under the optimal policy. This extensive dataset empowers us to devise two distinct categories of screening policies: history-dependent and history-independent policies, as elucidated in Figure 4. The feasibility of these simplified approximations of the optimal policy relies on the observation that the optimal policy screens patients on average two times over their lifetime.

Within the class of history-dependent policies, the decision-making process depends on two variables: number of past PEx's and the specific path of observations, i.e., the observation trajectory. These policies are tailored to provide patients with screening recommendations based on their unique medical history, thus optimizing the timing of screenings. History-dependent policies can be presented on a decision tree, as shown in Figure 4. In contrast, history-independent policies streamline decision-making by relying solely on the count of PEx's since the last observation, further simplifying the screening schedule. Specifically, in this scenario, the first three assessments occur after 6, 11, and 15 PEx's, corresponding to average patient ages of 17, 29, and 38 years.

To accommodate the requirements of real-world clinical settings, we provide flexibility by allowing the adjustment of the number of screenings, ranging from 2 to 3 within each policy class. This yields four distinct policies, each tailored for specific clinical situations. These policies provide practical guidance to healthcare providers, ensuring that screenings align with the unique requirements of individual patients.

Figure 4: Simulation-based Approximate Policies



Note. History-dependent (a) and history-independent (b) approximations. Approximate policies may use the number of PEx's elapsed since the last observation and observations trajectory (in the case of history-dependent policy) to advise screening decisions. In both policies, self-detection terminates the decision algorithm.

7.4. Policy Evaluation

We utilize a Monte Carlo simulation in R to assess the effectiveness of the proposed policies across various metrics. This simulation required approximately 5 minutes to complete. In addition to primary outcomes such as expected total cost, QALE, and our main objectives, we examine secondary outcomes critical for assessing the efficiency of hearing loss screening strategies. These include the expected hearing loss detection delay, lifetime screenings count, and the proportion of patients

missing hearing aids due to detection delays. The detection delay represents the time elapsed between a patient reaching severe hearing loss and either receiving a hearing aid, transitioning to profound hearing loss, or experiencing mortality, whichever occurs first.

To ensure robust results, we conduct 10,000 simulation replications, a number determined based on result convergence. Additionally, we employ the method of common random numbers to enhance precision and reduce variability (Glasserman and Yao 1992). This technique ensures that variables unaffected by screening policies, such as PEx state, mortality, hearing status transitions, etc., remain constant within each simulation replication and across all policies evaluated.

We conducted a thorough evaluation of a broad range of policies to offer valuable insights into different screening approaches, including their respective advantages, disadvantages, and practical feasibility. The policies are listed below in ascending order of implementation complexity.

- Extreme policies: (a) no-screening policy, 'W-All', a policy that abstains entirely from screening, and (b) the frequent screening policy, 'X-All', which advocates screening at each and all PEx's, a policy recommended by Huang et al. (2021).
- Fixed-interval strategies denoted by 'T=nY', which involve screenings scheduled at regular intervals such as every $n = 1, 2, 5$ years; annual screening was proposed by by Vijayasingam et al. (2020).
- History-dependent and history-independent policies, denoted by 'X-HD-n' and 'X-HI-n' for $n = 2$ and 3, respectively, where n is the maximum allowed number of screenings.
- Stationary threshold-based policy
- Optimal policies derived from the finite and infinite horizon, grid-based approximation models. Evaluation of grid-based policies requires a special nuanced approach. These policies operate under the assumption of transitions occurring exclusively between beliefs within the grid set. When the posterior does not belong to the grid set, we generate a sample posterior belief using the interpolation weights.

Based on simulation findings, patients exhibit an average life expectancy of approximately 45 years, with roughly 55% experiencing severe hearing loss at an average age of 28 years. The simulation results, presented in Table 2, underscore the notable advantages presented by the optimal screening policy. In comparison to a no-screening approach, the optimal strategy offers substantial enhancements in patient outcomes. Specifically, the optimal policy significantly reduces the detection delay from an average of 32.26 months to 17.66 months. Furthermore, the optimal policy reduces the proportion of patients who miss out on receiving hearing aids from 50% under the no-screening policy to 27%. This optimal approach ensures that patients with severe hearing loss receive hearing aids promptly, potentially leading to enhancements in their overall quality of life.

The cumulative impact of these enhancements on patient outcomes is reflected in QALE, which translates to approximately five additional Quality-Adjusted Life Days (QALDs) for each patient. However, it is essential to note that the magnitude of these health benefits, as indicated by the improved QALE, is moderated by the effects of discounting over more than two decades. It is worth highlighting that patients typically require a hearing aid around the age of 28 years, whereas the QALE is expressed with reference to a hypothetical three-year-old patient.

We provide a visual representation of the performance, including cost and QALE, of all the policies in Figure 5. These plots also feature the cost-effectiveness frontier, which will be discussed in the next section. This comprehensive visualization empowers policymakers with the necessary insights to make informed decisions by balancing policy performance and ease of implementation. As a result, it contributes to more effective and efficient healthcare decision-making.

Several notable insights emerge from this illustration. Firstly, fixed-interval policies are observed to be strictly dominated by other strategies, suggesting that they are not optimal choices for balancing QALE and cost. Secondly, the extreme policies of no screening and screening at every PEx are positioned on the cost-effectiveness frontier. These policies are associated with WTP values at the two ends of the WTP spectrum. Thirdly, the threshold-based and infinite horizon approximations virtually lie on the cost-effectiveness frontier. Finally, while the history-dependent and history-independent policies are slightly dominated by the optimal policies, they perform relatively well.

The proximity of the policy performances to the frontier in Figure 5 (b) indicates the quality of our approximation and a satisfactory optimality gap achieved with a grid resolution of 40. To illustrate this, we argue that the true frontier lies between the generated frontier and the vectors of policy performances. In Theorem 1, we showed that the generated frontier is positioned either below or on the true frontier. Furthermore, evaluating policy performance involves conducting simulation-based assessments on the true model, not one that is misspecified. This process yields unbiased vectors of policy performances, which may lie either above or on the true frontier, depending on whether the policy is inefficient or efficient, respectively.

7.5. ICER Algorithm

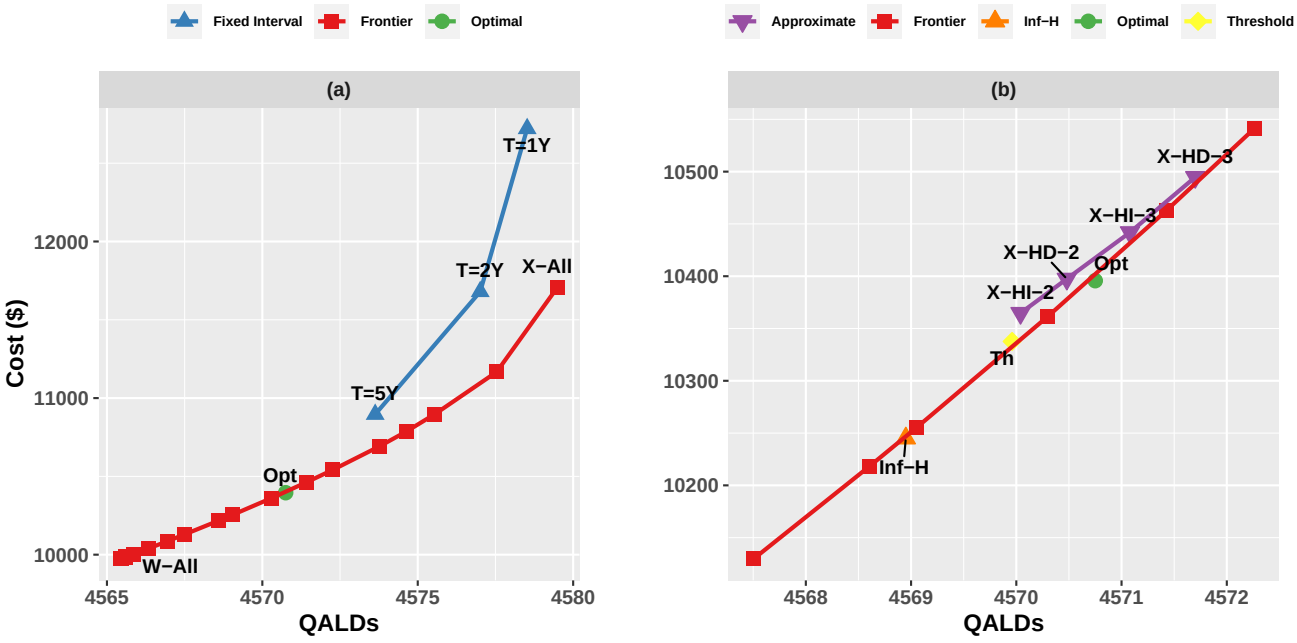
We employ Algorithm 1 to systematically identify all cost-effective policies, determine the valid range of WTP values for each policy, and evaluate their performance. Our findings for WTP $\leq$ \$100,000 per QALY are visually presented in Figure 6. Panel (a) displays the ICER values for a range of algorithm iterations, while panel (b) showcases the performance of the generated policies, collectively establishing the cost-effectiveness frontier.

Table 2: Performance of various screening strategies.

Policy	Cost ($)	QALDs	Delay (months)	#Screening	% Missed Hearing aid
X-All	11,709.06	4579.50	0.93	12.2	1.5
T=1Y	12,720.67	4578.52	2.8	29.6	4.34
T=2Y	11,680.06	4577.01	5.74	15.26	8.83
T=5Y	11,464.01	4573.6	12.66	6.57	19.42
X-HD-3	10,494.73	4571.69	18.86	1.89	30.64
X-HI-3	10,441.93	4571.08	19.7	1.83	32.05
Opt	10,395.72	4570.75	17.66	1.62	27.2
X-HD-2	10,397.45	4570.48	22.8	1.5	37.4
X-HI-2	10,364.72	4570.04	22.9	1.47	37.31
Th	10,337.63	4569.96	18.07	1.63	27.2
Inf-H	10,245.02	4568.95	21.54	1.06	32.9
W-All	9,985.05	4565.60	32.26	0	50.7

'W-All' and 'X-all' are no-screen and screen at all PEx's, 'T=nY' is the fixed interval screening every $n = 1, 2, 5$ years, 'X-HD-n' and 'X-HI-n' are history-dependent and history-independent policies with at most $n = 2, 3$ screenings, 'Opt' and 'Inf-H" are the optimal policies for the finite horizon and infinite horizon approximation, and 'Th' is the stationary threshold-based approximation. Note that policies are presented in decreasing order of QALDs.

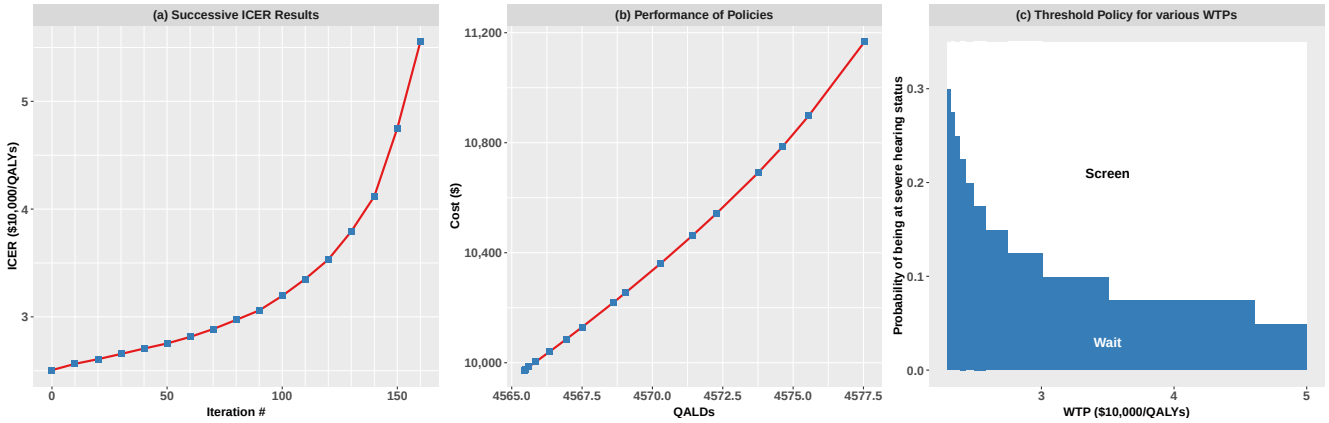
Figure 5: QALE and cost of various screening strategies.



Note. The policy abbreviations are given in Table 2. The optimal frontier is generated by Algorithm 1. Panel (a) shows the performance of fixed-interval policies, and panel (b) shows various approximate policies.

The proposed algorithm has the capability to generate the entire spectrum of cost-effective policies. Each policy can be represented by a set of ternary plots for each age and a single ternary plot for the finite and infinite horizon models, respectively. However, for enhanced clarity and visual-

Figure 6: Algorithm 1 Results.



Note. (a) Successive ICER results for a selection of iterations. (b) The cost-effectiveness frontier. (c) Menu of threshold policies for various WTP values. The threshold remains fixed within the WTP range of $50k-$100k.

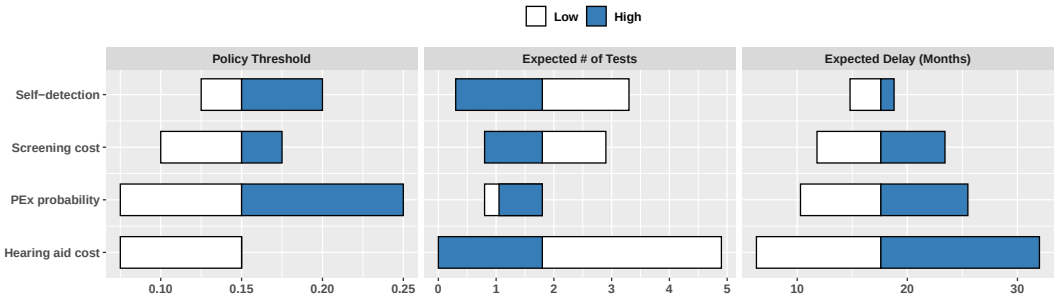
ization of the policy menu, we choose to present the threshold-based approximations outlined in Section 7.3.1. This strategic decision is in line with the insights discussed in Section 7.4, where it was observed that this category of easy-to-implement policies consistently exhibits robust performance and is positioned on the cost-effectiveness frontier. Panel (c) illustrates the threshold for the policy associated with various WTP values. Notably, the screening threshold diminishes as WTP increases, indicating a more intensive screening approach. This trend is expected, as increased screening frequency enhances patient outcomes, including QALE, which holds greater value at higher WTP levels.

7.6. Sensitivity Analysis

We conduct a series of scenario analyses to evaluate the sensitivity of the optimal policy and patient outcomes to variations in model parameters. In summary, our findings indicate that the model outcomes are most sensitive to changes in the PEx and self-detection probabilities, as well as the costs associated with screening and hearing aid. The sensitivity analysis results are depicted in a tornado diagram in Figure 7. We illustrate the impact of model inputs on the policy by analyzing both the threshold value of the corresponding threshold-based policy and the average number of screening tests performed. A lower threshold typically indicates more frequent screening. Additionally, we include the average detection delay as a representative patient outcome in our reporting.

- **Screening and hearing aid costs**: We examine the impact of a one-way, ±25% change in costs. We expect a less aggressive screening policy, leading to worse patient outcomes, as the cost of screening and medical device increase. As anticipated, we observe a less aggressive

Figure 7: Scenario analysis results.



Note. The tornado plot shows the changes in the optimal policy and its performance under alternative scenarios for input parameters. The bar charts are drawn relative to the base-case values. The threshold for the scenario involving an increase in hearing aid cost is 80%, and it has been omitted to maintain the plot scale. For costs, we consider a deviation of ±25% from the base-case values, while for PEx probability, we consider a deviation of ±37% as high/low scenarios. For self-detection, we consider scenarios of ±100%, representing doubling the chance and removing it altogether.

screening policy and consequently poorer patient outcomes with increasing costs of screening and medical devices. An increase in cost of the hearing aid device diminishes the cost-effectiveness and the appeal of both the device and screening, especially when the device's cost approaches the threshold of being not cost-effective. In the scenario with increased hearing aid cost, the policy effectively ceases screening any patients, leading to an expected detection delay exceeding two and a half years. The results indicate that the model outcomes are more sensitive to changes in the cost of the hearing aid compared to the cost of screening.

- **Self-detection probability**: Self-detection serves as a substitute for screening and can help decrease the detection delay associated with any given policy. We explore scenarios involving doubling the chance of self-detection and completely removing it as the low and high scenarios, respectively. Note that screening, whether through self-detection or actual screening, exhibits diminishing returns. Therefore, increased self-detection frequency reduces the value and cost-effectiveness of screening. We observe that decreasing (increasing) the probability of self-detection increases (decreases) the average number of screenings done, aligning with the concept of diminishing returns and the substitutability of the two modes of screening. In terms of detection delay, the net impact of self-detection frequency and the number of screenings leads to an increase (decrease) in detection delay as the self-detection probability increases (decreases).
- **Quarterly PEx probability**: With advances in medicine, new classes of medications may potentially reduce the frequency of PEx's. For example, recent combination therapy has been shown to reduce the probability of PEx by 37% for a specific genotype of CF patients (Tice

et al. 2020). Conversely, environmental factors and patient characteristics may contribute to higher rates of infection and PEx. We explore a change of ±37% in the probability of PEx in our scenario analysis. Note that the novel combination therapy and its clinical benefits are applicable only to a subset of CF patients.

In the scenario where the probability of PEx is reduced, the life expectancy and the average age of reaching severe hearing loss increase to 49 and 35 years. Patients progress through hearing loss states at a slower pace, resulting in only 18% of patients ever reaching the severe hearing loss state. However, for those who do reach it, they stay in it for longer (14 vs. 9 years in the base case). Since a hearing aid has a one-time setup cost, the extended interval of usage makes it more cost-effective. Therefore, we anticipate a more aggressive screening strategy, as evidenced by the decreased screening threshold of the optimal policy. Despite the policy being more aggressive, the lower frequency of PEx events results in fewer tests on average. However, the policy conducts more frequent tests at each PEx event (12% vs. 8% in the base case), resulting in an improved detection delay.

Conversely, in the scenario with an increased chance of PEx, the life expectancy and the average age of reaching severe hearing loss decrease to 42 and 22 years, respectively. Approximately 75% of patients reach the severe hearing loss state. The patient's duration of severe hearing loss is shorter (6.5 years vs. 9 years in the base case), leading to reduced cost-effectiveness of the hearing aid. As a result, we anticipate less aggressive screening, as confirmed by the decreased threshold of the optimal policy. The policy conducts fewer tests at each PEx event (3.5% vs. 8% in the base case), resulting in worsened detection delay.

8. Discussion and Conclusion

In cost-effectiveness analysis, maximizing NMB is a common approach for designing interventions. The UK's NICE and the US's Institution for Clinical Economic Review (ICER) rely heavily on NMB to assess the cost-effectiveness of treatments (Institute for Clinical and Economic Review 2020, National Institute for Health and Care Excellence 2023). However, some studies take budget constraints into account when designing cost-effective interventions, as seen in the optimization of breast cancer screening by Ayvaci et al. (2012). This approach is less common in the cost-effectiveness literature because, in practice, budgets are not usually allocated to individual interventions. Instead, they are fluid and subject to constant negotiation and periodic revision (Drummond et al. 2015). Given the uncertainty surrounding budgets, scenario analysis at the budget level is typically conducted in these studies.

CEA utilizes the Cost-Effectiveness Threshold (CET) as a crucial input parameter to quantify the monetary value of intervention benefits. The methodology for establishing the CET can vary

widely and remains a topic of considerable debate within health economics, as discussed in McCabe et al. (2008). One alternative, as seen in models aimed at maximizing NMB, is to determine the CET based on society's willingness to pay. Another approach involves using shadow prices in budget-constrained problems (Epstein et al. 2007). However, the latter approach encounters resistance within the health economics community due to its practical limitations. Additionally, CET is influenced by factors extending beyond mere budgetary constraints such as the population's mortality rate and the broader economic context of a society, as emphasized by Glassman et al. (2017). This complexity underscores the challenge of establishing a straightforward relationship between the CET and budget constraints.

It is customary for both the UK's NICE and the US's ICER to consider a range of CET, as described in Rawlins and Culyer (2004). As a result, constructing a cost-effectiveness frontier that encompasses the performance of cost-effective policies across a spectrum of CET values becomes essential. The cost-effectiveness frontier is a vital tool for evaluating and comparing healthcare interventions or policies. It helps policymakers understand the trade-offs between costs and health outcomes, ensuring they consider a wide range of effective strategies. This facilitates more informed resource allocation decisions in healthcare. Although the frontier may include numerous policies, the number of policies that prove to be cost-effective for a reasonable range of CET values, particularly those close to the base-case value, is likely to be small and manageable. In our paper, we introduce a novel algorithm for the efficient construction of the cost-effectiveness frontier. The algorithm operates by iteratively solving the ICER minimization problem and updating the comparator policy in each iteration.

It is common in the literature to conduct scenario analysis and solve the NMB problem with varying values of WTP as an alternative to the developed algorithm. While the computation times of both methods are comparable, our developed algorithm efficiently discovers a minimal set of WTPs that fully characterize the cost-effectiveness frontier. It also provides simple instructions to select a cost-effective policy from the set of generated policies for any desired WTP. Conducting scenario analysis without a carefully selected set of WTP values fails to provide any indication of the optimal policy associated with WTP values outside the considered set. Moreover, some of the generated policies may be redundant since the NMB maximizing policies are valid for a range of WTP values.

Drummond et al. (2015) and Neumann et al. (2016) underscore the strong appeal of non-dominated policies to policymakers, while cautioning against all forms of dominance, including weak and extended dominance. The former can be strictly improved in cost or QALE, while the latter are strongly dominated in both cost and QALE by a combination of non-dominated policies.

Following the recommendation of the domain experts, we have chosen to exclude both weakly and extended dominated policies from consideration.

The efficient frontier, a widely used concept in different disciplines, may utilize the weighted sum, epsilon-constraint, or other criteria to define efficiency. In the context of cost-effectiveness analysis, the epsilon-constraint method translates to formulating either a budget-constrained QALE maximization problem or a QALE-constrained cost minimization problem. When exclusively focusing on deterministic policies in MDP and POMDP problems, epsilon-constraint and weighted sum approaches yield different frontiers. However, when mixed strategies are allowed, the two concepts become equivalent. The utilization of weighted sums simplifies the solution of multi-objective problems and enhances their tractability, as noted by (Chiandussi et al. 2012).

Although we have emphasized the significance of the cost-effectiveness frontier, it is essential to recognize a potential limitation: If the policymaker's primary goal is to maximize QALE within a budget constraint, policies focused solely on maximizing NMB may be suboptimal, particularly when mixed strategies are not desirable and therefore excluded. Combining strategies can address this limitation. However, it is worth noting that this issue is less significant when the cost-effectiveness frontier is dense, as is often observed in MDP/POMDP problems with long horizons or large state spaces. Furthermore, if the policymaker's goal is to maximize the NMB of an intervention, they can choose policies along the cost-effectiveness frontier to optimize NMB while either minimizing costs or maximizing QALE.

The algorithm developed in this study demonstrates potential for broader applications beyond the specific research context. It can be extended for identifying efficient frontiers in various bi-objective MDP and POMDP problems. For instance, in the context of optimizing patient outcomes, the algorithm can facilitate the trade-off between treatment benefits and side-effects. Alternatively, when focusing on balancing patient and system outcomes, it can be utilized to balance between treatment benefits and the utilization of scare healthcare resources. One immediate application might involve balancing the early detection of cancer as the intended outcome of the screening intervention, alongside either the utilization or costs of imaging services, or the adverse effects of radiation from imaging. While the literature of multi-objective MDP problems is extensive, it is noteworthy that existing algorithms frequently prove inadequate for MDP or POMDP problems with infinite or long finite horizon lengths. Our proposed algorithm effectively bridges this gap.

Since screening methods within our clinical case may produce inaccurate results, we have introduced POMDP and CPOMDP models to address the NMB maximization and ICER minimization problems, respectively. In general, the optimal policy for CPOMDPs may be non-deterministic. To ensure the model generates deterministic policies, we may employ integer programming techniques.

We established conditions under which CPOMDPs with non-linear objective functions admit deterministic policies, eliminating the necessity for integer programming. Furthermore, we demonstrate that this finding extends to the specific context of the ICER minimization problem. To numerically solve our C/POMDP problems, we employed grid-based approximation. This method enables the generation of both lower and upper bounds within linear CPOMDPs, facilitating the calculation of the optimality gap and assessment of the accuracy of the approximation. We extend the applicability of this result to CPOMDPs with non-linear objective functions and general feasibility sets.

In our numerical study, we emphasize the use of highly interpretable threshold policies, which exhibit robust performance. This approach allows us to showcase a variety of threshold policies along with the corresponding WTP ranges, effectively addressing concerns related to interpretability. When easily implementable policies are not readily available, leveraging visualization software such as Tableau, Shiny, or similar tools can be valuable. By creating a dashboard populated with data generated by the developed algorithm, we can offer insights into cost-effective policies and their performance. Incorporating a visual user interface can greatly enhance the accessibility of these findings for policymakers and clinicians alike.

Our study has limitations. Due to the unavailability of a comprehensive set of patient trajectories, our clinical model relied on parameters extracted from the literature. However, we validated our model by leveraging evidence from the literature and consulting with our clinical collaborator. With access to patient data, we can enhance the clinical model by incorporating factors such as the potential seasonality of the PEx's. There are antibacterial treatments that are less effective yet have a milder impact on hearing compared to intravenous AGs. The inclusion of these treatments in the optimization model and developing a joint treatment/screening model is a promising avenue for future research. We employed regular grids and Freudenthal triangulation to solve the developed POMDPs. However, we did not compare the efficiency and performance of the proposed algorithm against alternative interpolation and grid generation methods, leaving this for future research.

References

Abrams HB, Chisolm TH, McArdle R (2005) Health-related quality of life and hearing aids: A tutorial. *Trends in Amplification* 9(3):99–109.

Altman E (1999) *Constrained Markov Decision Processes: Stochastic Modeling* (Routledge).

American Speech-Language-Hearing Association (1994) Audiologic management of individuals receiving cochleotoxic drug therapy. Technical report.

American Speech Language Hearing Association (2019) Hospital outpatient prospective payment system for audiologists and speech language pathologists. URL https://www.asha.org/siteassets/uploadedFiles/

2019-Hospital-Outpatient-Prospective-Payment-System-for-Audiologists-and-SLPs.pdf, accessed November 3, 2021.

Ayer T, Alagoz O, Stout NK, Burnside ES (2016) Heterogeneity in women's adherence and its role in optimal breast cancer screening policies. *Management Science* 62(5):1339–1362.

Ayvaci MU, Alagoz O, Burnside ES (2012) The effect of budgetary restrictions on breast cancer diagnostic decisions. *Manufacturing & Service Operations Management* 14(4):600–617.

Barton GR, Bankart J, Davis AC, Summerfield QA (2004) Comparing utility scores before and after hearing-aid provision. *Applied health economics and health policy* 3(2):103–105.

Bazaraa MS, Sherali HD, Shetty CM (2013) *Nonlinear programming: theory and algorithms* (John Wiley & Sons).

Bereanu B (1974) On the global minimum of a quasi-concave functional. *Archiv der Mathematik* 25(1):391–393.

Cevik M, Ayer T, Alagoz O, Sprague BL (2018) Analysis of mammography screening policies under resource constraints. *Production and Operations Management* 27(5):949–972.

Charnes A, Cooper W (1973) An explicit general solution in linear fractional programming. *Naval Research Logistics Quarterly* 20(3):449–467.

Chen Q, Ayer T, Chhatwal J (2018) Optimal M-switch surveillance policies for liver cancer in a hepatitis C–infected population. *Operations Research* 66(3):673–696.

Chiandussi G, Codegone M, Ferrero S, Varesio FE (2012) Comparison of multi-objective optimization methodologies for engineering applications. *Computers & Mathematics with Applications* 63(5):912–942.

Choi N, Pietrangelo A (2019) Cystic Fibrosis by the Numbers: Facts, Statistics, and You. https://www.healthline.com/health/cystic-fibrosis-facts.

Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW (2015) *Methods for the Economic Evaluation of Health Care Programmes* (Oxford University Press).

Durrant J, Campbell K, Fausti S, Guthrie O, Jacobson G, Lonsbury-Martin B, Poling G (2009) American academy of audiology position statement and clinical practice guidelines: ototoxicity monitoring. *Wahington: American Academiy of Audiology* .

Epstein DM, Chalabi Z, Claxton K, Sculpher M (2007) Efficiency, equity, and budgetary policies: informing decisions using mathematical programming. *Medical Decision Making* 27(2):128–137.

Farzal Z, Kou YF, St John R, Shah GB, Mitchell RB (2016) The role of routine hearing screening in children with cystic fibrosis on aminoglycosides: A systematic review. *The Laryngoscope* 126(1):228–235.

GAMS Development Corporation (2021) *General Algebraic Modeling System (GAMS) Release 36.1.0*. Fairfax, VA,USA, URL https://www.gams.com/download/.

Gan K, Scheller-Wolf AA, Tayur SR (2019) Personalized treatment for opioid use disorder. *Available at SSRN 3389539* .

Garinis A, Gleser M, Johns A, Larsen E, Vachhani J (2021) Prospective cohort study of ototoxicity in persons with cystic fibrosis following a single course of intravenous tobramycin. *Journal of Cystic Fibrosis* 20(2):278–283.

Garinis AC, Cross CP, Srikanth P, Carroll K, Feeney MP, Keefe DH, Hunter LL, Putterman DB, Cohen DM, Gold JA, et al. (2017) The cumulative effects of intravenous antibiotic treatments on hearing in patients with cystic fibrosis. *Journal of Cystic Fibrosis* 16(3):401–409.

Gillard DM, Harris JP (2020) Cost-effectiveness of stapedectomy vs hearing aids in the treatment of otosclerosis. *JAMA Otolaryngology–Head & Neck Surgery* 146(1):42–48.

Glasserman P, Yao DD (1992) Some guidelines and guarantees for common random numbers. *Management Science* 38(6):884–908.

Glassman A, Giedion U, Smith PC (2017) *What's in, what's out: designing benefits for universal health coverage* (Brookings Institution Press).

Gleser MA, Zettner EM (2018) Negative hearing effects of a single course of IV aminoglycoside therapy in cystic fibrosis patients. *International Journal of Audiology* 57(12):923–930.

Hauskrecht M (2000) Value-function approximations for partially observable Markov decision processes. *Journal of Artificial Intelligence Research* 13:33–94.

Helmeczi RK, Kavaklioglu C, Cevik M (2023a) Linear programming-based solution methods for constrained partially observable Markov decision processes. *Applied Intelligence* 1–27.

Helmeczi RK, Kavaklioglu C, Cevik M, Pirayesh Neghab D (2023b) A multi-objective constrained partially observable Markov decision process model for breast cancer screening. *Operational Research* 23(2):30.

Holder A (2010) Parametric LP Analysis. *Wiley Encyclopedia of Operations Research and Management Science* .

Huang SP, McKinzie CJ, Tak CR (2021) Cost-effectiveness of implementing routine hearing screening using a tablet audiometer for pediatric cystic fibrosis patients receiving high-dose IV aminoglycosides. *Journal of Managed Care & Specialty Pharmacy* 27(2):157–165.

Institute for Clinical and Economic Review (2020) A guide to ICER's methods for health technology assessment. https://icer.org/wp-content/uploads/2021/01/ICER_HTA_Guide_102720.pdf, accessed: 2024-02-16.

Kavaklioglu C, Cevik M (2022) Scalable grid-based approximation algorithms for partially observable Markov decision processes. *Concurrency and Computation: Practice and Experience* 34(5):e6743.

Keogh RH, Szczesniak R, Taylor-Robinson D, Bilton D (2018) Up-to-date and projected estimates of survival for people with cystic fibrosis using baseline characteristics: A longitudinal study using UK patient registry data. *Journal of Cystic Fibrosis* 17(2):218–227.

Laske RD, Dreyfuss M, Stulman A, Veraguth D, Huber AM, Röösli C (2019) Age dependent cost-effectiveness of cochlear implantation in adults. Is there an age related cut-off? *Otology & Neurotology* 40(7):892–899.

Li W, Denton BT, Morgan TM (2023) Optimizing active surveillance for prostate cancer using partially observable Markov decision processes. *European Journal of Operational Research* 305(1):386–399.

Liu Z, Khojandi A, Li X, Mohammed A, Davis RL, Kamaleswaran R (2022) A machine learning–enabled partially observable Markov decision process framework for early sepsis prediction. *INFORMS Journal on Computing* .

Lovejoy WS (1991) Computationally feasible bounds for partially observed Markov decision processes. *Operations Research* 39(1):162–175.

Lusena C, Goldsmith J, Mundhenk M (2001) Nonapproximability results for partially observable Markov decision processes. *Journal of artificial intelligence research* 14:83–103.

Macones GA, Goldie SJ, Peipert JF (1999) Cost-effectiveness analysis: an introductory guide for clinicians. *Obstetrical & gynecological survey* 54(10):663–672.

Mason JE, Denton BT, Shah ND, Smith SA (2014) Optimizing the simultaneous management of blood pressure and cholesterol for type 2 diabetes patients. *European Journal of Operational Research* 233(3):727–738.

McCabe C, Claxton K, Culyer AJ (2008) The NICE cost-effectiveness threshold. *Pharmacoeconomics* 26(9):733–744.

McMahon CM, Gopinath B, Schneider J, Reath J, Hickson L, Leeder SR, Mitchell P, Cowan R (2013) The need for improved detection and management of adult-onset hearing loss in Australia. *International Journal of Otolaryngology* 2013.

Michael SY, Britto MT, Wilmott RW, Kotagal UR, Eckman MH, Nielson DW, Kociela VL, Tsevat J (2003) Health values of adolescents with cystic fibrosis. *The Journal of pediatrics* 142(2):133–140.

National Institute for Health and Care Excellence (2023) NICE health technology evaluations: the manual. https://www.nice.org.uk/process/pmg36/resources/nice-health-technology-evaluations-the-manual-pdf-72286779244741, accessed: 2024-02-21.

Nazareth D, Walshaw M (2013) Coming of age in cystic fibrosis–transition from paediatric to adult care. *Clinical Medicine* 13(5):482.

Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG (2016) *Cost-effectiveness in health and medicine* (Oxford University Press).

Neumann PJ, Weinstein MC, et al. (2010) Legislating against use of cost-effectiveness information. *New England Journal of Medicine* 363(16):1495–1497.

O'Brien M, Murphy D, Plant B (2014) A 76 year old female diagnosed with cystic fibrosis. *Irish Medical Journal* 107(8):240–241.

Ong T, Ramsey BW (2015) Update in cystic fibrosis 2014. *American Journal of Respiratory and Critical Care Medicine* 192(6):669–675.

Piri H, Huh WT, Shechter SM, Hudson D (2022) Individualized dynamic patient monitoring under alarm fatigue. *Operations Research* 70(5):2749–2766.

Pistikopoulos EN, Diangelakis NA, Oberdieck R (2020) *Multi-parametric Optimization and Control* (John Wiley & Sons).

Poupart P, Malhotra A, Pei P, Kim KE, Goh B, Bowling M (2015) Approximate linear programming for constrained partially observable Markov decision processes. *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29.

Prayle A, Smyth AR (2010) Aminoglycoside use in cystic fibrosis: Therapeutic strategies and toxicity. *Current Opinion in Pulmonary Medicine* 16(6):604–610.

Puterman ML (2014) *Markov decision processes: discrete stochastic dynamic programming* (John Wiley & Sons).

R Core Team (2022) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.

Ratjen F, Bell SC, Rowe SM, Goss CH, Quittner AL, Bush A (2015) Cystic fibrosis. *Nature Reviews Disease Primers* 1(1):1–19.

Rawlins MD, Culyer AJ (2004) National Institute for Clinical Excellence and its value judgments. *BMJ* 329(7459):224–227.

Roijers DM, Vamplew P, Whiteson S, Dazeley R (2013) A survey of multi-objective sequential decision-making. *Journal of Artificial Intelligence Research* 48:67–113.

Ross SM (2014) *Introduction to stochastic dynamic programming* (Academic press).

Saliba J, Al-Reefi M, Carriere JS, Verma N, Provencal C, Rappaport JM (2017) Accuracy of mobile-based audiometry in the evaluation of hearing loss in quiet and noisy environments. *Otolaryngology–Head and Neck Surgery* 156(4):706–711.

Sanders DB, Bittner RC, Rosenfeld M, Redding GJ, Goss CH (2011) Pulmonary exacerbations are associated with subsequent FEV_1 decline in both adults and children with cystic fibrosis. *Pediatric Pulmonology* 46(4):393–400.

Solem CT, Vera-Llonch M, Liu S, Botteman M, Castiglione B (2016) Impact of pulmonary exacerbations and lung function on generic health-related quality of life in patients with cystic fibrosis. *Health and quality of life outcomes* 14(1):1–9.

Stephenson AL, Tom M, Berthiaume Y, Singer LG, Aaron SD, Whitmore G, Stanojevic S (2015) A contemporary survival analysis of individuals with cystic fibrosis: a cohort study. *European Respiratory Journal* 45(3):670–679.

Suen Sc, Goldhaber-Fiebert JD (2016) An efficient, noniterative method of identifying the cost-effectiveness frontier. *Medical Decision Making* 36(1):132–136.

Tice JA, Kuntz KM, Wherry K, Chapman R, Seidner M, Pearson SD, Rind DM (2020) Modulator treatments for cystic fibrosis: effectiveness and value. *Institute for Clinical and Economic Review* .

United States Bureau of Labor Statistics (2019) Occupational outlook handbook. URL https://www.bls.gov/ooh/healthcare/pharmacists.htm, accessed November 3, 2021.

United States Government (2022a) CPI Inflation Calculator. https://www.bls.gov/data/inflation_calculator.htm, accessed: 2022-12-12.

United States Government (2022b) Foreign Currency and Currency Exchange Rates. https://www.irs.gov/individuals/international-taxpayers/foreign-currency-and-currency-exchange-rates, accessed: 2022-12-02.

Vijayasingam A, Frost E, Wilkins J, Gillen L, Premachandra P, Mclaren K, Gilmartin D, Picinali L, Vidal-Diez A, Borsci S, et al. (2020) Tablet and web-based audiometry to screen for hearing loss in adults with cystic fibrosis. *Thorax* .

Wagner HM (1960) On the optimality of pure strategies. *Management Science* 6(3):268–269, ISSN 00251909, 15265501.

Waters V, Stanojevic S, Klingel M, Chiang J, Sonneveld N, Kukkar R, Tullis E, Ratjen F (2015) Prolongation of antibiotic treatment for cystic fibrosis pulmonary exacerbations. *Journal of Cystic Fibrosis* 14(6):770–776.

Weisstein EW (2021) Ternary Diagram. https://mathworld.wolfram.com/TernaryDiagram.html, accessed: 2022-12-12.

Westerman EM, Le Brun PP, Touw DJ, Frijlink HW, Heijerman HG (2004) Effect of nebulized colistin sulphate and colistin sulphomethate on lung function in patients with cystic fibrosis: A pilot study. *Journal of Cystic Fibrosis* 3(1):23–28.

Yeung J, Javidnia H, Heley S, Beauregard Y, Champagne S, Bromwich M (2013) The new age of play audiometry: Prospective validation testing of an iPad-based play audiometer. *Journal of Otolaryngology-Head & Neck Surgery* 42(1):1–7.

Yeung JC, Heley S, Beauregard Y, Champagne S, Bromwich MA (2015) Self-administered hearing loss screening using an interactive, tablet play audiometer with ear bud headphones. *International journal of pediatric otorhinolaryngology* 79(8):1248–1252.

Zhang J, Denton BT, Balasubramanian H, Shah ND, Inman BA (2012) Optimization of prostate biopsy referral decisions. *Manufacturing & Service Operations Management* 14(4):529–547.

Appendix A Table of notations

Table 3: Symbols used in the paper.

Symbol	Description
λ	Willingness to pay
γ	Discount factor
π	Policy
π_0	Comparator policy
q_0	QALE of comparator policy π_0
c_0	Cost of comparator policy π_0
X	Set of all acceptable policies
Y	Set of all occupancy measures
Ψ	Set of rewards for all feasible policies
χ	The feasibility set of rewards
Π^{HR}	Set of random, history-dependent policies
Π^{MR}	Set of random, history-independent policies
Functions	
$\text{nmb}(q,c)$	NMB function defined as $\lambda q - c$
$r(c,q)$	ICER function defined as $\frac{c-c_0}{q-q_0}$
Horizon	
t	Decision epoch
T	Terminal horizon
State	
h_n	Normal hearing state
h_m	Moderate hearing loss state
h_s	Severe hearing loss state
$\bar{S}$	Set of partially observable states, $\{h_n, h_m, h_s\}$
h_p	Profound hearing loss state
h_a	Hearing aid received state
D	Death state
$\hat{S}$	Set of terminal states, $\{h_p, h_a, D\}$
Belief states	
b	belief state, probability distribution over hidden states $\{h_n, h_m, h_s\}$
$b(s)$	Probability of being at hidden state s
B	Belief simplex, set of all beliefs
$\tau_i[b,o]$	step i posterior update for prior belief b after observing o, $i=1,2$
Action	
w	Wait action
m	Screen action
A	Set of actions, $\{w,m\}$
Observation	
O_w	Observation set for action w
O_m	Observation set for action m
e_0	Absence of PEx
e_1	Presence of PEx
e	PEx status, $\{0,1\}$

Probabilities	
$k(o\|s)$	Probability of observing o when true state is s
$k(o\|b)$	Probability of observing o under belief b
$\mathbf{P}^e(s'\|s)$	Transition probability between core states s and s', conditional on PEx status e
$\mathbf{P}^e(b'\|b,a)$	Transition probability under action a between belief states b and b', conditional on PEx status e
$\mathbf{P}^e(h_p\|b)$	Transition probability to terminal state h_p under belief b
$\bar{\mathbf{P}}^e(h_p\|b)$	Complement probability of $\mathbf{P}^e(h_p\|b)$, i.e., $1-\mathbf{P}^e(h_p\|b)$
d_t^e	Death probability at time t and PEx status e
$\bar{d}_t^e$	Survival probability, $1-d_t^e$
$p(e_i)$	Probability of absence/presence of PEx, $i=0,1$
$\delta(b)$	Probability the initial belief being b
$y_t(b,a)$	Discounted state-action occupancy measure for belief b and action a
$y_t(s)$	Discounted occupancy measure of terminal state s
$\pi^*(b,a)$	Chance of taking action a for belief b under optimal policy
Rewards	
$u_t^e(s)$	Per-period health utility for PEx status e and state s
$c_t(s,a)$	Expected immediate cost of action a under PEx status e and state s
c_m	Cost of mobile audiometry
c_f	Cost of formal audiometry
$R_t^u(s)$	Terminal health utility for absorbing state s
$R_t^c(s)$	Terminal cost for absorbing state s
$q_t^e(b)$	Expected per-period health utility for PEx status e and belief b
$c_t^e(b,a)$	Expected immediate cost for action a under PEx status e, belief b
$r_t^e(b,a)$	Expected per-period NMB reward
$\bar{r}_i$	Expected discounted total for reward i
$\mathcal{R}(\pi)$	Reward vector for policy π
Optimality equation	
$v_t(b)$	Value function for belief b at period t
$v_t^e(b)$	Value function for belief b at period t conditional on PEx status e
$z_t^e(b)$	Value function for belief b at period t conditional staying in the hidden states and under PEx status e
Grid-based approximation	
G	Set of all grid points
b_j^G	Belief state j in the grid set G
b_j^{ao}	Posterior belief corresponding to the initial belief b_j^G, action a, and observation o
θ_{ji}^{ao}	Weight of the grid point b_i^G corresponding to the posterior b_j^{ao}
$\check{\mathcal{R}}(\pi)$	Reward vector for policy π in the approximation problem

Appendix B Proofs

B.1 Solution to WTP Ambiguity

Proof of Lemma 1: We prove the result in a few steps. We first establish the following relationship between the two concepts of cost-effectiveness; the graph of function c^*, the QALE-constrained cost-minimization function, is equal to the efficiency frontier for the NMB maximization function. To that end, let L'' be the graph of c^*. We show $L = L''$ as follows.

- $L \subset L''$. Fix $(q,c) \in L$ and λ, the associated WTP, arbitrarily. Since (q,c) is on the frontier, for any feasible reward vector $(q',c') \in \Psi$, we have $\lambda q - c \geq \lambda q' - c'$. Therefore, for $(q, c^*(q)) \in \Psi$, we have: $\lambda q - c \geq \lambda q - c^*(q)$, which implies $c^*(q) \geq c$. Since (q,c) is a feasible point in Problem 2, we have $c \geq c^*(q)$. Therefore, $c = c^*(q)$, and hence $(q,c) \in L''$.
- $L'' \subset L$. Fix $(q,c) \in L''$, i.e., $c = c^*(q)$ and $q \in [q_{\min}, q_{\max}]$. It suffices to prove that for a certain λ, there exists an NMB maximizer with QALE q, i.e., for some cost c', $(q,c') \in L$. To show the sufficiency, assume this is correct, and therefore, for any $(q'',c'') \in \Psi$ we have $\lambda q - c' \geq \lambda q'' - c''$. Since $(q,c) \in \Psi$, we have $\lambda q - c' \geq \lambda q - c$, which implies $c' \leq c$. Since c is the cost-minimizer for QALE q, we have $c \leq c'$. Therefore, $c = c'$, and therefore $(q,c) = (q,c') \in L$.

  We now show the existence of one such λ by noting the continuity of the frontier. It is known that the optimal solution to parametric LP problems is continuous (e.g., see Theorem 3.1 in Pistikopoulos et al. (2020)). Therefore, $x(\lambda)$, the optimal solution to the NMB optimization problem, is continuous in λ, which implies that the optimal QALE and cost parametrized by λ, $q = ax(\lambda)$ and $c = bx(\lambda)$, are continuous in λ. Note that $q_{\min}$ and $q_{\max}$ are achieved in NMB optimization by setting $\lambda = 0$ and $\lambda \to \infty$. The existence then follows the continuity of q, the optimal QALE, and that $q \in [q_{\min}, q_{\max}]$.

Recall that $\phi^{-1}(Z)$ for any arbitrary set Z contains all feasible policies whose performance (q,c) belong to Z. Therefore, once $L(\lambda) \subset L$, the section of the optimal frontier associated with λ, has been constructed, we can characterize $V(\lambda)$, the set of all NMB maximizers for a specific WTP λ, by noting $V(\lambda) = \phi^{-1}(L(\lambda))$. We will show that $L(\lambda)$ is either a single breakpoint of L'', or a line segment connecting two successive breakpoints. Therefore, $V(\lambda)$ can be characterized by determining the (1 or 2) breakpoints of c^*'s map that optimize the NMB.

Note that $c^*(z)$ is continuous, convex, and piece-wise affine (e.g., see Theorem 2.1 in Pistikopoulos et al. (2020)). Therefore, the optimal frontier L can be constructed by successively connecting the breakpoints of $c^*(z)$ (a fact we use to develop and prove the correctness of Algorithm 1). Since NMB is affine in q and c, and the frontier is continuous and piece-wise affine, $L(\lambda)$ can be constructed by successively connecting the breakpoints of L that maximize the NMB. Below, we further demonstrate that NMB is maximized at a maximum of two breakpoints.

For any arbitrary λ, let $\text{nmb}_i = \lambda q_i - c_i$ for any $(q_i, c_i) \in L$. We now show how λ_i's, the slope between two consecutive breakpoints of $c^*(z)$ defined by $\lambda_i := \frac{c_{i+1}-c_i}{q_{i+1}-q_i}$, can determine which breakpoints of L is associated with NMB maximizers for an arbitrary λ. Noting that $\lambda \geq 0$ and using simple algebra, we readily obtain the following.

- For $\lambda = \lambda_j$, we have $\text{nmb}_j = \text{nmb}_{j+1}$.
- For any j with $\lambda > \lambda_j$, we have $\text{nmb}_{j+1} > \text{nmb}_j$.
- For any j with $\lambda < \lambda_j$, we have $\text{nmb}_{j+1} < \text{nmb}_j$.

We have $c^*(z)$ is increasing in z since the feasible set in Problem 2 shrinks as z increases. As a result, λ_i, slopes of segments of $c^*(z)$, strictly increases with i and $\lambda_i \geq 0$. Combining this observation and identities above we have the following.

- $\lambda_i < \lambda < \lambda_{i+1}$: nmb_j strictly increases for $j \leq i+1$ and then strictly decreases for $j \geq i+1$.
- $\lambda = \lambda_i$: nmb_j strictly increases for $j \leq i$, stays constant for $j = i, i+1$ and then strictly decreases for $j \geq i+1$.

We now show which breakpoint(s) of c^*'s map maximize the NMB, for each of the following conditions.

- $\lambda_i < \lambda < \lambda_{i+1}$: Based on the first observation, we have $\text{nmb}_{i+1} > \text{nmb}_j$ for all $j \neq i+1$.
- $\lambda < \lambda_1$: Similarly, $\text{nmb}_1 > \text{nmb}_j$ for all $j \neq 1$.
- $\lambda > \lambda_{m-1}$: Similarly, $\text{nmb}_m > \text{nmb}_i$ for all $i \neq m$.
- $\lambda = \lambda_i$: Based on the second observation, $\text{nmb}_i = \text{nmb}_{i+1} > \text{nmb}_j$ for any $j \notin \{i, i+1\}$. ■

Proof of Proposition 1: We can transform Problem 3 to the following equivalent problem.

$$\begin{aligned} r^*(u,v) = \inf_{q,c} \quad & (c-v)/(q-u) \\ \text{s.t.} \quad & q > u, \\ & (q,c) \in \Psi. \end{aligned} \tag{20}$$

We first show that without loss of optimality, we can restrict our optimization to points that satisfy $(q, c^*(q)) \in \Psi$. Since $v = c^*(u)$, by the definition of c^*, we have $c \geq v$ for any point $(q,c) \in \Psi$ with $q > u$. Therefore, the objective function has non-negative numerator and positive denominator at any feasible point. For any point (q,c) with $c > c^*(q)$, the point $(q, c^*(q))$ has a strictly better objective value since the numerator is strictly smaller in the latter point, and the denominator is identical at both points.

Consequently, we want to find the smallest slope between $(u, c^*(u))$ and $(q, c^*(q))$ for all $q > u$. As shown in the proof of Lemma 1, we know $c^*(q)$ is convex in q. As a result, the slope is increasing in q. Since the function is affine between q_{i-1} and q_i, the slope is constant in this region. As a result,

(q_i, c_i) is a minimizer for Problem 20, and consequently, $\phi^{-1}(q_i, c_i)$ optimizes Problem 3. Moreover, employing Lemma 1 part (d), $\phi^{-1}(q_i, c_i)$ also optimizes Problem 1. ■

Before proving Proposition 2, we show a relationship between the extreme points of the set resulting from a linear map applied to a set and the original set. We leverage this result to further prove that linear maps do not increase the number of extreme points when applied to a polytope.

LEMMA 3. Let g be a linear map from $\Re^n \to \Re^m$. Let $X \subset \Re^n$ be an arbitrary polytope and $\Psi = g(X)$. The extreme points of Ψ are a subset of the image of extreme points of X, and hence are (not necessarily strictly) fewer than the extreme points of X.

Proof. Let ψ be any arbitrary point in Ψ and pick $x \in g^{-1}(\psi)$ arbitrarily. Since X is a polytope, we can write $x \in X$ as a convex combination of its finite extreme points, i.e., there are some $\gamma_i \geq 0$ such that $\sum_i \gamma_i = 1$ and $x = \sum_i \gamma_i x_i$, where x_i's are extreme points of X. We have:

$$g(x) = g\big(\sum_i \gamma_i x_i\big) = \sum_i \gamma_i g(x_i),$$

where the second equality follows the linearity of g. As a result, any arbitrary point of Ψ can be written as a convex combination of $\psi_i = g(x_i)$. Therefore, the set of $g(x_i)$ for all x_i is a superset to the set of extreme points of Ψ. ■

Proof of Proposition 2:

- We will show that in each iteration of the algorithm, we produce a new breakpoint of c^*, and hence the algorithm runs in exactly $m-1$ iterations. Recall that m is the number of breakpoints of c^*. Note that the first breakpoint is produced in the initialization step. We show the claim by induction.

  Assume $d_i = (u, v)$ is a breakpoint of c^*. Note that the assumption holds by the algorithm's design for $i = 1$, the induction base case. In Proposition 1, we showed that $r(x^*; u, v) = \lambda_i$, and $\psi = \phi(x^*)$ must lie on L_i, the line segment between $(q_{i-1}, c^*(q_{i-1}))$ and $(q_i, c^*(q_i))$. By Lemma 1, for $\lambda = \lambda_i$, we have $V(\lambda) = \phi^{-1}(L_i)$. As a result, $x^* \in V(\lambda)$.

  Any feasible point in Problem 6 must satisfy $(\lambda_i^* a - b)x = (\lambda_i^* a - b)x^*$ (by re-arranging terms). Therefore, Problem 6 is equivalent to solving $\max_x ax$ subject to $x \in V(\lambda_i)$. Using change of variables $(q, c) = \phi(x)$, this problem transforms to $\max_{q,c} q$ subject to $(q, c) \in L_i$, which solves at $q = q_i$ and $c = c_i$. Therefore, Problem 6 is solved at $x_{i+1} \in \phi^{-1}(q_i, c_i)$, which proves our claim that d_{i+1} is a new breakpoint of c^*.
- We now prove that the breakpoints of the piece-wise affine function $c^*(z)$ are finite. This helps us to establish that the algorithm terminates in a finite number of iterations. Recall that Ψ

is a linear map of X, and set X is a polytope and has a finite number of extreme points. Therefore, by Lemma 3, Ψ has a finite number of extreme points. If we prove that points (q_i, c_i)'s are extreme points of Ψ, we immediately obtain $m < \infty$. Towards a contradiction, assume for some j, $\psi = (q_j, c_j)$ is a breakpoint but not an extreme point of Ψ, and hence, for some $\gamma \in (0,1)$ and $\psi_i = (u_i, v_i) \in \Psi$, $i = 1,2$, we have $\psi = \psi_1 + \gamma(\psi_2 - \psi_1)$. Let $c_i^* = c^*(u_i)$. We have $c_i^* \leq v_i$ since $(u_i, v_i) \in \Psi$. It suffices to show that $v_i = c_i^*$ for $i = 1,2$ since it contradicts the fact that ψ is a breakpoint of c^*. Assume $c_i^* < v_i$ for at least one $i \in \{1,2\}$. We have:

$$c_j \leq c_1^* + \gamma(c_2^* - c_1^*) < v_1 + \gamma(v_2 - v_1) = c_j,$$

where the first inequality follows the convexity of c^*, the second follows the assumption that $c_i^* \leq v_i$ for $i \in \{1,2\}$ and $c_i^* < v_i$ for at least one $i \in \{1,2\}$, and the equality follows $\psi = \psi_1 + \gamma(\psi_2 - \psi_1)$. Since we arrived at a contradiction ($c_j < c_j$), we must have $c_i^* = v_i$ for $i = 1,2$.

- Our goal now is to demonstrate that set W is the smallest set containing at least one solution to Problem 1 for any given $\lambda \geq 0$. Note that c^* is strictly increasing over $[q_{\min}, q_{\max}]$ and hence $\lambda_i > 0$ for all i. It follows Lemma 1 that $x_1 \in V(\lambda)$ for $\lambda \in I_1 = [0, \lambda_1]$ and $x_i \in V(\lambda)$ for $\lambda \in I_i = [\lambda_{i-1}, \lambda_i]$ for $i \geq 2$, and $x_m \in V(\lambda)$ for $\lambda \in I_m = [\lambda_{m-1}, \infty)$. As a result, W contains a solution to Problem 1 for any arbitrary $\lambda \geq 0$. Moreover, by the same lemma, $V(\alpha) \cap V(\beta) = \emptyset$ for any arbitrary α and β in the interior of I_i and I_j, for $i \neq j$. Therefore, $|V| \geq m$. Since $|W| = m$, we can conclude that it is the smallest set satisfying the requirements.

■

B.2 CPOMDP Result

Proof of Theorem 1: We prove the result by showing a series of intermediate results. We first show that we can restrict ourselves without loss of generality to history-independent (Markovian), random policies, here denoted by Π^{MR}. The result allows us to transform the problem into a tractable mathematical programming model.

LEMMA 4. We have $\mathcal{A} = \mathcal{R}(\Pi^{MR})$ and $\mathcal{B} = \hat{\mathcal{R}}(\hat{\Pi}^{MR})$.

Proof. Consider the belief state MDP. Let H_t be the history of the systems at time t, i.e., actions and states observed up until time t before making a decision in that period. The most general form of policies, history-dependent random policies, maps H_t to a distribution over permissible actions. Any such policy induces a joint probability distribution (known as state/action occupancy measure) $y_t(b,a)$ over states b and actions a for $t < N$ and a probability distribution (known as state occupancy measure) $y_N(b)$ over states b for the terminal period. Note that $\sum_a y_t(b,a)$ is equal

to the total probability of occupying state b at t. Let $\delta(b)$ be the probability distribution over the initial belief. We can show that y satisfies the following system of equations.

$$\begin{aligned}
&\sum_a y_1(b,a) = \delta(b), \quad \forall b \in B \\
&\sum_a y_t(b',a) = \sum_{a,b} p^a_{b,b'} y_{t-1}(b,a), \ \ \forall b' \in B, \ \ \forall t = 2,\dots,N-1, \\
&y_N(b') = \sum_{a,b} p^a_{b,b'} y_{N-1}(b,a), \quad \forall b' \in B \\
&y \geq 0.
\end{aligned}$$

where $p^a_{b,b'}$ is the transition probability from b to b' under action a.

We can conversely show that for any $y \geq 0$ satisfying these equations, there exists a history-independent (possibly random) policy $\pi_t(b,a)$ by letting $\pi_t(b,a) = y_t(b,a)/\sum_a y_t(b,a)$ for any b with $\sum_a y_t(b,a)) > 0$. Note that for states with $\sum_a y_t(b,a) = 0$, one can choose any arbitrary policy, since these states are never visited by the induced policy. It is easy to verify that this is a valid policy and that the associated occupancy measure coincides with y. Therefore, there is a many-to-one mapping from policies to the space of state/action and state occupancy measures. Using the law of total expectation, we can show that the expected cumulative reward $\bar{r}_i$ satisfies the following.

$$\bar{r}_i = \sum_{t<N} \gamma^t \sum_{b,a} y_t(b,a) r_i(b,a) + \gamma^N \sum_{b,a} y_N(b) R_i(b).$$

Therefore, for any history-dependent (random) policy, there is a history-independent (random) policy that performs exactly the same in terms of the expected cumulative reward. Hence, restricting to random, history-independent policies does not lead to loss of generality. This finding has frequently been utilized in the development of mathematical programming models for addressing C/MDP and C/POMDP problems (Ross 2014, Cevik et al. 2018).

We can then transform the search over policies to the space of occupancy measures, as follows.

$$\begin{aligned}
\inf_{y_t, y_N} \quad & f(\bar{r}_1,\dots,\bar{r}_m) \\
\text{s.t.} \quad & \sum_a y_1(b,a) = \delta(b), \quad \forall b \in B \\
& \sum_a y_t(b',a) = \sum_{a,b} p^a_{b,b'} y_{t-1}(b,a), \ \ \forall b' \in B, \ \ \forall t = 2,\dots,N-1, \\
& y_N(b') = \sum_{a,b} p^a_{b,b'} y_{N-1}(b,a), \quad \forall b' \in B, \\
& \bar{r}_i = \sum_{t<N} \gamma^t \sum_{b,a} y_t(b,a) r_i(b,a) + \gamma^N \sum_{b,a} y_N(b) R_i(b), \\
& (\bar{r}_1,\dots,\bar{r}_m) \in \chi.
\end{aligned} \tag{22}$$

Note that the same applies to the grid-based approximation problem. ■

We now prove that sets $\mathcal{A}$ and $\mathcal{B}$ are compact and convex. We use this in conjunction with results in convex analysis to prove the claim.

PROPOSITION 3. Sets $\mathcal{A}$ and $\mathcal{B}$ are compact and convex.

Proof.

We prove this result for both finite and infinite horizon models as follows:

- **Finite horizon:**

  Let Y be the set of occupancy measures y's satisfying Eq. 22. Note that Y is a polyhedron since it is generated by a system of linear equalities and inequalities. Therefore, Y is closed and convex. Also, note that y's are probability distributions, and hence $y \leq 1$. As a result, Y is bounded. Therefore, Y is compact (closed and bounded) and convex.

- **Infinite horizon:**

  A result similar to Lemma 4 applies to infinite horizon problems. Any arbitrary stationary random policy π generates discounted occupancy measures for each state/action pair, which represents the expected number of times we visit the state and take the action, discounted over time. Let $y(b,a)$ denote the discounted occupancy measure for the state/action pair (b,a), formally defined below.

$$y(b,a) = \mathbb{E}\left[\sum_t \gamma^{t-1} \mathbb{1}_{[b_t=b, a_t=a]} \middle| \pi\right].$$

  We can use $y(b,a)$ to calculate the expected total discounted rewards as follow.

$$\bar{r}_i = \mathbb{E}\left[\sum_t \gamma^{t-1} r_i(b_t,a_t) \middle| \pi\right] = \sum_{b,a} r_i(b,a) y(b,a) \tag{23}$$

  We can show that the occupancy measures satisfy the following system of linear equations.

$$\sum_a y(b',a) = \delta(b') + \gamma \sum_{b,a} p^a_{b,b'} y(b,a), \quad \forall b' \in B \qquad\qquad y \geq 0,$$

  where $\delta(b')$ is the distribution of the initial beliefs.

  Conversely, one can always construct a policy for any y satisfying the system by allowing $\pi(a'|b) = y(b,a')/\sum_a y(b,a)$ for any belief with $\sum_a y(b,a) > 0$. We utilize this one-to-many relationship to show our result.

  We need to show that Y is convex and compact. Note that Y satisfies a system of linear equations, and hence it is convex. Also, it is the intersection of hyper-planes and closed half-spaces. Hence, Y is closed as well. It remains to show Y is bounded. We can show $0 \leq y(b,a) \leq 1/(1-\gamma)$ as follows.

$$\begin{aligned} y(b,a) &= \mathbb{E}\left[\sum_t \gamma^{t-1} \mathbb{1}_{[b_t=b, a_t=a]} \middle| \pi\right] \\ &\leq \mathbb{E}\left[\sum_t \gamma^{t-1} \middle| \pi\right] = \sum_t \gamma^{t-1} = 1/(1-\gamma). \end{aligned}$$

Therefore, Y is compact as it is closed and bounded. ■

We now prove Theorem 1 as follows:

Proof. We show $\mathcal{A} \subset \mathcal{B}$. Toward contradiction, let $r' \in \mathcal{A} \setminus \mathcal{B}$. Since $r' \notin \mathcal{B}$ and set $\mathcal{B}$ is compact and convex by Proposition 3, we can properly separate r' from $\mathcal{B}$ by a hyperplane (Bazaraa et al. 2013). In other words, there exists a vector ω such that $\omega r' > \omega r$ for any $r \in \mathcal{B}$. This implies that $\omega r' > \max_{r\in\mathcal{B}} \omega r$. Additionally, we have $\max_{r\in\mathcal{A}} \omega r \geq \omega r'$. Combining these two inequalities, we conclude that $\max_{r\in\mathcal{A}} \omega r > \max_{r\in\mathcal{B}} \omega r$.

Note that the problem $\max_{r\in\mathcal{A}} \omega r$ (and similarly for $\max_{r\in\mathcal{B}} \omega r$) essentially involves maximizing a weighted sum of the expected total rewards $\bar{r}_i$ with weights ω_i. This is equivalent to maximizing an unconstrained POMDP with the same stochastic process and immediate/terminal rewards defined as $r(s,a) = \sum \omega_i r_i(s,a)$ and $R(s) = \sum \omega_i R(s)$. Lovejoy (1991) proved that the grid-based approximation approach produces an upper bound in maximization POMDPs. Consequently, we must have $\max_{r\in\mathcal{A}} \omega r \leq \max_{r\in\mathcal{B}} \omega r$, which contradicts our earlier assertion. ■

B.3 ICER Minimization Results

Proof of Lemma 2: We show this result separately for finite and infinite horizon models. Please note that in this proof, the variable x has been redefined and does not denote policy as in Section 4.

- **Finite horizon:**

  Consider the occupancy measure y^0 associated with an arbitrary non-deterministic policy, i.e., for some period j and state i we have $y^0_j(i,a) > 0$ for m actions a, with $m \geq 2$. We prove y^0 cannot be an extreme point by showing it is a (non-trivial) convex combination of m distinct $y^k \in Y$ for $k = 1, \ldots, m$. To that end, without loss of generality, reorder actions such that for actions indexed $k = 1, \ldots, m$, we have $y^0_j(i,k) > 0$. Let $\pi^0_t(s,a)$ be the original policy and $x^0_t(s) = \sum_a y^0_t(s,a)$ be the occupancy measure of state s at time t, both corresponding to occupancy measure y^0. Define m new policies π^k_t, $k = 1, \ldots, m$ according to $\pi^k_t(s,k) = 1$ for $t = j$ and $s = i$, and $\pi^k_t(s,i) = \pi^0_t(s,i)$, otherwise. Note that these policies differ from the original policy only for period j and state i: Policy π^k deterministically selects the kth action at state i during period j, otherwise following π^0.

  In this context $\pi_t(s,a)$ represents the probability of selecting action a at state s during period t. Let y^k and x^k be the state/action and state occupancy measures corresponding to policy k. Also, let $w_k = \pi^0_j(i,k)$ for $k \geq 1$. Note that w_k's constitute a set of non-trivial convex combination weights since $y^0_j(i,a) > 0$ for $m > 2$ actions, which implies $w_k > 0$ for all $k \geq 1$. It remains to show that $y^0 = \sum w_k y^k$ and that y^k are indeed distinct. We first use induction to show that $x^0 = \sum w_k x^k$ as follows.

Note that $x_t^0 = x_t^k$ and $y_t^0 = y_t^k$ for $t < j$ since policies π^k and π^0 concur for periods $t < j$. The former holds additionally for $t = j$ since the policy differences do not impact the states occupancy in periods $t \leq j$. Therefore, the results hold trivially for $t \leq j$ since $\sum w_k = 1$. It remains to show the result for $t > j$. For $t = j+1$ we have:

$$\begin{aligned}
\sum_k w_k x_t^k(s') &= \sum_{s,a,k} w_k x_{t-1}^k(s)\pi_{t-1}^k(s,a)\mathbf{P}[s'|s,a] \\
&= \sum_{a,k} w_k x_{t-1}^k(i)\pi_{t-1}^k(i,a)\mathbf{P}[s'|i,a] + \sum_{s\neq i,a,k} w_k x_{t-1}^k(s)\pi_{t-1}^k(s,a)\mathbf{P}[s'|s,a] \\
&= \sum_{k} x_{t-1}^0(i)\pi_{t-1}^0(i,k)\mathbf{P}[s'|i,k] + \sum_{s\neq i,a,k} w_k x_{t-1}^0(s)\pi_{t-1}^0(s,a)\mathbf{P}[s'|s,a] \\
&= \sum_{a} x_{t-1}^0(i)\pi_{t-1}^0(i,a)\mathbf{P}[s'|i,a] + \sum_{s\neq i,a} x_{t-1}^0(s)\pi_{t-1}^0(s,a)\mathbf{P}[s'|s,a] \\
&= \sum_{s,a} x_{t-1}^0(s)\pi_{t-1}^0(s,a)\mathbf{P}[s'|s,a] = x_t^0(s').
\end{aligned} \tag{26}$$

Here, we have used the following for the first and last equality.

$$x_t^k(s') = \sum_{s,a} x_{t-1}^k(s)\pi_{t-1}^k(s,a)\mathbf{P}[s'|s,a].$$

The third equality follows the fact that $w_k\pi_{t-1}^k(i,a)$ is equal to $\pi_{t-1}^0(i,a)$ for $a = k$ and 0 otherwise, and that $x_{t-1}^k = x_{t-1}^0$ and $\pi_{t-1}^k(i,a) = \pi_{t-1}^0(s,a)$ for $s \neq i$. The fourth equality follows the third equality by noting that that the $\pi_{t-1}^0(s,a) = 0$ for $a > m$, and hence the two first terms are equal, and that the $x_{t-1}^0(s)\pi_{t-1}^0(s,a)\mathbf{P}[s'|s,a]$ does not depend on k and $\sum w_k = 1$. The next two equalities follow directly. We use $t = j+1$ as our induction basis. For $t > j+1$ we have:

$$\begin{aligned}
\sum_k w_k x_t^k(s') &= \sum_{s,a,k} w_k x_{t-1}^k(s)\pi_{t-1}^k(s,a)\mathbf{P}[s'|s,a] \\
&= \sum_{s,a,k} w_k x_{t-1}^k(s)\pi_{t-1}^0(s,a)\mathbf{P}[s'|s,a] \\
&= \sum_{s,a} \pi_{t-1}^0(s,a)\mathbf{P}[s'|s,a] \sum_k w_k x_{t-1}^k(s) \\
&= \sum_{s,a} \pi_{t-1}^0(s,a)\mathbf{P}[s'|s,a] x_{t-1}^0(s) = x_t^0(s').
\end{aligned} \tag{27}$$

Here, the first equality follows directly. The second equality follows the fact that $\pi_{t-1}^k = \pi_{t-1}^0$ for $t \neq j$. The third equality follows the fact that $\pi_{t-1}^0(s,a)\mathbf{P}[s'|s,a]$ does not depend on k. The fourth equality follows $\sum_k w_k x_{t-1}^k(s) = x_{t-1}^0(s)$, which is implied by the induction assumption. The last equality follows directly.

We now show $y^0 = \sum w_k y^k$ for $t \geq j$. For period $t = j$, we have:

$$\begin{aligned}
\sum_k w_k y_t^k(s,a) &= \sum_k w_k x_t^k(s)\pi_t^k(s,a) = x_t^0(s) \sum_k w_k \pi_t^k(s,a) \\
&= x_t^0(s)\pi_t^0(s,a) = y_t^0(s,a).
\end{aligned} \tag{28}$$

Here, the first and last equations follow $y_t^k(s,a) = x_t^k(s)\pi_t^k(s,a)$, which holds by definition. The second equality follows since $x_t^k(s) = x_t^0(s)$ for $t = j$. The third equality can be explained as follows. If $s \neq i$, then $\pi_t^k(s,a) = \pi_t^0(s,a)$, which implies the result by noting $\sum_k w_k = 1$. If $s = i$ and $a > m$, then $\pi_t^k(s,a) = \pi_t^0(s,a) = 0$. If $a \leq m$, then $\pi_t^k(s,k) = 1$ and $\pi_t^k(s,a) = 0$ for $k \neq a$. Therefore, $\sum_k w_k \pi_t^k(s) = w_a$. Since we chose $w_a = \pi_t^0(s,a)$, the result follows. Finally, for $t > j$ we have.

$$\begin{aligned}\sum_k w_k y_t^k(s,a) &= \sum_k w_k x_t^k(s)\pi_t^k(s,a) = \pi_t^0(s,a)\sum_k w_k x_t^k(s,a) \\ &= x_t^0(s)\pi_t^0(s,a) = y_t^0(s,a).\end{aligned} \tag{29}$$

Here, the first and last equality follows directly. The second equality follows since we designed $\pi_t^k(s,a) = \pi_t^0(s,a)$ for $t \neq j$. The third equality follows $x_t^0(s) = \sum_k w_k x_t^k(s)$, which we just proved.

Finally, we show that y^k's are distinct. Fix period at j and state at i, i.e., focus on cases where policies differ. For any policy $k \geq 1$, we have $\pi_j^k(i,a) = 1$ for $a = k$ and $\pi_j^k(i,a) = 0$ for $a \neq k$. By definition $y_t^k(i,a) = \pi_t^k(i,a)x_t^k(i)$. Since $x_j^k = x_j^0$, we obtain $y_j^k(i,k) = x_t^0(i) > 0$ and $y_j^k(i,a) = 0$ for $a \neq k$. Therefore, $y_j^l(i,k) \neq y_j^k(i,k)$ for all $l \neq k$, and hence the claim.

- **Infinite horizon:**

To show this result for the infinite horizon model, we should first show an intermediate result. Let τ_k^j be the time between the jth and $(j+1)$th passage to state i under policy $k = 0, \ldots, m$. We let τ_k^1 denote the time until the first passage to state i. Note that τ_k^1 does not depend on the policy for i and hence, is identical for the original and constructed policies. Also, note that for a fixed policy k, τ_k^j are i.i.d. for $j \geq 2$.

Consider an arbitrary action/state dependent reward stream and for policy k. Let R_k^j be the total discounted reward of visiting state i between jth and $(j+1)$th passage to state i under policy $k = 0, \ldots, m$. Note that the discounting is with respect to the time of the jth passage. Similarly, R_k^1 denotes the discounted total reward collected until the first passage to state i, R_k^1 is identical for the original and constructed policies, and for a fixed policy k, R_k^j are i.i.d. for $j \geq 2$. Let R_k be the total discounted reward under policy $k = 0, \ldots, m$. We have the following result.

Lemma 5. The following identities hold.

(a) $\mathbb{E}R_k = \mathbb{E}R_k^1 + \mathbb{E}R_k^2 \mathbb{E}\gamma^{\tau_k^1}/(1 - \mathbb{E}\gamma^{\tau_k^2})$,

(b) $\mathbb{E}\gamma^{\tau_0^2} = \sum_{k\geq 1} \pi^0(i,k)\mathbb{E}\gamma^{\tau_k^2}$,

(c) $\mathbb{E}R_0^2 = \sum_{k\geq 1} \pi^0(i,k)\mathbb{E}R_k^2$.

Proof.

(a) Let T_j^k be the time until the jth visit to state i, i.e., let $T_j^k = \sum_{l\le j} \tau_l^k$. Since for all l, τ_l^k are independent and for $l \ge 2$ are i.i.d, we have;

$$\mathbb{E}\gamma^{T_j^k} = \mathbb{E}\gamma^{\sum_{l\le j}\tau_l^k} = \mathbb{E}\prod_{l\le j}\gamma^{\tau_l^k} = \prod_{l\le j}\mathbb{E}\gamma^{\tau_l^k} = \mathbb{E}\gamma^{\tau_1^k}(\mathbb{E}\gamma^{\tau_2^k})^{(j-1)}.$$

We also have the following by definition:

$$R_k = R_k^1 + \sum_{j\ge 2} R_j^k \gamma^{T_{j-1}^k}.$$

Therefore,

$$\begin{aligned}\mathbb{E}R_k &= \mathbb{E}R_k^1 + \sum_{j\ge 2}\mathbb{E}R_j^k\mathbb{E}\gamma^{T_{j-1}^k} = \mathbb{E}R_k^1 + \mathbb{E}R_k^2\sum_{j\ge 2}\mathbb{E}\gamma^{T_{j-1}^k}\\ &= \mathbb{E}R_k^1 + \mathbb{E}R_k^2\sum_{j\ge 2}\mathbb{E}\gamma^{\tau_1^k}(\mathbb{E}\gamma^{\tau_2^k})^{(j-2)} = \mathbb{E}R_k^1 + \mathbb{E}R_k^2\mathbb{E}\gamma^{\tau_1^k}/(1-\mathbb{E}\gamma^{\tau_2^k}).\end{aligned}$$

Here, the first equality follows since R_j^k and T_{j-1}^k are independent, the second holds since R_j^k are i.i.d, for $j \ge 2$, and the last equality follows the geometric series infinite-sum.

(b) Let t be the time of the first visit to state i and a_t be the action taken in that period. If action k is taken under policy π^0, we have τ_0^2 equals τ_k^2 since, until the next passage to state i, policies π^0 and π^k behave the same. In other words, $\tau_k^2 = [\tau_0^2 | a_t = k]$, where the equality is in distribution. Therefore, we have:

$$\begin{aligned}\gamma^{\tau_0^2} &= \sum_{k\ge 1}\mathbb{1}_{[a_t=k]}\gamma^{\tau_k^2},\\ \Longrightarrow \mathbb{E}\gamma^{\tau_0^2} &= \sum_{k\ge 1}\pi^0(i,k)\mathbb{E}\gamma^{\tau_k^2}.\end{aligned}$$

(c) The same line of proof as in part (b) can be used to show this result. ■

Now we can prove the result for the **infinite horizon** model as follows:

We approach the proof similarly to what we did for the finite horizon problems. Consider the occupancy measure $y^0 \in Y$ associated with an arbitrary random policy, i.e., for some state i we have $y^0(i,a) > 0$ for $m > 1$ actions a. We prove y^0 is not an extreme point of Y by showing that it is a (non-trivial) convex combination of m distinct $y^k \in Y$ for $k = 1,\dots,m$. To that end, without loss of generality, reorder actions such that for actions indexed $k = 1,\dots,m$, we have $y^0(i,k) > 0$. Let $\pi^0(s,a)$ be the policy associated with occupancy measure y^0. Define policies $\pi^k(s,a)$, $k = 1,\dots,m$ according to $\pi^k(i,k) = 1$ and $\pi^k(s,j) = \pi^0(s,j)$ for any state $s \ne i$. Note that these constructed policies differ from the original policy only for state i, at which the kth action is taken with certainty in π^k.

Let y^k be the state/action occupancy measure associated with policy k. Also, let $w_k = y^0(i,k)/y^k(i,k)$. We show that $y^k(i,k) > 0$ and hence w_k's are well-defined. We also show that

$\sum_{k\geq 1} w_k = 1$ and $\sum_{k\geq 1} w_k y^k = y^0$. Note that for any $k \geq 1$, $y^k(i,j) = 0$ for any $j \neq k$ since in π^k we never take action $j \neq k$ at state i. Also, we assumed $y^0(i,j) > 0$ for $j \leq m$. As a result, $y^k \neq y^0$ for all $k \geq 1$, i.e., y^k are distinct from y^0. Therefore, y^0 is a non-trivial convex combination of y^k's.

We now calculate $y^k(i,k)$ as follows. Consider a setting where the reward is 1 when we visit the state/action pair (i,k) and zero otherwise. For this reward process, the total discounted reward equals the occupancy measure for the state/action pair (i,k), i.e., $\mathbb{E}R_k = y^k(i,k)$. Note that by the reward definition $R_k^1 = 0$ since we start collecting rewards after we visit state i for the first time. Also, $R_k^2 = 1$ if we take action k when we visit i and zero otherwise. Therefore, $\mathbb{E}R_k^2 = \pi^k(i,k)$. Using Lemma 5 (a), we have

$$y^k(i,k) = \pi^k(i,k)\mathbb{E}\gamma^{\tau_k^1}/(1-\mathbb{E}\gamma^{\tau_k^2}).$$

Based on the definition of w_k and that τ_k^1 is identical for all k, we have

$$w_k = \pi^0(i,k)(1-\mathbb{E}\gamma^{\tau_k^2})/(1-\mathbb{E}\gamma^{\tau_0^2}).$$

Note that $w_k > 0$ since all terms on the right side are positive. We next show $\sum_{k\geq 1} w_k = 1$ as follows.

$$\begin{aligned}\sum_{k\geq 1} w_k &= \sum_{k\geq 1} \pi^0(i,k)[(1-\mathbb{E}\gamma^{\tau_k^2})/(1-\mathbb{E}\gamma^{\tau_0^2})] \\ &= 1/(1-\mathbb{E}\gamma^{\tau_0^2})[\sum_{k\geq 1}(\pi^0(i,k) - \sum_{k\geq 1}\pi^0(i,k)\mathbb{E}\gamma^{\tau_k^2}) \\ &= (1-\mathbb{E}\gamma^{\tau_0^2})/(1-\mathbb{E}\gamma^{\tau_0^2}) = 1.\end{aligned}$$

Here, the first equality follows the above result, the second follows re-arranging terms, and the third follows Lemma 5 (b) and that $\sum_{k\geq 1}\pi^0(i,k) = 1$.

It remains to show $y^0(j,l) = \sum_{k\geq 1} w_k y^k(j,l)$. For any state/action pair (j,l), the state/action occupancy measure is the total discounted reward for a setting where the reward is one when we visit the state/action pair (j,l) and zero otherwise. To show the result, it remains to show the following for any arbitrary reward process.

$$\mathbb{E}R_0 = \sum_{k\geq 1} w_k \mathbb{E}R_k.$$

We have:

$$\begin{aligned}\mathbb{E}R_0 &= \mathbb{E}R_0^1 + \mathbb{E}R_0^2\mathbb{E}\gamma^{\tau_0^1}/(1-\mathbb{E}\gamma^{\tau_0^2}) \\ &= \sum_{k\geq 1} w_k \mathbb{E}R_k^1 + \sum_{k\geq 1}\pi(i,k)\mathbb{E}R_k^2[\mathbb{E}\gamma^{\tau_0^1}/(1-\mathbb{E}\gamma^{\tau_0^2})]\end{aligned}$$

$$= \sum_{k\geq 1} w_k \mathbb{E}R_k^1 + \sum_{k\geq 1} w_k \mathbb{E}R_k^2 \mathbb{E}\gamma^{\tau_0^1}/(1-\mathbb{E}\gamma^{\tau_k^2})$$
$$= \sum_{k\geq 1} w_k [\mathbb{E}R_k^1 + R_k^2 \mathbb{E}\gamma^{\tau_0^1}/(1-\mathbb{E}\gamma^{\tau_k^2})] = \sum_{k\geq 1} w_k \mathbb{E}R_k.$$

■

Proof of Corollary 1: Let λ^* be the optimal ICER value. In Proposition 1, we show that policies optimizing NMB with a WTP of λ^* also optimize ICER. The NMB problem is an unconstrained POMDP problem with a linear objective, which is quasi-linear (both quasi-convex and quasi-concave). Therefore, by Lemma 2, it admits a deterministic, optimal policy. As a result, the ICER problem also admits a deterministic, optimal policy. ■

Appendix C Finite Horizon Ternary Plots

The plot in Figure 8 depicts the optimal policies for the finite horizon problem with WTP=£20,000 for various ages from 10 to 60 years. This visualization indicates that the optimal policy remains relatively consistent throughout this age range, covering the typical life expectancy of a patient.

Figure 8: Optimal policy for various ages for WTP=£20,000