Políticas Aproximadas e Parciais Sensíveis a Risco para o Controle da Propagação de Doenças Infecciosas

Henrique Dias Pastor[1],
Karina Valdivia Delgado[1], **Valdinei Freire** [1], **Leliane Nunes de Barros** [2]

[1]Escola de Artes, Ciências e Humanidades – Universidade de São Paulo (USP)
Rua Arlindo Bettio, 1000 – 03828-000 – São Paulo – SP – Brasil

[2]Instituto de Matemática e Estatística – Universidade de São Paulo (USP)
Rua do Matão, 1010 – 05508-090 – São Paulo – SP – Brasil

Abstract. *Markov Decision Processes (MDPs) can be used for controlling the spread of infectious diseases and finding an optimal vaccination control policy. However, since this is a problem involving lives, it is necessary to take into account the decision-making agent's attitude towards the risk. Thus, in this work, we use risk-sensitive MDPs with SIR compartmental model and propose two efficient algorithms to find optimized vaccination policies that allow controlling the spread of an infectious disease, i.e. to select the number of individuals who should be vaccinated at each time period considering a parameter that represents the attitude towards the risk. The first proposed solution finds a vaccination policy that is partial and optimal w.r.t. a given risk attitude. The second proposed solution is approximated and thus can solve even larger problems. The results show that: (i) the vaccination policies depend not only on the baseline reproduction rate R_0, as expected, but also on the cost and attitude towards risk of a decision-making agent; and (ii) both solutions obtain a great gain in execution time and little loss in the quality when compared with the complete and non-approximate policies.*

Resumo. *Os Processos de Decisão de Markov (MDPs) podem ser usados para controlar a propagação de doenças infecciosas e encontrar uma política ótima de controle de vacinação. No entanto, por se tratar de um problema que envolve vidas, é necessário levar em consideração a atitude do agente em relação ao risco. Assim, neste trabalho, são usados MDPs sensíveis ao risco com o modelo compartimental SIR e são propostos dois algoritmos eficientes para encontrar políticas de vacinação otimizadas que permitam controlar a propagação de uma doença infecciosa, ou seja, selecionar o número de indivíduos que devem ser vacinados a cada período considerando um parâmetro que representa a atitude frente ao risco. A primeira solução proposta encontra uma política de vacinação que é parcial e ótima dada uma determinada atitude de risco. A segunda solução proposta é aproximada e assim pode resolver problemas ainda maiores. Os resultados mostram que: (i) as políticas de vacinação dependem não apenas da taxa básica de reprodução R_0, como esperado, mas também do custo e da atitude em relação ao risco de um agente; e (ii) ambas as soluções obtêm um grande ganho de tempo de execução e pouca perda de qualidade quando comparadas com as políticas completas e não aproximadas.*

1. Introdução

As doenças infecciosas contagiosas representam uma ameaça à segurança pública uma vez que, sem um controle apropriado, podem atingir o status de pandemia. Entre as pandemias que tiveram um grande impacto temos a pandemia de H1N1 (gripe espanhola) de 1918, a pandemia de H3N2 (gripe de Hong Kong) de 1968, a pandemia de HIV/AIDS, a pandemia de influenza-A H1N1 (gripe suína) de 2009 e a pandemia que vivemos atualmente, a COVID-19 [Chandak et al. 2020].

Vários trabalhos da literatura investigam o efeito de diferentes medidas de combate às pandemias, entre elas, a vacinação, vigilância de fronteiras com testes para a doença, e intervenções não-farmacêuticas como rastreamento de contato, distanciamento social ou lockdown. Entre essas medidas, a vacinação (quando disponível) continua sendo a intervenção mais eficaz para reduzir a carga de doentes e mitigar surtos futuros [Usherwood et al. 2021]. Dentre os trabalhos que investigam o efeito de medidas de combate, existem os que comparam um número limitado de políticas pré-definidas através de simulações para predição dos efeitos dessas medidas previamente definidas, e os que usam abordagens de otimização automática para encontrar políticas ótimas. A maioria destes trabalhos representam o problema através de modelos compartimentais conhecidos, com pequenas adaptações, como SIR ou SEIR [Diekmann and Heesterbeek 2000], e os métodos de otimização mais utilizados são: otimização Bayesiana [Chandak et al. 2020], controle ótimo [Charpentier et al. 2020, Gatto and Schellhorn 2021], Processo de Decisão Markoviano (*Markov Decision Process* – MDP) [Yaesoubi and Cohen 2011, Yaesoubi and Cohen 2016, Nasir and Rehman 2017, Libin et al. 2021], análise de equilíbrio da teoria de jogos [Elie et al. 2020] e aproximação estocástica [Yaesoubi et al. 2021].

Dentre os trabalhos que usam MDPs para modelar o problema de propagação de epidemias e encontrar políticas ótimas de controle existem aqueles que especificam as probabilidades e resolvem o problema com algoritmos clássicos de Programação Dinâmica [Yaesoubi and Cohen 2011, Yaesoubi and Cohen 2016, Nasir and Rehman 2017] e os que consideram que essas probabilidades não são conhecidas e usam algoritmos de Aprendizado por Reforço [Libin et al. 2021]. Todos estes trabalhos que usam MDPs clássicos tem como objetivo encontrar uma política que minimiza o custo acumulado esperado e que seja neutra ao risco, isto é, *que não leve em conta a variância desse custo*. Porém, em problemas de controle epidêmico, se faz ainda mais necessário encontrar soluções para MDPs que levem em conta o risco, uma vez que lidamos com vidas humanas.

MDPs sensíveis a risco (*Risk Sensitive Markov Decision Process*– RSMDP) podem ser usados para o controle de ações de combate a pandemias, mais especificamente, através da vacinação de diferentes porcentagens da população ao longo de um horizonte de tempo [Pastor et al. 2020b]. Existem algoritmos para resolver RSMDPs, entre eles RSTL-VI [Mihatsch and Neuneier 2002] e RSTL-PI [Pastor et al. 2020a]. Porém, eles podem ter um consumo computacional alto permitindo apenas resolver problemas pequenos. Assim, neste trabalho são propostos dois algoritmos para lidar com essa limitação. O primeiro, o RSTL-ILAO*, encontra políticas ótimas parciais, sensíveis a risco e é baseado em um algoritmo considerado estado-da-arte de Programação Dinâmica assíncrona [Hansen and Zilberstein 2001]. O segundo, resolve o problema de maneira aproximada.

2. RSMDPs e o modelo SIR estocástico

Os modelos compartimentais foram projetados para descrição da dinâmica de transmissão de doenças epidemiológicas, que têm potencial não apenas explicativo, mas também preditivo. Neles são considerados, o deslocamento da população entre grupos (compartimentos) de indivíduos: *Suscetíveis* (S), *Infectados* (I), *Recuperados* (R) e *latentes* (E) (infectados mas que ainda não transmitem a doença). Dentre os modelos compartimentais estão [Diekmann and Heesterbeek 2000]: SI, SIS, SIR, SIRS, SEIR, SEIRS. A diferença entre os modelos é basicamente a classe de indivíduos que são considerados.

Nos modelos compartimentais básicos o tempo é contínuo e a dinâmica é determinística. No modelo SIR estocástico (caso em que a dinâmica é estocástica) com vacinação, usado como base neste trabalho, são conhecidos: o número de indivíduos suscetíveis (X_S), infectados (X_I) e recuperados (X_R). Assume-se que o tamanho da população N é constante, ou seja não são considerados nascimento, morte ou imigração de indivíduos. Ademais, assume-se que os indivíduos adquirem imunidade permanente por infecção ou vacinação. No SIR são dadas duas taxas: γ, *taxa de recuperação espontânea* e β, *taxa de infecção*. A evolução do modelo SIR pode ser analisada segundo a taxa de reprodução basal $R_0 = \beta/\gamma$ e é definida como o número esperado de infecções secundárias (indivíduos infectados por outros) decorrentes de um único indivíduo durante todo o seu período infeccioso.

O SIR estocástico com vacinação e tempo discreto pode ser modelado como um MDP $\langle \mathcal{S}, \mathcal{A}, Pr, \mathcal{C}, S_g \rangle$, em que: $\mathcal{S}$ é o conjunto de estados; $\mathcal{A}$ é o conjunto de ações; $Pr : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$ é uma função que define as probabilidades de transição, sendo que $Pr(s'|s, a)$ representa a probabilidade de transitar para o estado $s' \in S$, dado que o sistema está no estado $s \in \mathcal{S}$ e uma ação $a \in \mathcal{A}$ foi escolhida; $\mathcal{C} : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ é a função de custo; e $S_g \subseteq S$ é um conjunto de estados meta os quais são absorventes.

2.1. Modelagem dos estados, ações e custos

Seja $X_S(t) \in \mathbb{N}$ o número de indivíduos suscetíveis, $X_I(t) \in \mathbb{N}$ infectados e $X_R(t) \in \mathbb{N}$ recuperados no tempo t. Considerando uma população de tamanho N, em todo tempo t tem-se $X_S(t)+X_I(t)+X_R(t) = N$. O estado do espalhamento da doença no tempo t pode ser representado por $s_t = \langle X_S(t), X_I(t) \rangle$, pois $X_R(t) = N-(X_S(t) + X_I(t))$. O conjunto de estados $\mathcal{S}$ consiste de todas as combinações possíveis de X_S e X_I tal que $X_S+X_I \leq N$. Em cada estágio o tomador de decisão está em um estado $s_t = \langle X_S(t), X_I(t) \rangle$ e deve escolher a fração de indivíduos de $X_S(t)$ que devem ser vacinados as quais definem o conjunto de ações de vacinação $\mathcal{A}$. Neste trabalho são consideradas 11 frações de $X_S(t)$ distribuídas uniformemente entre 0 e 1, isto é, $\mathcal{A} = \{0, 0.1, 0.2, \ldots, 0.9, 1\}$. Por exemplo, se é escolhida a ação $a = 0.7$ no estado $\langle X_S = 30, X_I = 2 \rangle$, a ação indica vacinar 70% dos suscetíveis, isto é, 21 pessoas. O custo de se aplicar uma ação de vacinação $a \in \mathcal{A}$ no estado $\langle X_S, X_I \rangle$ é dado por: $C(\langle X_S, X_I \rangle, a) = (a \cdot X_S)^{1.5} \cdot cost_{vac} + X_I \cdot cost_{inf}$, em que $cost_{vac}$ é o custo de vacinação (que inclui o custo de compra da vacina e da logística necessária para aplicá-la) e $cost_{inf}$ é o custo de infecção (que inclui o custo do diagnóstico, custo do tratamento e outros custos indiretos). Essa equação é similar a função custo usada na literatura de controle ótimo de tempo contínuo para problemas modelados com SIR [Liu et al. 2017] a qual é geralmente quadrática. A não linearidade no custo da vacina modela que o custo de produzir e distribuir vacinas aumentam de forma

superlinear. Além disso, neste trabalho o conjunto de estados meta S_g é formado pelos estados em que $X_S = 0$ ou $X_I = 0$.

2.2. Modelagem das transições estocásticas

O modelo estocástico de SIR com vacinação permite que o número de indivíduos que se deslocam entre os compartimentos da população variem de forma estocástica. Porém, para fins didáticos, definiremos primeiro o modelo SIR determinístico. Considere o estado atual $\langle X_S(t), X_I(t)\rangle$ e as seguintes variáveis: $D(t)$ é a quantidade de pessoas suscetíveis que foram vacinadas no passo t; $E(t)$ é a quantidade de pessoas suscetíveis que foram infectadas no passo t; e $F(t)$ é a quantidade de pessoas que se tornaram recuperadas no passo t, após estarem infectadas. Assim, o próximo estado é: $X_S(t+1) = X_S(t) - D(t) - E(t)$ e $X_I(t+1) = X_I(t) - F(t) + E(t)$. Na formulação determinística tem-se: $D(t) = a \cdot X_S(t)$, $E(t) = \beta \cdot \frac{X_I(t)}{N} \cdot (X_S(t) - D(t))$ e $F(t) = \gamma \cdot X_I(t)$.

A formulação estocástica usada neste trabalho considera as mesmas dependências entre variáveis, porém com aleatoriedade e é a mesma utilizada em [Pastor et al. 2020b]. A primeira fonte de incerteza está relacionada ao fato de não ser possível vacinar uma fração de um indivíduo. Sendo $m = \text{floor}(a \cdot X_S(t))$, a quantidade de pessoas suscetíveis que foram vacinadas é dada por:

$$\begin{aligned} &\Pr(D(t) = m) = (m+1) - a \cdot X_S(t), \\ &\Pr(D(t) = m+1) = a \cdot X_S(t) - m. \end{aligned} \tag{1}$$

Outra fonte de incerteza é associada à taxa de infecção β no cálculo de $E(t)$, modelada por uma distribuição binomial:

$$\Pr(E(t) = y | D(t) = x, X_S(t), X_I(t)) = \text{Binomial}\left(y; n = X_S(t) - x, \beta \cdot \tfrac{X_I(t)}{N}\right). \tag{2}$$

Finalmente, a terceira fonte de incerteza é devido à taxa de recuperação espontânea γ. Sendo $q = \text{floor}(\gamma \cdot X_I(t))$, F(t) é dada pela distribuição:

$$\begin{aligned} &\Pr(F(t) = q) = (q+1) - \gamma \cdot X_I(t), \\ &\Pr(F(t) = q+1) = \gamma \cdot X_I(t) - q. \end{aligned} \tag{3}$$

Assim, no SIR estocástico, para especificar a transição probabilística $\Pr(s_{t+1} = s' | s_t = s, a_t = a)$ modificamos as equações para $X_S(t+1)$ e $X_I(t+1)$ considerando as distribuições de probabilidades para $D(t), E(t)$ e $F(t)$.

2.3. Políticas de vacinação sensíveis a risco

Considerando o problema SIR estocástico com vacinação modelado como um MDP $\langle \mathcal{S}, \mathcal{A}, Pr, \mathcal{C}, S_g\rangle$, pode-se escolher uma política de controle, isto é, a porcentagem de indivíduos que devem ser vacinados para um dado estado $\langle X_S(t), X_I(t)\rangle$. Por se tratar de um problema que envolve vidas, é importante que políticas de vacinação sejam sensíveis a risco. Uma extensão de MDPs clássicos é chamada de MDPs sensíveis ao risco (RSMDP). Diferente dos MDPs clássicos, RSMDPs utilizam critérios capazes de considerar a variância do custo acumulado das execuções de uma política permitindo assim levar em consideração a atitude frente ao risco dos agentes tomadores de decisões.

Neste trabalho, a proposta do modelo SIR sensível a risco é baseada no modelo RSMDP proposto em [Mihatsch and Neuneier 2002], que é baseada em funções lineares

por partes e considera um fator de risco k. O agente tomador de decisão é averso ao risco se $k \in (0, 1)$, o agente é neutro se $k = 0$ e o agente é propenso ao risco se $k \in (-1, 0)$. Além disso, neste modelo são considerados um conjunto de estados meta, o que torna um MDP em um em SSP MDP (*Stochastic Shortest Path* MDP). SSP RSMDP usa uma função de transformação $\chi^{(k)}$ baseada no valor de entrada $z \in \mathbb{R}$, chamada de diferença temporal, e no fator de risco k, sendo essa transformação dada por [Mihatsch and Neuneier 2002]:

$$\chi^{(k)}(z) = \begin{cases} (1-k)z, & \text{se } z < 0. \\ (1+k)z, & \text{caso contrário.} \end{cases} \tag{4}$$

A solução de um SSP RSMDP é uma política estacionária $\pi : S \rightarrow A$. Para avaliar uma política π é usada a função valor $V_\pi^k(s)$, a qual usa a função de transformação $\chi^{(k)}$ sobre a diferença temporal e pode ser obtida resolvendo o seguinte sistema de equações para todo $s \in S$ [Mihatsch and Neuneier 2002]: $\sum_{s \in S} Pr(s'|s, \pi(s))\chi^{(k)}\left(C(s, \pi(s)) + V_\pi^k(s') - V_\pi^k(s)\right) = 0$. Como para os MDPs tradicionais que são neutros ao risco, existem políticas ótimas estacionárias para SSP RSMDPs e sua correspondente função valor ótima é única [Mihatsch and Neuneier 2002]. Para cada $k \in (1, -1)$ existe uma única função valor ótima, $V_k^*(s) = \min_{\pi \in \Pi} V_k^\pi(s), \forall s \in S$, que satisfaz a seguinte equação: $\min_{a \in A} \sum_{s' \in \mathcal{S}} Pr(s'|s, a)\mathcal{X}^{(k)}\Big(C(s, a) + V_k^*(s') - V_k^*(s)\Big) = 0, \forall s \in S$. A partir de V_k^*, uma política π^* ótima pode ser obtida.

2.4. Algoritmos de Programação Dinâmica síncrona que devolvem uma política ótima completa para SSP RSMDPs

O algoritmo RSTL-VI [Mihatsch and Neuneier 2002]: atualiza a função qualidade $Q^i(s, a)$ para todos os pares estado-ação (s, a) em cada iteração i da seguinte forma:

$$Q^i(s,a) = \; Q^{i-1}(s,a) + \alpha \sum_{s \in S} Pr(s'|s,a) \cdot \chi^{(k)}\left(C(s,a) + \min_a Q^{i-1}(s',a) - Q^{i-1}(s,a)\right),$$

em que α é o tamanho do passo. A partir de $Q^i(s, a)$ pode ser obtida a função valor $V^i(s) = \min_{a \in A}\{Q^i(s, a)\}$ e a política gulosa $\pi(s) = \arg\min_{a \in A}\{Q^i(s, a)\}$. O algoritmo RSTL-VI tem garantia de convergência se $0 \leq \alpha \leq (1 + |k|)^{-1}$.

O algoritmo RSTL-PI [Pastor et al. 2020a]: envolve duas etapas: (1) avaliação da política e (2) melhoria da política. Essas duas etapas são executadas até que a política não mude. Dada uma política π e $0 < \alpha \leq (1 + |k|)^{-1}$, a etapa de avaliação de política pode ser executada usando o operador de contração:

$$T_{\alpha k}^\pi[V](s) = \; V(s) + \alpha \sum_{s' \in S} Pr(s'|s,\pi(s)) \cdot \chi^{(k)}\left(C(s,\pi(s)) + V(s') - V(s)\right).$$

Dada uma política estacionária π, a etapa de melhoria de política obtém uma política estacionária melhorada π' da seguinte forma: $\pi'(s) = \arg\min_{a \in A} \sum_{s' \in \mathcal{S}} Pr(s'|s, a)\mathcal{X}^{(k)}(C(s, a) + V_k^\pi(s') - V_k^\pi(s))$.

Exemplo de política completa: A Figura 1(b) ilustra um exemplo de política completa gerada pelos algoritmos RSTL-VI (que é igual ao RSTL-PI) para $N = 10$, $R_0 = 3$ e $k = -0.5$. A política é representada por uma matriz triangular $\mathbf{B}[i, j] = a_{ij}$ sendo a_{ij} a ação que corresponde à fração de indivíduos selecionada para vacinação para o estado correspondente ao par (i, j). As linhas da matriz são enumeradas de 0 a 10, representando o número de pessoas suscetíveis X_S e as colunas são enumeradas de 0

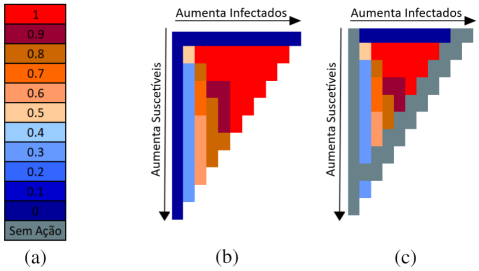


Figura 1. Legenda de porcentagens de vacinação e cores usadas neste artigo e exemplos de política completa e parcial para N=10.

a 10, representando o número de pessoas infectadas X_I. Note que o triângulo inferior não é mostrado uma vez que os estados correspondem a quantidades inválidas, isto é, $X_S + X_I > N$. Cada ação da política é representada por uma cor diferente de acordo com a legenda (Figura 1(a)), isto é, $a_{ij} \in \{0, 0.1, 0.2, \ldots, 1\}$. Por exemplo, para o estado $\langle X_S = 3, X_I = 2\rangle$ (quarta linha e terceira coluna da matriz), a ação indicada pela política é vacinar 70% dos suscetíveis, isto é, 2.1 pessoas. De acordo com a Equação 1, a probabilidade de vacinar 3 pessoas é 0.1 e de vacinar 2 pessoas é 0.9.

3. Políticas de vacinação sensíveis a risco: algoritmos eficientes

Os algoritmos baseados em Programação Dinâmica síncrona, devolvem uma política completa e precisam atualizar todos os estados a cada iteração. Por tal motivo, podem ter um consumo alto de tempo. Assim, é proposta a seguir uma versão adaptada do ILAO* [Hansen and Zilberstein 2001] e também é proposta uma solução aproximada.

3.1. Solução I: RSTL-ILAO*

O algoritmo RSTL-ILAO* é um algoritmo assíncrono que devolve uma política ótima parcial a qual é definida considerando um estado inicial em particular. Para isso, ele gera um grafo explícito G' a cada iteração i (a partir de uma política gulosa π_i) que armazena todos os estados visitados até então, as ações aplicáveis nesses estados e os estados sucessores alcançáveis por essas ações. Chamamos o grafo explícito G' de *grafo induzido pela política gulosa* π_i, sendo que cada nó representa um estado $s \in S$ e cada aresta $a \in A$ uma ação. Um nó folha em G' é uma folha terminal se é um estado meta, caso contrário é um estado não-terminal.

O algoritmo RSTL-ILAO* é dividido em duas etapas: *busca em profundidade* e *verificação de convergência*. A primeira etapa faz: (i) busca em profundidade gerando uma solução parcial gulosa; (ii) atualiza os custos; e (iii) realiza um novo reconhecimento das melhores ações criando um novo grafo da solução parcial chamado de G''. Na segunda etapa é feito o teste de convergência executando o RSTL-VI em todos os estados que estão em G'' e em seguida verifica se houve alguma alteração na política. Se ocorreu uma mudança, o algoritmo continua para a próxima iteração, caso contrário devolve a solução.

A Figura 1(c) mostra um exemplo de política parcial gerada pelo algoritmo RSTL-ILAO* para $N = 10$, $R_0 = 3$, $k = -0.5$ e estado inicial $X_S = 9, X_I = 1, X_R = 0$. A política é mostrada de forma similar à política completa, exceto que neste caso foi adicionado o rótulo cinza *sem-ação* para os estados que não pertencem à política parcial.

Por exemplo, a ação sugerida para o estado $X_S = 8, X_I = 1, X_R = 1$ (linha 9 e coluna 1 da matriz) foi *sem-ação* por não ser um estado alcançável a partir do estado inicial do problema. Note que a política encontrada pelo RSTL-ILAO* para k=-0.5 apresentam um comportamento exatamente igual ao apresentado pelos algoritmos RSTL-VI e RSTL-PI em termos de atitude a risco. A principal diferença é devido ao fato da política ótima ser parcial, o que implica que a política devolvida não é definida para todos os estados.

3.2. Solução II: políticas aproximadas

As políticas ótimas devolvidas pelos algoritmo RSTL-VI, RSTL-PI e RSTL-ILAO* para um problema menor W de tamanho w, chamada de π_W, é utilizada para gerar uma política aproximada para um problema maior B de tamanho b, chamada de π_B. Seja a razão $r = w/b$. A técnica de aproximação consiste em fazer um mapeamento de cada estado $s_B = \langle X_{S_B}, X_{I_B} \rangle$ com $X_{R_B} = N - X_{S_B} + X_{I_B}$ do problema B para um estado $s_W = \langle X_{S_W}, X_{I_W} \rangle$ do problema W da seguinte maneira: $X_{S_W} = round(X_{S_B} \cdot r)$, $X_{I_W} = round(X_{I_B} \cdot r)$ e $X_{R_W} = round(X_{R_B} \cdot r)$. Após realizar esse cálculo, é necessário verificar se $X_{S_W} + X_{I_W} + X_{R_W} = N$. Caso não, são feitos ajustes em X_{S_W} e em X_{I_W} considerando se foi feito o arredondamento para um número inteiro imediatamente maior ou imediatamente menor. Finalmente, após esses ajustes, a ação $\pi_W(s_W)$ é atribuída para $\pi_B(s_B)$. Note que essa técnica de aproximação também pode ser aplicada a políticas parciais. A única diferença é que são mantidas as atribuições de ações *sem-ação* aos estados não alcançáveis por $\pi(s)$.

4. Resultados empíricos

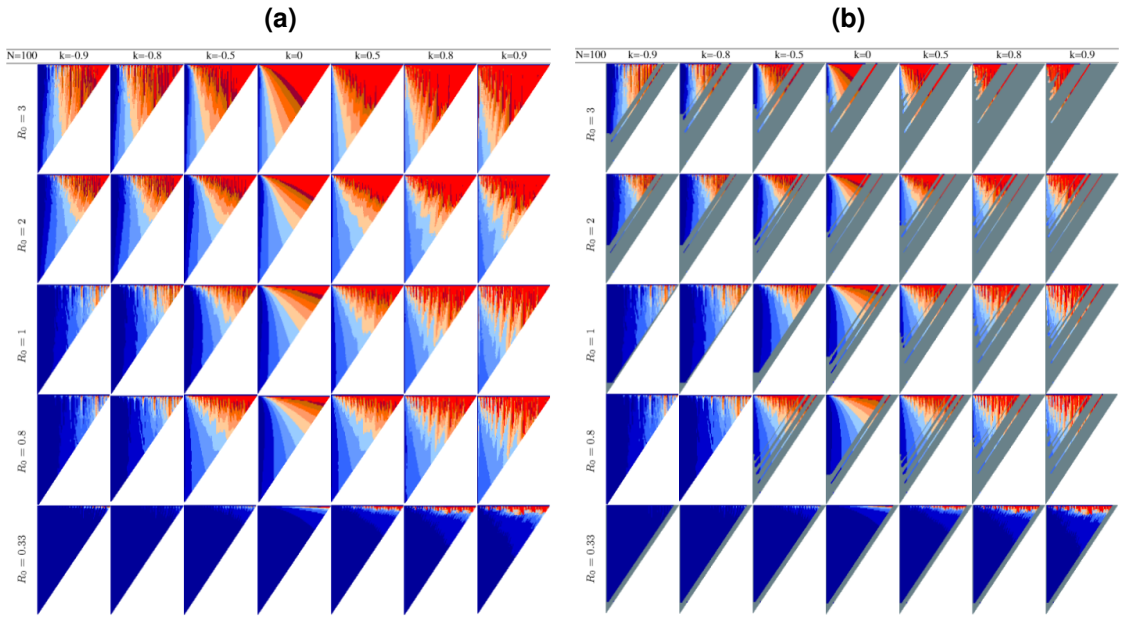
Os experimentos foram realizados em Java 8 SE em um processador AMD 8320e de 3.2GHz. Nas Seções 4.1, 4.2 e 4.3 é feita uma análise dos algoritmos variando R_0, k e N. Os cinco valores de $R_0 = \beta/\lambda$ considerados são: $R_0 = 0.75/0.25 = 3$, $R_0 = 0.8/0.4 = 2$, $R_0 = 0.25/0.25 = 1$, $R_0 = 0.2/0.25 = 0.8$ e $R_0 = 0.25/0.75 = 0.33$. Os sete valores considerados para k são: $k = \{-0.9, -0.8, -0.5, 0, 0.5, 0.8, 0.9\}$. Além disso, são utilizados $cost_{inf} = 4$ e $cost_{vac} = 1$, e os estados meta são aqueles com $X_S = 0$ ou $X_I = 0$. A proporção de 4:1 entre $cost_{inf}$ e $cost_{vac}$ é igual à utilizada em [Yaesoubi and Cohen 2011]. Finalmente, na Seção 4.4 são analisadas as políticas encontradas no contexto de um estudo de caso simplificado de COVID-19 para $N = 1000$.

4.1. Análise das políticas devolvidas por RSTL-VI, RSTL-PI e RSTL-ILAO*

Nesta seção são mostradas as políticas quando variamos R_0 e k. As políticas computadas podem ser interpretadas de acordo com as seguintes atitudes diante o risco:
Agente averso a risco ($k > 0$): Um agente de controle averso a risco tende a vacinar uma fração maior de susceptíveis. Trata-se de um agente pessimista, que espera que aconteça os piores cenários por não contar com a sorte.
Agente propenso a risco ($k < 0$): Um agente de controle propenso ao risco tende a vacinar o mínimo possível da população. Ao não vacinar, ele conta com a sorte acreditando que ocorrerão os melhores cenários, isto é, menos pessoas serão infectadas e os infectados terão uma rápida recuperação.

A Tabela 1a) mostra as políticas completas geradas pelo algoritmo RSTL-VI (que é igual ao RSTL-PI) para um problema com $N = 100$, variando R_0 (linhas) e k (colunas). Para cada coluna que varia de valores maiores para valores menores de R_0, observa-se que

Tabela 1. (a) Políticas completas encontradas por RSTL-VI e RSTL-PI e (b) parciais encontradas pelo RSTL-ILAO* para N=100.



as frações de vacinação também diminuem (i.e., aumentam as ações de cor azul). Para um fator $R_0 = 3$ (1a. linha da tabela), a política sugere vacinar mais do que 50% dos suscetíveis na maioria dos estados devido ao alto risco de surgimento de uma epidemia, mesmo para agentes propensos a risco ($k < 0$). Para $R_0 = 2$ (2a. linha), a fração de vacinados diminui para 30%. As políticas para $R_0 = 1$ e $R_0 = 0.8$ (3a. e 4a. linhas) são similares para cada um dos valores de k. Porém, para $R_0 = 0.8$ existe uma quantidade ligeiramente menor de ações que vacinam uma fração maior ou igual a 20%. Para $R_0 = 0.33$ (5a. linha da tabela) as políticas para $k \leq 0$ indicam vacinar 10% ou menos, na maioria dos estados, enquanto agentes mais aversos a risco ($k = 0.8$ e $k = 0.9$) indicam vacinar de 50% a 100% dos suscetíveis em alguns poucos estados.

A Tabela 1b) mostra as políticas parciais encontradas pelo RSTL-ILAO* para uma população $N = 100$, estado inicial $X_S = 90, X_I = 10, X_R = 0$, variando R_0 (linhas) e k (colunas). Este experimento mostra como a política parcial varia com a diminuição de R_0. Observando cada coluna, de cima para baixo, vemos que em geral a fração de vacinação também diminui (i.e., a quantidade de ações azuis aumenta). Também aumenta o número de estados com ações de vacinação que fazem parte da política parcial ótima, i.e., a quantidade de estados cinza diminui (vide explicação a seguir). Na tabela, quando $R_0 = 3$, mesmo com o agente propenso a risco, a política parcial sugere vacinar 50% ou mais da população na maioria dos estados. Note que têm ações definidas na política parcial para os estados mais próximos a meta, pois no estado inicial vacina-se 50% portanto os próximos estados visitados em geral terão menos suscetíveis que o estado anterior devido a vacinação. Por esse motivo existe um grande número de estados, principalmente próximos ao estado inicial, que não possuem política (ações cinzas). Conforme ocorreu com RSTL-VI e RSTL-PI, quando R_0 diminui, a taxa de vacinação também diminui (Tabela 1b). Como consequência, as políticas parciais obtidas pelo RSTL-ILAO* ficam

definidas para um número maior de estados: quanto menor a taxa de vacinação, menor será a variação do número de suscetíveis fazendo com que as políticas parciais visitem um número maior de estados.

4.2. Análise no tempo para encontrar políticas ótimas

Nesta seção é avaliado o tempo gasto por RSTL-VI, RSTL-PI e RSTL-ILAO* para encontrar a política ótima (sem usar aproximação) para os 5 valores de R_0, 4 valores de N $(30,\ 40,\ 50,\ 100)$ e os 7 valores de k. O RSTL-PI foi inicializado usando a política gerada pelo RSTL-VI com $k = 0$ (foi incluído o tempo desta inicialização). O tempo de executar o RSTL-VI e RSTL-PI nem sempre diminui ao diminuir R_0: observam-se comportamentos diferentes para $R_0 = 3$, $R_0 = 1$ e $R_0 = 0.8$ quando comparados com $R_0 = 2$ e $R_0 = 0.33$. O valor de γ nos 3 primeiros é 0.25 já nos outros 2 valores de R_0, γ é maior. Assim, observa-se que quanto menor o valor de γ, o RSTL-VI e RSTL-PI demoram mais tempo para encontrar a política ótima. Foi possível observar ainda que RSTL-ILAO* é mais rápido que RSTL-VI em todos os casos. Quando $R_0 = 3$ e $k = -0.9$, RSTL-ILAO* é 12.77 vezes mais rápido, quando $k = 0$ é 47.48 vezes mais rápido e quando $k = 0.9$ é 24.74 vezes mais rápido. A diferença diminui apenas quando $R_0 = 0.3$, ficando com ganhos entre 7.5 a 10.5. Comparando o RSTL-ILAO* com RSTL-PI há um ganho menor nos extremos e um ganho maior quando k é próximo de 0. Quando $R_0 = 3$ e $k = -0.9$, RSTL-ILAO* é 4.22 vezes mais rápido, quando $k = 0$ é 50.78 vezes mais rápido e quando $k = 0.9$ é 4.84 vezes mais rápido. Com relação à variação de ganho do RSTL-ILAO*, o ganho dele para valores menores de R_0 é menor pois as políticas são maiores e assim ele realiza mais atualizações. Além disso, visita uma quantidade diferente de estados para cada valor de k, o que faz com que o ganho seja diferente quando k varia.

4.3. Análise do ganho de tempo vs. porcentagem de perda na qualidade das políticas com o uso da técnica de aproximação

O objetivo desta seção é determinar se o ganho em termos de tempo ao realizar a aproximação compensa ou não o nível de perda na qualidade da política medida em termos da média $V(s)$. Foram executados os três algoritmos para um problema de tamanho $N = 100$ com aproximações de tamanho 30, 40 e 50, os cinco valores de R_0 e os sete valores de k. No lado esquerdo da Tabela 2 são mostradas a proporção de ganho de tempo (GT), a porcentagem de perda na qualidade (PQ) e nas últimas duas colunas a média do ganho (MG) e a média da perda (MP) apenas para $R_0 = 3$, $R_0 = 2$ e $R_0 = 1$, por causa do limite de espaço. Já no lado direito da Tabela 2 são mostradas a média do ganho geral (MGG) e a média da perda geral (MPG) considerando os cinco valores de R_0.

Primeiro, com relação ao ganho de tempo, foi calculada a proporção de ganho de tempo de cada algoritmo usando aproximações com tamanho 30, 40 e 50 com relação ao uso do mesmo algoritmo para N=100 (i.e, sem usar aproximação). Na Tabela 2 estão destacados em negrito os melhores valores de GT por linha. Os melhores valores de GTs para o RSTL-VI acontecem para $k = -0.5$ e 0, para RSTL-PI acontecem para diferentes valores de k e para o RSTL-ILAO* acontecem principalmente para $k = 0.9$. Para os três algoritmos percebe-se que quanto menor o tamanho do problema usado para realizar a aproximação, maior o ganho de tempo para qualquer R_0. O RSTL-VI usando uma aproximação com tamanho 30, 40 e 50 foi 77.76, 31.08 e 13.76 vezes mais rápido em média, respectivamente. Já o RSTL-PI usando uma aproximação com tamanho 30, 40 e

Tabela 2. Ganho de tempo e porcentagem de perda do RSTL-VI, RSTL-PI e RSTL-ILAO* a partir do mapeamento de $N = 30$, $N = 40$ e $N = 50$ para $N = 100$.

			k=-0.9		k=-0.8		k=-0.5		k=0		k=0.5		k=0.8		k=0.9			
	N	R_0	GT	PQ	GT	PQ	GT	PQ	GT	PQ	GT	PQ	GT	PQ	GT	PQ	MG	MP
RSTL-VI	30	3	84.16	19.91	78.19	13.95	**101.34**	4.89	69.66	0.04	84.42	1.50	73.59	4.60	73.23	5.48	80.66	7.20
	40	3	32.62	14.24	30.48	9.73	**37.43**	1.71	31.25	0.04	28.70	2.08	28.71	2.97	28.31	4.70	31.07	5.07
	50	3	13.83	11.05	13.59	7.39	**16.30**	2.00	14.40	0.04	13.08	1.21	13.10	1.00	13.31	1.97	13.94	3.52
	30	2	72.84	26.01	71.87	14.90	**110.43**	5.83	42.83	0.00	73.90	2.57	71.74	3.27	69.98	4.9	73.37	7.43
	40	2	29.87	14.61	27.85	9.74	**38.16**	2.90	28.12	0.00	28.86	2.29	27.57	3.76	28.10	4.51	29.79	5.41
	50	2	13.38	8.77	12.97	8.34	**16.11**	2.15	14.03	0.00	13.26	0.05	12.50	2.12	12.80	2.56	13.63	3.51
	30	1	84.18	0.00	81.88	0.14	89.53	16.83	**101.21**	0.05	70.73	2.17	73.65	9.81	70.79	16.02	81.71	6.43
	40	1	33.51	0.00	31.51	0.14	33.54	15.98	**43.09**	0.18	29.27	2.37	29.07	7.85	28.73	9.58	32.67	5.16
	50	1	14.51	0.00	13.43	0.14	**15.11**	14.97	14.61	0.18	13.16	0.96	13.11	4.21	13.18	6.87	13.87	3.90
RSTL-PI	30	3	**77.23**	19.91	68.55	13.95	68.55	4.89	73.66	0.04	66.97	1.50	71.37	4.60	72.57	5.48	71.27	7.20
	40	3	**19.79**	14.24	19.18	9.73	18.80	1.71	19.38	0.04	19.33	2.08	19.52	2.97	18.86	4.70	19.27	5.07
	50	3	12.49	11.05	12.51	7.39	12.18	2.00	**12.67**	0.04	12.59	1.21	12.43	1.00	12.57	1.97	12.49	3.52
	30	2	73.39	26.01	92.04	14.90	90.70	5.83	**92.46**	0.00	81.33	2.57	82.80	3.27	86.07	4.9	85.54	7.43
	40	2	19.47	14.61	22.13	9.74	22.44	2.90	22.31	0.00	**22.57**	2.29	21.98	3.76	22.29	4.51	21.88	5.41
	50	2	13.40	8.77	13.64	8.34	13.93	2.15	**13.94**	0.00	13.75	0.05	13.52	2.12	13.55	2.56	13.67	3.51
	30	1	94.68	0.00	100.86	0.14	102.92	16.83	92.30	0.05	75.95	2.17	**105.72**	9.81	100.72	16.02	96.16	6.43
	40	1	22.56	0.00	23.47	0.14	24.37	15.98	24.40	0.18	**24.64**	2.37	23.75	7.85	24.38	9.58	23.94	5.16
	50	1	13.62	0.00	14.08	0.14	14.29	14.97	**14.52**	0.18	14.26	0.96	14.19	4.21	14.04	6.87	14.14	3.90
RSTL-ILAO*	30	3	28.55	9.81	39.43	2.31	38.45	3.36	10.99	0.59	66.53	4.96	105.06	8.46	**113.59**	8.46	57.51	5.45
	40	3	16.52	5.91	20.17	7.99	23.20	8.27	8.69	0.04	39.39	4.57	69.80	8.33	**71.13**	10.80	35.56	6.56
	50	3	8.10	30.92	10.66	30.70	9.72	11.51	5.81	30.56	15.13	26.91	26.45	7.63	**28.29**	8.95	14.88	19.34
	30	2	38.37	4.82	25.53	3.37	33.84	2.22	31.01	1.82	38.61	9.33	42.72	11.68	**46.27**	11.48	36.62	6.40
	40	2	17.47	4.67	10.80	3.37	13.21	3.93	15.27	2.05	15.51	5.06	**20.65**	10.30	20.17	11.18	16.63	5.29
	50	2	9.01	0.08	5.81	1.07	8.01	2.14	8.96	0.01	9.11	0.05	13.64	2.11	**15.76**	2.55	10.05	1.33
	30	1	45.29	6.96	49.92	6.81	35.83	2.91	32.61	0.00	35.19	10.09	37.71	14.76	**50.71**	19.84	41.04	8.77
	40	1	19.96	0.00	20.73	0,0	15,63	2,56	18,97	3,61	16,13	8,72	20,59	12,41	**21,12**	12,12	19,02	5,63
	50	1	9,31	31,37	**10,17**	30,70	7,95	11,50	9,65	30,41	6,93	26,91	8,17	7,63	9,34	8,9	8,79	21,01

	N	MGG	MPG
RSTL-VI	30	77.76	4.90
	40	31.08	4.08
	50	13.76	2.66
RSTL-PI	30	81.54	4.90
	40	20.99	4.08
	50	13.09	2.66
RSTL-ILAO*	30	43.54	8.88
	40	21.69	5.79
	50	10.01	11.09

50 foi 81.54, 20.99 e 13.09 vezes mais rápido em média, respectivamente. Para o RSTL-ILAO* observa-se que com $R_0 = 3$ e $N = 30$ houve um ganho expressivo, principalmente para os valores de $k = 0.9$ e 0.8. Ao diminuir R_0, continua havendo ganho, porém menor. O RSTL-ILAO* usando uma aproximação com tamanho 30, 40 e 50 foi 43.54, 21.69 e 10.01 vezes mais rápido em média do que o RSTL-ILAO* para N=100, respectivamente.

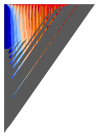
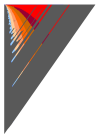
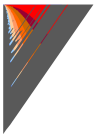
Segundo, para calcular a perda na qualidade, inicialmente, para cada política, executa-se a etapa de avaliação de política do RSTL-VI (Equação 5), usando o fator de risco $k = 0$ que calcula a média $V(s)$ para todos os estados. E posteriormente, calcula-se a porcentagem de perda entre V(s) usando as aproximações de tamanhos diferentes e V(s) sem aproximação para o estado inicial $X_S = 90, X_I = 10$ e $X_R = 0$. A percentagem de perda da qualidade em média é baixa para RSTL-VI (que é igual ao RSTL-PI) quando é utilizada a técnica de aproximação, e as maiores variações estão em $R_0 = 3$ e $R_0 = 2$, sendo a maior quando $k = -0.9$ e $R_0 = 2$ que chega a 26.01%. A porcentagem de perda desses dois algoritmos usando uma aproximação com tamanho 30, 40 e 50 foi 4.90%, 4.08% e 2.66% em média, respectivamente. A percentagem de perda da qualidade em média é também baixa para RSTL-ILAO* para $N = 30$ e 40, porém quando $N = 50$, ocorre um leve aumento. A porcentagem de perda do RSTL-ILAO* usando uma aproximação com tamanho 30, 40 e 50 foi 8.88%, 5.79% e 11.09% em média.

Considerando uma aproximação com tamanho 30, a perda de qualidade nas políticas é no máximo 8.88% e o ganho no tempo de execução é no mínimo 43.54%, para os três algoritmos. O ganho usando a aproximação é maior para RSTL-VI e RSTL-PI quando comparados com o RSTL-ILAO*. Porém, o tempo de execução do RSTL-ILAO* já era baixo e ao usar a técnica de aproximação, reduziu ainda mais o tempo.

4.4. Caso de estudo: COVID-19 simplificado

Nesta seção, são realizados experimentos combinando a técnica de aproximação e RSTL-ILAO* para uma população de tamanho $N = 1000$, com $\gamma = 0.1$ e $\beta = 0.2$ ($R_0 =$

Tabela 3. Políticas parciais geradas do mapeamento de $N = 300$ para $N = 1000$.

$k = -0.8$	$k = -0.5$	$k = 0.0$	$k = 0.5$	$k = 0.8$
				

2). O custo incorrido por um indivíduo infectado, $cost_{inf}$, é definido como 300. Esses valores são os mesmos utilizados em [Elie et al. 2020] em que são analisadas estratégias de isolamento no contexto do COVID-19 modelado com SIR para a cidade de Wuhan. $cost_{vac}$ é definido como 30 considerando o valor médio de duas doses das vacinas. A proporção de 10:1 entre $cost_{inf}$ e $cost_{vac}$ é igual à utilizada em [Nasir and Rehman 2017] e o valor de γ é o mesmo utilizado em [Usherwood et al. 2021] no contexto de controle da COVID-19, num estudo em diferentes cidades dos EUA. Note que o modelo usado neste artigo é muito simples e não pretende modelar realisticamente a disseminação dessa doença. Esse caso de estudo é apresentado apenas para ilustrar a abordagem proposta.

A Tabela 3 mostra as políticas parciais encontradas pelo RSTL-ILAO* com mapeamento de N=300 para N=1000 e estado inicial $X_S = 900, X_I = 100, X_R = 0$, para 5 valores de k. Note que o número total de estados para esse problema é 501501 porém, muitos desses estados não fazem parte da política parcial (cinza). Quando $k > 0$ (agente averso a risco) o número de estados com ações de vacinação que fazem parte da política parcial é baixo (menor que 45451) e a política parcial sugere vacinar 50-100% da população nestes estados. Já para $k < 0$ (agente propenso a risco), a política parcial sugere vacinar 40-100% ($k = -0.5$) ou 0-100% ($k = -0.8$).

5. Conclusão

Neste trabalho foi proposto um algoritmo assíncrono que devolve uma política parcial, RSTL-ILAO*, e uma solução aproximada baseadas na literatura de MDP sensível a risco para resolver problemas de controle de doenças infecciosas usando o modelo SIR estocástico com vacinação. Os algoritmos foram analisados para diferentes valores de R_0 e k. Os experimentos mostram que o RSTL-ILAO* é melhor em tempo de convergência quando comparado com RSTL-VI e RSTL-PI. Os resultados também sugerem que ao resolver problemas maiores de forma aproximada, é possível obter ganho em tempo de execução, tanto para políticas parciais como para políticas completas, com pouca perda em qualidade. Para um problema de tamanho 100, considerando uma aproximação com tamanho 30, a perda de qualidade nas políticas foi no máximo 8.88% e o ganho no tempo de execução foi no mínimo 43.54%, para os três algoritmos. Ademais, foram realizados experimentos combinando a técnica de aproximação e o RSTL-ILAO* num estudo de caso simplificado de COVID-19.

Agradecimentos

O presente trabalho foi realizado com apoio da CAPES (Código de Financiamento 001), da FAPESP (Processo #2018/11236-9) e do Centro C4AI-USP com apoio da FAPESP (Processo 2019/07665-4) e da IBM.

Referências

Chandak, A., Dey, D., Mukhoty, B., and Kar, P. (2020). Epidemiologically and socio-economically optimal policies via Bayesian optimization. *Transactions of the Indian National Academy of Engineering*, 5(2):117–127.

Charpentier, A., Elie, R., Laurière, M., and Tran, V.-C. (2020). COVID-19 pandemic control: balancing detection policy and lockdown intervention under ICU sustainability. *Mathematical Modelling of Natural Phenomena*, 15:57.

Diekmann, O. and Heesterbeek, J. (2000). *Mathematical Epidemiology of Infectious Diseases: model building, analysis and interpretation*. John Wiley & Son.

Elie, R., Hubert, E., and Turinici, G. (2020). Contact rate epidemic control of COVID-19: an equilibrium view. *Mathematical Modelling of Natural Phenomena*, 15:35.

Gatto, N. M. and Schellhorn, H. (2021). Optimal control of the SIR model in the presence of transmission and treatment uncertainty. *Mathematical Biosciences*, 333:108539.

Hansen, E. A. and Zilberstein, S. (2001). LAO*: A heuristic search algorithm that finds solutions with loops. *Artif. Intell.*, 129:35–62.

Libin, P. J. K., Moonens, A., Verstraeten, T., Perez-Sanjines, F., Hens, N., Lemey, P., and Nowé, A. (2021). Deep reinforcement learning for large-scale epidemic control. In *ECML PKDD. Applied Data Science and Demo Track*, pages 155–170.

Liu, L., Luo, X., and Chang, L. (2017). Vaccination strategies of an SIR pair approximation model with demographics on complex networks. *Chaos, Solitons & Fractals*, 104:282–290.

Mihatsch, O. and Neuneier, R. (2002). Risk-sensitive reinforcement learning. *Machine Learning*, 49(2):267–290.

Nasir, A. and Rehman, H. (2017). Optimal control for stochastic model of epidemic infections. In *IBCAST*, pages 278–284. IEEE.

Pastor, H. D., Borges, I. O., Freire, V., Delgado, K. V., and de Barros, L. N. (2020a). Risk-sensitive piecewise-linear policy iteration for stochastic shortest path Markov decision processes. In *19th MICAI*, pages 383–395. Springer.

Pastor, H. D., Freire, V., Barros, L., and Delgado, K. V. (2020b). Políticas sensíveis ao risco para o controle da propagação de doenças infecciosas. In *ENIAC*, pages 366–377.

Usherwood, T., LaJoie, Z., and Srivastava, V. (2021). A model and predictions for COVID-19 considering population behavior and vaccination. *Sci. Rep.*, 11(1):1–11.

Yaesoubi, R. and Cohen, T. (2011). Dynamic health policies for controlling the spread of emerging infections: influenza as an example. *PloS one*, 6(9):e24043.

Yaesoubi, R. and Cohen, T. (2016). Identifying cost-effective dynamic policies to control epidemics. *Statistics in medicine*, 35(28):5189–5209.

Yaesoubi, R., Havumaki, J., Chitwood, M. H., Menzies, N. A., Gonsalves, G., Salomon, J. A., Paltiel, A. D., and Cohen, T. (2021). Adaptive policies to balance health benefits and economic costs of physical distancing interventions during the COVID-19 pandemic. *Medical Decision Making*, 41(4):386–392.