

Núcleo Interinstitucional de Estudos Epidemiológicos (NIEE)

Nota Técnica NIEE - nº 07/2020

Estimativa do limiar da imunidade de rebanho e de cenários contrafactuais da epidemia de Covid-19 na RMGV/ES

Ms. Frederico Guilherme Nogueira (IJSN)

Ms. Antônio Rocha (IJSN)

Ms. Gustavo Ribeiro (IJSN)

Dr. Pablo Jabor (IJSN)

Dra. Ethel Maciel (UFES)

Dr. Etereldes Gonçalves Júnior (UFES)

Dr. Fabiano Petronetto do Carmo (UFES)

Ms. Hélio Gomes Filho (UFES)

Dr. Pablo Lira (IJSN)



GOVERNO DO ESTADO
DO ESPÍRITO SANTO
Secretaria de Economia
e Planejamento

1. Introdução

Nesta nota técnica será descrito o modelo de equações diferenciais ordinárias, para uma população heterogênea, utilizado na estimativa dos parâmetros associados a epidemia de Covid-19 na Região Metropolitana da Grande Vitória (RMGV) do Estado do Espírito Santo.

Complementarmente serão apresentadas as estimativas de óbitos, casos ativos e prevalência até o final da epidemia, de limiar de imunidade de rebanho, assim como, de cenários contrafactuais que evidenciam o que poderia ter ocorrido na ausência de medidas de mitigação e de imunidade prévia.

2. Motivação para criação do modelo

A razão para o desenvolvimento desse modelo em específico, representando a dinâmica da propagação da Covid-19 no Estado do Espírito Santo, surgiu do comportamento atípico observado na curva de casos ativos. Durante o monitoramento rotineiro realizado, usando o painel de indicadores descrito na Nota Técnica NIEE - nº 06/2020 [1], notou-se que o topo da curva de casos ativos foi alcançado no início de Junho de 2020 na RMGV. Muito antes do esperado.

Esse momento de chegada ao topo da curva de ativos está intimamente ligado a prevalência da doença na população. Tal valor é normalmente estimado utilizando a fórmula de cálculo de prevalência limiar para que a imunidade de rebanho (HIT - *Herd Immunity Threshold*) seja alcançada, condição também necessária para que a taxa de transmissão, R_t , atinja o valor 1,0. Usando dados preliminares do CDC americano [2] [3] da taxa de transmissão inicial, R_0 , para a Covid-19, foi estimado que a prevalência mínima de imunidade de rebanho esperada era entre 50% e 75%, como indicado nos cálculos abaixo:

Limite Inferior da Imunidade de Rebanho (IR_{inf})

$$IR_{inf} = 1 - \frac{1}{R_0} = 1 - \frac{1}{2,0} = 50\% \quad (1)$$

Limite Superior da Imunidade de Rebanho (IR_{sup})

$$IR_{sup} = 1 - \frac{1}{R_0} = 1 - \frac{1}{4,0} = 75\% \quad (2)$$

O que corresponderia a uma população ainda suscetível entre 25% a 50%. No entanto, na data 10 de Junho de 2020, próxima ao pico observado empiricamente, o valor da prevalência na população da RMGV estimada pela terceira fase do primeiro inquérito sorológico realizado pela SESA/ES (Secretaria de Saúde do Espírito Santo) com o apoio de IJSN/ES (Instituto Jones dos Santos Neves), OPAS/OMS (Organização Pan-Americana da Saúde), UFES (Universidade Federal do Espírito Santo) e HUCAM (Hospital Universitário Cassiano Antônio Moraes) era de apenas 9%.

Esse fato evidenciou que existiam fatores ainda desconhecidos que estavam influenciando a dinâmica da doença no estado e que precisavam ser estudados mais detalhadamente.

3. Hipóteses levantadas para explicar o fenômeno

Foram levantados dois tipos de hipóteses que poderiam explicar essa discrepância:

1. Erro de medição de prevalência na população;
2. Fatores desconhecidos que influenciam a dinâmica da doença de tal forma que, no momento do pico, efetivamente apenas de 25% a 50% da população ainda era suscetível.

Para analisar a primeira possibilidade foram utilizados inquéritos sorológicos realizados em outros países [4] [5]. Tais estudos apontaram a mesma tendência de prevalência abaixo do esperado e uma possível rápida perda de anticorpos por parte de algumas populações, o que talvez justificasse uma subnotificação em relação a soroprevalência. No entanto, a diferença encontrada na RMGV foi tão grande que apenas a influência de subnotificação não conseguiria justificar

tal discrepância. O que eliminaria nossa primeira possibilidade como a única razão para justificar o comportamento observado.

A segunda possibilidade contempla a existência de 3 fatores associados a heterogeneidade da população, referente a dinâmica da epidemia, listadas abaixo:

- H1: O percentual da população da RMGV seguindo rigorosamente as recomendações de isolamento e distanciamento social a ponto de temporariamente deixarem de ser suscetíveis em termos práticos. Esse fator se insere na dinâmica da epidemia via conscientização da população com os esforços das ações de comunicação governamental com o apoio da mídia e das redes sociais;
- H2: O percentual da população que possui imunidade prévia, não fazendo parte da população suscetível à doença. Tornando-se uma barreira a propagação da mesma. Essa hipótese depende de fatores relacionados ao sistema imunológico existentes a priori, como células T [6] capazes de identificar o SARS-CoV-2;
- H3: O percentual da população que ao ser exposta a uma baixa carga viral não se torna infecciosa [7], mas desenvolve imunidade ao SARS-CoV-2 [8], apesar de não produzir anticorpos em níveis detectáveis pelos testes rápidos usados no inquérito sorológico aplicado na RMGV. Já essa hipótese, além de depender do comportamento da população referente a uso de máscara e distanciamento social, também depende de fatores relacionados ao sistema imunológico.

Hip	Possível Causa	Efeito	Influencia	Duração do Efeito
H1	Informação	Mudança radical de comportamento	Grau de suscetibilidade	Temporário
H2	Células T	Imunidade prévia		Permanente

H3	Exposição a baixa carga viral	Foi exposto, não transmitiu, mas desenvolveu imunidade, apesar de não possuir anticorpos detectáveis	Grau de transmissão	
----	-------------------------------	--	---------------------	--

Tabela 1 - Três Hipóteses Levantadas

A tabela acima resume as três hipóteses levantadas cujo efeito combinado possivelmente produz o comportamento observado na dinâmica da doença. E devido a essa suposta interação entre elas, foi escolhido o *framework* de modelagem bayesiano na tentativa de explorar como elas estão relacionadas matematicamente.

4. Como modelar essas hipóteses?

Como essas hipóteses assumem uma heterogeneidade da população na dinâmica da propagação da Covid-19 dentro da RMGV, a melhor abordagem é realizar uma modelagem matemática que permita representar as mesmas de forma fidedigna e coesa.

O modelo matemático mais simples que permite tal estudo é composto por equações diferenciais ordinárias (EDOs), inicialmente proposto em 1927 por Kermack e McKendrick [9] para representar as leis epidêmicas da Peste Negra na Europa. Desde então esse modelo, chamado SIR (Suscetível - Infecioso – Recuperado, do inglês, *Susceptible - Infectious - Recovery*), se consolidou como uma ferramenta poderosa para estudar tais dinâmicas e foi escolhido como base para nosso modelo pela sua simplicidade, consistência e eficiência.

No entanto, o foco da utilização desses modelos tem sido na realização do ajuste/calibragem (*fitting*) de dados, seguido da projeção dos estados futuros do sistema durante uma epidemia, tomando como base parâmetros já conhecidos da mesma. E menos no problema inverso, que toma como base a definição da estrutura de um modelo e realiza uma estimativa de todos os parâmetros que

caracterizam a dinâmica da doença, utilizando apenas os dados observados de uma população afetada pela epidemia [10].

Como o propósito desse estudo é testar uma estrutura alternativa de modelo SIR para explicar os dados observados, a segunda abordagem é a mais adequada. Entretanto, tal exploração torna possível a identificação de diversas soluções para o mesmo conjunto de dados observados. E essa multitude de soluções não é tratada de forma adequada por técnicas tradicionais utilizadas em EDOs não lineares. Pois ela utiliza o princípio da maximização da verossimilhança para sistemas não lineares, normalmente associados a funções objetivas não convexas, cujas soluções, em sua grande maioria, representam máximos locais que dependem apenas dos valores iniciais do processo de otimização. O que não é adequado para uma exploração consistente dos intervalos de valores dos parâmetros que definem a dinâmica da doença

A alternativa para esse problema seria a utilização da técnica bayesiana MCMC (*Monte Carlo Markov Chain*) baseada na amostragem da distribuição posterior dos parâmetros. Dessa forma teríamos uma melhor estimativa dos intervalos de valores possíveis de cada parâmetro e como eles influenciam uns aos outros, permitindo explorar algum tipo de *trade off* entre as hipóteses discutidas anteriormente. Uma desvantagem dessa escolha é a complexidade inata da técnica bayesiana, quando comparada as tradicionais. No entanto, sua popularidade tem aumentado nos últimos anos, devido a surgimento de inúmeras ferramentas computacionais que permitem aplicá-la de forma prática. Por exemplo, a linguagem de programação probabilística Stan [11], selecionada para esse estudo.

5. Como implementar as hipóteses levantadas?

Normalmente ao se implementar o modelo SIR, implicitamente, assume-se como premissa a homogeneidade populacional. Por exemplo, duas características comuns a todos os indivíduos, que compõe a população estudada, seriam o grau de suscetibilidade e de transmissão. Essa simplificação é razoável para se

modelar a dinâmica de doenças infecciosas, em boa parte dos casos. No entanto, existem situações onde não é possível representar os dados observados dessa forma e a heterogeneidade não pode ser ignorada. Para o estudo da Covid-19 na RMGV, foi verificado que essa simplificação não é compatível com o processo natural de geração dos dados da epidemia na RMGV, devido a não convergência do procedimento de inferência bayesiana quando o SIR clássico é utilizado.

Não restou outra alternativa, a não ser aumentar a complexidade da estrutura do modelo, introduzindo as hipóteses discutidas anteriormente. E para isso duas abordagens são possíveis:

1. Discreta, criando novos compartimentos e alterando os valores iniciais dos compartimentos já existentes no modelo SIR. O que a torna uma opção simples de ser implementada;
2. Ou contínua, modelando os parâmetros de forma mais complexa, integrando sobre a distribuição dos parâmetros de suscetibilidade e de transmissão individuais. Conforme descrito em [12], que também evidencia a importância da heterogeneidade, afirmando que “conhecer a extensão da variação na suscetibilidade presente na população é essencial para os modelos serem preditivos”. Do contrário eles só poderão ser usados em um horizonte curto de previsão, para não produzir estimativas imprecisas.

Para esse estudo a primeira opção foi escolhida, por ser suficiente para aproximar a dinâmica apresentada pelos dados observados. Com isso, foi criado um compartimento novo (PNNT - Pessoas Que Naturalmente Não Transmitem) para representar H3 e foi alterado o valor inicial do compartimento S, para representar H1 e H2. A seguir será descrito detalhadamente a estrutura implementada para o modelo.

6. Descrição do modelo de equações diferenciais

Para representar adequadamente o percentual inicial da população suscetível, compartimento S, é preciso definir uma relação matemática entre a população estudada e a suscetível. Pode-se considerar que H1 e H2 influenciam simultaneamente a suscetibilidade da população, ou seja, uma provável sobreposição de seus efeitos. Isso ocorre, porque há pessoas que mudaram radicalmente seu comportamento, devido a epidemia, mas o efeito teria sido o mesmo na prática caso não o tivessem feito, por já serem imunes. Essa sobreposição é modelada e a população suscetível resultante é definida por

$$S_0 = (1 - H1)(1 - H2) \quad (3)$$

e é representada visualmente na figura abaixo.

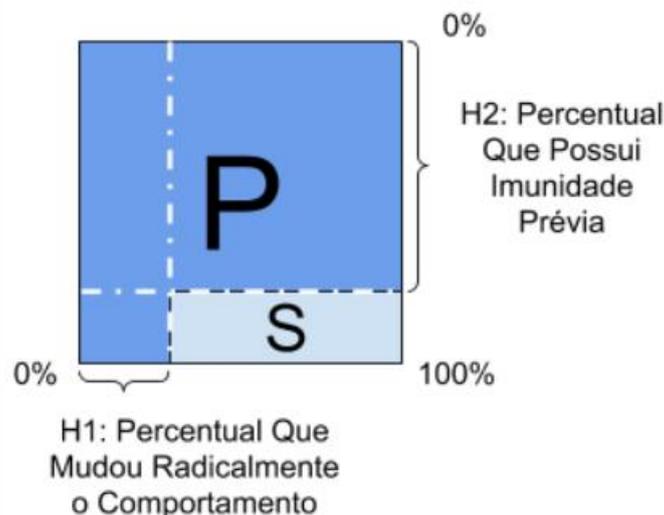


Figura 1 - Relação População x Suscetíveis

Uma vez definida a parametrização do valor inicial do compartimento de suscetíveis, agora o compartimento R será separado em três: R1 - PNNT, R2 - Inquérito Sorológico e R3 - Óbitos, para modelar H3 adequadamente e permitir a utilização de dados empíricos referentes aos inquéritos sorológicos realizados e os óbitos registrados na RMGV. Conforme indicado no diagrama abaixo.

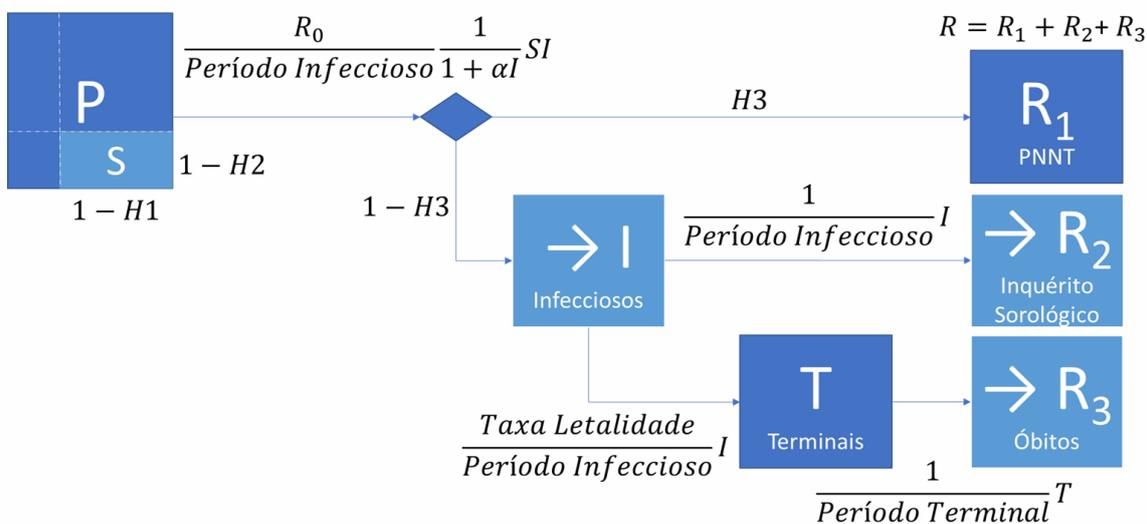


Figura 2 - Modelo SITR

O diagrama, na figura 2, representa a estrutura do modelo SITR de equações diferenciais ordinárias implementada nesse estudo. Resumidamente, ela é um modelo clássico SIR com algumas adaptações possibilitando a representação das hipóteses levantadas anteriormente e a inclusão dos dados empíricos dos casos ativos, óbitos e inquéritos sorológicos realizados.

Sua população (N) é fixa, onde $N = S + I + T + R$, onde $R = R_1 + R_2 + R_3$. No estado inicial, a população suscetível definida na equação (3), se encontra no compartimento S (Suscetíveis). O compartimento I (Infecciosos) possui um indivíduo contaminado e infeccioso, como valor inicial, para permitir que o processo de infecção se inicie. Os três compartimentos que representam R (Recuperados) e o compartimento T (Terminais) estão vazios inicialmente. Esse último modela a defasagem entre casos ativos e óbitos.

A tabela abaixo lista todos os compartimentos do modelo, onde a seta apontando para a direita indica compartimentos que possuem dados observados que devem ser usados como referência no processo de inferência dos parâmetros:

Compartimento	Significado
S (Suscetíveis)	População suscetível. Associado às hipóteses H1 e H2

→I (Infecciosos)	Pessoas Infectadas por Covid-19 que ainda podem infectar outros. Foi utilizada a curva de casos ativos subnotificados estimada em [1]
T (Terminais)	Não possui significado na realidade. Sua função é permitir o ajuste temporal entre a curva de ativos e de óbitos
R1 (PNNT)	Pessoas que naturalmente não transmitem a doença, apesar de terem sido infectadas, e não geraram anticorpos em quantidade detectável. Associado à hipótese H3
→R2 (Inquérito Sorológico)	Pessoas que já se curaram. Esses indivíduos apresentaram anticorpos detectáveis nos inquéritos sorológicos realizados
→R3 (Óbitos)	Pessoas que foram a óbito devido a Covid-19. São os óbitos extraídos do sistema e-Sus VS [1]

Tabela 2 - Compartimentos

A única forma de uma pessoa deixar o compartimento S é sendo exposta ao vírus e se infectando. Caso ela se torne contagiosa, ela vai para o compartimento I (Infecciosos), caso contrário, para o compartimento R1 (PNNT), onde as pessoas não transmitem o vírus. A taxa de saída do compartimento S é proporcional a quantidade de pessoas em S multiplicado pela quantidade de pessoas em I. Pessoas no compartimento I podem seguir dois caminhos, se recuperarem seguindo para R2 (Inquérito Sorológico), onde não podem ser mais reinfectadas, ou para o compartimento T (Terminais), onde inevitavelmente irão a óbito. Finalmente, chegando ao compartimento R3 (Óbitos). Os compartimentos T e R3 foram introduzidos para aumentar a precisão da estimativa dos parâmetros, usando os óbitos associados a Covid-19 registrados pelo sistema de saúde (e-Sus VS).

No diagrama, todos os parâmetros, utilizados para simular a dinâmica da doença, foram representados apenas usando variáveis que possuem significado na realidade, evitando explicitar parâmetros com β e γ normalmente utilizados, com o objetivo de facilitar a comunicação com o leitor.

Antes de executar o procedimento de inferência bayesiana é necessário definir a distribuição a priori de cada um dos parâmetros. Ela pode ter dois propósitos: ou dar mais ênfase ao conhecimento anterior já adquirido, definindo variâncias pequenas para as distribuições e centradas em valores anteriores que apareceram na literatura, ou dar mais ênfase aos dados, definindo uma variância maior que permite a distribuição abranger os valores que fazem sentido para os parâmetros. Como estamos propondo um novo modelo, utilizaremos a última. No lugar de usar a literatura para definir nossas distribuições a priori, a utilizaremos como forma de validação para os intervalos estimados. A lista completa dos parâmetros segue abaixo:

Parâmetros a Serem Estimados	Significado	Distribuição Normal a Priori dos Parâmetros
		$N(\text{Média; Desvio Padrão})$
H1 limites: 5% inferior e 15% superior	Percentual de pessoas que praticam o isolamento e/ou distanciamento social de forma rigorosa	$ N(1; 2) _{0,05 \text{ a } 0,15}$
H2	Percentual de pessoas que possuem imunidade prévia	$ N(1; 2) _{0,0 \text{ a } 1,0}$
H3	Percentual de pessoas que apesar de serem expostas ao vírus não transmitem	$ N(1; 2) _{0,0 \text{ a } 1,0}$
R_0 : beta * Período Infeccioso	Taxa de transmissão inicial	beta: $ N(1; 2) $

Período Infeccioso	Período no qual uma pessoa é capaz de infectar outra	$ N(12; 2) $
Subnotificação dos Casos Ativos	Representa a proporção de subnotificação ocorrida	$ N(14; 2) $
Distanciamento Social (α)	Representa o efeito do distanciamento social no fluxo de entrada do compartimento I. Quanto maior, mais achatada a curva de casos ativos fica [13]	$ N(13; 2) $
Taxa de Letalidade: $\mu * \text{Período Infeccioso}$	Percentual de pessoas infectadas pelo vírus que vão a óbito, estimado sem subnotificação	$\mu: N(1; 2) $
Período Terminal	Não tem significado na realidade, apenas permite a sincronização temporal entre a curva de casos ativos e de óbitos	$ N(1; 2) $

Tabela 3 - Parâmetros

O parâmetro distanciamento social (α) participa no termo $1/(1 + \alpha I)$ que tem como objetivo captar a mudança de comportamento da população que uma vez ciente do aumento dos casos ativos através das redes sociais e mídia, tende a aumentar proporcionalmente a prática do distanciamento social. E da mesma forma, diminuir essa prática, quando o número de casos ativos se reduz.

Das três hipóteses propostas, apenas o intervalo de possíveis valores de $H1$ é viável ser delimitado a priori, por estar relacionada diretamente ao comportamento da população. Seu intervalo teórico seria de 0% a 100%, mas imaginar que na média a maioria da população da RMGV mudou radicalmente seu comportamento durante toda epidemia não condiz com a realidade. Para

esse estudo será escolhido, de forma conservadora, um intervalo de 5 a 15% para retratar essa realidade.

7. Validação do modelo através da análise da evolução estimada do sistema comparada aos dados empíricos

Após a realizar a inferência dos parâmetros usando amostragem das distribuições posteriores do parâmetros baseada em MCMC para as 1500 possibilidades (500 por cadeia de Markov) que representam o cenário real, obtêm-se as distribuições de todos os parâmetros definidos no modelo que melhor representam os óbitos, inquérito sorológico e a forma da curva de ativos observados. Usando essas distribuições, podemos calcular o intervalo de credibilidade, versão bayesiana do intervalo de confiança que é frequentista.

Mas antes é preciso verificar se o modelo gerou curvas referentes a dinâmica da doença que são compatíveis aos dados observados. Abaixo estão três gráficos evidenciando visualmente a qualidade do *fitting* das informações.

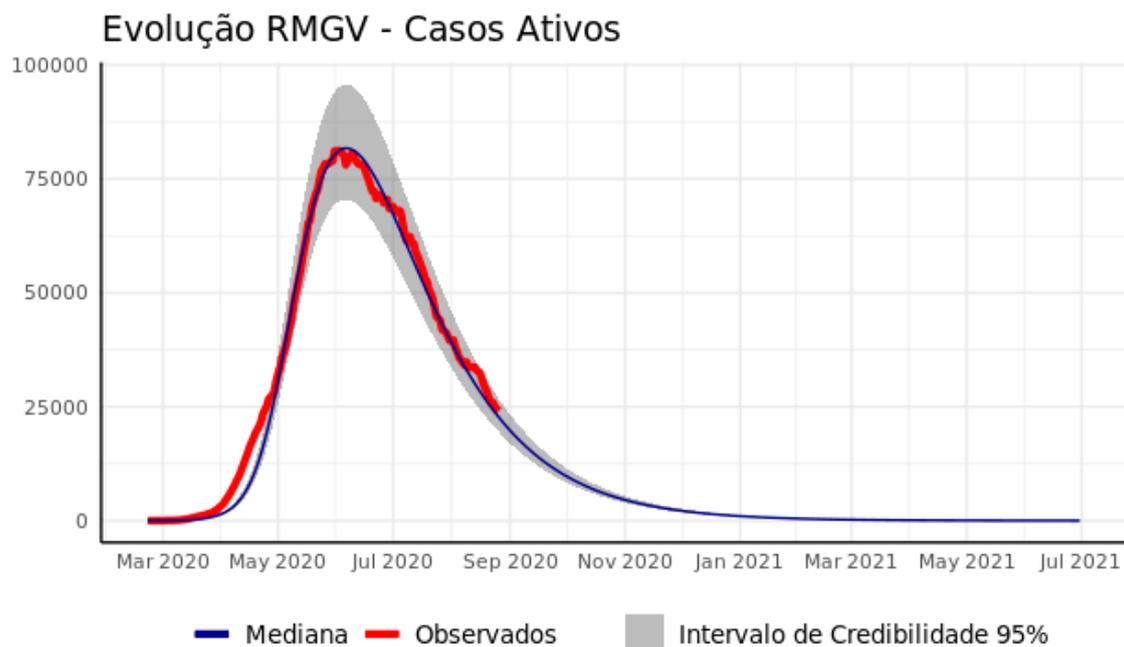


Figura 3 - Casos Ativos Ajustados

No gráfico acima, observa-se a estimativa mediana dos casos ativos ajustados na linha azul contínua e os intervalos de credibilidade de 2,5% a 97,5% nas linhas azuis pontilhadas. Os pontos vermelhos representam os casos ativos observados no painel discutido em [1], ajustado pelo modelo. O modelo estima que o valor mediano no pico da curva de casos ativos foi de 81.721 casos ativos, no dia 6 de Junho de 2020.

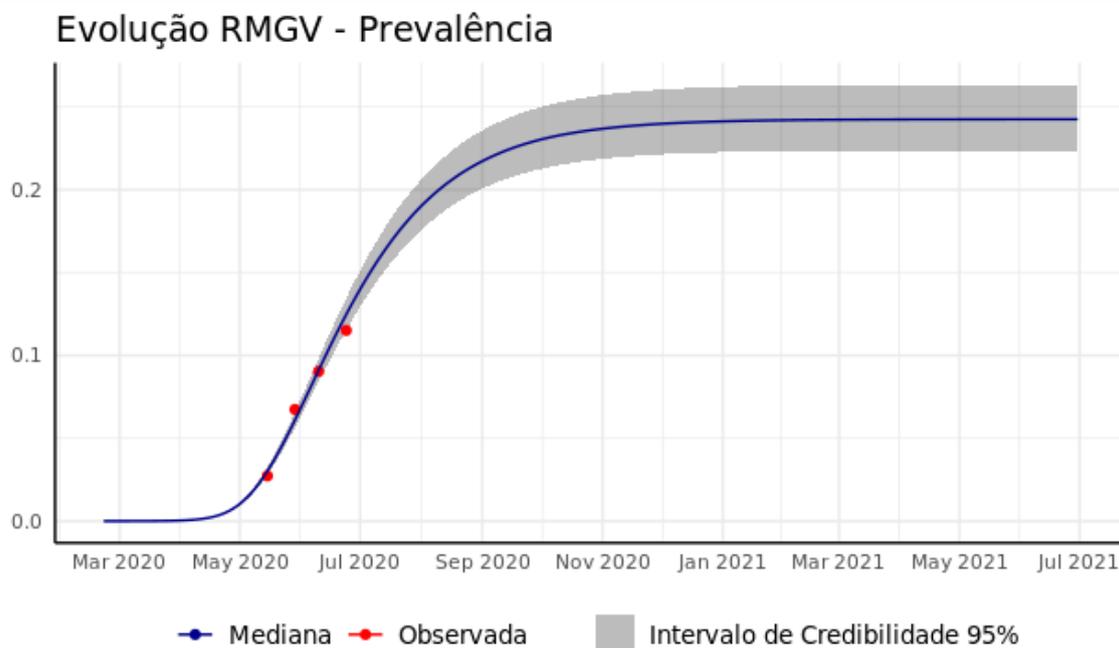


Figura 4 - Prevalência na População

Na figura 4, os pontos vermelhos representam as 4 fases do primeiro inquérito sorológico realizado na RMGV. As linhas azuis representam a evolução esperada do percentual de prevalência ao longo do tempo, com seu respectivo intervalo de credibilidade.

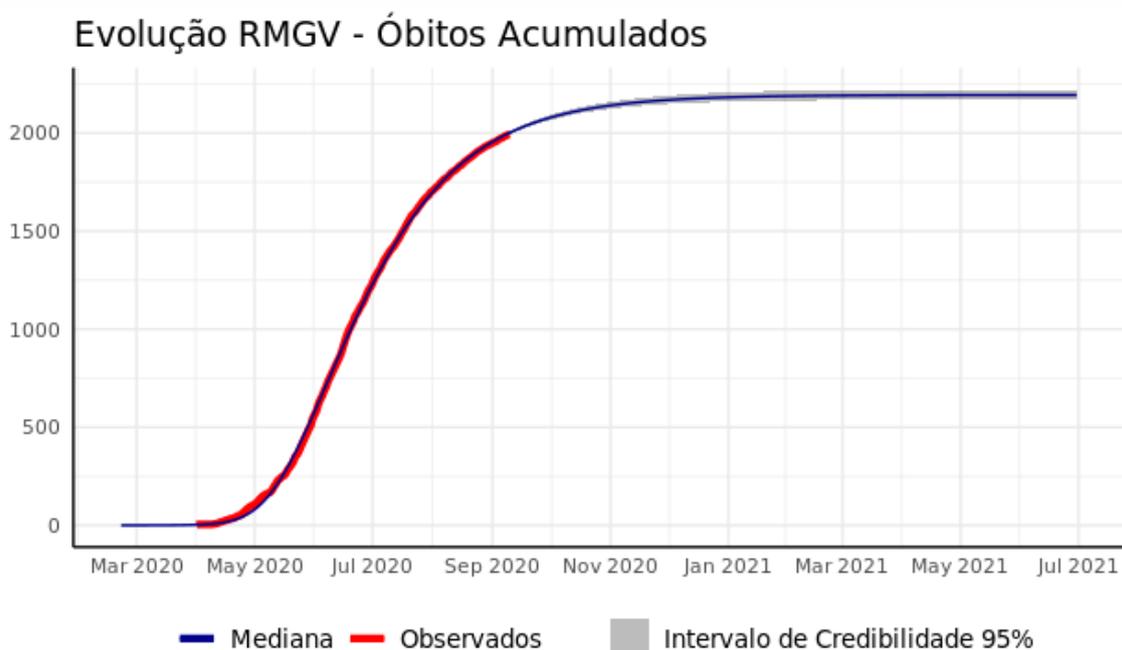


Figura 5 - Óbitos Notificados e-SUS VS

Na figura acima representando a evolução dos óbitos acumulados, percebemos que dentre as 1500 possibilidades de representação do cenário real, é o indicador que possui a menor variância, cujo valor mediano estimado é 2.182 óbitos acumulados para a data de 31 de dezembro de 2020.

Em todas as 3 figuras, os dados observados ficaram muito próximos a evolução dos indicadores estimados para as 1500 possibilidades de representação dos cenários real. Aparentemente, não havendo problema de compatibilidade entre o modelo e os dados observados. No entanto ainda é preciso analisar as distribuições dos parâmetros estimados para finalizar o processo de validação.

8. Validação do modelo através da análise dos parâmetros estimados

Os primeiros valores estimados a serem analisados serão os que motivaram esse estudo, o valor de dois compartimentos do modelo no exato momento que o pico de casos ativos foi alcançado, na data de 6 de Junho de 2020. A figura

abaixo, mostra as distribuições resultantes do processo de inferência para os dois compartimentos:

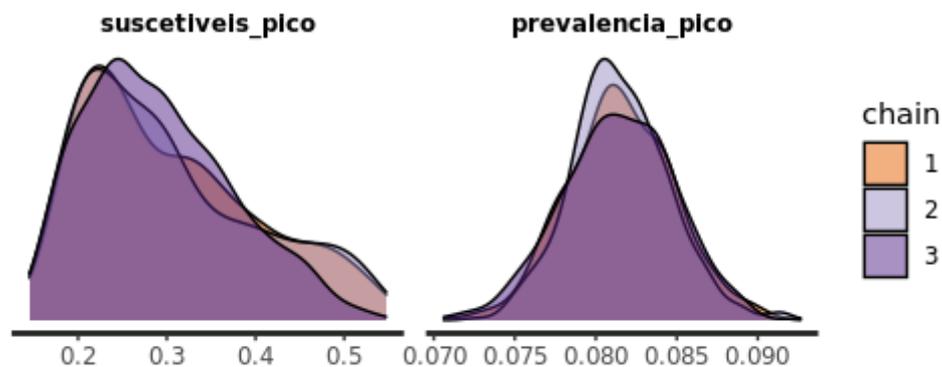


Figura 6 - Distribuição do Valor dos Suscetíveis 28% [17%; 51%] e da Prevalência 8% [8%; 9%] no Pico de Casos Ativos

A distribuição estimada dos suscetíveis, 28% [17%; 51%], é compatível ao valor teórico, calculado anteriormente, entre 25% e 50% no momento do pico e a estimativa da prevalência, 8% [8%; 9%] é compatível com o valor de 9% observado quatro dias após o pico, no inquérito sorológico do dia 10 de Junho de 2020.

Em seguida, serão analisados, utilizando a ampla literatura sobre Covid-19, os parâmetros estimados.

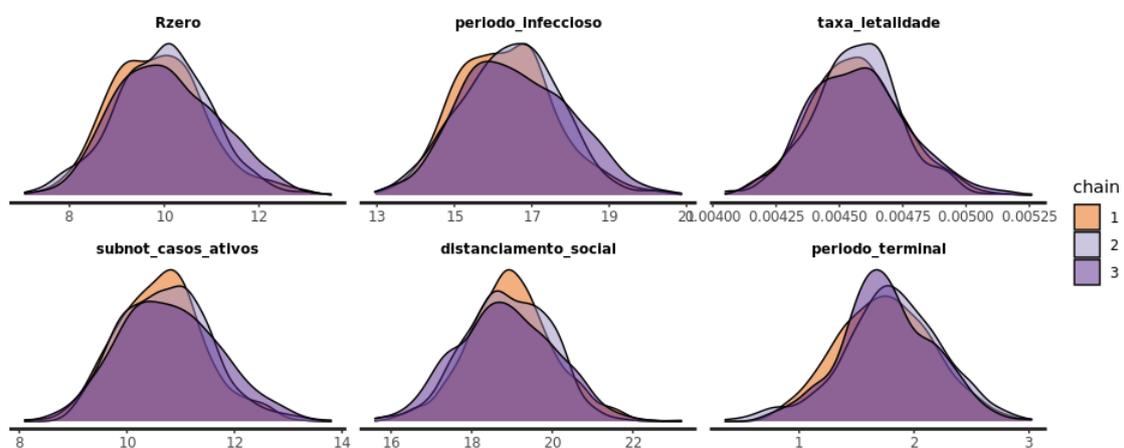


Figura 7 – Parâmetros Estimados

Parâmetro	Similar Na Literatura	Valor Médio e Intervalo de Credibilidade [2,5%; 97,5%]
R_0	2,5 [2; 4] [2]	9,9 [8,0; 12,1]
Período Infeccioso (Δ_i)	Pré-sintomático: [1; 4] Sintomático: 13,4 [10,9; 15,8] [14]	16 [14; 19]
Proporção de Subnotificação de Casos Ativos	11 [6; 24] [2]	10,7 [9,2; 12,5]
Taxa de Letalidade	0-19 anos: 0,003% 20-49 anos: 0,02% 50-69 anos: 0,5% 70+ anos: 5,4% Equivalente para RMGV: 0,43% [2] [15]	0,46% [0,42%; 0,49%]
Distanciamento Social (α)	Não Encontrado	19 [17; 21]
Período Terminal	Não Encontrado	1,8 [0,8; 2,7]

Tabela 4 - Comparação dos Parâmetros Estimados com a Literatura

Dentre todos os parâmetros estimados, apenas R_0 destoa do que normalmente se vê na literatura. No início desse documento foi utilizado a estimativa do CDC americano, R_0 entre 2,0 a 4,0, para estimar o HIT teórico. No entanto, o obtido pelo modelo, assumindo as três hipóteses, tem como valor estimado 9,9 [8,0; 12,1]. A melhor hipótese para justificar essa diferença é a utilização regular do modelo SIR clássico, supondo população homogênea, com abordagem frequentista via máxima verossimilhança. Para comparar da forma mais adequada ambos os valores, pode-se adaptar a equação de entrada do compartimento I (Infecciosos) para que se aproxime do possível modelo SIR clássico utilizado para gerar essa estimativa que se tem como referência.

Apenas o termo associado ao distanciamento social, $1/(1 + \alpha I)$, será ignorado, devido a sua dependência ao estado do compartimento I, o que faria a estimativa

de R_0 alternativo variar com o tempo. Assim, será atribuído zero ao parâmetro α para uma simplificação conservadora do exercício a seguir:

$$\text{Entrada } I(t_1) = \frac{R_0}{\Delta_i} [1 - H1][1 - H2][1 - H3] S_{100\%}(t_0) I(t_0) \quad (4)$$

onde Δ_i é o parâmetro Período Infeccioso.

Os fatores associados a H1, H2 e H3 podem ser omitidos de três formas distintas preservando a coerência do parâmetro Período Infeccioso entre as alternativas:

1. Influenciando apenas o valor estimado de R_0 , que pode ser interpretado como uma heterogeneidade no grau de transmissão da população, e mantendo a população suscetível homogênea e em 100%:

$$\text{Entrada } I(t_1) = \frac{R_{zero_h1_h2_h3}}{\Delta_i} S_{100\%}(t_0) I(t_0) \quad (5)$$

$$R_{zero_h1_h2_h3} = R_0 [1 - H1][1 - H2][1 - H3] \quad (6)$$

2. Preservando o valor de R_0 , mantendo a homogeneidade da conectividade, e reduzindo o valor da população inicialmente suscetível para menos que 100%, ou seja, assumindo uma população heterogênea em sua suscetibilidade:

$$\text{Entrada } I(t_1) = \frac{R_0}{\Delta_i} S_{<100\%}(t_0) I(t_0) \quad (7)$$

$$S_{<100\%}(t_0) = S_{100\%}(t_0) [1 - H1][1 - H2][1 - H3] \quad (8)$$

3. E finalmente supondo ambos heterogêneos:

$$\text{Entrada } I(t_1) = \frac{R_{zero_h3}}{\Delta_i} S_{<100\%}(t_0) I(t_0) \quad (9)$$

$$R_{zero_h3} = R_0 [1 - H3] \quad (10)$$

$$S_{<100\%}(t_0) = S_{100\%}(t_0) [1 - H1][1 - H2] \quad (11)$$

Na alternativa 1, supõe-se que 100% da população é suscetível inicialmente, mas ignoramos a existência da imunidade prévia e da parcela da população que apesar de ser exposta ao vírus não transmite a doença. Alternativa que supostamente se aproxima dos modelos referência do CDC americano que estimam o valor entre 2,0 a 4,0 para o R_0 .

Na alternativa 2, foi ajustada a população inicialmente suscetível para valores compatíveis as hipóteses H1, H2 e H3, preservando o valor de R_0 .

Na última alternativa, foram ajustados a população inicialmente suscetível com as hipóteses H1 e H2 e o valor de R_0 com a H3. O que equivaleria ao R_0 efetivo do modelo que proposto nesse estudo.

A figura a seguir, mostra os três valores de R_0 para as alternativas mencionadas. Valores estes calculados utilizando todas as 1500 possibilidades geradas compatíveis ao cenário real.

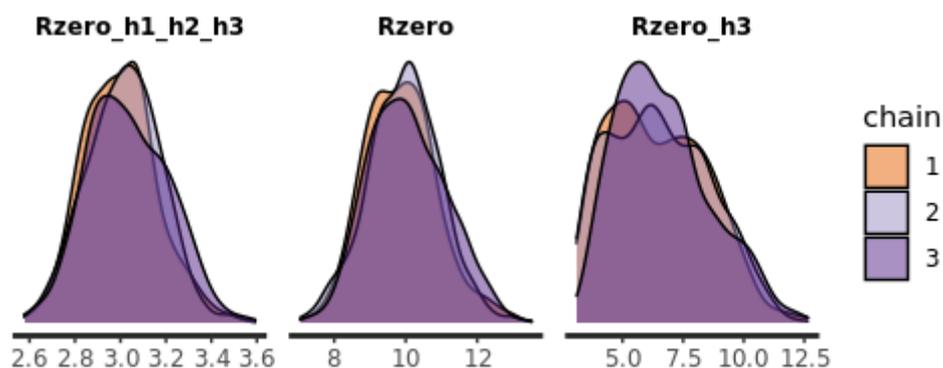


Figura 8 - R_0 's Alternativos

Parâmetro	Valor Médio e Intervalo de Credibilidade [2,5%; 97,5%]
Alternativa 1 - Rzero_h1_h2_h3	3,0 [2,7; 3,4]
Alternativa 2 - Rzero	9,9 [8,0; 12,1]
Alternativa 3 - Rzero_h3	6,3 [3,5; 10,6]

Tabela 5 - R_0 's Alternativos

Observa-se que o intervalo de credibilidade [2,7; 3,4] da distribuição de $R_{zero_h1_h2_h3}$ está contido no intervalo de referência [2,0; 4,0] do CDC americano. Por fim, justificando a aparente discrepância com a literatura.

Essa aparente discrepância surgiu devido a modelagem heterogênea da população que utilizamos para nosso estudo. Como mencionado em [12]: “em comparação com o modelo homogêneo, um R_0 mais alto é necessário para alcançar o mesmo nível endêmico.”

Assim encerramos o processo de validação. A figura a seguir representa um gráfico de dispersão entre H2 (eixo y) e H3 (eixo x), onde cada ponto representa uma das 1500 possibilidades do cenário real geradas pela inferência bayesiana, cuja cor indica o R_0 correspondente.

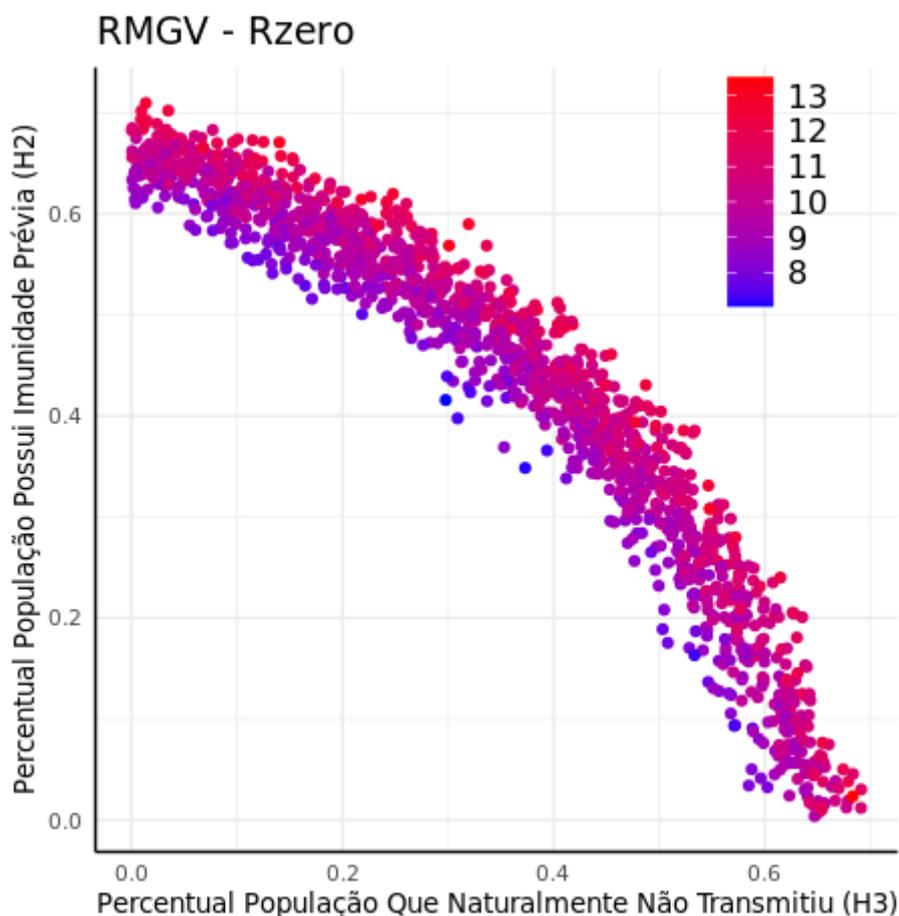


Figura 9 - Gráfico de Dispersão para H2 x H3 x R_0

Percebe-se claramente uma forte simetria entre os três parâmetros. A forma em arco evidencia um *trade-off* entre as hipóteses H2 e H3. Quando o percentual de uma é alterado, necessariamente o valor da outra deve sofrer modificação no sentido oposto, para manter as simulações condizentes com os dados observados. Como os 1500 pontos que representam possibilidades compatíveis com o cenário real do ponto de vista empírico, não é possível afirmar qual deles melhor representa o cenário real, apenas utilizando as informações empíricas disponíveis nesse estudo ou assumindo mais premissas sobre o fenômeno.

Além disso, percebe-se que os valores de R_0 estão alinhados em relação a origem do plano cartesiano, indicando que independente da linha a partir da origem que seja escolhida, a seção do arco que ela corta possui distribuições para R_0 similares.

9. Estimativa da data na qual a imunidade de rebanho é alcançada

Uma vez verificada a validade da estrutura e da inferência de parâmetros realizada, a próxima etapa é usar essas informações para estimar a data na qual o HIT será alcançado. Lembrando que o HIT é o percentual da população que já adquiriu ou está prestes a adquirir imunidade no momento do pico dos casos ativos em um cenário contrafactual, onde não houve mudança de comportamento por parte da população. Ou seja, a população continuaria se comportando como se a epidemia não estivesse ocorrendo, o que equivale a igualar a zero os parâmetros referentes ao comportamento. Utiliza-se o cenário contrafactual para estimar o HIT, pois, uma vez alcançado esse valor, o processo de retorno ao comportamento pré-epidemia, equivalente a H1, H3 e distanciamento social (α) iguais a zero, não causaria um aumento sustentável de casos ativos na população.

Iniciamos com a equação do HIT Mensurável representada pelo percentual da população nos compartimentos passíveis de mensuração no momento do pico, onde as pessoas são imunes ou se tornarão em breve:

$$HIT \text{ Mensurável} = (I + T + R2 + R3)_{no \ pico} \quad (12)$$

Como o sistema desenhado é fechado, não há inclusão, nem retirada de pessoas ao longo do tempo, tornando verdadeira a seguinte igualdade:

$$(S + I + T + R1 + R2 + R3)_{inicial} = (S + I + T + R1 + R2 + R3)_{no \ pico} \quad (13)$$

Sabendo que os compartimentos T, R1, R2 e R3 possuem valores iniciais iguais a zero e que o valor inicial de I é muito menor que o de S, temos:

$$S_0 \cong (S + I + T + R1 + R2 + R3)_{no \ pico} \quad (14)$$

$$S_0 - S_{no \ pico} - R1_{no \ pico} \cong (I + T + R2 + R3)_{no \ pico} \quad (15)$$

Em seguida, separamos a expressão acima introduzindo termos baseados na hipótese H3. Apesar de H3 ser igual a zero no cenário contrafactual, o objetivo é estimar um valor de HIT que seja a referência de comparação às medidas realizadas no cenário real, onde H3 pode ser diferente de zero, assim temos:

$$(S_0 - S_{no \ pico}) \times H3 + (S_0 - S_{no \ pico}) \times (1 - H3) -$$

$$R1_{no \ pico} \cong (I + T + R2 + R3)_{no \ pico} \quad (16)$$

Pela estrutura do modelo, todas as pessoas que saem do compartimento S devem ir para R1 ou para I. A divisão desse fluxo é determinada pelo valor de H3. Assim podemos afirmar que a variação do compartimento S multiplicado por H3, mais a variação do compartimento R1 é sempre igual a zero:

$$(S_0 - S_{no \ pico}) \times H3 + (R1_0 - R1_{no \ pico}) = 0 \quad (17)$$

Sabendo que $R1_0$ é igual a zero:

$$(S_0 - S_{no \ pico}) \times H3 = R1_{no \ pico} \quad (18)$$

Substituindo (18) em (16):

$$R1_{no \ pico} + (S_0 - S_{no \ pico}) \times (1 - H3) -$$

$$R1_{no \ pico} \cong (I + T + R2 + R3)_{no \ pico} \quad (19)$$

Removendo os termos que se cancelam:

$$(S_0 - S_{no\ pico}) \times (1 - H3) \cong (I + T + R2 + R3)_{no\ pico} \quad (20)$$

Usando (3) em (20), para H1 igual a zero:

$$\left((1 - H2) - S_{no\ pico} \right) \times (1 - H3) \cong (I + T + R2 + R3)_{no\ pico} \quad (21)$$

Substituindo (21) em (12), finalmente temos:

$$HIT\ Mensurável \cong \left((1 - H2) - S_{no\ pico} \right) \times (1 - H3) \quad (22)$$

Uma vez definida a equação que determina o HIT Mensurável, agora precisamos estimar o percentual de pessoas suscetíveis no pico do cenário contrafactual. Para calcular esse valor utilizaremos dois métodos:

1. Numérico, onde a simulação das possibilidades do cenário contrafactual irá fornecer diretamente esse valor. Uma vez que a mesma gera toda a evolução temporal para cada um dos compartimentos do modelo;

HIT Mensurável Método Numérico

$$\cong \left((1 - H2) - S_{no\ pico\ via\ simulação} \right) \times (1 - H3) \quad (23)$$

2. Analítico, onde usaremos a equação diferencial associada ao compartimento I (Infecciosos) para calcular esse valor, auferido no ponto de equilíbrio instável do sistema (pico).

HIT Mensurável Método Analítico

$$\cong \left((1 - H2) - S_{equilíbrio\ instável} \right) \times (1 - H3) \quad (24)$$

Para o método analítico, precisamos estimar o percentual de suscetíveis no equilíbrio instável, usando a equação diferencial do compartimento I, no cenário contrafactual, igualada a zero (ponto de equilíbrio):

$$\frac{dI}{dt} = \frac{R_o}{Período\ Infeccioso} SI - \frac{I}{Período\ Infeccioso} \quad (25)$$

$$\frac{R_o}{\text{Período Infecioso}} S_{\text{equilíbrio}} I - \frac{I}{\text{Período Infecioso}} = 0 \quad (26)$$

Multiplicando ambos os lados pelo Período Infecioso e dividindo por I, para obter-se apenas a solução desejada associada ao equilíbrio instável, temos:

$$R_o \times S_{\text{equilíbrio instável}} - 1 = 0 \quad (27)$$

Rearranjando a equação:

$$S_{\text{equilíbrio instável}} = \frac{1}{R_o} \quad (28)$$

E substituindo (28) em (24), tem-se finalmente:

$$\text{HIT Mensurável Método Analítico} \cong \left((1 - H2) - \frac{1}{R_o} \right) \times (1 - H3) \quad (29)$$

Com essa equação é possível estimar o HIT Mensurável via método analítico para cada uma das 1500 possibilidades representando o cenário real usando seus respectivos H2, H3 e R_o.

O gráfico de dispersão a seguir mostra como o valor estimado do HIT Mensurável nas duas metodologias, descritas acima, varia de acordo com H2 e H3 para cada uma das combinações desses parâmetros compatíveis com o cenário real:

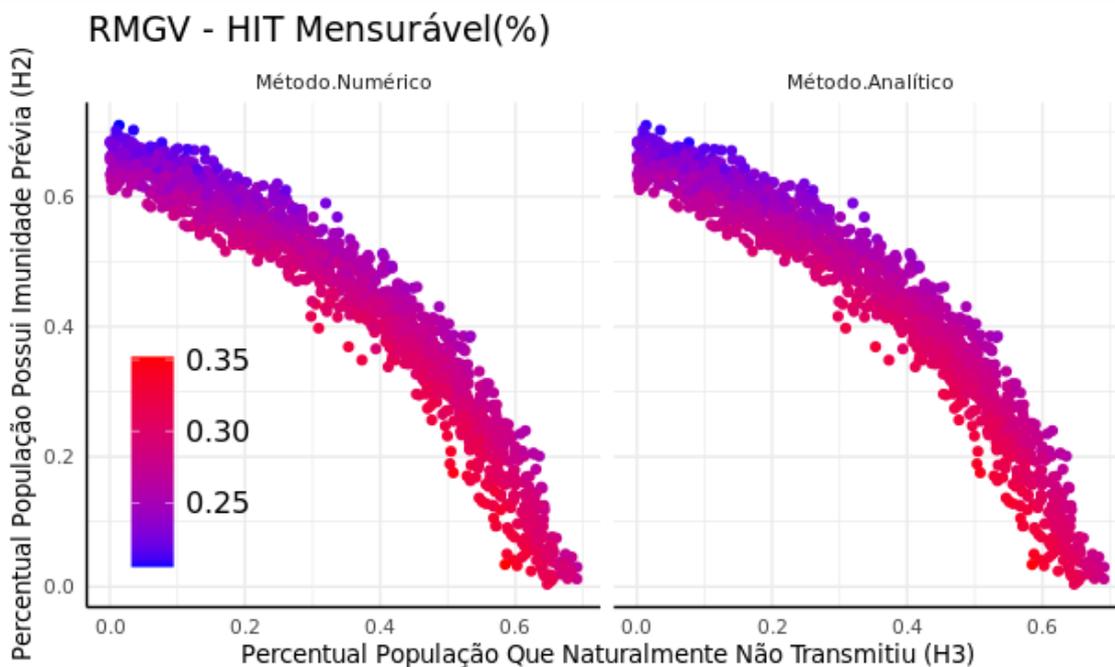


Figura 10 - HIT Mensurável x H2 x H3

Diferença entre Métodos	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
	0,00%	0,13%	0,25%	0,26%	0,37%	0,87%

Tabela 6 - Diferença Pareada do HIT Mensurável Entre Os Dois Métodos Utilizados

Nota-se resultados muito próximos para ambas metodologias, além disso percebe-se ao aumentar H3 o valor do HIT Mensurável também aumenta. Isso se deve ao *trade off* entre H2 e H3, uma vez que ter um percentual maior de pessoas que não transmitem depois de expostas (H3), implica em necessariamente ter um percentual menor de imunidade prévia, e ao analisar a população no cenário contrafactual, onde ela não mudou de comportamento, apenas H2 continua influenciando a dinâmica do sistema, uma vez que H3 é igualado a zero por depender do comportamento da população. Na próxima figura, mostramos a distribuição do valor de HIT para cada um dos métodos. E a tabela 7 resume suas estatísticas.

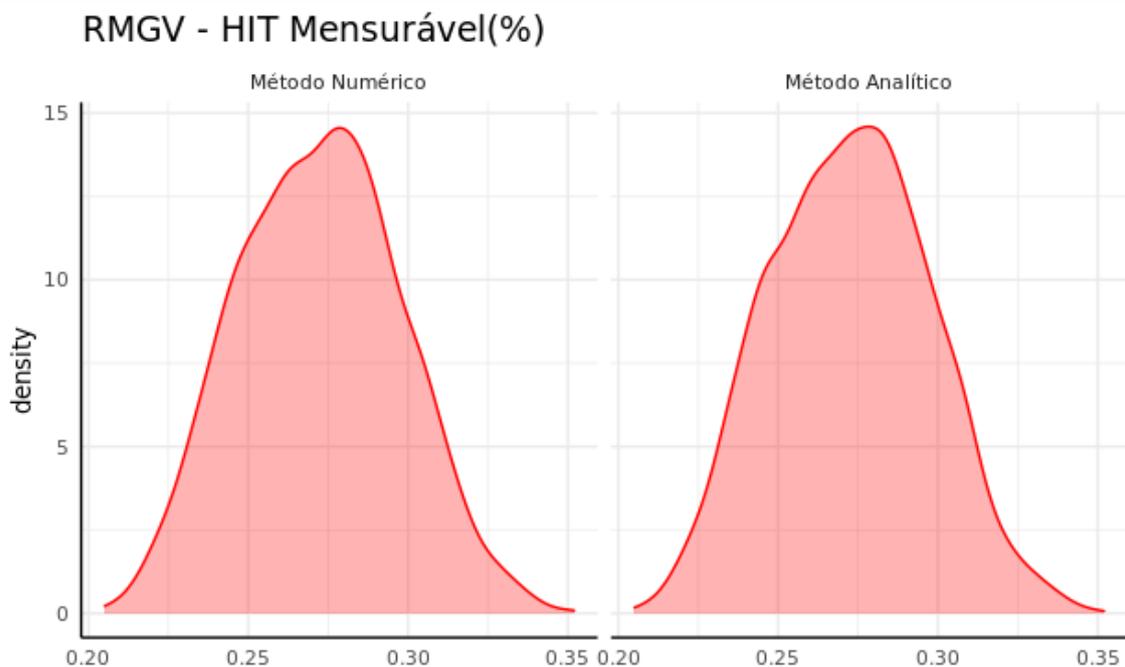


Figura 11 - Distribuição do Valor Estimado do HIT Mensurável

Método	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
Numérico	20,5%	25,4%	27,3%	27,2%	29,0%	35,2%
Analítico	20,8%	25,4%	27,3%	27,2%	29,0%	35,0%

Tabela 7 - Valor Estimado HIT Mensurável

Finalmente, o próximo gráfico exibe a distribuição dos dias nos quais o HIT será alcançado para as 1500 possíveis representações do cenário real:

RMGV - Distribuição da Data do HIT Mensurável

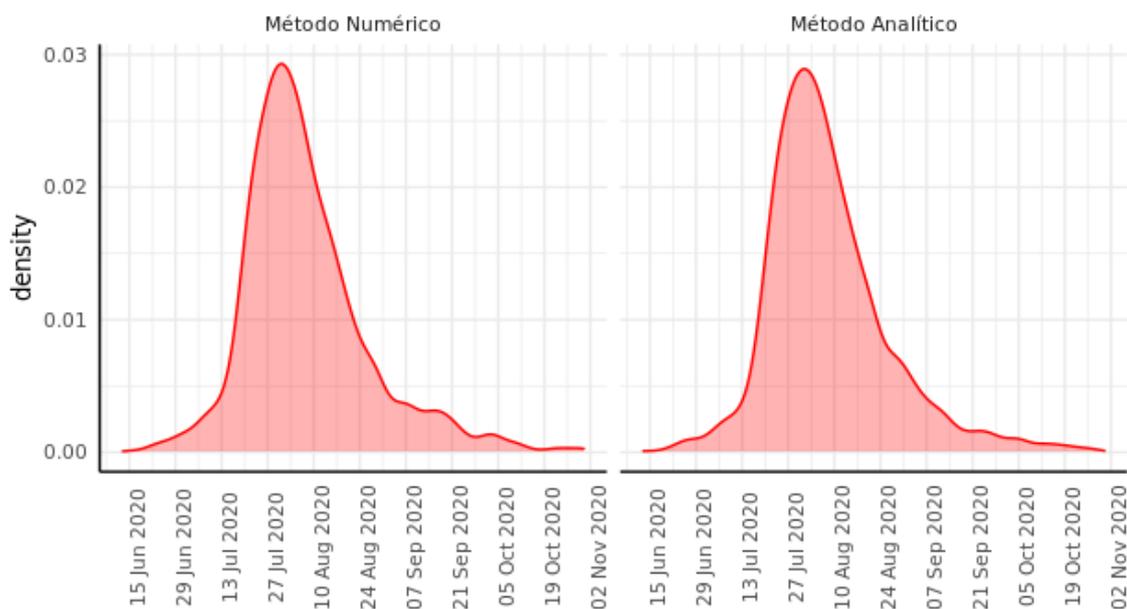


Figura 12 - Distribuição da Data Quando o HIT Mensurável Será Alcançado

Observe-se que as duas distribuições se concentram nos meses de Julho, Agosto e Setembro de 2020. A melhor forma de interpretar o gráfico acima é com o auxílio das duas seguintes tabelas, onde é evidenciado para cada mês a probabilidade do HIT já ter sido alcançado:

Mês no Método Numérico	# Possibilidades do Cenário Real	% Possibilidades do Cenário Real	Probabilidade do HIT já ter sido alcançado (%)
Junho 2020	14	0,9	0,9
Julho 2020	573	38,2	39,1
Agosto 2020	750	50,0	89,1
Setembro 2020	121	8,1	97,2
Outubro 2020	29	1,9	99,1
Novembro 2020	6	0,4	99,5

Dezembro 2020	3	0,2	99,7
Junho 2021	4	0,3	100,0

Tabela 8 - Método Numérico: Probabilidade do HIT Mensurável Já Ter Sido Alcançado No Mês Indicado

Mês no Método Analítico	# Possibilidades do Cenário Real	% Possibilidades do Cenário Real	Probabilidade do HIT já ter sido alcançado (%)
Junho 2020	14	0,9	0,9
Julho 2020	536	35,7	36,7
Agosto 2020	795	53,0	89,7
Setembro 2020	121	8,1	97,7
Outubro 2020	28	1,9	99,6
Novembro 2020	3	0,2	99,8
Dezembro 2020	2	0,1	99,9
Janeiro 2021	1	0,1	100,0

Tabela 9 - Método Analítico: Probabilidade do HIT Mensurável Já Ter Sido Alcançado No Mês Indicado

Em um processo de reabertura ideal, a data do HIT deve estar muito próxima a data de finalização da reabertura, ou seja, a data na qual toda sociedade volta ao seu comportamento normal pré-epidemia. Mas devido à natureza incerta da estimativa, a única alternativa é escolher um mês, dentro de um contexto socioeconômico, cuja probabilidade do HIT estar à frente é pequena.

10. Cenários contrafactuais

Em seguida, será analisado cenários contrafactuais nos quais alguns parâmetros estimados são igualados a zero no intuito de estimar a evolução de indicadores nesses cenários e compará-los ao cenário real.

Todos os gráficos a seguir possuem as seguintes séries temporais codificada em cores seguindo a seguinte lógica:

- Linha azul escura representa a estimativa da evolução do indicador no cenário real. Ela é a mediana dentre todas as 1500 possíveis representações do cenário real;
- Pontos vermelhos representam os dados observados que devem sempre estar alinhados com a linha azul descrita acima;
- Linha verde representa o cenário contrafactual 1, onde a população não mudou seu comportamento no dia a dia. Ela corresponde a igualar a zero os parâmetros: H1, H3 e distanciamento social (α). Ela é a mediana dentre todas as 1500 possíveis representações do cenário contrafactual 1;
- Linha amarela representa o cenário contrafactual 2, onde além de não ter ocorrido mudança de comportamento por parte da população a imunidade prévia também não existe. Ela corresponde a igualar a zero os parâmetros: H1, H2, H3 e distanciamento social (α). Ela é a mediana dentre todas as 1500 possíveis representações do cenário contrafactual 2.

O exercício de comparação entre do cenário real (linha azul) com o cenário contrafactual 1 (linha verde) é uma importante tentativa de mensurar como as medidas de mitigação adotadas pelo Governo do ES influenciaram o resultado prático obtido, com a redução da propagação do vírus SARS-CoV-2 na RMGV, conseqüentemente, salvando vidas.

A figura 12 mostra a evolução de casos ativos para a RMGV caso não houvesse mudança de comportamento da população. O pico de casos ativos do cenário real que ocorreu no dia 6 de Junho de 2020, estimado em 81.721 casos, se deslocaria para o dia 7 de Abril de 2020 e atingiria 429.020 casos. O que representa um valor 5 vezes maior. No cenário contrafactual 2, onde além de não haver mudança de comportamento da população, não existe imunidade prévia. O pico da linha amarela se localiza para 18 de Março de 2020 e com o

valor de 1.307.089 casos. O que equivaleria a 66% da população da RMGV estar infectada nessa data.

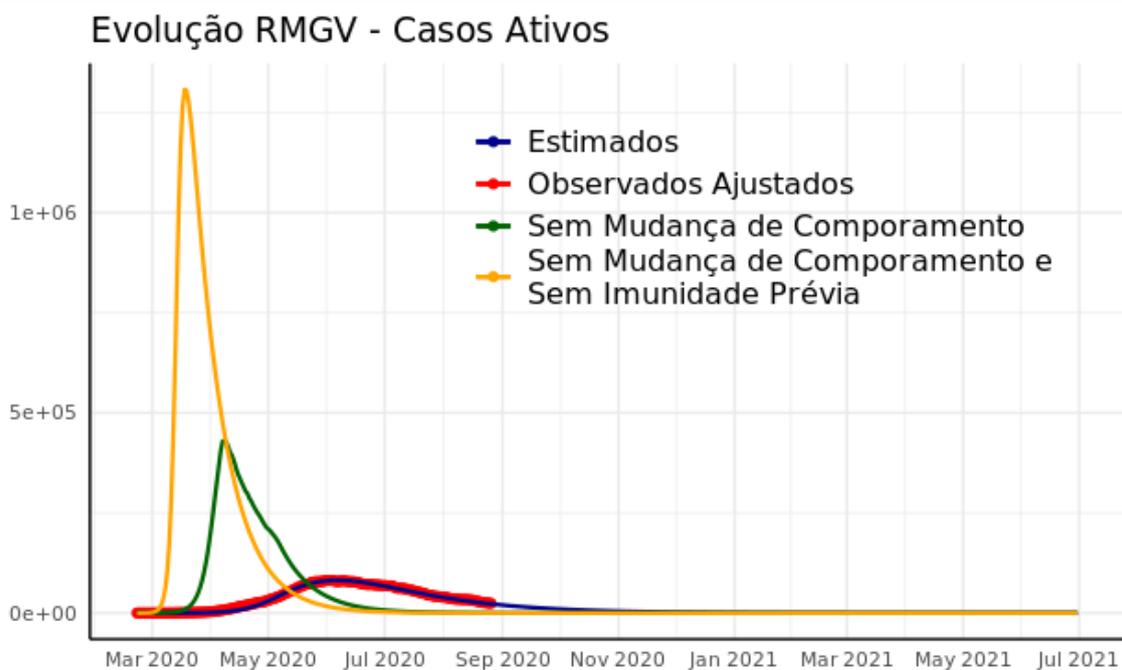


Figura 12 - Casos Ativos Sem Mudança de Comportamento e Sem Imunidade Prévia

O gráfico seguinte mostra a evolução dos óbitos para os cenários real e contrafactual 1. No dia 31 de Dezembro de 2020 estima-se que a quantidade de óbitos acumulados alcançará o valor de 2.182 óbitos para o cenário real e 4.718 para o cenário contrafactual 1, caso não houvesse mudança de comportamento. O que representa uma redução de 56% no número de óbitos atribuída a mudança de comportamento da população através do distanciamento social e da utilização de máscaras. O número de óbitos para o cenário contrafactual 2, alcançaria a marca de 8.994 óbitos no dia 31 de Dezembro de 2020.

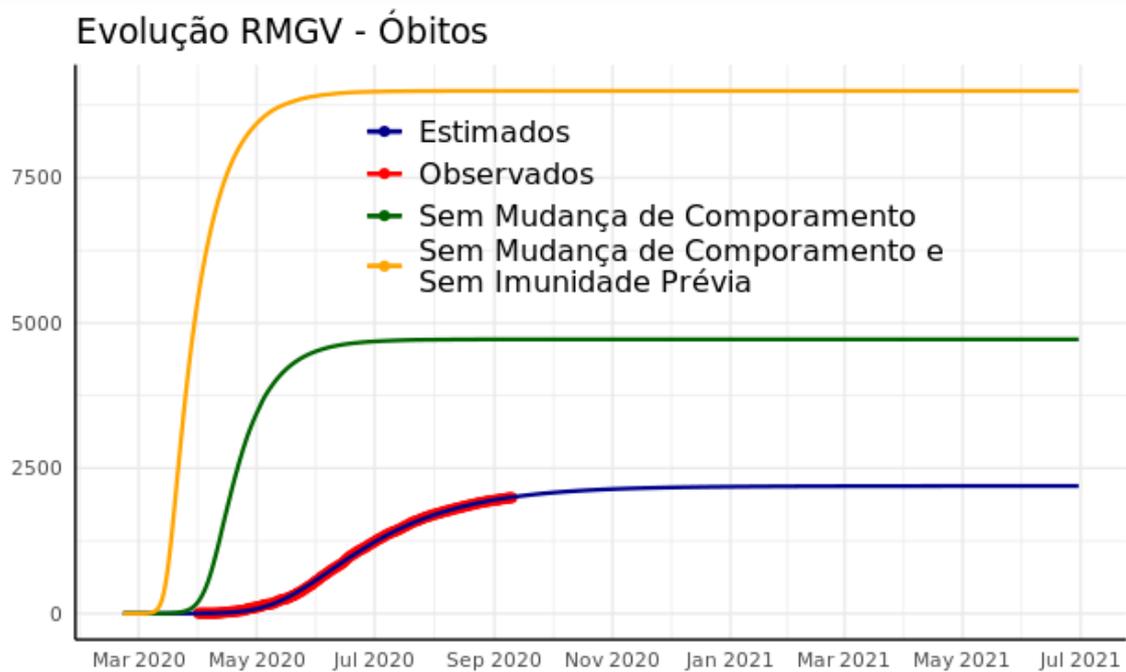


Figura 13 - Óbitos Acumulados Sem Mudança de Comportamento e Sem Imunidade Prévia

Por último os óbitos diários, para o cenário real, são estimados. Na figura 14, o pico da curva de óbitos diários foi alcançado no dia 8 de Junho de 2020 com 23 óbitos. E para o cenário contrafactual 1, dia 10 de Abril de 2020 com 125 óbitos. Percebe-se uma “nuvem” de pontos vermelhos acompanhando a linha azul no gráfico abaixo, isso se deve a variação estatística dos dados empíricos. A linha azul é a melhor representação sem ruído da informação relativa aos óbitos diários observados. No cenário contrafactual 2, o pico de óbitos diários para esse cenário seria dia 21 de Março de 2020 com 346 óbitos.

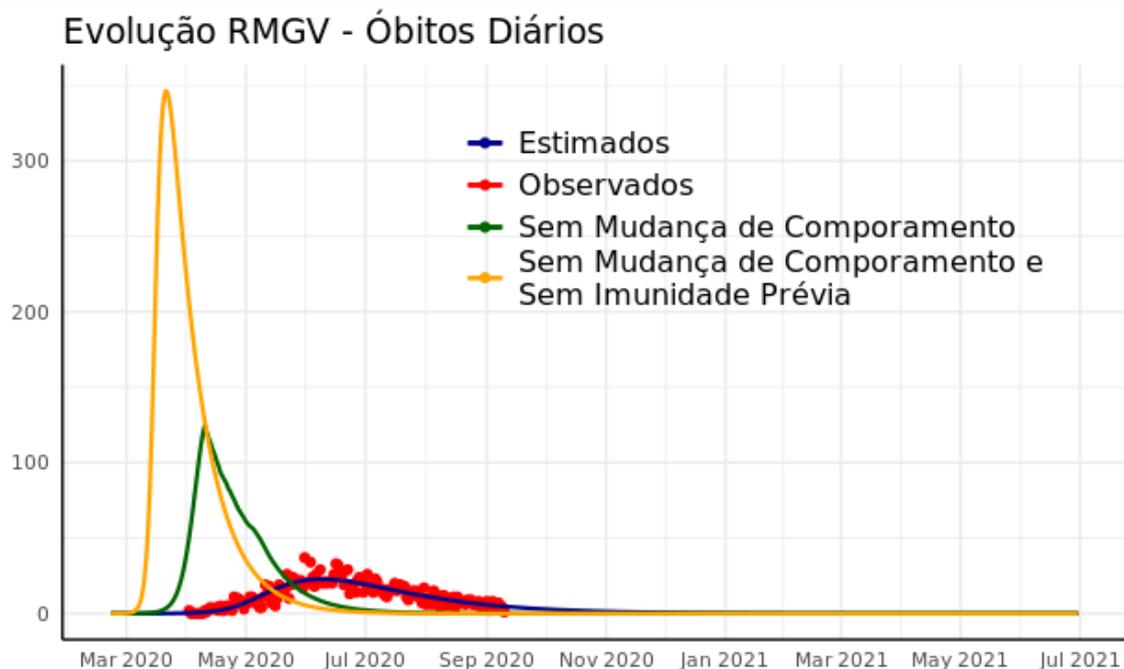


Figura 14 – Óbitos Diários Sem Mudança de Comportamento e Sem Imunidade Prévia

11. Considerações finais

Nesse estudo introduzimos novas hipóteses com o intuito de explicar o valor atípico da prevalência para o vírus SARS-CoV-2 na população da RMGV. Os parâmetros estimados também permitiram a estimativa da distribuição da data do HIT e de cenários contrafactuais que possibilitaram a mensuração aproximada do impacto das medidas de mitigação de propagação do Covid-19 na população.

No entanto, não foi possível desenvolver o mesmo estudo para a população do interior do Estado do Espírito Santo, que apesar de possuir população semelhante à da RMGV, dois fatores inviabilizam essa análise:

1. O fato do interior não possuir o mesmo grau de interconectividade, como os 7 municípios que compõem a RMGV, faz com que os municípios estejam em pontos diferentes no momento da epidemia. O que dificulta uma estimativa precisa dos parâmetros;

2. Os dados preenchidos pelos municípios do interior apresentarem problemas, desde dados faltando, a valor preenchido que não representa a realidade do paciente. O que cria ruído e atrapalha a realização das estimativas com a mesma precisão que a RMGV permite.

Bibliografia

- [1] NIEE, Núcleo Interinstitucional de Estudos Epidemiológicos, “Metodologia utilizada para estimar o número de casos ativos de Covid-19 no Estado do Espírito Santo,” 17 Julho 2020. [Online]. Available: <http://www.ijsn.es.gov.br/component/attachments/download/6989>. [Acesso em 23 Setembro 2020].
- [2] CDC, “COVID-19 Pandemic Planning Scenarios,” CDC, 10 Setembro 2020. [Online]. Available: <https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html>. [Acesso em 23 09 2020].
- [3] L. Y. X. C. R.-S. E. H. N. K. R. Sanche S, “High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2,” vol. 26, nº 7, pp. 1470-1477, 2020.
- [4] P.-G. B. P.-B. R. e. a. o. b. o. t. E.-C. S. G. Pollán M, “Prevalence of SARS-CoV-2 in Spain (ENE-COVID): A Nationwide, Population-Based Seroepidemiological Study,” vol. 396, nº 10250, pp. 535-544, 22 Agosto 2020.
- [5] R. C. L. T. e. a. Havers FP, “Seroprevalence of Antibodies to SARS-CoV-2 in 10 Sites in the United States,” *JAMA Intern Med*, 21 Julho 2020.

- [6] N. T. A. K. K. e. a. Le Bert, “SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls,” *Nature*, p. 457–462, 15 Julho 2020.
- [7] Y. a. H. X. a. T. Y. a. W. T. a. W. B. a. M. H. a. Z. W. a. Z. D. a. M. H. M. a. K. A. J. K. a. Z. M. a. X. Y. a. J. T. Li, “A Low Viral Dose in COVID-19 Patient: A Case Report,” *Frontiers in Public Health*, vol. 8, p. 339, 26 Junho 2020.
- [8] M. a. R. G. W. Gandhi, “Facial Masking for Covid-19 — Potential for “Variolation” as We Await a Vaccine,” *New England Journal of Medicine*, 08 Setembro 2020.
- [9] W. O. a. M. A. G. Kermack, “A contribution to the mathematical theory of epidemics,” *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, vol. 115, nº 772, pp. 700-721, 1927.
- [10] H. a. X. X. a. P. A. S. a. W. H. Miao, “On identifiability of nonlinear ODE models and applications in viral dynamics,” *SIAM review*, vol. 53, nº 1, pp. 3-39, 08 Fevereiro 2011.
- [11] B. a. G. A. a. H. M. D. a. L. D. a. G. B. a. B. M. a. B. M. a. G. J. a. L. P. a. R. A. Carpenter, “Stan: A probabilistic programming language,” *Journal of statistical software*, vol. 76, nº i01, 11 Janeiro 2017.
- [12] M. G. M. Gomes, “On the Mathematics of Populations,” *bioRxiv*, 26 Abril 2019.
- [13] R. Baker, “Reactive Social distancing in a SIR model of epidemics such as COVID-19,” *arXiv preprint arXiv:2003.08285*, 18 Março 2020.
- [14] A. W. a. M. D. a. C. A. B. a. H. K. a. C. M. a. B. A. a. B. F. a. G. J. a. L. E. A. a. M. C. a. O. K. a. W. P. a. W. K. Byrne, “Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available

evidence for asymptomatic and symptomatic COVID-19 cases,” *BMJ Open*, vol. 10, nº 8, 05 Agosto 2020.

[15] IBGE, “PNAD Contínua Anual 2019,” 2019. [Online]. Available: <https://www.ibge.gov.br/estatisticas/sociais/trabalho/9173-pesquisa-nacional-por-amostra-de-domicilios-continua-trimestral.html>. [Acesso em 23 09 2020].