



Universidade de Brasília — UnB
Departamento de Economia — ECO
Programa de Pós-graduação em Economia — PPGECO

JOÃO CARLOS NICOLINI DE MORAIS

Efeitos da pandemia do COVID-19:
demanda hospitalar por internações em leitos SUS e comportamento do paciente

Brasília
2021

JOÃO CARLOS NICOLINI DE MORAIS

Efeitos da pandemia do COVID-19:
demanda hospitalar por internações em leitos SUS e comportamento do paciente

Tese apresentada à Pró-reitoria de Pós-graduação e Pesquisa da Universidade de Brasília, como parte dos requisitos para obtenção do título de Doutor em Economia Aplicada no Departamento de Economia.

Área de Concentração: Análise Econômica

Linha de Pesquisa: Economia Aplicada

Orientador: Prof. Dr. Victor Gomes e Silva

Brasília
2021

Dados Internacionais de Catalogação na Publicação (CIP)

Divisão Biblioteca Central da UnB

Morais, João Carlos Nicolini de
Efeitos da pandemia do COVID-19: demanda hospitalar por internações em leitos SUS e comportamento do paciente / João Carlos Nicolini de Moraes.
Brasília, 2021.
____f.

Tese de Doutorado – Curso de Departamento de Economia. Área de Economia Aplicada –
Universidade de Brasília, 2020. Orientador: Prof. Dr. Victor Gomes e Silva.

1. Construção dos Dados. 2. Referencial Teórico. 3. Descrição do Modelo. 4. Resultados:
Efeitos da Pandemia Covid-19. 5. Resultados: variação regionais dos parâmetros estimados 6.
Conclusão. I. Departamento de Economia.
Universidade de Brasília. Divisão de Economia.

REFERÊNCIA BIBLIOGRÁFICA

MORAIS, João Carlos Nicolini de. **Efeitos da pandemia do COVID-19: demanda hospitalar por internações em leitos SUS e comportamento do paciente**. 2021. ____f. Tese de Doutorado
– Universidade de Brasília, Brasília.

CESSÃO DE DIREITOS

NOME DO AUTOR: João Carlos Nicolini de Moraes

TÍTULO DO TRABALHO: Efeitos da pandemia do COVID-19: demanda hospitalar por internações em leitos SUS e comportamento do paciente.

TIPO DO TRABALHO/ANO: Tese / 2021

É concedida à Universidade de Brasília permissão para reproduzir cópias desta Tese e emprestá-las ou vendê-las somente com propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta Tese pode ser reproduzida sem a autorização do autor.

João Carlos Nicolini de Moraes

Campus Universitário Darcy Ribeiro - Prédio da FACE

CEP 70910-900 – Brasília-DF

JOÃO CARLOS NICOLINI DE MORAIS

Efeitos da pandemia do COVID-19:

demanda hospitalar por internações em leitos SUS e comportamento do paciente

Tese apresentada à Pró-Reitoria de Pós-graduação e Pesquisa da Universidade de Brasília, como parte dos requisitos para obtenção do título de Doutor em Economia Aplicada no Departamento de Economia.

Tese aprovada em sua versão final pelos abaixo-assinados:

Presidente, Prof. Dr. Victor Gomes e Silva — ECO/UnB

Membro externo, Prof.^a Dr.^a Tatiana de Macedo Nogueira Lima — SG/CADE

Membro externo, Prof.^a Dr.^a Tainá Leandro — ANS

Membro interno, Prof. Dr. Roberto Ellery Jr — ECO/UnB

Campus Universitário Darcy Ribeiro

Brasília, DF - Brasil

2021

AGRADECIMENTOS

Antes formalizar e direcionar os meus agradecimentos, é importante destacar a simples sensação de cumprimento de uma jornada e constatar que esta tese, que parecia certa feita tão remota em sua conclusão, finalmente foi concluída. O desejo de agradecer não só forças terrenas, mas outras forças, incognoscíveis ou místicas, também me ocorre. Talvez seja fora do comum um economista pensar assim, mas é o que me veio enquanto redigia estes agradecimentos — essa sensação me parece particularmente boa.

Dado o escopo deste trabalho, eu gostaria de dar os meus agradecimentos iniciais aos profissionais da saúde, atuantes nas linhas de frente ou consolidando as informações que permitiram a análise efetuada por este trabalho. Certamente este trabalho teria demonstrado cenários mais problemáticos ou catastróficos nas internações hospitalares não fosse pela atividade desses profissionais. Ao mesmo tempo, os profissionais que atuaram na gestão e organização das informações dos dados que foram utilizados neste trabalho permitiram a existência desta tese. Em função da atividade crucial desses funcionários atuantes nas esferas federais, estaduais e municipais do governo, não houve um apagão de dados durante a pandemia do COVID-19.

Agradeço ao meu orientador, Victor Gomes e Silva, pela atenção e compreensão na construção deste trabalho. Durante o ano de 2020 e 2021, como a maioria dos brasileiros, atravessei momentos difíceis e pude contar com um apoio fundamental para a sua conclusão.

Eu também gostaria de agradecer à LCA Consultores, empresa à qual atualmente estou vinculado como associado, por me trazer discussões extremamente relevantes no tema de saúde paralelamente aos encaminhamentos desta tese — certamente isso enriqueceu o meu conhecimento sobre o setor de saúde. A empresa também me permitiu acesso a um “ultra computador” com capacidades de processamento e memória excepcionais. Esse “ultra computador” foi crucial e absolutamente essencial para a conclusão deste trabalho. Agradeço também, nominalmente, a Silvia Fagá, minha chefe direta e detentora de conhecimentos inigualáveis sobre o setor de saúde; a Anna Olimpia, também dotada de grande experiência e conhecimentos específicos do setor de saúde, além da parcimônia ao

me chefiar e da parceria no trabalho do dia-a-dia efetuado na LCA; a Gabriel Takahashi, especialista em bases de dados do setor de saúde e com quem o aprendizado também foi profícuo; e ao Leonardo Lima, recentemente integrado à empresa, que investigou o modelo de demanda implementado neste trabalho comigo. Aos outros integrantes da LCA, também não posso esquecer de Paulo Oliveira e Leonardo Makoto, que me hospedaram em um momento sensível da minha vida pessoal recente e enquanto ainda concluía este trabalho: levo a amizade e o carinho dessas novas amizades para a vida. Também não poderia esquecer de outras pessoas que me enriqueceram profissionalmente na LCA e tenho em grande conta — seguem os meus agradecimentos a André de Almeida Júnior, Jéssica Caieiro, Fernando Silveira, Matheus Escobet, Verônica Cardoso, Verônica Lameira, Lucas Mauro, Júlia Burle, João Marchi, Fernando Sakon, Henrique Vicente e Raytza Yoshimura; e aos diretores Cláudia Viegas, Gustavo Madi e Bernardo Macedo.

De outras experiências profissionais e que me acompanharam neste período de doutorando, não posso esquecer de ex-colegas do Conselho Administrativo de Defesa Econômica (CADE), como Lucas Collacino (que também posteriormente foi um parceiro na LCA Consultores), Luiza Jacobsen, Talita Miotto, Paula Bogossian, João Aurélio, Gérson Berna, Felipe Bispo e Bernardo Sordo; também destaco Luiz Esteves e Guilherme Resende, o economista-chefe antecessor e o atual da instituição CADE, com os quais aprendi imensamente e devo parte sensível de minha formação, além de Simone Cuiabano e Patrícia Sakowski, economistas-chefe adjuntas em dois momentos diferentes. Destaco também as interações que tive com a ex-conselheira do CADE, Cristiane Alkmin, e atual Secretária de Estado de Economia de Goiás, onde fui posteriormente assessorá-la, e o ex-conselheiro João Paulo Resende, com o seu esforço para pensar fora do lugar-comum dentro da autoridade. Fora do CADE, destaco a relação e aprendizado profícuo que tive com Juan Ferres, meu primeiro chefe fora de Brasília, além da amizade e enriquecimento profissional alcançados com Felipe Alvarenga, Felipe Ito Anuatti, Luiz Paulo Jorge, Igor Statonato, Ligia Lopes, Fernando Parmagnani e Andres Rojas.

Destaco também as amizades e aprendizados que obtive na Secretaria de Estado de Goiás, com a amizade especial de Danilo Caetano, Márcio Silva, Gilvan Cândido, Everton Chaves, Anderson Teixeira e Francisco Sérvulo. Todos estiveram presentes nesta trajetória, afetando as minhas posições direta ou indiretamente.

“It all means more than I can tell you. So you must not judge what I know by what I find words for.”
— **Marilynne Robinson, Gilead**

RESUMO

A dinâmica de escolha de pacientes hospitalares no Brasil não foi investigada em grande detalhe pela literatura econômica, sobretudo em uma análise ampla (nacional e regional) e de vários resultados. Assim, utilizando uma enorme massa de dados do SIH-RD (Sistema de Informações Hospitalares), estima-se uma demanda hospitalar através de um modelo de escolha discreta (regressão logística condicional ou modelo *logit* condicional), considerando também o impacto da pandemia do COVID-19. Com o modelo, encontra-se uma relação intuitiva do *odds ratio* associado à distância percorrida e uma mudança sensível do *odds ratio* associado estruturas hospitalares e diagnósticos específicos (e.g. serviços de cardíacos e diagnóstico cardíacos), entre outros resultados similares. Do ponto de vista de variação geográfica, algumas das relações se mantêm aproximadamente estáveis entre as regiões, enquanto outras estimativas demonstram-se mais instáveis regionalmente. O presente trabalho abre espaço para uma agenda de investigação de demanda hospitalar através de modelos de escolha discreta, além de variações regionais.

Palavras-Chave: Hospitais; SUS; Estimação de Demanda; Regressão Logística Condicional.

ABSTRACT

The dynamics of hospital choice by patients in Brazil has not been investigated thoroughly in the economic literature or in a broad analysis of regional or national patient behavior. Thus, using the SIH-RD dataset (*Sistema de informações Hospitalares* or “Hospital Information System”) from the Brazilian public healthcare system (*Sistema Único de Saúde – SUS*), this work estimates a discrete choice model (conditional logistic regression) in order to address such behavior, considering the period of the COVID-19 pandemic. The estimated model shows a intuitive value for the odds ratio associated with the distance covered by the patient, a change in the odds ratio associated with hospital services and specific diseases (e.g. cardiac services and cardiac diagnosis), among other unique results with the aforementioned approach. From the point of view of geographic variation, some of the relationships remain approximately stable across regions, while some other estimates are more unstable regionally. The present work makes room for a research agenda for investigating hospital demand through discrete choice models, in addition to regional variations.

Keywords: Hospitals; SUS; Demand Estimation; Conditional Logistic Regression.

LISTA DE FIGURAS

Figura 1 – Inscrições de Doenças Respiratórias no Brasil – Comparativo entre 2019 e 2020.....	25
Figura 2 - Inscrições “B34” no Brasil – Comparativo entre 2019 e 2020.....	26
Figura 3 – Casos de Covid-19 confirmados no ano de 2020 – janeiro a agosto.....	27
Figura 4 – Dados SIH-RD 2015 a 2020 - Proporção de pacientes atendidos nas próprias Regiões de Saúde (“Razão de Atendimento Local %”).....	33
Figura 5 - Exemplo Ilustrativo para motivar o uso de um filtro do tipo boxplot.....	43
Figura 6 -- Gráfico de Antes e Durante da Pandemia do COVID-19 – Coeficientes de interação com Renda (<i>odds ratio</i>).....	51
Figura 7 – Gráfico de Antes e Durante a Pandemia do COVID-19 – Coeficientes de interação com Serviços de Imagem (<i>odds ratio</i>).....	Erro! Indicador não definido.
Figura 8 Gráfico de Antes e Durante a Pandemia da COVID-19 – Coeficientes com o comportamento consistente (<i>odds ratio</i>).....	56
Figura 9 - Gráfico de Antes e Durante a Pandemia da COVID-19 – Outros coeficientes com o comportamento consistente (<i>odds ratio</i>).....	57
Figura 10 – Gráfico de Antes e Durante a Pandemia do COVID-19 por Região do Brasil – interações com Renda (Odds Ratio).....	62
Figura 11 – Gráfico de Antes e Durante a Pandemia do COVID-19 por Região do Brasil – Coeficientes de interação com Serviços de Imagem.....	67
Figura 12- Gráfico de Antes e Durante a Pandemia do COVID-19 por Região do Brasil – Coeficientes com o comportamento consistente ou de interesse.....	70
Figura A.1 – Gráfico de Antes e Durante a Pandemia do COVID-19 – Todos os valores estimados (<i>odds ratio</i>).....	84
Figura A.2 – Gráfico de Antes e Durante a Pandemia do COVID-19 – Todos os coeficientes estimados.....	85
Figura A.3 – Gráfico de Antes e Durante a Pandemia do COVID-19 – Coeficientes de interação com Renda	87

Figura A.4 – Gráfico de Antes e Durante a Pandemia do COVID-19 – Coeficientes de interação com Serviços de Imagem.....	88
Figura A.5 – Gráfico de Antes e Durante a Pandemia do COVID-19 – Coeficientes com o comportamento consistente.....	89
Figura A.6 – Gráfico de Antes e Durante a Pandemia do COVID-19 – Coeficientes adicionais.....	90
Figura A.7 – Gráfico de Antes e Durante a Pandemia do COVID-19 por Região do Brasil – Coeficientes de interação com Renda.....	93
Figura A.8 – Gráfico de Antes e Durante a Pandemia do COVID-19 por Região do Brasil – Coeficientes de interação com Serviços de Imagem.....	94
Figura A.9 – Gráfico de Antes e Durante a Pandemia do COVID-19 por Região do Brasil – Coeficientes com o comportamento consistente.....	95
Figura A.10 – Gráfico de Antes e Durante a Pandemia do COVID-19 por Região do Brasil – Coeficientes Adicionais.....	96
Figura A.11– Gráfico de Antes e Depois da Pandemia do COVID-19 por Região do Brasil – Coeficientes Adicionais.....	97
Figura A.12 – Análise com pareando pacientes – Análise antes e depois com os mesmos tipos de Pacientes com o mês julho.....	115

LISTA DE TABELAS

Tabela 1 - Dados de CEPs em que foi possível assinalar uma informação	21
Tabela 2 - Definição dos Grupos de Serviços	23
Tabela 3 - Tabelas com os diagnósticos utilizados.....	24
Tabela 4 - Estatísticas de Sumário das variáveis utilizadas para a regressão.....	28
Tabela 5 – Média de Distâncias Percorridas por Pacientes Internados – 2015-202032	
Tabela 6 – Percentual de Regiões de Saúde que tiveram pelo menos um resultado convergente antes e depois da Pandemia do COVID-19	47
Tabela 7 – Regressões que convergiram <i>antes</i> e <i>depois</i> nas Regiões de Saúde	48
Tabela 8 – Percentual de Regiões de Saúde que tiveram pelo menos um resultado convergente antes e depois da Pandemia do COVID-19	49
Tabela 9 – Antes e Durante a Pandemia do COVID-19 por Região do Brasil – interações com Renda (Odds Ratio) – Coeficientes, Desvio Padrão e Odds Ratio ...	52
Tabela 10 – Antes e Durante a Pandemia do COVID-19 – Coeficientes de interação com Serviços de Imagem – Coeficientes, Desvio Padrão e <i>Odds Ratio</i>	55
Tabela 11 Antes e Durante a Pandemia do COVID-19 – Coeficientes com o comportamento consistente - Coeficientes, Desvio Padrão e Odds Ratio.....	59
Tabela 12 – Antes e Durante a Pandemia do COVID-19 por Região do Brasil – interações com Renda (Odds Ratio) – Coeficientes, Desvio Padrão e Odds Ratio ...	63
Tabela 13 –Antes e Durante a Pandemia do COVID-19 por Região do Brasil – Coeficientes de interação com Serviços de Imagem – Coeficientes, Desvio Padrão e <i>Odds Ratio</i>	68
Tabela 14 - Antes e Durante a Pandemia do COVID-19 por Região do Brasil – Coeficientes com o comportamento consistente e outros coeficientes – Coeficientes, Desvio Padrão e <i>Odds Ratio</i>	71
Tabela A. 1- Resultados completos – análise nacional	81
Tabela A. 2 – Coeficientes Significantes estimados.....	99

Tabela A.3 – Variáveis Disponíveis para acesso com a base SIH-RD	116
Tabela A. 4 – Resultados da Regressão Logística com os dados obtidos para a Cidade de Fortaleza	121

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Objetivo	14
1.2	Organização do trabalho	15
2	CONSTRUÇÃO DOS DADOS.....	17
2.1	Dados do Sistema DATASUS	18
2.1.1	Uso do Sistema de Informações Hospitalares (SIH).....	18
2.2	Construção de informações georreferenciadas.....	19
2.3	Construção final dos dados.....	21
2.3.1	Construção dos dados com base no exercício proposto.....	22
2.3.2	Definição geográfica da abrangência da estimação de demanda.....	29
3	DESCRIÇÃO DOS MODELOS	35
3.1	Descrição do modelo <i>logit</i> para n alternativas.....	35
3.1.1	A descrição de um modelo <i>logit</i> generalizado.....	35
3.1.2	O modelo <i>logit</i> multinomial	37
3.1.3	O modelo <i>logit</i> condicional.....	37
3.2	Breve resumo da maximização de utilidade do paciente e o modelo utilizado para a estimação de demanda de hospitais	38
3.3	O modelo utilizado para estimação de demanda de hospitais	39
3.4	Problemas de convergência no modelo <i>logit</i> condicional.....	40
3.5	Aplicação de um filtro <i>boxplot</i>	41
3.6	Meta-análise nacional e regional.....	44
3.7	Escolha do paciente e interpretação dos resultados	44
4	RESULTADOS: EFEITOS DA PANDEMIA DO COVID-19.....	47
4.1	Convergência e capacidade explicativa dos modelos	47
4.2	Análise dos parâmetros estimados	49

5	RESULTADOS: VARIAÇÕES REGIONAIS NOS PARÂMETROS ESTIMADOS.....	60
5.1	Análise do impacto regional	60
6	CONSIDERAÇÕES FINAIS E CONCLUSÃO.....	73
	REFERÊNCIAS	77
	APÊNDICE A – DETALHES DAS VARIÁVEIS UTILIZADAS: TODOS RESULTADOS E COEFICIENTES ESTIMADOS	81
A.1	Análise completa	81
A.2	Análise agregada dos coeficientes	86
A.3	Análise regional	92
A.4	Significância estatística dos coeficientes estimados	98
	APÊNDICE B – ROTINAS	102
B.1	Rotina da regressão em <i>loop</i>	102
B.2	Extração dos coeficientes e valores relevantes.....	108
B.3	Construção dos resultados	111
	APÊNDICE C – ANÁLISE <i>ANTES E DEPOIS</i> DE BEM-ESTAR	113
C.1	Resultados preliminares.....	114
	APÊNDICE D – DETALHAMENTO SOBRE A BASE SIH-RD.....	116
	APÊNDICE E – ANÁLISE DE FORTALEZA	119
E.1	Exceções da Análise de Fortaleza.....	119

1 INTRODUÇÃO

1.1 Objetivo

A demanda por serviços hospitalares é um tópico de considerável relevância no Brasil por diversas razões, a julgar pelo envelhecimento da população,¹ impactos no orçamento² e capacidade da infraestrutura do SUS (Sistema Único de Saúde).³ Ainda que existam questões particulares a uma análise com dados do SUS⁴, há pouca exploração do tema de escolha de pacientes do hospital e as características que afetam essa seleção, sobretudo com dados brasileiros e por meio da estimação de um modelo econométrico de demanda. Resumidamente, essa relativa ausência é a motivação da investigação implementada neste trabalho.

A escolha de pacientes, associada a características hospitalares (e.g. número de leitos, número de médicos), é um importante parâmetro para identificação de características de bem-estar e potenciais gargalos do setor de saúde. Ainda, uma investigação ampla — para diferentes localizações do país, por exemplo — poderá informar detalhes importantes sobre a heterogeneidade regional da demanda hospitalar. Novamente, reitera-se que há poucos trabalhos, no Brasil, de investigação regional no setor de saúde, sobretudo em relação às preocupações de disparidades ou heterogeneidades de escolha. O presente trabalho também busca suprir parcialmente essas lacunas na produção científica brasileira.

Ademais, a pandemia do COVID-19 revela um *stress particular* da estrutura hospitalar, o que motiva também a investigação deste trabalho. Embora seja possível discutir que a pandemia forneça uma variação exógena para uma identificação econométrica por meio de diferentes estratégias, este trabalho busca meramente a

¹ Reis, Noronha e Wajnman (2016) demonstram um impacto relevante dos gastos no SUS (Sistema Único de Saúde).

² Saúde compõe uma parcela grande do orçamento do governo brasileiro, seja nas esferas da União, Estados ou municípios. Conforme o Art. 198 da Constituição Federal de 1988, há previsões constitucionais de valores mínimos que devem ser aplicados em todas essas esferas e que compõem uma parte substancial do orçamento — no caso da União, por exemplo, o valor mínimo era de 15% da Receita Corrente Líquida da União. Em dados mais gerais, os gastos públicos no Brasil com o SUS giram em torno de 3,9% do PIB, sendo possivelmente insuficientes conforme Figueiredo et al (2018). Trata-se de valores expressivos e que necessitarão de informação para a melhor gestão no futuro.

³ O SUS (Sistema Único de Saúde) foi criado a partir da Constituição Federal de 1988, previsto pelo art. 195. Em suma, o sistema engloba todo o conjunto de ações do sistema público de saúde do Brasil.

⁴ Este trabalho desenvolve uma pequena subseção sobre o espaço de escolha do paciente e que de fato pode ser interpretado assim com dados do SUS — “Escolha do paciente e interpretação dos resultados”, no capítulo de descrição dos modelos.

investigação de um modelo estrutural de demanda, observando o comportamento do modelo antes da pandemia e após o seu início⁵. Contudo, a análise desse comportamento não deixa de ser informativa e permite auxílio para outras investigações.

Por fim, momentos como o da atual pandemia, embora raros, poderão se repetir no futuro e gestores devem estar preparados em relação à conduta de pacientes e infraestrutura hospitalar. Este trabalho objetiva oferecer um conjunto amplo de resultados, que descrevam tendências de comportamento do paciente à procura de internação hospitalar.

À frente, detalha-se a organização deste trabalho.

1.2 Organização do trabalho

O trabalho está separado nos seguintes tópicos: esta introdução (“Introdução”), o capítulo 2 (“Construção dos dados”), o capítulo 3 (“Descrição dos modelos”) o capítulo 4 (“Resultados: Efeitos da pandemia do COVID-19”), o capítulo 5 (“Resultados: Variações regionais nos parâmetros estimados”), a conclusão desses resultados (“Considerações finais e conclusão”) e um Apêndice⁶ contendo informações e resultados adicionais, além do histórico de investigações implementadas em outras fases da análise, como uma estimação do modelo de demanda para a cidade de Fortaleza-CE e outras investigações adicionais, como uma análise de bem-estar.

O capítulo 2 descreve a coleta e a construção da base de dados efetuada para este trabalho, considerando todo o desafio do processo: análise da base SIH-RD do DATASUS, a construção de informações georreferenciadas de pacientes apropriadas para o presente exercício e a consolidação final dos dados de forma a adaptá-los adequadamente para o presente exercício. Trata-se da parte mais intensiva e

⁵ A demanda dos pacientes é referenciada pela mudança nos protocolos de atendimento e referência do sistema e dos médicos. Ou seja, o modelo de demanda captura mudança nas preferências, mas também mudanças na organização do referenciamento do sistema de saúde. Portanto, o choque da pandemia não permite identificar de fato e isoladamente efeitos na demanda. Ver Ho e Pakes (2014).

⁶ Dividido em 5 grupos: Apêndice A, B, C, D e E.

desafiadora deste trabalho: a construção da informação e adaptação da base demorou meses, com o fim de adequar os dados à literatura disponível.⁷

O capítulo 3 descreve os modelos de escolha discreta, utilizados no âmbito de estimação de um modelo de demanda. O capítulo descreve brevemente um modelo *logit* genérico para eventualmente detalhar o modelo *logit* condicional, que é o aplicado no presente trabalho. Em seguida o capítulo trata das particularidades do modelo aplicado no contexto de uma estimação de demanda hospitalar e, eventualmente, os problemas associados à convergência do modelo. O capítulo termina discutindo filtros para eliminar esses problemas de convergência e as agregações implementadas, com o fim de apresentar os parâmetros médios em um contexto nacional ou regional.

O capítulo 4 descreve os resultados em função da pandemia do COVID-19, detalhando os resultados antes e durante a pandemia dos coeficientes obtidos. O foco é nos coeficientes que se demonstraram significantes, e os resultados são demonstrados apenas para esses casos. Alguns coeficientes apresentam pouca mudança durante a pandemia, como os coeficientes associados à renda. Outros coeficientes, como as interações entre diagnósticos específicos e estruturas hospitalares (e.g. diagnóstico de cardíacos) sofrem alterações substanciais.

O capítulo 5 descreve a análise regional dos parâmetros estimados, pelos mesmos aspectos destacados no capítulo anterior deste trabalho. Algumas variações se sustentam regionalmente, o que é uma validação das tendências encontradas para a agregação nacional do exercício.

O capítulo 6 conclui com os resultados aqui apresentados e análises que poderão seguir este presente trabalho. O Apêndice final deste trabalho segue adiante, com informações adicionais dos resultados, dados sobre a robustez dos resultados e outros exercícios adicionais, como a análise preliminar para a cidade de Fortaleza-CE e uma breve investigação sobre efeitos de bem-estar.

⁷ O conjunto de dados disponível em Ho (2006) ou Ho e Lee (2017) não estava imediatamente disposto para a estimação efetuada neste trabalho e, no contexto de dados brasileiros, necessitava da “junção” de um conjunto relevante de dados. Especificamente dados de distância entre o hospital e o paciente internado podem ser bastante desafiadores.

2 CONSTRUÇÃO DOS DADOS

Este trabalho construiu uma grande base de dados e que necessitou de diversas fases para chegar — de maneira consistente — na estrutura da presente análise. Em resumo, esse período envolveu identificar as fontes de dados mais adequadas para a análise, validá-las e consolidá-las de acordo com o exercício proposto para esta tese. É possível separar esse período de construção dos dados em três fases diferentes, que servirão para apresentar as hipóteses utilizadas para a construção dos dados e a natureza deles.

A primeira fase da análise consistiu na construção e seleção dos dados do sistema pelo Departamento de Informática do Sistema Único de Saúde ou DATASUS, criado três anos após a surgimento do Sistema Único de Saúde (SUS) a partir da Constituição Federal de 1988⁸ e pensado essencialmente como um bando de dados administrativo.⁹ Buscou-se identificar quais dados disponíveis, dentro do sistema DATASUS, seriam adequados para a estimação de uma demanda hospitalar.

A segunda fase buscou complementar a base de dados com outras bases, ligando os dados do sistema DATASUS a outras bases de dados, como o CENSO (2010) georreferenciado e a estruturação de informações de latitude e longitude com base no CEP obtido das bases do sistema DATASUS, o que permitiu a construção de informações de renda e distância.

A terceira fase foi finalmente a manipulação dos dados, de forma a se encaixar da maneira mais adequada à presente análise os dados construídos, compreendendo a natureza do objetivo deste trabalho descrito na seção anterior. A fase dividiu-se em duas etapas: (i) a adaptação dos dados à estrutura dos trabalhos de Ho (2006) e Ho e Lee (2017); e (ii) a definição geográfica da análise, considerando o desejo de investigar possíveis variações regionais dos modelos.

⁸ Previsto conforme o art. nº 195 da Constituição Federal de 1988.

⁹ A base mais antiga do sistema são as informações constituídas a partir dos laudos das AIHs (Autorização de Internação Hospitalar). Tais dados permitiram a construção de um banco de dados administrativo, que resultou na SIH-RD disponibilizada pelo DATASUS (BRASIL, 2021a).

2.1 Dados do Sistema DATASUS

Os dados brasileiros que permitem a presente análise são os dados processados pelo Departamento de Informática do Sistema Único de Saúde ou DATASUS.¹⁰ O sistema foi efetivamente criado a partir do decreto nº 100 de 16/04/1991, em sequência à criação do SUS (Sistema Único de Saúde). A missão DATASUS, descrita com grande extensão na apresentação do departamento no site do Ministério da Saúde,¹¹ envolve processar e manter as bases necessárias para gestão do sistema de saúde além de definir diretrizes, normas e procedimentos na consolidação desses dados — entre outras competências.

Há duas principais fontes de dados hospedadas pelo sistema DATASUS e utilizadas neste trabalho: os dados SIH-RD (ou SIH reduzida)¹² e as variadas bases CNES (Cadastro Nacional de Estabelecimentos de Saúde).¹³ Ambas as bases possuem uma riqueza amplamente detalhada de informações e merecem uma justificativa exaustiva do porquê de seu uso neste trabalho.

2.1.1 Uso do Sistema de Informações Hospitalares (SIH)

O SIH contém as informações produzidas pelos formulários AIH (Autorização de Internações Hospitalares),¹⁴ que são todas internações financiadas pelo SUS. O sistema é o mais antigo de registros do DATASUS,¹⁵ com a criação da AIH com uma motivação puramente administrativa: trata-se da geração de relatórios para que gestores possam pagar os estabelecimentos de saúde conveniados ao SUS.

Dentre as variáveis relevantes que foram utilizadas direta ou indiretamente em Ho (2006) e Ho e Lee (2017), a base SIH-RD descreve parte relevante delas e reúne

¹⁰ O site do DATASUS possui uma extensa descrição do histórico e atribuições do sistema (BRASIL, 2021b).

¹¹ Ver nota 3.

¹² Um detalhamento geral da base SIH-RD, baseada nas autorizações de internação hospitalares (AIH), pode ser encontrada no site do Sistema de Informações Hospitalares do DATASUS (MINISTÉRIO DA SAÚDE, 2021a).

¹³ O Ministério da Saúde hospeda diversas bases relacionadas ao Cadastro Nacional de Estabelecimentos de Saúde ou CNES. Há bases hospitalares (CNES-ST), há bases de profissionais (CNES-PF), bases de equipamentos (CNES-EQ), entre outras bases (BRASIL, 2021c).

¹⁴ É a mais antiga base de dados do DATASUS, possuindo todo o detalhamento necessário para a cobertura financeira das internações (BRASIL, 2010a).

¹⁵ Ver nota 7.

características dos hospitais, o CEP dos pacientes, a data de admissão do paciente, o diagnóstico dos pacientes, idade e gênero.

A extração da base neste trabalho foi efetuada com o pacote *microdatasus*, disponível no software R (SALDANHA, 2019). No Apêndice B deste trabalho providenciamos o código de extração dos dados e um dicionário das variáveis disponíveis na base na SIH-RD.

2.2 Construção de informações georreferenciadas

O presente trabalho identificou as localizações dos hospitais com base nos CEPs dos estabelecimentos hospitalares e as localizações dos pacientes internados com os seus CEPs — em ambos os casos, utilizando as informações da base SIH-RD. Assim, foi possível definir um conjunto de CEPs para obtenção de dados de longitude e latitude que potencialmente poderia ligar a outras bases de dados, sobretudo para identificação de parâmetros associados à renda dos pacientes.

Em uma análise inicial e preliminar do presente trabalho, construiu-se uma estimativa exclusiva para a cidade de Fortaleza, observando apenas quatro meses do ano de 2019. Nessa análise foi possível utilizar dados da RAIS (Registro Anual de Informações Sociais, 2020) e dados de uma API (*Application Programming Interface*):¹⁶ a API CEP aberto.¹⁷ Com essa análise, foi possível realizar uma análise preliminar que se encontra no Apêndice¹⁸ deste trabalho — mais detalhes da análise serão cobertos mais à frente. Contudo, trata-se de uma análise mais restrita, compreendendo um espaço mais limitado de dados.

A presente análise é consideravelmente mais abrangente. Logo, algumas bases, como a própria RAIS, não poderiam ser utilizadas como aproximação da renda dos trabalhadores.¹⁹ Assim, fez-se necessária a ligação com outras bases, como o

¹⁶ Uma API é uma interface que permite uma interação entre diversos aplicativos: trata-se de um *software* intermediário (HUBSPIRE, 2021). O uso de APIs permitiu a construção de parte das bases de dados utilizadas.

¹⁷ Os dados da API podem ser acessados mediante registro no site “CEP Aberto” (CEP ABERTO, 2021).

¹⁸ Ver Apêndice E deste trabalho.

¹⁹ O site PDET (Programa de Disseminação de Estatísticas do Trabalho) disponibiliza os dados da RAIS somente com a abertura até o nível municipal, com exceção das cidades de Fortaleza (CE), Rio de Janeiro (RJ) e São Paulo (SP). Nesses casos há a abertura por bairros e foi possível realizar uma aproximação dos dados de renda a partir dos dados da RAIS — no caso, assumindo que os salários dos estabelecimentos de trabalho nesses bairros representavam aproximadamente a renda das famílias locais (BRASIL, 2021d).

CENSO 2010, que possui dados de renda com a disponibilização do georreferenciamento dos setores censitários. Em maior detalhe, a “renda” utilizada é o valor nominal e médio mensal das pessoas acima de 10 anos de idade.²⁰ Os dados de georreferenciamento foram obtidos no site do IBGE.²¹

A abrangência dos dados também necessitou da identificação de estratégias alternativas para a complementação de dados de localização. No caso, considerando a abrangência nacional da análise, a obtenção dos dados georreferenciados de CEP se demonstraram consideravelmente intensivos em tempo,²² ponderando que a maioria das APIs existentes para a tarefa disponibilizam um número limitado de captura de observações sem cobrança.²³ Dessa forma, foi necessário a aquisição de uma base auxiliar²⁴ com dados georreferenciados de CEP, que foram complementados com os dados API CEP Aberto sempre que possível para cobrir eventuais lacunas de dados.

Com os dados de CEP georreferenciados e os dados do Censo 2010 também georreferenciados, foi finalmente possível realizar uma ligação da base de dados SIH-RD com os dados renda do Censo 2010, atingindo um dos principais objetivos na construção da informação para informar o modelo que será aplicado neste trabalho.

Os dados adquiridos ou complementados não cobriram todos os CEPs na base SIH-RD, em virtude de possíveis erros de inscrição ou inexistência de inscrição, mas cobriram parte sensível da base — ver Tabela 1 abaixo para observar os CEPs em que foi possível assinalar informações. Há diversas razões: erros de inscrição na base, inscrição do CEP geral do município, inexistência de inscrição adequada, entre outras. Na maioria dessas localizações aplicou-se a média municipal da renda como

²⁰ Conforme o dicionário de dados do Censo 2010, a descrição da variável é explicitamente: “Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento)” (BRASIL, 2011).

²¹ Após a publicação do Censo 2010, o IBGE passou a disponibilizar o georreferenciamento dos setores censitários, de modo que fosse possível realizar uma ligação das variáveis do censo a dados georreferenciados. Esse detalhe é absolutamente crucial para a construção da base de dados com a variável renda (BRASIL, 2021e).

²² Considerando o enorme número de CEPs na base SIH-RD, que ultrapassa mais de um milhão CEPs, a extração de dados pela maioria das APIs se demonstrou extremamente intensiva em tempo.

²³ A API de matriz de distâncias do google, acessível por meio do pacote *googleway* do R, é apenas acessível mediante registro no site da empresa e inscrição dos dados de cartão de crédito. Assim, a partir de um certo limite ultrapassado, a plataforma cobra valores devidos no cartão. Considerando tais questões, este trabalho buscou outras alternativas.

²⁴ Considerando a existência de uma restrição de tempo e dinheiro, optou-se pela aquisição da base já consolidada do site “Qual o CEP?”. O site disponibiliza dados em diferentes formatos para aquisição e é regularmente utilizado por empresas que buscam soluções de dados para a arquitetura dos seus sistemas ou em complementação ao seu *Business Intelligence*

uma aproximação suficientemente adequada para a presente análise, considerando sobretudo que a maioria desses casos ocorrem em município pequenos (em relação à população) e que estão integrados a regiões mais abrangentes das administrações de Saúde.²⁵

Tabela 1 - Dados de CEPs em que foi possível assinalar uma informação

Base de Dados	Número de CEPs
Base de dados SIH integral – abril a agosto – 2015 a 2020	800.787
Base de dados do trabalho – abril a agosto – 2015 a 2020	799.247

Fonte: Elaboração própria.

2.3 Construção final dos dados

Esta subseção trata da consolidação final dos dados e das alterações implementadas na base, considerando o exercício proposto, de investigação de um modelo de demanda hospitalar antes e durante a pandemia do COVID-19. A primeira parte diz dos ajustes às variáveis da base SIH-RD, com as classificações que foram implementadas de forma a capturar características relevantes para a seleção de um hospital. O ajuste dessas variáveis é importante para recuperar a estimação de demanda proposta em Ho (2006) e Ho e Lee (2017), além de buscar um ajuste para análise particular da pandemia do COVID-19.

A segunda parte trata da dimensão geográfica do processo de escolha do paciente, considerando a abrangência escolhida para analisar a demanda hospitalar por internações: a Região de Saúde do SUS. A Região de Saúde, além de ser de fato o espaço geográfico pelo qual o atendimento conveniado ao SUS é organizado, cobre a maioria dos seus atendimentos: há poucos “vazamentos” de atendimento para outras Regiões de Saúde.

²⁵ Conforme se verá à frente, as Regiões de Saúde são tipicamente mais abrangentes em municípios menores em termos populacionais, o que não afeta a estratégia de análise que será aplicada a este trabalho.

2.3.1 Construção dos dados com base no exercício proposto

Esta seção trata da adaptação das variáveis ao modelo de Ho (2006) e Ho e Lee (2017). Algumas aproximações foram efetuadas de forma a enquadrar o modelo na melhor aproximação possível das variáveis construídas para a estimação de demanda.

2.3.1.1 Construção de variáveis adicionais – variáveis de estrutura hospitalar

As variáveis que compõem outras informações sobre o hospital foram obtidas através de dados CNES (“Cadastro Nacional de Estabelecimentos de Saúde”): número de leitos, números de enfermeiros, detalhamento se o hospital é também uma instituição de ensino e se o hospital é privado.²⁶ Assim, foi possível criar uma razão entre leitos e enfermeiros (“Número de enfermeiros por leito”), uma variável indicadora binária variando entre 0 e 1 para instituições de ensino (“Ensino”) e uma variável indicadora binária variando entre 0 e 1 para instituições privadas (“Hospitais privados”).²⁷

2.3.1.2 Construção de variáveis adicionais – variáveis de serviços hospitalares

A construção das variáveis de serviços hospitalares seguiu a lógica apresentada em Kate Ho (2006): se um hospital provê o serviço menos comum no mercado, ele ganha um score de 1; se não provê nenhum dos serviços do grupo, ganha o score de 0. Para as demais alternativas, que consistem nos demais serviços comuns que um hospital poderia ofertar, a variável é calculada da seguinte maneira:

$$score = (1 - x)(1 - y) \quad (1)$$

²⁶ A ligação entre as bases SIH-RD e CNES é efetuada com base no CNPJ dos hospitais.

²⁷ Nesse último caso, utilizamos a proxy das classificações CONCLA para chegar a hospitais privados. Não foi possível indicar se o hospital é *for profit*, tal qual consta no trabalho de Kate Ho (2006), por falta de informação nas bases utilizadas.

Onde x é a percentagem de hospitais que oferecem o serviço mais comum e y é a percentagem de hospitais ofertando o serviço menos comum. A construção foi possível de ser realizada, exceto para serviços cardíacos,²⁸ onde se manteve a variável como uma variável binária variando entre 0 e 1.

As variáveis foram inferidas com base em extrações de dados: a extração dos procedimentos que constam na base SIH-RD analisada e da base CNES — de estabelecimentos e de equipamentos. Da base SIH-RD, extraiu-se a identificação dos seguintes procedimentos: angioplastia, serviços oncológicos, serviços associados a obstetrícia e radioterapia.

Da base de equipamentos e estabelecimentos, extraiu-se as informações sobre equipamentos de ressonância magnética, tomografia computadorizada, ultrassom e leitos específicos para o parto.

Tabela 2 - Definição dos Grupos de Serviços

Serviços para Câncer	Serviços Cardíacos	Serviços de Obstetrícia	Serviços de imagem
Serviços de Oncologia	Angioplastia	Serviços de Obstetrícia	Tomografia Computadorizada
Serviços de Radioterapia		Salas de parto	Ultrassom
			Ressonância Magnética

Fonte: Elaboração própria em adaptação de Ho (2006)

Com essa definição, construiu-se as variáveis para os 33 hospitais observados em Fortaleza. A Tabela 2 demonstra, conforme discutido anteriormente, que a variável “Serviços Cardíacos” possui uma natureza binária. As demais variáveis seguem a construção das variáveis com base na distribuição de serviços entre os hospitais.

2.3.1.3 Construção de variáveis adicionais – variáveis de diagnósticos

A construção das variáveis de diagnóstico também foi realizada com base na Tabela CID-10²⁹ para morbidades hospitalares, considerando a seguinte classificação:

²⁸ A variável foi aproximada pela verificação dos hospitais efetuam angioplastia nos pacientes. Os demais serviços listados no trabalho original de Ho (2006) não foram identificados nas bases CNES e nas bases de procedimentos da SIH-RD.

²⁹ Ver Tabela CID-10 nas grandes agregações (BRASIL, 2010b).

Tabela 3 - Tabelas com os diagnósticos utilizados

Diagnósticos	Código CID-10
Diagnósticos de doenças cardíacas	I00 a I25 e I27 a I52
Diagnósticos associados à gravidez	O00 a O99
Diagnóstico de doenças associadas ao aparelho digestivo	K00 a K93
Diagnósticos associados ao câncer	C00 a C96
Diagnósticos associados a doenças neurológicas	G00 a 99
Diagnósticos associados a doenças respiratórias	B34, J9 a J22 e J40 a J99

Fonte: Tabela Morbidade Hospitalar do SUS CID-10.

Os códigos CID-10 constam em grandes grupos de classificação. Os “Diagnósticos de Doenças Cardíacas” são extraídos do grande grupo de doenças circulatórias (códigos I00 a I99) e, desse grande grupo, focam apenas em doenças associadas a problemas cardíacos. Os “Diagnósticos associados ao câncer” focam apenas nas neoplasias (tumores) malignas, excluindo demais tipos de tumores. Os demais diagnósticos são efetivamente grandes grupos da codificação CID-10 ao qual pertencem — “Diagnósticos associados à gravidez” condizem ao grande grupo “Gravidez, parto e puerpério” (codificação CID-10 de O00 a O99); “Diagnósticos associados ao aparelho digestivo” condizem a “Doenças do aparelho digestivo” (codificação CID-10 de K00 a K93) e “Diagnósticos associados a doenças neurológicas” condizem com “Doenças do sistema nervoso” (codificação CID-10 de G00 a G99).

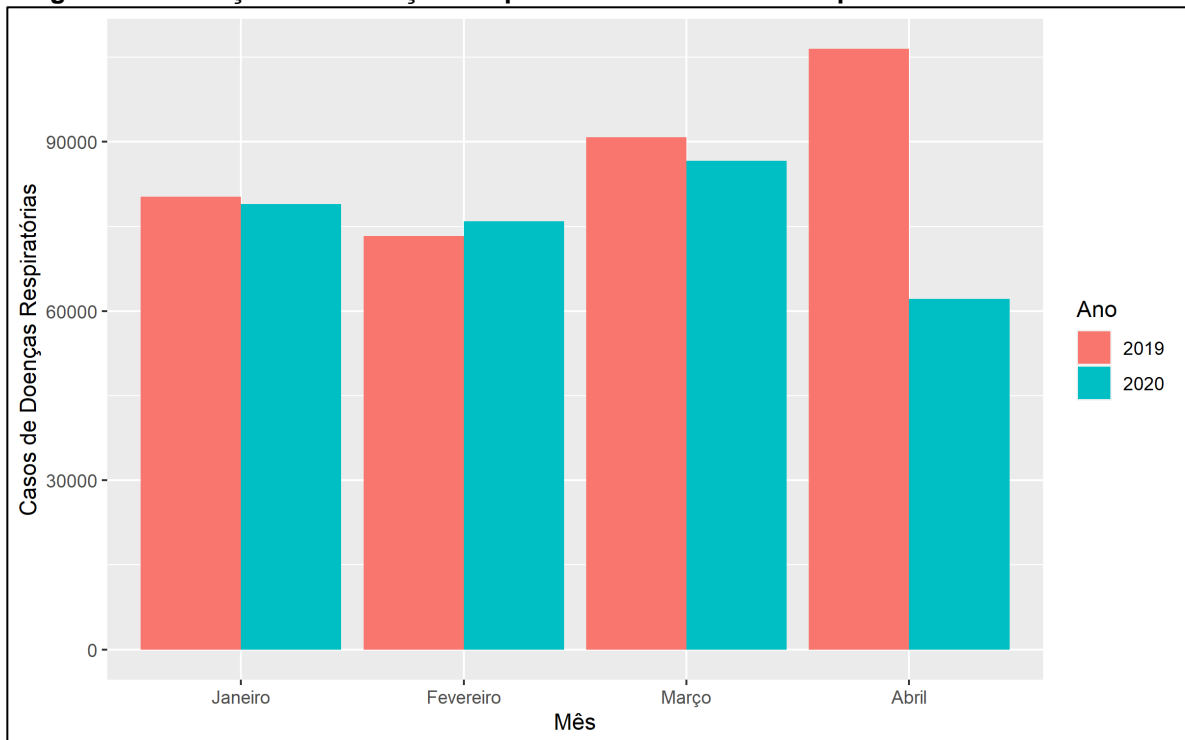
Para os códigos associados às doenças respiratórias partiu-se do grande grupo de doenças do aparelho respiratório (as codificações de J00 a J99), mas desconsiderando as doenças respiratórias das vias aéreas superiores e outras doenças da região da garganta, como faringites e laringites. Adicionalmente, foi considerada codificação “B34”, que compreende “Doenças Por Vírus, de Localização Não Especificada” para compreender o impacto do COVID-19 (*Corona Virus Disease 2019*) entre as doenças respiratórias.³⁰ A hipótese agregação considera a possibilidade de inscrição de casos da doença entre doenças respiratórias e vice-versa, considerando que o diagnóstico de COVID-19 pode ser confundido com o de

³⁰ A especificação, em detalhe, pode ser consultada no site MedicinaNET, que oferece um detalhamento mais específico da tabela CID-10. É possível verificar que a especificação para o COVID-19 se dá na codificação “B342”, isto é, “Infecção por coronavírus de localização não especificada”.

uma pneumonia viral. Considerou-se inclusão da codificação para a “Síndrome Respiratória Aguda Grave” (SRAG), isto é, codificações iniciadas em “U04”, mas não foram encontradas inscrições na base analisada SIH-RD no período relevante desta análise.

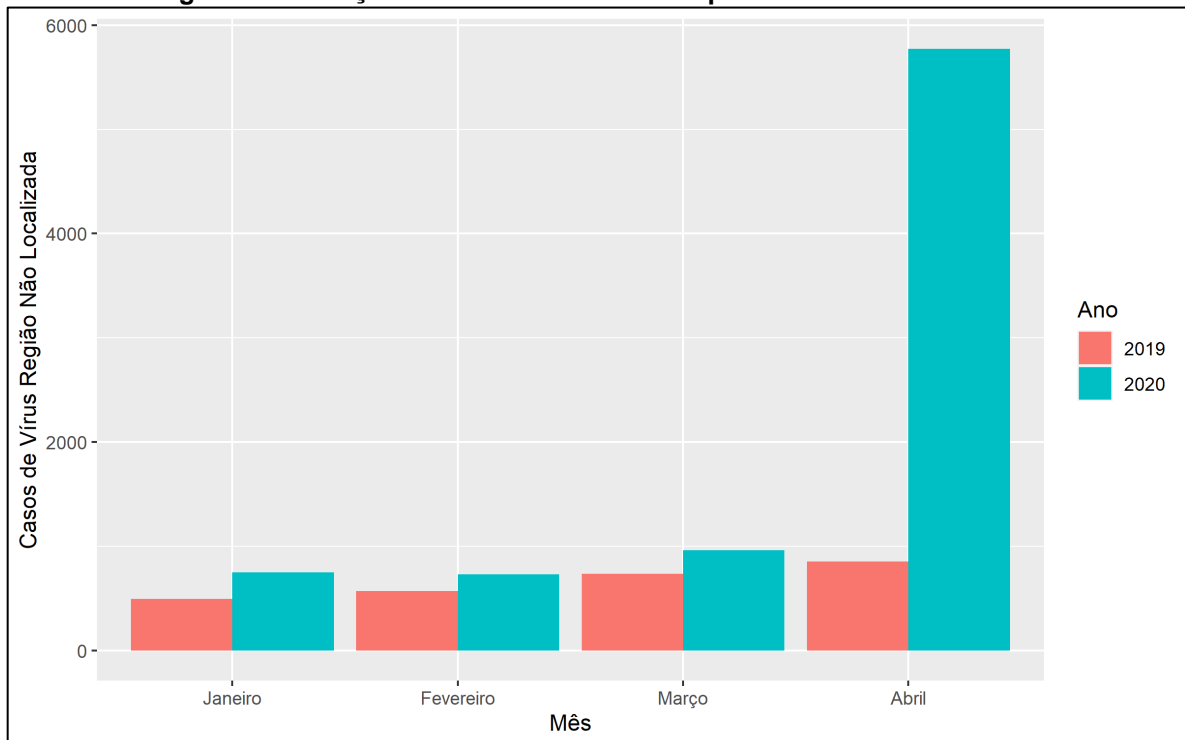
Abaixo, na Figura 1, descrevemos a dinâmica do número de casos, nos primeiros quatro meses de 2019 e 2020, com a classificação destacada neste trabalho.

Figura 1 – Inscrições de Doenças Respiratórias no Brasil – Comparativo entre 2019 e 2020



Fonte: SIH-RD. Elaboração própria.

Figura 2 - Inscrições “B34” no Brasil – Comparativo entre 2019 e 2020



Fonte: SIH-RD. Elaboração própria.

A Figura 1 utiliza-se apenas da base SIH-RD, isto é, internações realizadas pelo SUS. Conforme se observa, houve aparentemente uma redução na inscrição de casos de doenças respiratórias no comparativo entre 2019 e 2020 — pela inscrição adotada na Tabela 2. Mas houve uma expansão na classificação de doenças do tipo “B34”, de acordo com a Figura 2. Assim, o mês de abril marca o início da inscrição relativamente mais adequada — com dados SIH-RD — das internações associadas ao COVID-19 e que, neste trabalho, estão agregadas ao conjunto de doenças respiratórias.

Por fim, construiu-se uma variável de identificação para os pacientes com base no gênero, idade, localização e diagnóstico — nesse último caso, os diagnósticos discutidos anteriormente. Estabeleceu-se oito níveis para a idade analisada³¹ e, com base na classificação de gênero e idade que consta na base SIH-RD, construiu-se essa identificação.

Juntando essas características — gênero e idade — com os dados de diagnóstico gerais e localização (através do CEP do paciente), criou-se um identificador para o paciente. Sob a hipótese de que um mesmo paciente possua o

³¹ Os níveis são os respectivos: 0-19 anos, 20-34 anos, 35-44 anos, 45-54 anos, 55-64 anos, 65-74 anos, 75-84 anos e maiores do que 84 anos.

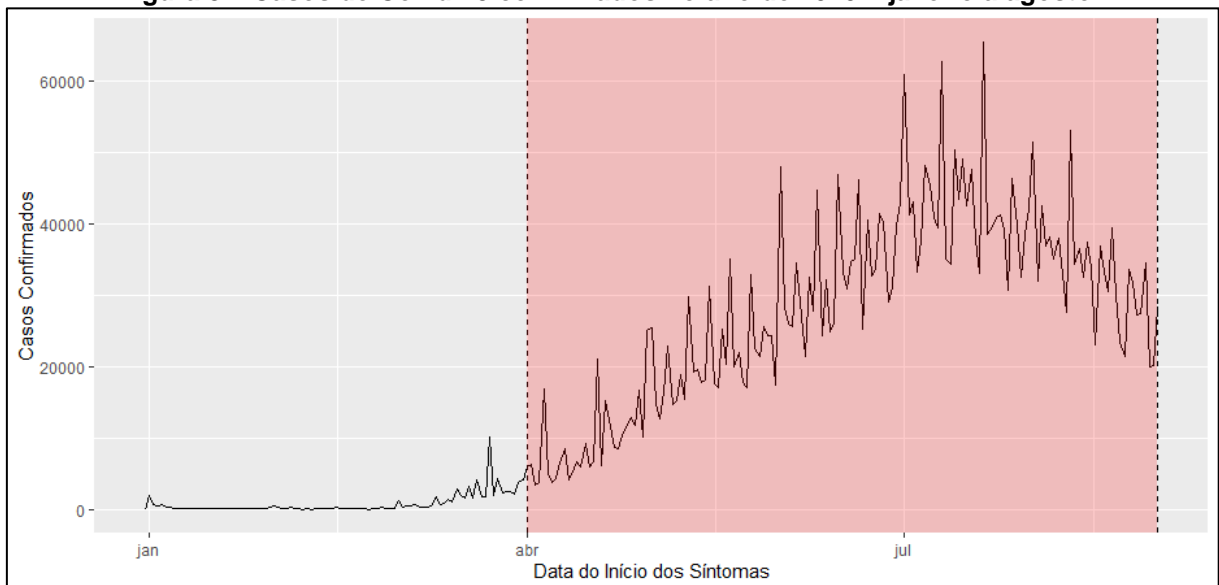
mesmo gênero, idade, localização e diagnóstico geral,³² este trabalho definiu o paciente como sendo o mesmo paciente representativo.

2.3.1.4 Ajuste na base de dados, considerando a pandemia do COVID-19

Considerando que este trabalho realiza uma análise de impacto da pandemia do COVID-19, realizou-se uma média das características dos pacientes dos últimos 5 anos (2015 a 2019) e uma análise posterior para o ano de 2020. Essa média cobre os meses de abril a agosto, os meses da amostra do ano de 2020. Explica-se à frente.

Considerando a Figura 2, que demonstra o início das inscrições com os dados SIH-RD de maneira relativamente adequada em meados de abril e o início da pandemia do Covid-19 e o seu pico em 2020 em meados de julho (ver Figura 3 abaixo), a análise cobre apenas os meses de abril a agosto, tanto de 2015 a 2019 como em 2020.

Figura 3 – Casos de Covid-19 confirmados no ano de 2020 – janeiro a agosto



Fonte: Painel FioCruz Covid-19, Dados ESUS-VE. Elaboração própria.

Dessa forma, a análise pretende capturar o início do *stress* do sistema hospitalar e o seu pior momento em 2020.

³² Tal como definido nas agregações de diagnóstico da Tabela XX.

Assim, considerando as alterações aqui descritas na consolidação dos dados SIH-RD para o presente exercício, chegou-se a um total de 16.583.709 eventos no período. Com esses dados, foi possível construir as variáveis finais de análise que constam na Tabela 4 abaixo — com seus valores descritivos: média (“Méd.”), Desvio Padrão (“DP”), valor mínimo (“Mín.”) e valor máximo (“Máx.”).

Tabela 4 - Estatísticas de Sumário das variáveis utilizadas para a regressão

Variáveis	Méd.	DP	Mín.	Máx.
Diagnóstico associado a doenças digestivas ("Digestivos")	0,1	0,3	0,0	1,0
Diagnóstico associado à gravidez ("Gravidez")	0,2	0,4	0,0	1,0
Diagnóstico de doenças cardíacas ("Cardíacos")	0,1	0,3	0,0	1,0
Diagnóstico de outras condições	0,5	0,5	0,0	1,0
Diagnóstico de tumores ("Câncer")	0,1	0,3	0,0	1,0
Diagnóstico para doenças neurológicas ("Neurológico")	0,0	0,2	0,0	1,0
Diagnósticos associados ao aparelho respiratório ("Respiratório")	0,1	0,3	0,0	1,0
Distância (km)	3,7	13,9	0,0	349,5
Distância (km) ao quadrado	205,3	2615,0	0,0	122157,8
Emergência	0,3	0,5	0,0	1,0
Ensino	0,4	0,5	0,0	1,0
Hospitais Privados	0,5	0,5	0,0	1,0
Número de enfermeiros por leito	3,7	6,4	0,0	160,0
Número de Respiradores	39,6	49,2	0,0	399,6
Renda Censo	1,2	0,9	0,0	27,2
Serviços cardíacos	0,7	0,4	0,0	1,0
Serviços de imagem	0,5	0,4	0,0	1,0
Serviços de Obstetria	0,7	0,4	0,0	1,0
Serviços para câncer	0,4	0,4	0,0	1,0

Fonte: SIH-RD, CNES, CENSO, Qual o Cep.

Legenda: Méd: média; DP: desvio padrão; Mín: Mínimo; Máx: máximo.

As variáveis da Tabela 4 serão utilizadas considerando a descrição de mais detalhes do exercício no tópico “Descrição dos modelos”. A grande maioria das variáveis serão utilizadas em interações entre as variáveis de serviço (e.g. serviços de imagem) e diagnósticos específicos (e.g. diagnóstico associado a doenças cardíacas). Trata-se, pois, da modelagem adotada em Ho (2006) e subsequentemente em Ho e Lee (2017).

Cabe um último detalhe sobre uma das variáveis utilizadas: a razão de número de enfermeiros e leitos pode estar superestimada em função de erros em inscrições no número de leitos assim como na estimação do número de vínculos ativos em um

determinado hospital,³³ com alguns casos extremamente altos. Como a exclusão dos valores implica na exclusão de um hospital da base, os valores foram mantidos, apesar de se reconhecer potenciais problemas. Ressalta-se, porém, que se trata um montante pouco substancial na base as razões que ultrapassam 15 enfermeiros por leito, sendo aproximadamente 4% dos valores.³⁴

2.3.2 Definição geográfica da abrangência da estimacão de demanda

Com a definição do escopo de variáveis e a possibilidade de análise com dados públicos hospitalares (foco em internacões), foi igualmente necessário definir a dimensão geográfica ou área de influência das unidades hospitalares. Para tanto, há um conjunto de referências que foram investigadas e consideradas em outros estágios da análise deste trabalho.

Em particular, quando se avalia a dimensão geográfica dos mercados privados hospitalares (neste caso, com um foco concorrencial), há um conjunto bem estabelecido de metodologias para a adequação dessa análise: raios de distância e tempo, o método Elzinga-Hogarty,³⁵ entre outras análises.³⁶

Contudo a presente análise foca exclusivamente na escolha de hospitais considerando apenas leitos de internacão do SUS. Assim, uma aproximacão da dimensão geográfica ou área de influências dos hospitais que se julgou suficientemente adequada são as Regiões de Saúde, criadas com o Decreto nº 7.508, de 2011 (BRASIL, 2011) e que regulamenta a lei nº 8080 de 1990 — a lei trata da organizacão do SUS, entre outras questões voltadas ao sistema de Saúde (BRASIL, 1990).

É possível admitir que exista uma variabilidade no espaço de escolha do paciente que não esteja completamente incorporada pela Região de Saúde, com o

³³ Na investigacão implementada, até o momento não se encontrou um dicionário para identificar se a base CNES-PF possui variáveis que identificam vínculos ativos ou inativos. A classificacão foi montada em cima da classificacão CBO de enfermeiros.

³⁴ Os valores excessivamente grandes, que ultrapassam 50, são menos do que 1%.

³⁵ Ver Elzinga e Hogarty (1973).

³⁶ É possível analisar a dimensão geográfica do mercado, sobretudo no setor privado, olhando outras características, como a relacão com planos de saúde e a rede credenciada de hospitais. Outra análise é o uso do Teste do Monopolista Hipotético – ver, por exemplo, Coate e Fischer (2008) -- aplicado regularmente em casos concorrenciais. Contudo, o teste não faz sentido para ser aplicado no presente trabalho.

deslocamento sendo às vezes mais restrito ou menos restrito (a Região não contempla o conjunto de fatos de escolhas do paciente, incorporando um conjunto a mais ou a menos de hospitais) e que esse espaço possa ser dinâmico (as escolhas realmente disponíveis se alteram com o tempo). Este trabalho, em estágios preliminares, investigou mais de uma dimensão de análise — encontrando, por vezes, alguns problemas.³⁷ As Regiões de Saúde, embora possam passar por refinamentos, tiveram não só benefício de permitir uma análise mais adequada do ponto de vista conceitual, mas também facilitaram soluções computacionais. À frente, este trabalho dá mais detalhes e motiva a escolha.

2.3.2.1 Regiões de Saúde

Conforme Santos (2017) destaca, Região de Saúde é a dimensão geográfica na qual o SUS organiza a sua gestão hospitalar. Em maior detalhe:

A região é recorte territorial, administrativo-sanitário que permite integrar o que a descentralização supostamente teria fracionado, definindo para a população um espaço sanitário de serviços, constituído pelas redes de atenção à saúde, dotadas de inteligência sanitária que permita à pessoa o acesso ao itinerário terapêutico adequado à sua necessidade. A região de saúde é pré-requisito para a ordenação sanitária, com o fim específico de garantir o acesso às ações e serviços de saúde dentro de um território delimitado e disciplinado podendo ser inter-regional, conforme forem as necessidades de saúde. É na região que o SUS deve garantir às pessoas suas necessidades de saúde em acordo às referências interfederativas e à gestão compartilhada, definidas em acordos e consagradas no contrato. (SANTOS, 2017, p. 1283)

A Região de Saúde é, portanto, a área em que o SUS se organiza para prover serviços médico-hospitalares aos pacientes. O Decreto nº 7.508, de 2011, que regulamenta as Regiões de Saúde, também destaca as ações de serviços mínimas, *in verbis*:

“Art. 5º Para ser instituída, a Região de Saúde deve conter, no mínimo, ações e serviços de:
I - atenção primária;
II - urgência e emergência;
III - atenção psicossocial;
IV - atenção ambulatorial especializada e hospitalar; e

³⁷ Este trabalho investigou tanto a dimensão municipal como da microrregião, encontrando dificuldades no tratamento dos dados acima como nas soluções computacionais para o modelo utilizado mais à frente neste trabalho.

V - vigilância em saúde.” (BRASIL, 2011, grifo nosso)

Verifica-se, pois, uma necessidade de um requerimento mínimo de atendimento nas regiões e com foco em atendimentos que compreendem o escopo deste trabalho: a demanda por serviços de internação hospitalar. Assim, foram desenhadas 450 Regiões de Saúde, agregando diversos municípios em uma única região. Há sentido nessa definição geográfica da organização da saúde em virtude da escala necessária para a realizar alguns investimentos no setor, com as regiões sendo bastante heterogêneas em tamanho.³⁸

Outro aspecto relevante é o uso das Regiões de Saúde pela própria ANS, em nova norma que consta na Consulta Pública nº 82, sobre o uso das Regiões de Saúde como a região geográfica de referência em caso de substituição de prestadores hospitalares em determinadas circunstâncias (BRASIL, 2020). Embora se trate da lógica do mercado privado, o uso da aproximação pela ANS talvez esteja focado no lado da demanda e disponibilidade de um paciente para se deslocar regionalmente, o que as Regiões de Saúde potencialmente incorporam.³⁹

Deve-se ressaltar, porém, que a ida a um hospital pode ser emergencial e que nas análises concorrenciais do setor privado considera-se distâncias de 10 km ou de até 20 minutos de tempo de deslocamento para identificar uma área de concorrência. Embora tal natureza seja inegável para alguns tipos de atendimentos, há indícios claros de que pacientes viajam distâncias maiores para internações e procedimentos mais complexos.⁴⁰ No município de São Paulo especificamente, onde há uma grande densidade populacional e um conjunto considerável de leitos no mesmo município, há indícios de que 60% dos pacientes viajam mais de 10 km, incluindo atendimentos ambulatoriais, embora existam incertezas sobre a real magnitude dessa informação disponibilizada à época em virtude de problemas de inscrição de endereços.⁴¹

³⁸ Ver Figura 4 mais à frente para uma visualização dessa característica.

³⁹ Este estudo não está familiarizado se a ANS realizou investigações mais detalhadas para motivar essa escolha.

⁴⁰ A viagem a pacientes não-municípios é considerável e compreendem uma parte não desprezível do conjunto de atendimentos de uma região, considerando a gestão pela Região de Saúde. Há fluxos nacionais e focos em diferentes especialidades para essas viagens, conforme pode se ver no agrupamento de estudos como o de Xavier et al (2017).

⁴¹ Em 2019, o portal UOL publicou uma matéria discutindo os deslocamentos de pacientes para o município de São Paulo. Inicialmente, a autora da matéria alegou que cerca de 60% dos deslocamentos estavam acima de 10 kms e depois alegou que não era possível afirmar isso em virtude de uma incerteza no registro dos endereços (MEGA, 2019).

Contudo, há autores como Xavier et al (2017), focados na distribuição de fluxo de pacientes, que compreendem que há uma movimentação considerável de pacientes por diferentes serviços, considerando a complexidade. O dado desagregado por unidade federativa mostra também como os valores podem variar radicalmente, a depender do nível de cobertura hospitalar e densidade populacional (ver Tabela 5, abaixo).

Tabela 5 – Média de Distâncias Percorridas por Pacientes Internados – 2015-2020

UF	Média - Distância Percorrida (Km)
MA	84,17
CE	81,64
MT	66,67
AC	64,80
PI	63,40
TO	61,86
RO	59,10
PE	56,18
BA	50,52
RN	48,58
PA	47,78
GO	46,95
PB	44,76
MS	42,73
PR	41,84
DF	38,51
RR	37,60
SE	36,61
AL	35,74
RJ	33,80
AP	33,35
AM	32,25
MG	31,79
ES	31,52
RS	26,28
SC	24,84
SP	24,18

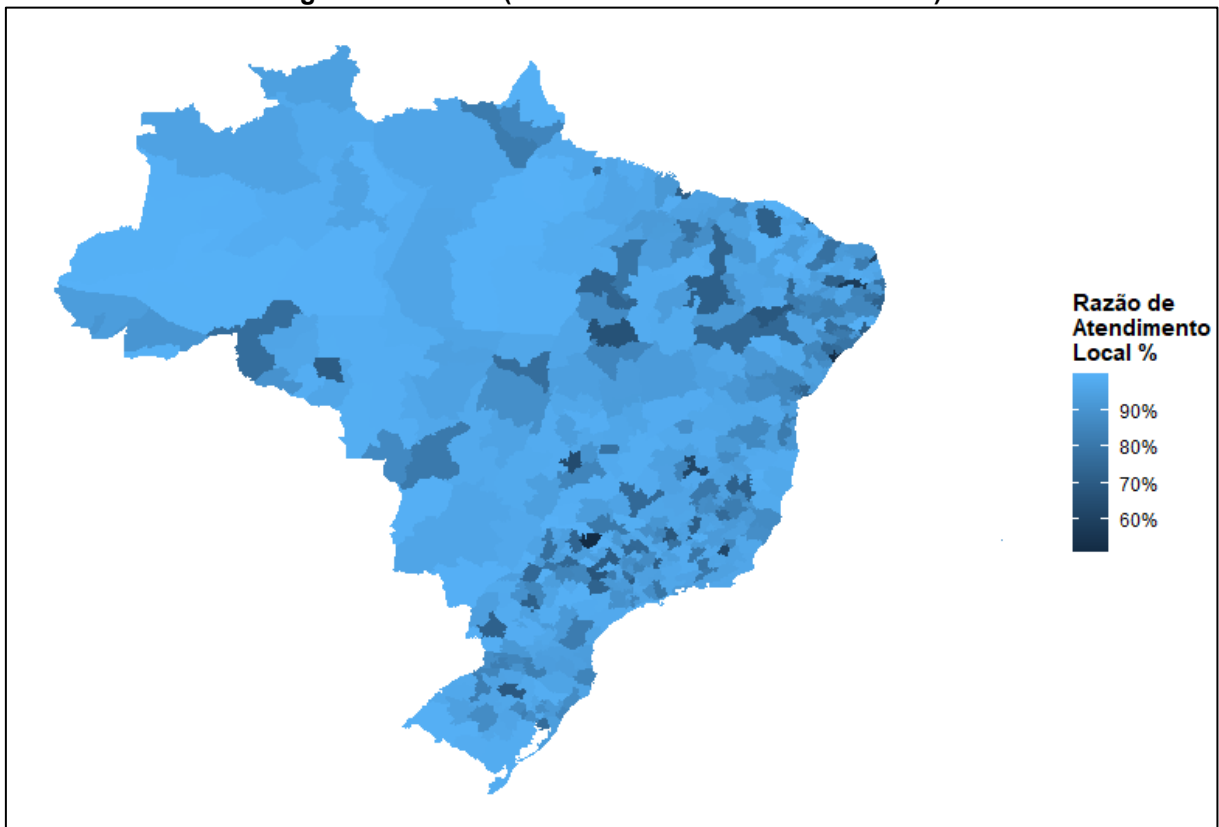
Fonte: SIH-RD, dados latitude e longitude da base “qual o cep?”. Elaboração própria.

Assim, em média, pacientes podem viajar bastante, especialmente para internações e procedimentos mais complexos, o que pode exigir um atendimento hospitalar inexistente no município de residência do paciente.

Como evidência adicional para sustentar a escolha da Região de Saúde como espaço geográfico suficientemente adequado para a investigação deste trabalho, o trabalho de Silva et al (2021) encontra uma taxa de evasão de 11,9% das Regiões de Saúde para internações do Covid-19. Portanto, aproximadamente 90% das internações para a COVID-19 são de pacientes que residem nas Regiões de Saúde.⁴²

Esses resultados estão aproximadamente de acordo com as investigações implementadas com a base SIH-RD neste trabalho. Reproduzimos abaixo, na Figura 4, as proporções de atendimento considerando as Regiões de Saúde e a base SIH-RD utilizada neste trabalho — no caso, considerando todas as internações da base de dados.

Figura 4 – Dados SIH-RD 2015 a 2020 - Proporção de pacientes atendidos nas próprias Regiões de Saúde (“Razão de Atendimento Local %”)



Fonte: SIH-RD. Elaboração própria.

⁴² Embora seja alvo de críticas consideráveis, o método Elzinga-Horgarty tipicamente considera um *threshold* de 75% de “vendas” dentro do mercado relevante para defini-lo como tal. Ver Capps, Dranove e Zabinski (2017). Assim, por esse critério, as Regiões de Saúde seriam um espaço geográfico de escolha conservador (já que aproximadamente 90% das internações se dão nas Regiões de Saúde) e seria possível admitir, hipoteticamente, uma segmentação mais restrita desse espaço de escolha do paciente.

Assim, embora este trabalho reconheça que sejam possíveis heterogeneidades e mudanças temporais no espaço de escolha dos pacientes, entende-se que as Regiões de Saúde oferecem uma agregação suficientemente adequada para discutir o rol de escolhas disponíveis para um paciente em busca de internação pelo SUS.

3 DESCRIÇÃO DOS MODELOS

Este capítulo trata especificamente das teorias e referências técnicas para o método implementado neste trabalho, cobrindo os modelos de regressão logística, os métodos associados a uma estimação de demanda hospitalar e outros aspectos relevantes para a presente análise.

Em maior detalhe, após a descrição da literatura relevante e o modelo de fato aplicado neste trabalho, apresenta-se uma discussão técnica sobre os problemas de convergência tipicamente encontrados na literatura de regressões logística — ver Allison (2008) —, a aplicação de um filtro *boxplot* para retirar casos limites das regressões implementadas e uso de métodos associados à meta-análise de regressões para identificar padrões nas mudanças de coeficientes de maneira agregada.

Em uma análise no Apêndice C deste trabalho também são desenvolvidos alguns resultados de análise de bem-estar.

3.1 Descrição do modelo *logit* para n alternativas

Para melhor qualificar a discussão sobre o modelo, esta seção subdivide-se em três subtópicos descrevendo de maneira geral e simples o modelo padrão de regressão logística (o tópico “A descrição de um modelo *logit* generalizado”, seguido pela descrição do “modelo *logit* multinomial” mais comum) para depois identificar o método implementado neste trabalho, efetuando uma regressão logística condicional (“O modelo *logit* condicional”).

3.1.1 A descrição de um modelo *logit* generalizado

Para iniciar uma descrição das formas de estimação de demanda para muitas alternativas através de modelos probabilísticos no formato *logit*, apresenta-se um formato generalizado de utilidade para escolhas discretas — tal estrutura irá permitir

uma diferenciação dos modelos de *logit* multinomial “padrões” e o *logit* condicional descrito por Mcfadden (1973).⁴³

Esta exposição se beneficia das notas de Rodriguez (2020) e Maddala (1983). Por suposição, considere uma utilidade u_{ij} para um indivíduo i com j alternativas com o seguinte formato funcional:

$$u_{ij} = \eta_{ij} + \epsilon_{ij} \quad (2)$$

Assumindo que os indivíduos i agem de maneira economicamente racional, assume-se que o indivíduo i escolherá a alternativa j^* em que $u_{ij^*} \geq u_{ij}$. Assumindo que U descreve o conjunto de utilidades com as diferentes alternativas j , a probabilidade de que o indivíduo i escolherá a alternativa j^* pode ser descrita com a seguinte equação:

$$\pi_{ij^*} = Pr(\text{Máx}\{U\} = U_{ij^*}) \quad (3)$$

Assim, é possível demonstrar que o termo ϵ_{ij} segue a uma distribuição de valores extremos Tipo I ou distribuição de Gumbel,⁴⁴ com uma função de densidade de probabilidade que pode ser descrita da seguinte forma, conforme Maddala (1983):

$$f(\epsilon) = \exp\{-\epsilon - \exp\{-\epsilon\}\} \quad (4)$$

Conforme Maddala (1983), é possível descrever que termo ϵ_{ij} possui uma distribuição de valores extremos tipo I, o que permite a escrever as equações de um *logit* multinomial tal como descrita em (5):

$$\pi_{ij} = \frac{\exp\{\eta_{ij}\}}{\sum_k \exp\{\eta_{ik}\}} \quad (5)$$

⁴³ A seção é construída considerando as notas de aula de Rodriguez (2020), que compreendem uma sintetização que facilita a introdução aos conceitos dos modelos *logit* com um número maior de escolhas que a escolha binária.

⁴⁴ As variáveis inferidas de um modelo de regressão logística tipicamente seguem uma distribuição do tipo Gumbel.

3.1.2 O modelo *logit* multinomial

A característica distintiva, em termos funcionais, do *logit* multinomial são os parâmetros variando de acordo com as j alternativas disponíveis. O modelo simplificado, em detalhe, é construído essencialmente com base nas características dos indivíduos. De maneira genérica, é possível reescrever o termo η_{ij} tal como:

$$\eta_{ij} = v_i' \beta_j + e_{ij} \quad (5)$$

Onde o termo β_j descreve reflete as chances e a magnitude relacionadas a uma determinada escolha j , com a limitação de que as escolhas são modeladas desconsiderando as características associadas às escolhas, isto é, modela-se utilizando apenas os vetores de características individuais v_i .

3.1.3 O modelo *logit* condicional

De maneira alternativa, Mcfadden (1973) propôs que as seleções dos indivíduos fossem modeladas de acordo com as características das possíveis escolhas. Dessa maneira, reescreve-se o termo η_{ij} da seguinte maneira:

$$\eta_{ij} = x_j' \beta + e_{ij} \quad (6)$$

Onde β não varia mais de acordo com as escolhas e onde x_j , um vetor de características das j alternativas, é utilizado para modelar as escolhas. Tipicamente, quando se há muitas alternativas, utiliza-se o modelo de *logit* condicional. De maneira alternativa, também é possível incluir variações que incorporam características individuais, tal como através de interações com características individuais. Reescrevendo (4), o modelo pode ser descrito como:

$$\Pr_i(\text{escolher } j) = \Pr_i(j \text{ oferece a maior utilidade para o indivíduo } i) = \pi_{ij} = \frac{\exp\{x_j' \beta + e_{ij}\}}{\sum_k \exp\{x_k' \beta + e_{ik}\}} \quad (7)$$

O modelo é tipicamente estimado através de uma função de máximo-verossimilhança, com o seguinte formato:

$$LL = \sum_i (x'_{c_i} \beta - \log (\sum_j \exp \{x'_j \beta\})) \quad (8)$$

Onde $c_i \in \{1, \dots, j\}$ é a escolha que o indivíduo i realiza. E os betas β são obtidos como os argumentos que maximizam a função (8). Escrito outra forma:

$$\beta = \operatorname{argmax}_{\beta} \sum^N LL_n \quad (9)$$

A função de máximo-verossimilhança nem sempre atinge o seu ótimo, com os problemas de *separação* ou *quasi-separação* descritos por Alisson (2008). Na subseção 3.4 este trabalho descreverá em amplo detalhe esse aspecto.

3.2 Breve resumo da maximização de utilidade do paciente e o modelo utilizado para a estimação de demanda de hospitais

A análise aqui aplicada tem por base os trabalhos implementados em Town e Vistnes (2001), Capps et al (2003), Ho (2006) e Ho e Lee (2017). Especificamente, aproveita-se da descrição apresentada em Ho (2006), omitindo uma discussão sobre o indexador m para mercados, considerando que a estimação é realizada para cada um dos mercados em separado. Assim, supõe-se que um indivíduo i no tempo t fica doente com alguma probabilidade e a sua utilidade, considerando a escolha de um hospital h e um diagnóstico l , pode ser descrita pela seguinte equação:

$$u_i = u(x_{ht}, \delta_h, v_{il} | \alpha, \beta) \quad (7)$$

Onde x_{ht} e δ_h são vetores de variáveis que descrevem características de hospitais (h) que mudam com o tempo (t) e não mudam com o tempo — no caso de δ_h , especificamente, trata-se do vetor de efeitos fixos dos hospitais. O vetor v_{il} define característica do indivíduo i com o diagnóstico l . Com esse formato funcional da utilidade dos indivíduos internados, um indivíduo i escolhe o hospital h^* com probabilidade quando:

$$u_{ih^*l} = u(x_{h^*t}, \delta_{h^*}, v_{il} | \alpha, \beta) \geq u_{ihl} = u(x_{ht}, \delta_h, v_{il} | \alpha, \beta) \quad (8)$$

A maximização para todos os indivíduos i pela lógica descrita em (3), conduz aos conjuntos A_h para todos os hospitais ora em análise, o que permite escrever a equação de *share* tal qual como apresentada em Ho (2006):

$$s_h(x, \delta | \alpha, \beta) = P(v \in A_h) \quad (9)$$

3.3 O modelo utilizado para estimação de demanda de hospitais

O modelo parte da ideia de que o consumidor visita o hospital que maximiza a sua utilidade, condicional ao diagnóstico e ao tipo de consumidor conforme discutido na seção anterior. Em detalhe e explicitamente, o modelo parte da seguinte função de utilidade descrita abaixo:

$$u_{ihl} = \delta_h + x_h v_{il} \beta + e_{ihl} \quad (10)$$

Onde δ_h representam os efeitos fixos de hospitais — omitidos nas equações estimadas —, x_h as variáveis associadas aos hospitais (serviços de hospitais, como serviços cardíacos, ou variáveis da estrutura, como número de enfermeiros por leito) em interação com v_{il} , que representa as variáveis no nível do indivíduo, como salário e diagnóstico associado à ida ao hospital. O modelo de escolha do paciente pode ser estimado sob o formato de regressão logística.

A estrutura, com a proposta de modelo *logit* para a demanda, também permite construir a equação de participações de mercados descrita em (4) de maneira mais explícita:

$$S_h = \sum_{i,l} \frac{N_{il}}{N} \left(\frac{\exp(\delta_h + x_h v_{il} \beta)}{\sum_{p \in h} \exp(\delta_h + x_h v_{il} \beta)} \right) \quad (11)$$

Onde S_h reflete a probabilidade de escolha agregada entre os pacientes, o que resulta, finalmente, nas participações de mercado que podem ser potencialmente estimadas.

Como o modelo é condicional em valores demográficos e diagnóstico, o modelo é implementado utilizando o pacote “Survival” no *software* estatístico R, onde é possível estimar o modelo de *logit* condicional através da função “clogit”, isto é, condicional à construção dos identificadores dos pacientes.

Os resultados são obtidos por meio do resultado máximo-verossimilhança e pela aproximação sugerida em Breslow (1974),⁴⁵ para facilitar a computação dos resultados com um número significativo de observações.

3.4 Problemas de convergência no modelo *logit* condicional

Nas regressões implementadas, algumas apresentaram problemas para convergência ou obtiveram valores explosivos nos coeficientes, indicando um problema relevante em regressões logísticas, conforme Allison (2008).

Problemas dessa natureza ocorrem tipicamente em função da estrutura dos dados. Em algumas regiões, foi possível identificar um número considerável de distâncias de tamanho 0 km, em virtude da inscrição do paciente com a mesmo CEP geral do município. Em outros casos, em função da construção das variáveis, é possível que um conjunto de serviços possa estar localizado em um único hospital na região, o que potencialmente gera situações de *separação completa* e *quase-separação*, também descritas por Allison (2008).

Em maior detalhe, a separação completa ocorre quando uma função linear pode gerar previsões perfeitas da variável binária de um modelo do tipo *logit* (*separação completa*). Assumindo um modelo *logit* com uma variável binária $y_i \in \{0,1\}$ explicada por um coeficiente β e uma variável x_i , uma separação completa ocorreria se para todo $\beta x_i > 0$, $y_i = 1$ e para todo $\beta x_i < 0$, tivéssemos $y_i = 0$. Conforme demonstrado por Allison (2008), a função de máximo-verossimilhança não alcança um valor máximo nesse caso. A quase-separação é similar e ocorreria sob as seguintes condições: para todo $\beta x_i \geq 0$, $y_i = 1$ e para todo $\beta x_i \leq 0$, $y_i = 0$. A convergência para um valor máximo da função de máximo-verossimilhança, nesse caso, pode ser

⁴⁵ A aproximação é tipicamente utilizada em *matchings* de 1 um para 1 ou 1 para n observações alternativas na literatura médica. No caso específico de modelos de demanda, como os de Mcfadden, não se realiza um *matching*, mas abre-se um conjunto de alternativas para cada escolha — as observações de “controle” são escolhas não observadas do paciente, assumindo que o paciente realiza a escolha que otimiza o seu bem-estar.

improvável. Um indicativo de que há problemas dessa natureza são os coeficientes desproporcionalmente grandes que alguns pacotes produzem

Além dessas questões, há a possibilidade de o modelo não atingir uma convergência dos valores por outras questões computacionais, considerando que a base de dados utilizada é consideravelmente grande e o modelo possui um conjunto limitado de tentativas para convergência.⁴⁶

Alisson (2008) e outros manuais⁴⁷ enumeram algumas soluções: (i) inferência exata, que corresponde à estimação do parâmetro mediano dos coeficientes; (ii) otimização de máximo verossimilhança penalizada, conforme apresentada em Firth (1993); (iii) outras soluções envolvem uma hipótese à distribuição dos coeficientes, o que seria uma aplicação bayesiana frente ao problema de convergência.

Considerando que essas soluções não estão presentes para os pacotes de regressão logística condicional (modelo *logit* condicional) ou necessitariam de uma capacidade computacional enorme, este trabalho utiliza a solução mais simples proposta por Alisson (2008): a exclusão dos coeficientes que aparentemente não convergiram, considerando que os demais coeficientes são estatisticamente válidos.

Contudo, em virtude do tamanho da base de dados deste trabalho e da dificuldade da análise detalhada de cada um dos coeficientes estimados nas várias regressões implementadas⁴⁸, adota-se um critério mais simples para a exclusão dos possíveis coeficientes problemáticos da análise: a exclusão por meio de um filtro de um gráfico de dispersão, como o *boxplot*.

3.5 Aplicação de um filtro *boxplot*

A metodologia da identificação visual de *outliers* com o gráfico do tipo *boxplot* é antiga, simples e amplamente usada em diversos trabalhos — não só para a

⁴⁶ Ver o trabalho dos autores do pacote *survival* — Theneau e Lumley (2014) — sobre questões de tempo de convergência do modelo implementado pela função *clogit*, considerando o limite iterações implementado.

⁴⁷ O Manual da UCLA, disponível online, exemplifica situações de *separação* ou *quase separação* dos dados e indica potenciais soluções. Uma delas é simples *do nothing*, considerando que as demais estimativas dos coeficientes são suficientemente adequadas para a análise individual.

⁴⁸ Como se verá à frente, em tese há 900 regressões que podem ser potencialmente estimadas e cada uma com 40 regressores, desconsiderando os efeitos fixos hospitalares. Ao todo são 36.000 coeficientes que podem ser estimados na presente análise.

identificação de *outliers*, mas para se ter uma dimensão visual da dispersão de valores, muitas vezes pouco clara através de parâmetros simples, como uma média. Para maiores detalhes, Stryjewski (2010) faz um *survey* da metodologia de visualização, considerando as suas primeiras implementações e a evolução prática adotada por Tukey (1977), que é a mais comumente usada (STRYJEWSKI, 2010). Neste trabalho aplicamos a metodologia de Tukey (1977), a mais amplamente utilizada em *softwares* de estatística, como um filtro para a exclusão dos casos *limites* e que descrevem pouco da regularidade dos resultados.

Explicitamente, a metodologia gráfica define os limites das barras como sendo os quartis de 25% (primeiro quartil ou Q_1) e 75% (terceiro quartil ou Q_3): o chamado *interquartile range* ou *IQR*; e os *whiskers*, as linhas acima e abaixo das barras, se estendem 1,5 vezes pela diferença entre o valor do segundo quartil e o terceiro quartil.

Dessa forma, é possível estabelecer dois limites que poderiam ser aplicados aos coeficientes que destoam do padrão identificado na maioria dos casos: exclui-se os casos que ultrapassem os limites inferiores e superiores. As expressões abaixo deixam o procedimento mais explícito e seguem a explicação providenciada em Stryjewski (2010):

$$IQR = Q_3 - Q_1 \quad (12)$$

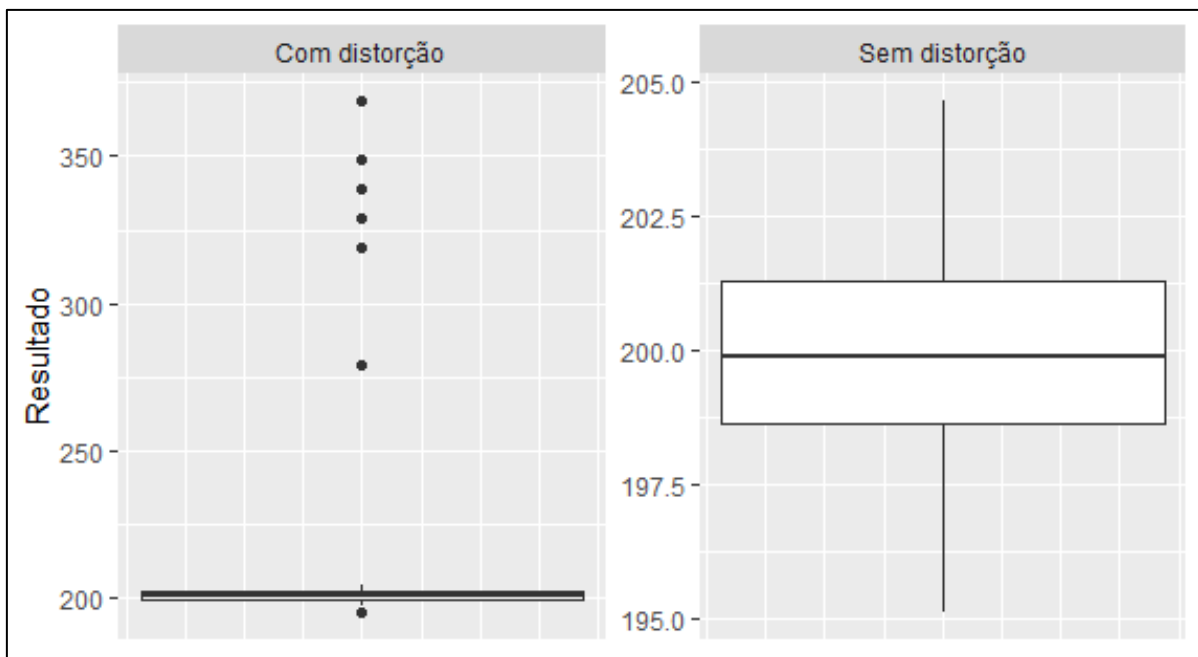
$$Limite Inferior = Q_1 - 1.5IQR \quad (13)$$

$$Limite Superior = Q_3 + 1.5IQR \quad (14)$$

É possível buscar uma intuição visual da motivação. Em um exemplo hipotético introduzindo sete *outliers* em uma distribuição de duzentos valores⁴⁹, é possível facilmente identificar as distorções (Figura 4). A remoção dos valores que excedem os limites do *boxplot* permite identificar onde a maior parte das distribuição se encontra.

⁴⁹ Simulação com dados imputados, construída como mero exemplo.

Figura 5 - Exemplo Ilustrativo para motivar o uso de um filtro do tipo *boxplot*



Fonte: Elaboração própria

Assim, a exclusão de valores limites ou *outliers*, assumindo que a estimação produziu um conjunto sensível de valores, é uma opção conservadora e não deve afetar sensivelmente as conclusões da presente investigação. Deve-se ressaltar, também, que os coeficientes explosivos (evidência de falha de convergência) no contexto de uma regressão logística são notavelmente grandes e destacam-se enormemente em relação aos demais coeficientes estimados.

Contudo, este filtro somente providencia uma forma de identificar os coeficientes suficientemente adequados para a análise, mas ainda pode ser insuficiente para informar o leitor deste trabalho sobre os principais movimentos dos coeficientes. Dito de outra forma, é interessante verificar padrões nacionais ou regionais desses coeficientes e isso requer um passo adicional na análise.

Nesse caso, é necessário utilizar recursos de *meta-análise* como apresentados em trabalhos como Hartung, Knapp e Sinha (2008). Abaixo, este trabalho produz mais detalhes.

3.6 Meta-análise nacional e regional

A presente investigação também enxerga como potencialmente interessante uma análise dos coeficientes regional e nacionalmente, assim como antes e durante a pandemia do COVID-19. A análise padrão para agregar coeficientes é o uso da propriedade do *inverse variance weighting*, conforme apresentado em sua forma mais simples em Hartung, Knapp e Sinha (2008). Em maior detalhe a fórmula da metodologia é a seguinte:

$$\hat{\beta} = \frac{\sum_i \beta_i / \sigma_i^2}{\sum_i 1 / \sigma_i^2} \quad (15)$$

Onde $\hat{\beta}$ é média dos β_i estimados e de menor variância, ponderada pelas variâncias estimadas de cada coeficiente (σ_i^2). Ou seja, a proporcionalidade dos pesos é dada pela precisão das estimativas.

Como a variância dos betas é uma função do tamanho das amostras, a metodologia também tende a dar uma maior proporção aos coeficientes de amostras maiores — isto é, as Regiões de Saúde de maior tamanho tendem a possuir um maior impacto nos coeficientes estimados.

3.7 Escolha do paciente e interpretação dos resultados

O sistema de informações hospitalares do SUS fornece elementos relevantes para informar o processo de escolha do paciente. Na investigação realizada por este trabalho, é possível inferir que ocorreram possíveis alterações na abertura para a seleção hospitalar de pacientes. Por exemplo, no caso de internações eletivas, constava foi frisado que o paciente teria o seguinte direito:

“[...] de posse da AIH o paciente terá direito à escolha do Hospital de sua preferência, dentre os prestadores de serviços para SUS e a

internação se dará de acordo com a disponibilidade de vaga.” (BRASIL, 2004, pág 13).

Porém, em versões mais recentes do manual o detalhamento da escolha do paciente não é mais frisado: ressalta-se a necessidade de um atendimento ambulatorial prévio, que resultará na solicitação da AIH através do laudo da consulta (BRASIL, 2007). Deve-se ressaltar, porém, que a internação também pode ocorrer no estabelecimento de escolha do paciente: um paciente informado pode selecionar o estabelecimento de acordo com as suas preferências antes de uma eventual autorização. Outros trabalhos, que realizaram questionários com pacientes do SUS, também compreendem que há espaço para a escolha do hospital, pois pacientes ainda elencam credibilidade no estabelecimento ou distância como critério para seleção de uma alternativa hospitalar (DUBEAUX, FREESE e FELISBERTO, 2013).

É possível, portanto, assumir que há espaço para escolha do paciente, mas há limitações ao processo de escolha que precisam ser destacadas. É possível destacar alguns elementos: i) internações de urgência ou fruto de acidentes, onde não necessariamente há uma escolha do paciente; ii) a redução drástica de procedimentos eletivos durante a pandemia de COVID-19, que aconteceu de maneira independente e desigual em várias partes do país; iii) e situações em que a internação não foi o estabelecimento de primeira escolha do paciente, mas ocorreu por um direcionamento de outro estabelecimento vinculado ao SUS.

Uma investigação detalhada desses possíveis vieses foge do escopo deste trabalho. Mas é importante compreender que algumas métricas na presente análise podem ser afetadas: a análise da distância entre um paciente ou o direcionamento de um paciente em virtude de uma especialidade associada a um diagnóstico pode gerar particularidades interpretativas.

No caso do período que antecede a pandemia do COVID-19, é possível estabelecer que a variável de distância é um componente que de fato faz parte do processo de escolha, considerando a presença relevante de procedimentos eletivos. No período da pandemia do COVID-19, com uma dinâmica diferente de internações em virtude do cancelamento de um conjunto substancial de procedimentos, conforme CFM (2021), é possível assinalar que a variável de distância possui uma lógica

potencialmente distinta no processo de seleção de um hospital – antes a seleção estava mais associada à escolha do paciente, posteriormente (com o início da Pandemia) possivelmente menos associada. Trata-se, portanto, de uma dinâmica que junta os itens “i” e “ii” listados acima: a composição das internações se altera durante a pandemia e afeta a interpretação dos resultados.

No caso da escolha de um hospital associada a algum tipo específico de diagnósticos, é possível supor que há novamente uma redução do espaço de escolha do paciente e há um “direcionamento”, em função da possibilidade de atendimento emergenciais prevalecerem.

Por último, as escolhas dos pacientes também podem ser afetadas por uma busca inicial de um estabelecimento hospitalar ou posto de saúde que direcionou o paciente para um outro estabelecimento. Esse viés potencialmente afeta os dois períodos de estimação dos dados indistintamente e, em verdade, gera um problema interpretativo dos resultados, já que estabelecimento não é necessariamente o estabelecimento final de escolha do paciente: a seleção está afetada pelo direcionamento dado um profissional em um outro estabelecimento.

Contudo, considerando a redução de atendimentos eletivos, a especificação econométrica utilizada e o interesse mais generalista da investigação deste trabalho, não se buscou uma análise que buscasse separar os potenciais efeitos associados aos diferentes tipos de internação, considerando as possíveis mudanças de composição antes e durante a pandemia do COVID-19 e a separação dos efeitos de escolha do paciente em relação ao de uma orientação profissional em outro estabelecimento.

4 RESULTADOS: EFEITOS DA PANDEMIA DO COVID-19

Este capítulo está dividido em três grandes tópicos: em uma primeira subseção, o capítulo discute a convergência e capacidade explicativa do modelo, além da distribuição regional dos resultados; uma segunda subseção trata sobre as mudanças de magnitude e diferenças entre os coeficientes antes e durante a pandemia do COVID-19.

A seguir, a terceira subseção apresenta as mudanças de magnitude e direção dos parâmetros, utilizando média ponderadas pelo inverso da variância e separa-se os coeficientes que mantiveram a sua direção consistente e mudaram a sua magnitude durante após o início da pandemia do COVID-19.

4.1 Convergência e capacidade explicativa dos modelos

A estratégia comparativa deste trabalho procurou analisar somente regiões que tiveram resultados que convergiram antes e durante a pandemia, mantendo uma base comparativa estável. A Tabela 6 abaixo resume, de acordo com a estratégia de regressão implementada (a análise em painel antes e durante a pandemia do COVID-19) e o número de Regiões de Saúde que tiveram regressões que convergiram antes e depois da pandemia.

Tabela 6 – Percentual de Regiões de Saúde que tiveram pelo menos um resultado convergente antes e durante a Pandemia do COVID-19

Análise	Total de Regiões	Total de Regiões com convergência	Percentual de Convergência
Painel de 5 Meses - Antes e Depois	450	384	85,3%

Fonte: Elaboração própria.

A subseção “Problemas de convergência do modelo *logit*”, discutida anteriormente neste documento, já debateu os potenciais problemas associados a um modelo do tipo *logit* não convergir. Há os problemas de separação ou quase-separação, discutidos em Alisson (2008), que podem surgir nesses modelos, além de problemas de capacidade computacional ou de tempo para convergência do modelo.⁵⁰

⁵⁰ As regressões implementadas pelo pacote *survival* de Therneau e Lumley (2014).

Na maior parte das regressões implementadas, o número de observações ultrapassa os milhões, considerando a expansão das observações para cada opção hospitalar em uma Região de Saúde. Outro fator é que se observa somente as Regiões de Saúde que convergiram antes e depois, com as regressões tipicamente falhando em convergir após o início da pandemia do COVID-19⁵¹.

De todo modo, a maior parte das Regiões de Saúde está representada na análise. Assim, do ponto de vista de uma análise *antes e depois* geral ou uma análise regional, este trabalho entende que as regressões aqui implementadas são suficientes para os propósitos deste trabalho. A Tabela 7 abaixo demonstra as convergências dos resultados por região geográfica:

Tabela 7 – Regressões que convergiram nas Regiões de Saúde

Região	Regiões de Saúde com Análises que Convergiram Antes e Depois da Análise	Total de Regiões de Saúde	Proporção
Centro-Oeste	35	39	89,7%
Nordeste	106	133	79,7%
Norte	40	45	88,9%
Sudeste	142	165	86,1%
Sul	61	68	89,7%

Fonte: Elaboração própria.

Contudo, algumas capitais, após o início do período da pandemia, não convergiram na integralidade dos dados – como Rio de Janeiro, São Paulo, Recife, Porto Alegre, Belo Horizonte e Curitiba. Os dados se revelaram de proporções enormes para a capacidade computacional e a análise foi efetuada com apenas um único mês – o mês de junho, antes e durante a pandemia do COVID-19. Como os valores estimados à frente não mudam substancialmente com a inclusão dessas capitais em apenas um mês, optou-se pela inclusão da análise com apenas esses meses.

Do ponto de vista da qualidade de ajuste das regressões, entende-se que um *score* entre 0,2 e 0,4 (ou maior) do Pseudo-R² de Mcfadden revela um ajuste suficientemente adequado de uma regressão logística. A Tabela 8 abaixo detalha os *scores* das Regressões de todos os resultados que convergiram.

⁵¹ Cerca de 421 Regiões de Saúde convergem observando apenas o resultado anterior ao início da pandemia. Uma outra fonte de possível falha de convergência é a mudança no perfil de atendimentos, com a queda substancial para alguns atendimentos – impactando na estratégia adotada.

Tabela 8 – Percentual de Regiões de Saúde que tiveram pelo menos um resultado convergente antes e durante a Pandemia do COVID-19

Período	Pseudo-R2 de McFadden	Nº de Regressões
Antes - Média de 5 Anos (2015-2019)	$R^2 < 0,1$	7
Antes - Média de 5 Anos (2015-2019)	$R^2 \geq 0,1$ & $R^2 < 0,2$	17
Antes - Média de 5 Anos (2015-2019)	$R^2 \geq 0,2$ & $R^2 < 0,4$	194
Antes - Média de 5 Anos (2015-2019)	$R^2 \geq 0,4$	166
Depois - abril a agosto de 2020	$R^2 < 0,1$	1
Depois - abril a agosto de 2020	$R^2 \geq 0,1$ & $R^2 < 0,2$	4
Depois - abril a agosto de 2020	$R^2 \geq 0,2$ & $R^2 < 0,4$	78
Depois - abril a agosto de 2020	$R^2 \geq 0,4$	301

Fonte: Elaboração própria.

Conforme a Tabela 8 destaca, a grande maioria das regressões possuem um ajuste suficientemente adequado conforme o *rule of thumb* para a regressão logística e o Pseudo-R² de Mcfadden. Em todos os cenários, não menos que 95% das regressões revelam um valor suficientemente adequado, isto é, dentre as que não tiveram problemas de convergência.

Cabe ressaltar que de um universo de cerca de possíveis 30.720 coeficientes (de 768 regressões que convergiram de 900 possíveis), o que contempla coeficientes que não foram possíveis de serem estimados, 7.310 coeficientes se demonstraram significantes e não foram excluídas da análise em virtude do filtro adotado.

4.2 Análise dos parâmetros estimados

Este trabalho efetuou uma análise antes e durante a pandemia de todos os parâmetros estimados, conforme destacados na seção de dados. As análises efetuaram as interações similares a Ho (2006) e Ho e Lee (2017) passíveis de serem efetuadas com dados da SIH-RD e dados CNES, além das demais incorporações discutidas na seção de dados deste trabalho. Mais detalhes sobre as variáveis podem ser encontrados no Apêndice A (Tabela A.1, com os resultados completos).

Antes de tudo, destaca-se uma característica particular do presente modelo e da apresentação dos dados: o modelo *logit* condicional não permite a obtenção de

parâmetros de efeitos marginais, em razão dos efeitos fixos não serem estimados na presente análise, mas excluídos. Assim, os coeficientes obtidos neste trabalho são interpretados pelos seus respectivos *odds ratio* – ou seja, os aumentos das chances (*likelihood*). Os valores de *odds ratio* podem ser interpretados da seguinte forma: um *score* de *odds ratio* de 1,15 pode ser interpretado como um aumento de 15% nas chances de que um determinado evento ocorra (seleção de um hospital) em função da variação positiva de uma unidade de um determinada variável (e.g. distância percorrida), assim como um *score* de 0,85 pode ser interpretado como uma redução de 15% nas chances em função de uma variação positiva de um unidade da variável. Dessa maneira, valores de *odds ratio* muito próximos a 1 revelam efeitos pouco substanciais, enquanto *odds ratios* menores do que 1 ou maiores que 1 em montas mais substanciais podem refletir efeitos mais relevantes, seja em um aumento da chance (maiores do que 1) ou em reduções da chance (menores do que 1). No caso de variáveis de interações (e.g. característica do paciente e característica hospitalar), a variação de uma unidade positiva está associada ao fato de se ter duas características reunidas, que pode ser uma *dummy* (assume os valores 0 ou 1) ou uma variável assumindo diversos valores, como as variáveis de serviços ou renda.

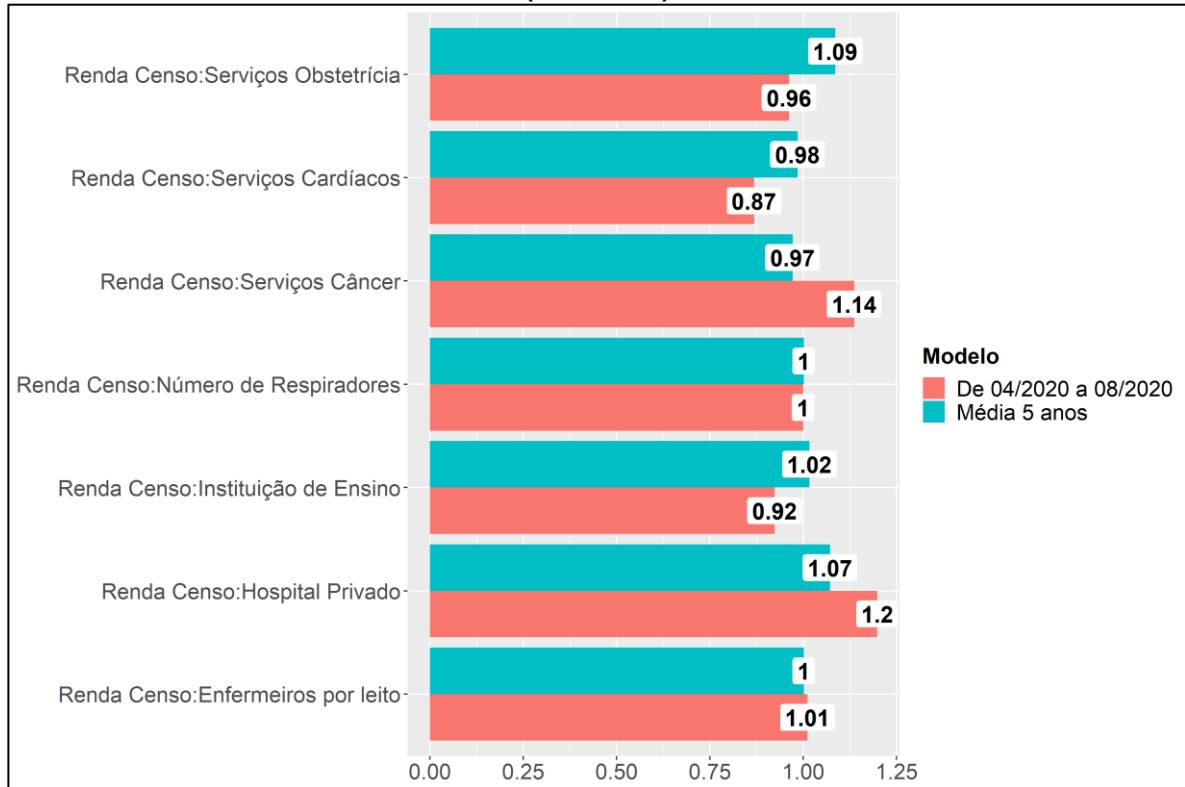
Também focamos apenas o grupo de coeficientes que se demonstraram significantes, bem-comportados ou interessantes – o resultado completo pode ser observado no Anexo A deste trabalho.

Alguns destes parâmetros adquiriram uma direção distinta em relação ao trabalho de Ho (2006) e Ho e Lee (2017)⁵². É importante destacar, porém, que a análise efetuada em Ho (2006) e Ho e Lee (2017) contempla não somente pacientes internados, mas se trata de uma base mais ampla, com pacientes do sistema privado. Ademais, os dados tratam de um país distinto e um período também consideravelmente distinto que o período deste trabalho (os anos de 2015 e 2020).

Não obstante esses aspectos, vejamos a Figura 6, que destaca a relação entre os parâmetros de interação com a renda antes e durante a pandemia:

⁵² É o caso de serviços de imagem e diagnósticos de gravidez, por exemplo.

Figura 6 — Gráfico de Antes e Durante da Pandemia do COVID-19 –Interações com Renda (odds ratio)



Fonte: Elaboração própria.

É possível observar que boa parte dos valores é próxima a um score de um 1 ou ligeiramente abaixo ou acima do valor. Ou seja, no geral, as interações das variáveis de renda possuem uma relação inexistente entre a renda do paciente e os serviços prestados pelas unidades hospitalares analisadas — conforme se pode observar nos coeficientes obtidos no Anexo A. Dito de outra forma, as chances de escolha da unidade hospitalar associada às demais qualidades (e.g. instituições de ensino, número de respiradores, serviços cardíacos e de obstetrícia) não governam a escolha do paciente. Há pouca diferença de magnitude em termos *odds ratio*, nas alterações das chances de escolha, mas a Tabela 9 demonstra que os coeficientes estimados para os casos significantes possuem um erro padrão (“EP”) consideravelmente baixo e um teste hipótese rejeitaria a semelhança entre todos os coeficientes antes e durante a pandemia,⁵³ mesmo apesar da baixa diferença em magnitude.

Tabela 9 – Antes e Durante a Pandemia do COVID-19 por Região do Brasil – interações com Renda (Odds Ratio) – Coeficientes, Desvio Padrão e Odds Ratio

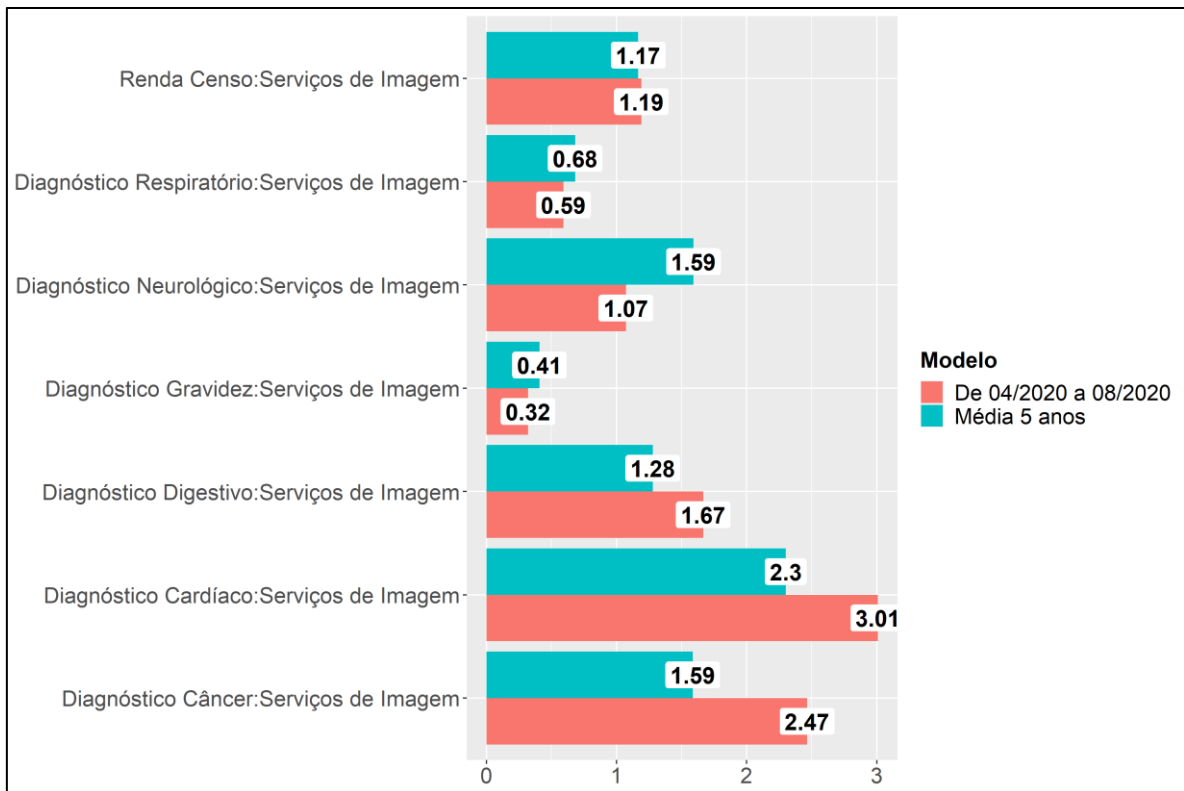
Variáveis	04/2020 a 08/2020			Média 5 anos - 2015-2019		
	Coef.	EP	OR	Coef.	EP	OR
Renda Censo:Enfermeiros por leito	0,0112	0,0001	1,0112	0,0014	0,0001	1,0014
Renda Censo:Hospital Privado	0,1805	0,0020	1,1978	0,0691	0,0010	1,0715
Renda Censo:Instituição de Ensino	-0,0798	0,0020	0,9233	0,0162	0,0010	1,0164
Renda Censo:Número de Respiradores	-0,0007	0,0000	0,9993	0,0012	0,0000	1,0012
Renda Censo:Serviços Câncer	0,1277	0,0026	1,1362	-0,0292	0,0012	0,9712
Renda Censo:Serviços Cardíacos	-0,1414	0,0023	0,8682	-0,0154	0,0011	0,9847
Renda Censo:Serviços Obstetrícia	-0,0389	0,0020	0,9618	0,0826	0,0010	1,0861

Fonte: Elaboração própria. Nota: Coef.: Coeficiente, DP: Erro Padrão. OR: *Odds Ratio*.

A Figura 6 abaixo destaca, por sua vez, os valores observados para exames de imagem, que adquiriram valores maiores e menores do que 1 – em termos de coeficientes, direção positivas e negativas (ver Anexo A). Em alguns casos, há uma direção distinta dos resultados de primeiro estágio observados em Ho (2006) e Ho e Lee (2017)⁵⁴. Mas é importante ressaltar como parte desses valores adquire uma direção diferente antes e durante a pandemia.

⁵⁴ Diagnósticos de gravidez e serviços de imagem, por exemplo.

Figura 7 – Gráfico de Antes e Durante a Pandemia do COVID-19 – Interação com Serviços de Imagem (*odds ratio*)



Fonte: Elaboração própria.

Notadamente, há diferenças mais pronunciadas de magnitude em três casos específicos de interação com serviços de imagem: diagnósticos de digestivos, cardíacos e de câncer. Nestes três casos o valor do *odds ratio* é acima de 1 antes do início da pandemia e aumenta a sua magnitude após o início da pandemia em quase duas vezes em alguns casos (diagnósticos digestivos e diagnósticos de câncer). Em detalhe, observa-se um aumento das chances associadas à seleção de serviços de imagem quando o paciente possui o diagnóstico de problemas digestivos de 28% antes do início da pandemia e 67% após o início da pandemia; no caso de diagnósticos cardíacos associados ao mesmo serviço, 130% antes da pandemia e 201% após o início da pandemia; com o diagnóstico de câncer, 59% antes da pandemia e 147% durante a pandemia.

No caso da interação com renda, em termos de *odds ratio* o valor obtido fica acima de 1, mas com poucas diferenças marcantes entre um período e outro – chances de 17% (antes da pandemia) e 19% (durante a pandemia). No caso da interação com diagnósticos de gravidez, a relação também muda pouco, mas é menor

do que 1 (coeficientes negativos), o que indica um resultado pouco intuitivo. No caso das interações de imagem com diagnósticos respiratórios, o *score* de *odds ratio* também é menor do que 1 e com uma queda durante a pandemia – o *score* sai de 0,68 e vai para 0,59, ou seja, reduções de chances 32% antes da pandemia e reduções de chances de 41% durante a pandemia. No caso de diagnósticos neurológicos, a interação se mantém positiva, mas cai durante a pandemia (o *score* sai de 1,59 e vai para 1,07, ou seja, os aumentos de chances 59% que caem para 7%).

Cabem algumas considerações. É bem estabelecida a redução de procedimentos eletivos, como cirurgias (CFM, 2021). Considerando que a natureza das internações e procedimentos realizados com internações mudaram, é razoável supor que parte da mudança pode estar associada à redução de procedimentos postergáveis. A fonte exata dessa variação, porém, não entra no escopo deste trabalho.

Cabe também um esclarecimento para outros valores da Figura 7. No caso de serviços de imagem, a base trata de internações pelo SUS e não cobre a ida ao hospital para um atendimento ambulatorial, exceto se o atendimento ambulatorial foi sucedido por uma internação – trata-se de uma questão não controlada na presente análise, conforme discutido anteriormente. De fato, atendimentos para procedimentos de imagem e outros exames específicos não precisam estar associados à escolha do estabelecimento em que ocorrerá a internação do paciente.⁵⁵ Trata-se de razões que potencialmente explicam o *odds ratio* abaixo de 1 para a interação com diagnósticos de gravidez. Outras razões em potencial podem estar associadas a particularidades específicas de hospitais especializados⁵⁶ que não atendem determinados tipos de diagnósticos. Por fim, cabe ressaltar que o modelo analisado é um modelo de efeitos fixos, controlando características do indivíduo e do hospital, assim cabe considerar que as interações adquiram direções inesperadas.

A Tabela 10, à maneira da tabela anterior, demonstra novamente a baixa dispersão alcançada (baixo desvio padrão), mas demonstra a diferença de magnitude já discutida.

⁵⁵ Os ambulatórios, com os AMES (Ambulatórios Médicos de Especialidades) presentes no estado de São Paulo, prestam serviços de imagem e outros diagnósticos conveniados ao SUS e não estão associados a uma estrutura hospitalar necessariamente.

⁵⁶ É possível que um hospital pediátrico possua serviços de imagem, entre outros serviços destacados, e ao mesmo tempo não atenda pacientes adultos com câncer, problemas cardíacos ou digestivos.

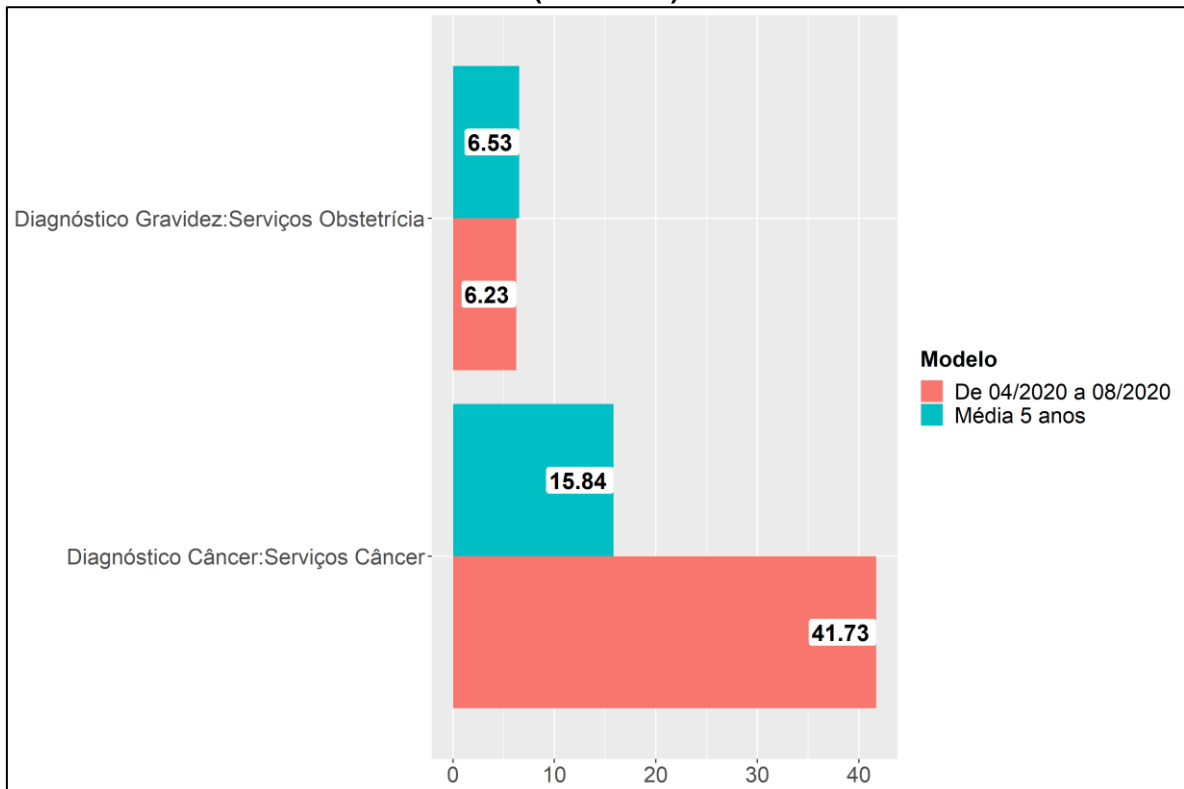
Tabela 10 – Antes e Durante a Pandemia do COVID-19 – Coeficientes de interação com Serviços de Imagem – Coeficientes, Desvio Padrão e Odds Ratio

Variáveis	04/2020 a 08/2020			Média 5 anos - 2015-2019		
	Coef.	EP	OR	Coef.	EP	OR
Diagnóstico Câncer:Serviços de Imagem	0,9023	0,0165	2,4653	0,4613	0,0070	1,5862
Diagnóstico Cardíaco:Serviços de Imagem	1,1016	0,0171	3,0089	0,8331	0,0066	2,3004
Diagnóstico Digestivo:Serviços de Imagem	0,5106	0,0131	1,6663	0,2460	0,0048	1,2790
Diagnóstico Gravidez:Serviços de Imagem	-1,1403	0,0106	0,3197	-0,8966	0,0052	0,4080
Diagnóstico Neurológico:Serviços de Imagem	0,0695	0,0268	1,0720	0,4649	0,0102	1,5919
Diagnóstico Respiratório:Serviços de Imagem	-0,5230	0,0105	0,5927	-0,3817	0,0057	0,6827
Renda Censo:Serviços de Imagem	0,1746	0,0026	1,1908	0,1546	0,0013	1,1672

Fonte: Elaboração própria. Nota: Coef.: Coeficiente, DP: Erro Padrão. OR: Odds Ratio.

Para o fim desta subseção, na Figura 8 e Figura 9 abaixo, apresenta-se os coeficientes que tiveram direções bem fortes (Figura 8) e outros coeficientes que demonstraram intuição no resultado, incluindo um caso com pouca diferença em magnitude no antes e durante a pandemia do COVID-19 (número de respiradores interagido com diagnóstico de doenças respiratórias), mas que este trabalho julgou interessante frisá-lo (Figura 9). Na Figura 9, observa-se uma direção intuitiva e esperada para os dois demais coeficientes: as distâncias percorridas e as interações de serviços cardíacos com doenças cardíacas – mais detalhes interpretativos serão dados à frente. Esses coeficientes validam também os resultados de outros trabalhos, como o de Cruppé e Geraedts (2011), que elenca distância, familiaridade e especialidade do hospital como as características mais relevantes para internação em seu *survey*. Os coeficientes estão separados em duas Figuras por uma motivação de visualização.

Figura 8 Gráfico de Antes e Durante a Pandemia da COVID-19 – Valores mais expressivos (odds ratio)

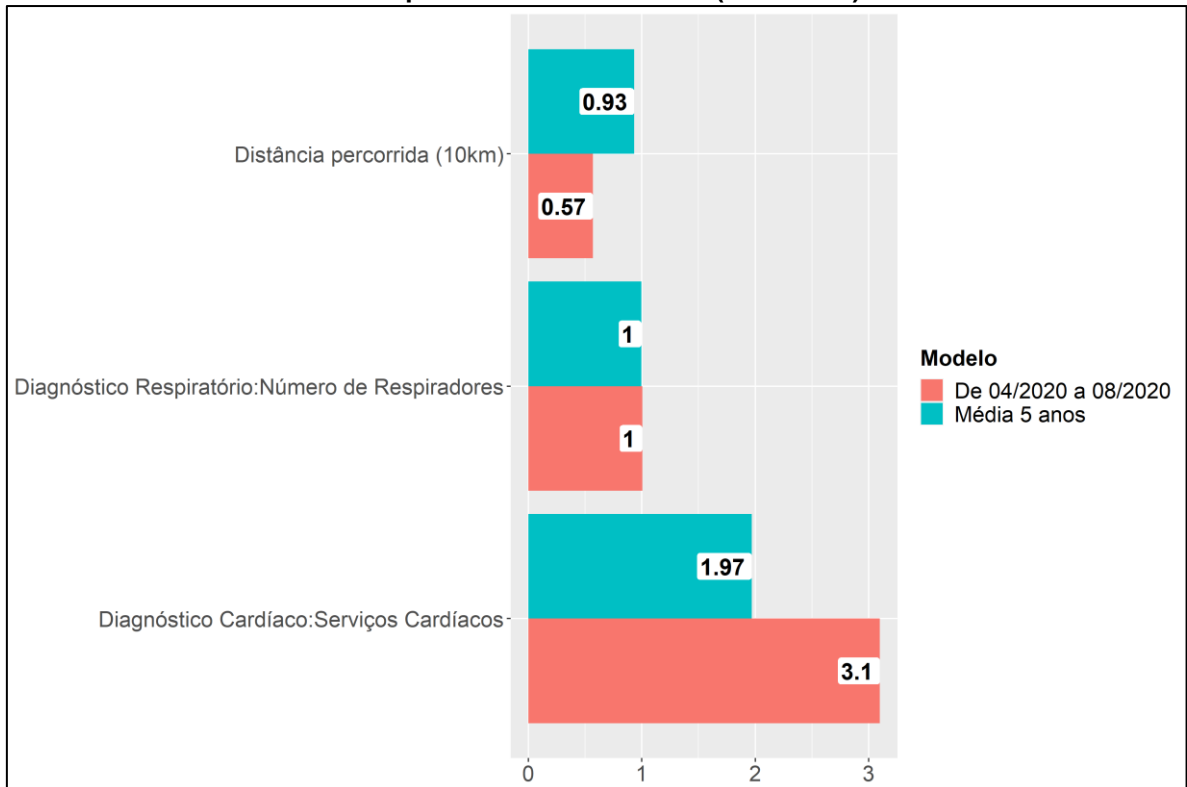


Fonte: Elaboração própria.

O valor de *odds ratio* de serviços de obstetrícia e diagnóstico de obstetrícia adquire uma direção esperada e consistente com outras estimativas – ver Ho (2006). O *odds ratio* de serviços de Câncer e diagnóstico de câncer também adquire uma direção positiva e esperada, mas de magnitude consideravelmente maior do que outras estimativas analisadas. É importante ressaltar que um *odds ratio* de tal magnitude já foi encontrado, por exemplo, nas relações entre serviços de obstetrícia e o diagnóstico associado a este serviço no exercício construído por Ho (2006), mas não se deve descartar que o problema de separação não foi inteiramente resolvido. De fato, nas análises regionais discutidas em seção posterior, há evidência de valores substancialmente altos – assim, julga-se que há uma certa limitação com as conclusões do resultado da Figura 8.

Neste último caso (interações de serviços de câncer com o diagnóstico de câncer) há também um aumento de magnitude durante a pandemia do COVID-19. Trata-se, também, de resultado possivelmente associados à discussão trazida sobre procedimentos eletivos, mas que fogem do escopo deste exercício.

Figura 9 - Gráfico de Antes e Durante a Pandemia da COVID-19 – Valores com o comportamento consistente (odds ratio)



Fonte: Elaboração própria.

No primeiro resultado em específico, em relação às distâncias percorridas, é possível identificar uma redução do *odds ratio*: possivelmente um aumento da desutilidade no quilômetro percorrido pelo paciente até o hospital no período da pandemia. Antes da pandemia, 10 km percorridos adicionais estavam associados a uma diminuição das chances de seleção de um hospital de 7% - com a pandemia, 10 km adicionais passaram a estar associados a uma redução de chance de 43%.

É importante ressaltar que há possíveis características não controladas que talvez afetem esse comportamento: os pacientes que buscaram internações talvez sejam de um perfil que viaja distâncias maiores, a julgar pelas limitações de atendimentos durante a pandemia do COVID-19. De todo modo, o coeficiente apresentou uma direção esperada: quanto maior a distância percorrida, maior o desinteresse pela escolha dos hospital.

O coeficiente de diagnósticos respiratórios, conforme a Figura 8 (ou a Tabela 11 abaixo), interagindo com o número de respiradores obteve um *odds ratio* levemente menor que 1 em períodos anteriores à pandemia e ligeiramente acima de 1 durante a pandemia, mas os valores arredondados são iguais. É importante ressaltar mais uma

vez que a construção da variável de diagnósticos respiratórios inclui o diagnóstico associado à COVID-19 da tabela CID-10, na abrangência do terceiro dígito (“B34”), de magnitude consideravelmente menor antes da pandemia — a motivação dessa escolha é feita no segundo capítulo deste trabalho,⁵⁷ pois assume-se a possibilidade de inscrições erradas de COVID-19 entre o maior grupo de doenças respiratórias. Contudo, em termos práticos um diagnóstico de doença respiratória e o número de respiradores não está particularmente associado ao processo de escolha do paciente, controlados efeitos fixos individuais e hospitalares. É importante ressaltar que os valores são estatisticamente distintos de zero, em virtude dos erros padrões extremamente baixos, mas em termos práticos a diferença é inexistente.

O coeficiente de interação serviços cardíacos adquiriu uma tendência intuitiva e de aumento durante a pandemia, em movimento similar à interação de diagnósticos de imagem – portanto, o mesmo ocorre na Figura 8 com o *odds ratio*. Ao que tudo indica, pacientes com o diagnóstico de doenças cardíacas possuem forte associação à escolha por hospitais com serviços cardíacos e há um aumento dessa relação durante a pandemia do COVID-19. Especificamente, o diagnóstico cardíaco e serviços cardíacos hospitalares estiveram associados a um aumento de chances de 97% de seleção do hospital (antes de pandemia) e 210% (durante a pandemia).

A Tabela 11 abaixo demonstra alguns resultados adicionais de distância percorrida que foram controlados nas regressões: a distância ao quadrado e a interação da distância com emergência. Em ambos os casos, não há diferença sensível na magnitude dos coeficientes antes e durante a pandemia do COVID-19. Para os demais coeficientes já discutidos, a Tabela 10 descreve que os resultados são significantes.

⁵⁷ Reforçamos que a escolha da inclusão da variável entre os diagnósticos respiratórios se deve às possíveis dificuldades iniciais de reconhecimento da doença e, portanto, sendo possível a sua confusão com outras doenças respiratórias, como pneumonia, influenza, entre outras.

Tabela 11 Antes e Durante a Pandemia do COVID-19 – Coeficientes com o comportamento consistente - Coeficientes, Desvio Padrão e Odds Ratio

Variáveis	2020			Média 5 anos - 2015-2019		
	Coef.	EP	OR	Coef.	EP	OR
Diagnóstico Câncer:Serviços Câncer	3,7311	0,0184	41,7250	2,7627	0,0074	15,8423
Diagnóstico Cardíaco:Serviços Cardíacos	1,1311	0,0190	3,0991	0,6777	0,0072	1,9694
Diagnóstico Gravidez:Serviços Obstetrícia	1,8293	0,0096	6,2294	1,8768	0,0043	6,5323
Diagnóstico Respiratório:Número de Respiradores	0,0047	0,0001	1,0047	-0,0042	0,0001	0,9958
Distância percorrida (10km)	-0,5625	0,0008	0,5698	-0,0688	0,0000	0,9335
Distância percorrida (10km) ao quadrado	0,0012	0,0000	1,0012	0,0002	0,0000	1,0002
Distância percorrida (10km)Emergência	0,0005	0,0002	1,0005	0,0012	0,0001	1,0012

Fonte: Elaboração própria. Nota: Coef.: Coeficiente, EP: Erro Padrão. OR: *Odds Ratio*.

Por fim, reitera-se que os resultados não indicam relações necessariamente causais da pandemia do COVID-19, mas apresentam o comportamento de um modelo estrutural antes e durante o período da Pandemia do COVID-19.

Outros coeficientes não foram destacados ou porque adquiriram diferenças pouco substanciais ou não adquiriram direções interessantes para uma discussão aprofundada, como as interações da razão entre enfermeiros e leitos e diagnósticos. Um aprofundamento dessas relações poderão ser fruto de uma investigação futura com a base de dados deste trabalho.

5 RESULTADOS: VARIAÇÕES REGIONAIS NOS PARÂMETROS ESTIMADOS

Esta seção trata dos coeficientes discutidos anteriormente, mas com uma ótica de variabilidade regional. A investigação implementada nesta seção possui valor próprio na investigação das variações regionais, com o fim de identificar possíveis disparidades, mas há também um valor na validação dos resultados anteriores. Dito de outra forma, se os coeficientes regionais seguirem aproximadamente a mesma tendência do “antes e durante” discutido na seção anterior, há uma validação da tendência observada na dimensão nacional.

Sobre as disparidades regionais, este trabalho entende que nem sempre é possível identificar com clareza as razões dessas diferenças, considerando as afirmações existentes em trabalhos como Skinner (2002), que discutem a variabilidade regional nos EUA e compreendem uma dificuldade para estabelecer com clareza relação de causa. Este trabalho, portanto, busca antes de tudo evidenciar essas diferenças e possivelmente discutir essas disparidades em uma futura agenda de pesquisa.

5.1 Análise do impacto regional

Novamente se implementa com os resultados da seção anterior, mas agregando os valores por Regiões do Brasil, a saber: as regiões Sul, Sudeste, Centro-Oeste, Nordeste e Norte.

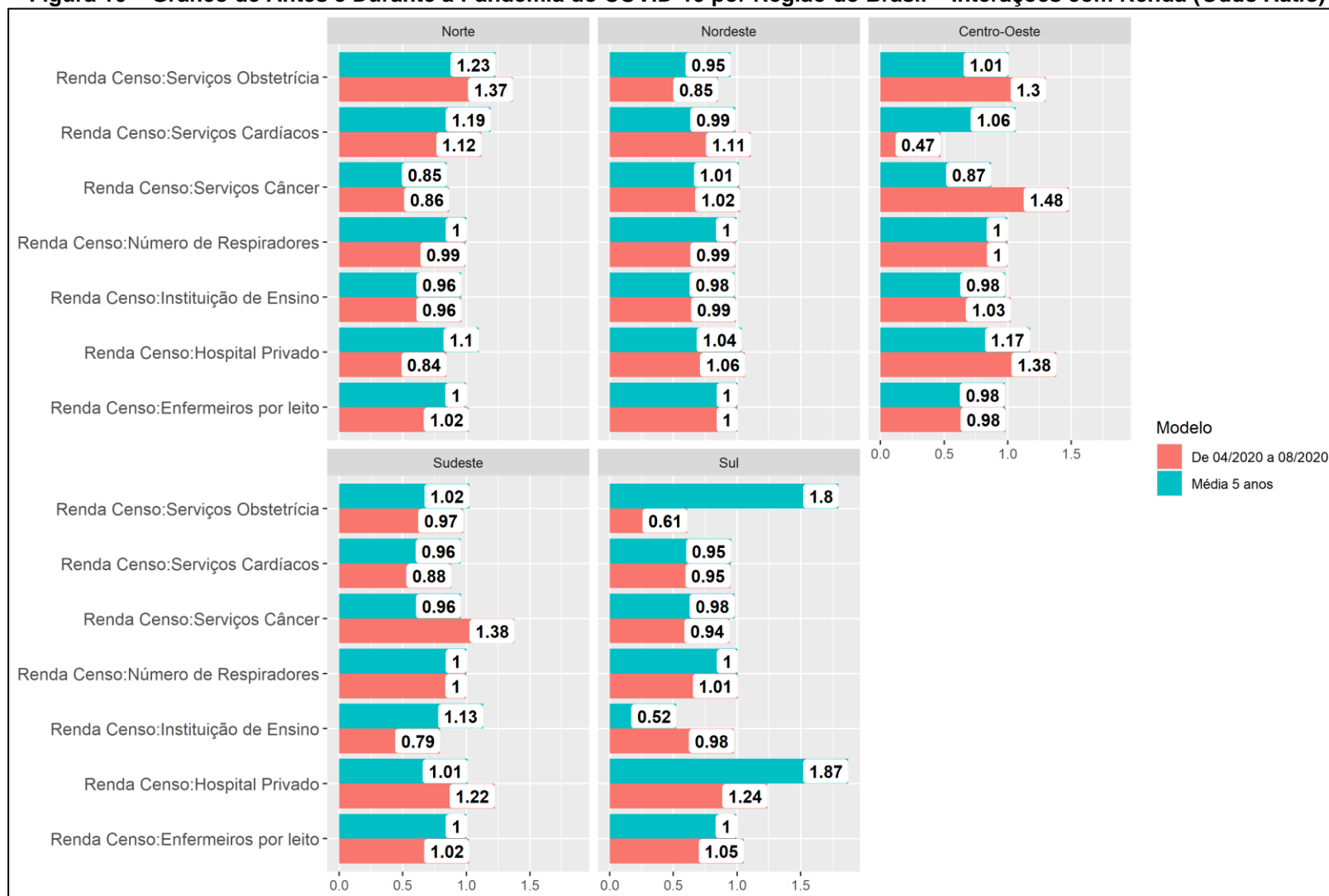
Também se implementa uma análise gráfica, de forma a destacar as diferenças entre os valores de *odds ratio*, assim como uma tabela destacando com maior precisão os valores estimados e o erro padrão. As tabelas demonstram que os resultados obtidos são estatisticamente significantes.

A seção segue a ordem de análise implementada na seção anterior. Primeiro discute-se os coeficientes associados à renda; em seguida, interações com serviços de imagem; e por último, os coeficientes que este trabalho tinha identificado inicialmente como potencialmente consistentes, em intuição, direção e significância estatística — para maiores detalhes de significância, ver a Tabela A.2 no Apêndice A.4 deste trabalho.

A ideia, reforça-se, é verificar como os resultados descritos na seção anterior estão se comportando por região — e, possivelmente, validar alguns dos resultados encontrados como tendências nacionais.

A Figura 10 abaixo descreve o comportamento das interações de renda, mas com a abertura regional: as regiões Sul, Sudeste, Centro-Oeste, Nordeste e Norte. Novamente, os coeficientes são próximos do *score* de 1, o que indica uma consistência regional em relação ao resultado nacional e em termos de *odds ratio*.

Figura 10 – Gráfico de Antes e Durante a Pandemia do COVID-19 por Região do Brasil – Interações com Renda (Odds Ratio)



Fonte: Elaboração própria.

Tabela 12 – Antes e Durante a Pandemia do COVID-19 por Região do Brasil – interações com Renda (Odds Ratio) – Coeficientes, Desvio Padrão e Odds Ratio

Região	Variáveis	2020			Média 5 anos - 2015-2019		
		Coef.	EP	OR	Coef.	EP	OR
CO	Renda Censo:Enfermeiros por leito	-0,0176	0,0005	0,9826	-0,0226	0,0007	0,9776
CO	Renda Censo:Hospital Privado	0,3240	0,0036	1,3826	0,1610	0,0025	1,1747
CO	Renda Censo:Instituição de Ensino	0,0249	0,0040	1,0252	-0,0209	0,0019	0,9793
CO	Renda Censo:Número de Respiradores	-0,0001	0,0000	0,9999	-0,0026	0,0000	0,9974
CO	Renda Censo:Serviços Câncer	0,3911	0,0128	1,4787	-0,1402	0,0076	0,8692
CO	Renda Censo:Serviços Cardíacos	-0,7536	0,0060	0,4706	0,0598	0,0032	1,0616
CO	Renda Censo:Serviços de Imagem	0,1210	0,0047	1,1286	-0,0301	0,0025	0,9704
CO	Renda Censo:Serviços Obstetrícia	0,2624	0,0043	1,3000	0,0061	0,0021	1,0061
NE	Renda Censo:Enfermeiros por leito	0,0028	0,0002	1,0028	0,0029	0,0001	1,0030
NE	Renda Censo:Hospital Privado	0,0613	0,0050	1,0632	0,0351	0,0019	1,0358
NE	Renda Censo:Instituição de Ensino	-0,0081	0,0041	0,9920	-0,0197	0,0018	0,9805
NE	Renda Censo:Número de Respiradores	-0,0139	0,0001	0,9862	-0,0015	0,0000	0,9985
NE	Renda Censo:Serviços Câncer	0,0239	0,0051	1,0242	0,0146	0,0023	1,0147
NE	Renda Censo:Serviços Cardíacos	0,1032	0,0045	1,1087	-0,0141	0,0024	0,9860
NE	Renda Censo:Serviços de Imagem	0,1160	0,0059	1,1230	0,1951	0,0025	1,2154
NE	Renda Censo:Serviços Obstetrícia	-0,1603	0,0052	0,8519	-0,0531	0,0024	0,9483
N	Renda Censo:Enfermeiros por leito	0,0194	0,0004	1,0196	-0,0013	0,0002	0,9987
N	Renda Censo:Hospital Privado	-0,1710	0,0062	0,8428	0,0937	0,0092	1,0983
N	Renda Censo:Instituição de Ensino	-0,0408	0,0061	0,9600	-0,0403	0,0043	0,9605
N	Renda Censo:Número de Respiradores	-0,0082	0,0001	0,9918	0,0007	0,0001	1,0007
N	Renda Censo:Serviços Câncer	-0,1453	0,0111	0,8648	-0,1663	0,0079	0,8468
N	Renda Censo:Serviços Cardíacos	0,1109	0,0171	1,1173	0,1754	0,0045	1,1917
N	Renda Censo:Serviços de Imagem	-0,2361	0,0080	0,7897	-0,1069	0,0049	0,8986
N	Renda Censo:Serviços Obstetrícia	0,3129	0,0077	1,3674	0,2063	0,0091	1,2292
SE	Renda Censo:Enfermeiros por leito	0,0211	0,0003	1,0214	0,0019	0,0001	1,0019
SE	Renda Censo:Hospital Privado	0,2008	0,0034	1,2224	0,0107	0,0014	1,0108
SE	Renda Censo:Instituição de Ensino	-0,2315	0,0035	0,7933	0,1236	0,0016	1,1316
SE	Renda Censo:Número de Respiradores	-0,0011	0,0000	0,9989	0,0020	0,0000	1,0020
SE	Renda Censo:Serviços Câncer	0,3215	0,0040	1,3792	-0,0435	0,0017	0,9574
SE	Renda Censo:Serviços Cardíacos	-0,1248	0,0035	0,8827	-0,0451	0,0014	0,9559
SE	Renda Censo:Serviços de Imagem	-0,3665	0,0056	0,6932	0,1284	0,0025	1,1370
SE	Renda Censo:Serviços Obstetrícia	-0,0253	0,0031	0,9750	0,0231	0,0014	1,0233
S	Renda Censo:Enfermeiros por leito	0,0511	0,0006	1,0525	-0,0041	0,0003	0,9959
S	Renda Censo:Hospital Privado	0,2133	0,0061	1,2378	0,6284	0,0049	1,8745
S	Renda Censo:Instituição de Ensino	-0,0246	0,0078	0,9757	-0,6497	0,0065	0,5222
S	Renda Censo:Número de Respiradores	0,0057	0,0000	1,0057	0,0049	0,0000	1,0049
S	Renda Censo:Serviços Câncer	-0,0624	0,0051	0,9395	-0,0204	0,0028	0,9798
S	Renda Censo:Serviços Cardíacos	-0,0539	0,0056	0,9476	-0,0480	0,0034	0,9532
S	Renda Censo:Serviços de Imagem	1,2249	0,0061	3,4039	0,5802	0,0034	1,7863
S	Renda Censo:Serviços Obstetrícia	-0,4947	0,0048	0,6098	0,5865	0,0027	1,7976

Fonte: Elaboração própria. Nota: Coef.: Coeficiente, EP: Erro Padrão. OR: Odds Ratio.

Como se observa nas Figura 10 e na Tabela 12, os achados não são muito distintos em magnitude em relação aos resultados nacionais (no geral), o que provavelmente corrobora os achados anteriores e não prejudica as conclusões já trazidas. Isto é, a variável de renda tende a explicar pouco a seleção por serviços hospitalares específicos entre internações do SUS⁵⁸. Em praticamente todas as regiões, analisadas os valores de *odds ratio* de interação com renda são próximos ao valor de 1 – o que indica pouco ou nenhum efeito da interação em grande parte dos casos. Mas é possível destacar alguns efeitos regionais.

Na região Norte é possível destacar a interação entre renda e serviços de obstetrícia, com um score maior que 1 antes e após o início da pandemia do COVID-19 – no caso, um valor de 1,23 (média de 5 anos) ou 23% de aumento de chances e outro de 1,37 (durante a pandemia), com 37% de aumento de chances. Na região centro-oeste é observável uma queda substancial na interação entre renda e serviços cardíacos após o início da pandemia: o *score* de *odds ratio* foi de 1,06 (média de 5 anos), ou 6% de aumento de chances, para 0,47 (após o início da pandemia), o que indica 43% de redução de chances. Há, portanto, uma queda substancial no *odds ratio* e reflete uma menor chance de escolha do serviço em função da renda. Na região Sudeste é possível destacar o aumento da interação entre renda e serviços de câncer – o *score* sai de 0,96 (média de 5 anos) ou 2% de redução de chances e vai para 1,38 (após o início da pandemia), 38% de aumento de chances – e a interação entre renda e instituições de ensino, que cai: o *score* sai de 1,13 (média de 5 anos) ou 13% de aumento de chances e vai para 0,79 (durante a pandemia) ou 21% de redução de chances.

A região Sul, porém, destaca-se na interação entre renda e obstetrícia, com um *score* substancialmente alto antes do início da pandemia de 1,80 e um novo *score* de 0,61 com a pandemia. Trata-se de um caso em que a relação se inverte completamente durante a pandemia: a interação entre renda e serviços obstetrícia antes tinha capacidade de aumentar a chance de escolha na região Sul, com um aumento nas chances de seleção do hospital com o serviço em 80% antes do início

⁵⁸ É importante ressaltar que se trata da renda observada de pacientes internados pelo SUS. Não se observa pacientes que não são internados pelo SUS entre as internações. Assim, a renda de pacientes tipicamente internados pelo SUS tem baixa capacidade explicativa da escolha dos serviços, o que é diferente de dizer que renda não tem potencial para explicar uma internação realizada pelo SUS – por exemplo, se fosse modelada a escolha da saúde suplementar e o atendimento por hospitais privados.

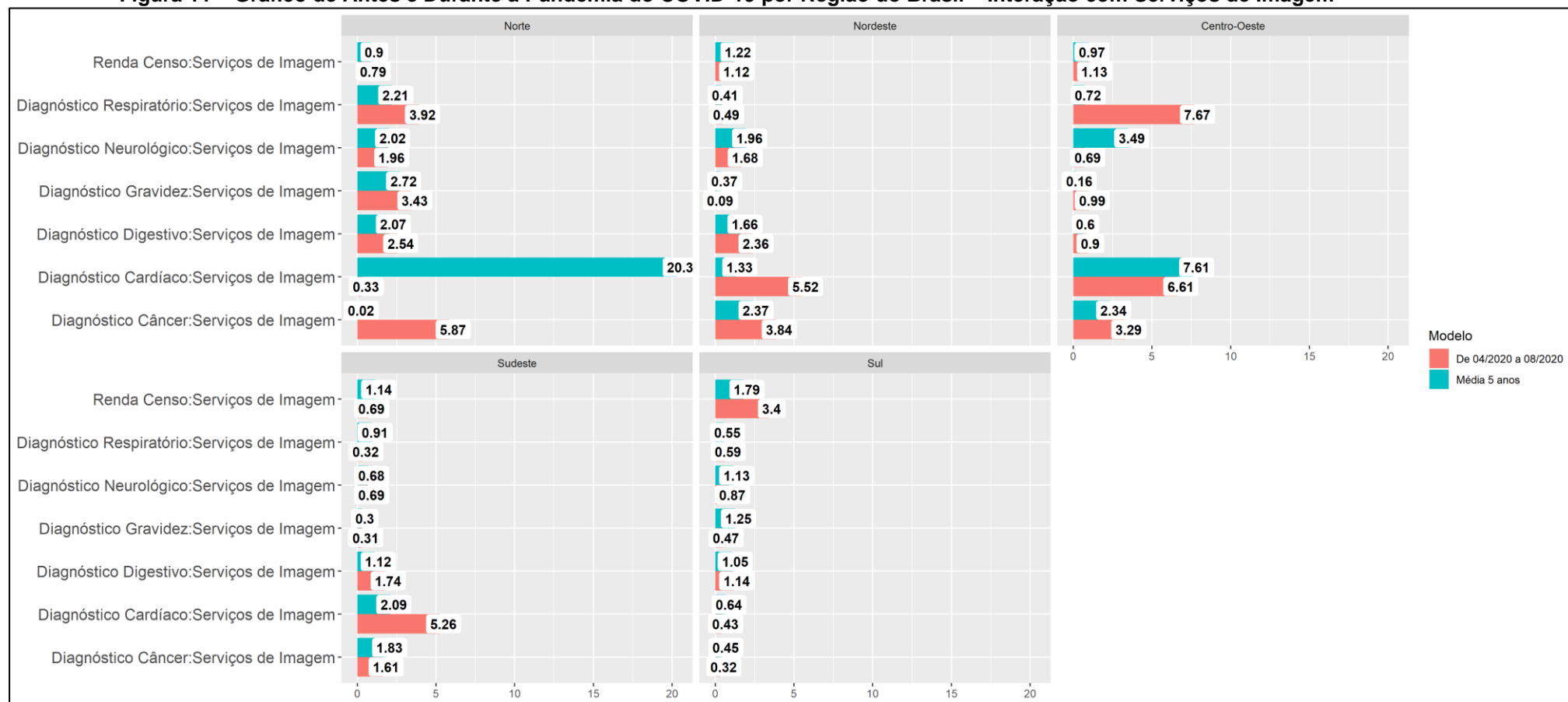
da pandemia e uma redução 39% durante a pandemia. Ainda na região Sul, outro destaque é a interação entre renda e hospitais privados (leitos SUS), com um *score* de *odds ratio* de 1,87 (antes do início da pandemia) e 1,24 (após o início da pandemia) – ou um aumento de 87% antes do início da pandemia nas chances de seleção que foi para 24% durante a pandemia. Por último, a região Nordeste não apresenta grandes variações de magnitude e parece não fugir muito da tendência nacional da seção anterior.

De resto, é possível verificar pela Figura 10 e pela Tabela 12 que há poucas oscilações em termos de magnitude e os valores tendem a ser próximos ao *score* de 1, o que reflete os valores encontrados na seção anterior e coeficientes estimados muito próximos de zero. Há, reitera-se, pouca evidência de que renda esteja associada a um aumento da escolha.

Seguindo novamente para os coeficientes de imagem, a Figura 11 abaixo demonstra também os resultados associados às interações de serviços de imagem. Nesse caso, porém, há interações com variações significantes regionalmente. Há mudança substanciais e particularidades que podem ser investigadas em detalhe. Um exemplo é o elevado *odds ratio* estimado antes do início da pandemia (*score* elevadíssimo de 20,3) para a interação entre serviços de imagem e o diagnóstico cardíaco na região Norte – uma queda pronunciada após o início da pandemia (0,33 de *odds ratio*, que implica em uma redução de 67% nas chances de seleção). Outro detalhe substancial é a interação entre renda e serviços de imagem: na seção anterior, a relação se observou como sendo bastante próxima (valores de *odds ratio* um pouco acima de 1 e próximos), mas na análise regional desta seção os resultados se demonstraram bastante diversos: na região Sul o valor é maior no geral e praticamente dobra durante a pandemia (sai do *score* de 1,79 e vai para 3,4 durante a pandemia, que refletem aumentos de chances de 79% antes da pandemia e 240% durante a pandemia), enquanto na região sudeste a direção é diversa e o valor cai praticamente pela metade (de 1,14 para 0,69 durante a pandemia, que refletem aumentos de chances de 14% antes da pandemia e reduções de 69% durante a pandemia). Outro ponto relevante é diferença regional da interação entre serviços de imagem e diagnósticos respiratórios: nas regiões Norte, Nordeste e Sul há aumento no *odds ratio* em relação ao período de antes da pandemia, ainda que se mantenha abaixo de 1 na região Sul. O coeficiente de interação entre serviços de imagem e diagnóstico

neurológico se manteve estável em três regiões (Norte, Sudeste e Sul), mas nas demais regiões seguiu tendência de possuir um *score* maior de *odds ratio* antes do início da pandemia. O diagnóstico de gravidez interagido com serviços de imagem atingiu um valor acima de 1 na região Norte, assim como também atingiu um valor de 1 na região Sul antes do início da pandemia. Trata-se, portanto, de alguns resultados, mas que demonstram que nas interações de serviços de imagem há grande diversidade regional.

Figura 11 – Gráfico de Antes e Durante a Pandemia do COVID-19 por Região do Brasil – Interação com Serviços de Imagem



Fonte: Elaboração própria.

Tabela 13 –Antes e Durante a Pandemia do COVID-19 por Região do Brasil – Coeficientes de interação com Serviços de Imagem – Coeficientes, Desvio Padrão e Odds Ratio

Região	Variáveis	2020			Média 5 anos - 2015-2019		
		Coef.	EP	OR	Coef.	EP	OR
CO	Diagnóstico Câncer:Serviços de Imagem	1,1922	0,0407	3,2943	0,8520	0,0209	2,3443
CO	Diagnóstico Cardíaco:Serviços de Imagem	1,8885	0,0428	6,6091	2,0301	0,0192	7,6146
CO	Diagnóstico Digestivo:Serviços de Imagem	-0,1028	0,0374	0,9023	-0,5134	0,0158	0,5984
CO	Diagnóstico Gravidez:Serviços de Imagem	-0,0137	0,0339	0,9863	-1,8458	0,0193	0,1579
CO	Diagnóstico Neurológico:Serviços de Imagem	-0,3693	0,0607	0,6912	1,2512	0,0233	3,4947
CO	Diagnóstico Respiratório:Serviços de Imagem	2,0371	0,0307	7,6686	-0,3308	0,0184	0,7183
NE	Diagnóstico Câncer:Serviços de Imagem	1,3447	0,0260	3,8372	0,8612	0,0105	2,3659
NE	Diagnóstico Cardíaco:Serviços de Imagem	1,7081	0,0352	5,5183	0,2887	0,0124	1,3347
NE	Diagnóstico Digestivo:Serviços de Imagem	0,8592	0,0264	2,3612	0,5081	0,0084	1,6621
NE	Diagnóstico Gravidez:Serviços de Imagem	-2,4030	0,0200	0,0904	-0,9937	0,0099	0,3702
NE	Diagnóstico Neurológico:Serviços de Imagem	0,5192	0,0431	1,6808	0,6720	0,0170	1,9581
NE	Diagnóstico Respiratório:Serviços de Imagem	-0,7213	0,0177	0,4861	-0,8885	0,0107	0,4113
N	Diagnóstico Câncer:Serviços de Imagem	1,7700	0,0653	5,8707	-3,7679	0,0339	0,0231
N	Diagnóstico Cardíaco:Serviços de Imagem	-1,1050	0,0683	0,3312	3,0107	0,0219	20,3023
N	Diagnóstico Digestivo:Serviços de Imagem	0,9324	0,0414	2,5407	0,7259	0,0138	2,0667
N	Diagnóstico Gravidez:Serviços de Imagem	1,2338	0,0358	3,4344	1,0015	0,0207	2,7224
N	Diagnóstico Neurológico:Serviços de Imagem	0,6743	0,1059	1,9627	0,7042	0,0337	2,0222
N	Diagnóstico Respiratório:Serviços de Imagem	1,3663	0,0872	3,9207	0,7941	0,0239	2,2124
SE	Diagnóstico Câncer:Serviços de Imagem	0,4737	0,0309	1,6058	0,6057	0,0123	1,8325
SE	Diagnóstico Cardíaco:Serviços de Imagem	1,6603	0,0280	5,2609	0,7394	0,0106	2,0947
SE	Diagnóstico Digestivo:Serviços de Imagem	0,5560	0,0221	1,7438	0,1098	0,0081	1,1161
SE	Diagnóstico Gravidez:Serviços de Imagem	-1,1701	0,0168	0,3103	-1,2074	0,0076	0,2990
SE	Diagnóstico Neurológico:Serviços de Imagem	-0,3749	0,0565	0,6874	-0,3803	0,0204	0,6837
SE	Diagnóstico Respiratório:Serviços de Imagem	-1,1550	0,0165	0,3151	-0,0945	0,0091	0,9099
S	Diagnóstico Câncer:Serviços de Imagem	-1,1372	0,0591	0,3207	-0,8095	0,0267	0,4451
S	Diagnóstico Cardíaco:Serviços de Imagem	-0,8525	0,0411	0,4263	-0,4423	0,0193	0,6426
S	Diagnóstico Digestivo:Serviços de Imagem	0,1291	0,0308	1,1378	0,0450	0,0136	1,0460
S	Diagnóstico Gravidez:Serviços de Imagem	-0,7608	0,0297	0,4673	0,2197	0,0158	1,2456
S	Diagnóstico Neurológico:Serviços de Imagem	-0,1402	0,0752	0,8692	0,1259	0,0317	1,1341
S	Diagnóstico Respiratório:Serviços de Imagem	-0,5328	0,0332	0,5870	-0,6059	0,0132	0,5456

Fonte: Elaboração própria. Nota: Coef.: Coeficiente, EP: Erro Padrão. OR: Odds Ratio.

Contudo, é possível verificar que alguns coeficientes mantiveram a sua tendência nacional aproximadamente. Na seção anterior, os valores de interações com serviços de imagem foram maiores durante a pandemia e de alta magnitude em três casos: diagnósticos digestivos, cardíacos e de câncer. Na presente seção é possível verificar que essa dinâmica se mantém em duas regiões (Sudeste e Nordeste) e, particularmente, diagnósticos digestivos e de câncer seguem a mesma tendência de aumento na maioria das regiões, com a exceção da região Sul.

Abaixo seguimos para os coeficientes que este trabalho identificou como os mais intuitivos ou de maior interesse para investigação — conforme já mencionados na seção anterior. Contudo, apresentemos apenas em Apêndice os resultados em gráfico associados às interações de serviços de câncer e serviços de obstetrícia e os seus respectivos diagnósticos, em virtude de potenciais problemas de convergência dos coeficientes regionais, o que afeta a segurança na interpretação dos resultados (Figura A.11 do Apêndice A.3).

Vejamos abaixo, portanto, os coeficientes associados à Figura 9 da seção anterior, mas agora na nova Figura 12 com a discriminação regional.

Figura 12- Gráfico de Antes e Durante a Pandemia do COVID-19 por Região do Brasil – Valores com o comportamento consistente ou de interesse



Fonte: Elaboração própria.

Tabela 14 - Antes e Durante a Pandemia do COVID-19 por Região do Brasil – Coeficientes com o comportamento consistente e outros coeficientes – Coeficientes, Desvio Padrão e Odds Ratio

Região	Variáveis	2020			Média 5 anos - 2015-2019		
		Coef.	EP	OR	Coef.	EP	OR
CO	Diagnóstico Cardíaco:Serviços Cardíacos	1,0954	0,0400	2,9904	1,1038	0,0202	3,0156
CO	Diagnóstico Respiratório:Número de Respiradores	0,0040	0,0002	1,0040	-0,0062	0,0002	0,9938
CO	Distância percorrida (10km)	-0,5639	0,0020	0,5690	-0,5557	0,0009	0,5737
CO	Distância percorrida (10km) ao quadrado	0,0024	0,0000	1,0024	0,0033	0,0000	1,0033
CO	Distância percorrida (10km) Emergência	-0,0292	0,0038	0,9712	-0,0316	0,0014	0,9689
NE	Diagnóstico Cardíaco:Serviços Cardíacos	1,4819	0,0308	4,4015	0,8148	0,0105	2,2588
NE	Diagnóstico Respiratório:Número de Respiradores	0,0052	0,0001	1,0052	0,0029	0,0001	1,0029
NE	Distância percorrida (10km)	-0,3120	0,0011	0,7320	-0,0643	0,0000	0,9378
NE	Distância percorrida (10km) ao quadrado	0,0006	0,0000	1,0006	0,0002	0,0000	1,0002
NE	Distância percorrida (10km) Emergência	0,0004	0,0002	1,0004	0,0011	0,0001	1,0011
N	Diagnóstico Cardíaco:Serviços Cardíacos	0,9515	0,0776	2,5896	0,2690	0,0207	1,3087
N	Diagnóstico Respiratório:Número de Respiradores	0,0105	0,0004	1,0106	-0,0119	0,0004	0,9882
N	Distância percorrida (10km)	-0,6056	0,0025	0,5457	-0,9015	0,0009	0,4060
N	Distância percorrida (10km) ao quadrado	0,0022	0,0000	1,0022	0,0029	0,0000	1,0029
N	Distância percorrida (10km) Emergência	0,0147	0,0049	1,0148	-0,0411	0,0019	0,9598
SE	Diagnóstico Cardíaco:Serviços Cardíacos	0,6430	0,0443	1,9022	0,4772	0,0174	1,6115
SE	Diagnóstico Respiratório:Número de Respiradores	0,0044	0,0001	1,0044	-0,0075	0,0001	0,9925
SE	Distância percorrida (10km)	-0,9976	0,0019	0,3688	-0,4635	0,0006	0,6291
SE	Distância percorrida (10km) ao quadrado	0,0017	0,0000	1,0017	0,0036	0,0000	1,0037
SE	Distância percorrida (10km) Emergência	0,0043	0,0008	1,0043	0,0066	0,0003	1,0066
S	Diagnóstico Cardíaco:Serviços Cardíacos	0,9685	0,0488	2,6340	0,3541	0,0220	1,4248
S	Diagnóstico Respiratório:Número de Respiradores	0,0034	0,0002	1,0034	-0,0071	0,0001	0,9930
S	Distância percorrida (10km)	-0,9427	0,0022	0,3896	-0,6554	0,0009	0,5192
S	Distância percorrida (10km) ao quadrado	0,0126	0,0001	1,0126	0,0102	0,0000	1,0102
S	Distância percorrida (10km) Emergência	-0,0786	0,0063	0,9244	-0,0692	0,0017	0,9332

Fonte: Elaboração própria. Nota: Coef.: Coeficiente, EP: Erro Padrão. OR: Odds Ratio.

Resumidamente, os valores de *odds ratio* de distância percorrida caíram (em magnitude) nas regiões Sudeste (0,63 para 0,37, reduções de chances mudando de 37% para 63% a cada 10 km adicionais), Sul (0,52 para 0,39, reduções de chances mudando de 48% para 61% a cada 10 km adicionais), Nordeste (0,94 para 0,74, reduções de chances mudando de 6% para 26% a cada 10 km adicionais) e se mantiveram iguais na região Centro-Oeste – estatisticamente, o coeficientes são iguais, inclusive⁵⁹. Na região Norte, houve a exceção: o valor de *odds ratio* produziu um aumento após o início da pandemia do COVID-19 (de 0,41 para 0,55, reduções

⁵⁹ O coeficiente estimado de -0,564 (após o início da pandemia) está a menos de dois desvios padrões do coeficiente de antes da pandemia, de -0,566 (antes do início da pandemia).

de chances mudando de 69% para 45% a cada 10 km adicionais). Novamente, os demais controles de distância (distância ao quadrado e distância interagida com um atendimento de emergência) não sofreram alterações substanciais e omitimos em gráfico, mas podem ser observados na Tabela 14. Em suma, trata-se de uma evidente heterogeneidade regional para as distâncias percorridas por pacientes em busca de uma internação, indicando possíveis lógicas distintas do paciente com o início da pandemia. Mas a fonte dessa heterogeneidade escapa a investigação deste trabalho.

A interação entre o número de respiradores e o diagnóstico de doença respiratória também seguiu em um ligeiríssimo aumento entre os dois períodos analisados, mas a relação continuou praticamente inexistente ou extremamente baixa para identificar uma alteração substancial.

Por fim, a mudança no *odds ratio* da interação entre o diagnóstico de doenças cardíacas e serviços hospitalares cardíacos apresentou uma pequena heterogeneidade: na região centro-oeste, o coeficiente adquiriu a direção positiva teoricamente esperada, mas obteve pouca ou nenhuma variação entre o antes e durante a pandemia do COVID-19 – especificamente, os valores de *odds ratio* foram de 3,02 (média de 5 anos, antes da pandemia, um aumento de 202% nas chances de seleção) e 2,99 (durante a pandemia, um aumento de 199% nas chances de seleção) na regiões Centro-oeste. Nas demais regiões (Norte, Nordeste, Sul e sudeste) o valor de *odds ratio* se demonstrou com a direção esperada (acima de 1) e dobrando de valor praticamente.

6 CONSIDERAÇÕES FINAIS E CONCLUSÃO

O presente trabalho buscou realizar uma estimação de demanda hospitalar com um volume considerável de informações: partindo de uma base de dados inicial com mais de 16 milhões de observações e de construção laboriosa. A análise também se torna consideravelmente desafiadora no âmbito de modelos de escolha discreta, em que há necessidade de associar essas observações com um conjunto de escolhas disponíveis para a estimação de demanda. Em maior detalhe, as estimações tomam um tempo sensível e algumas análises fogem do escopo deste trabalho em virtude do tempo que seria consumido.

Contudo, utilizando um modelo de escolha discreta de regressão logística condicional (ou o modelo *logit* condicional, como já chamado antes neste trabalho), efetuou-se uma análise para as 450 Regiões de Saúde do Brasil separando os efeitos antes e durante a pandemia do COVID-19. A regressão para cada região foi efetuada com cerca de 40 variáveis explicativas.⁶⁰ Contudo, apenas 384 regiões convergiram antes e durante a pandemia do COVID-19 em virtude de uma série de fatores abordados neste trabalho, como os problemas de separação de uma regressão logística e a tratabilidade desses resultados no âmbito de uma enorme massa de resultados. Não se observou, porém, diferenças substanciais entre as regiões do Brasil em percentual de regiões que convergiram, permitindo inferir que não há vieses substanciais nos resultados por uma composição regional.

Ademais, foram analisadas diversas variáveis, observáveis no Apêndice A deste trabalho, mas foram elencados alguns conjuntos de resultados para dar destaque e discutir as suas direções. Em detalhe, o trabalho deu destaque às interações entre serviços de imagem e renda, além de destacar um conjunto de coeficientes que foram identificados como bem-comportados, intuitivos ou de forte de associação com escolha hospitalar.

No primeiro caso, destacou-se as interações entre diagnósticos e serviços de imagem e as interações de renda e estrutura hospitalar. No segundo caso, em relação aos coeficientes que se apresentaram como intuitivos ou que este trabalho julgou interessante apresentar, como a interação entre diagnósticos e os seus serviços

⁶⁰ Desconsiderando os efeitos fixos associado às unidades hospitalares. Ver Tabela A.1 do Apêndice A para maiores detalhes sobre os coeficientes estimados e todas as variáveis utilizadas.

específicos (e.g. serviços cardíacos e diagnósticos cardíacos) ou os efeitos os efeitos em relação à distância percorrida pelos pacientes em busca de internação.

As interações de renda são interessantes porque demonstram pouca capacidade de governar a escolha dos pacientes, além da diferença no antes e durante a pandemia do COVID-19, na maioria das interações com a estrutura hospitalar, serem pequenas. Isto pode refletir também que a estrutura hospitalar já era adequada para a estratégia da crise da Covid-19, ao menos em parte (não significa que a resposta do sistema de saúde tenha sido adequada). Na análise regional, porém, há alguma heterogeneidade: na região Norte é possível destacar a interação entre renda e serviços de obstetrícia, que diverge do resultado nacional durante a pandemia (com um score maior que 1 antes e após o início da pandemia do COVID-19); na região centro-oeste é observável uma queda substancial do *odds ratio* na interação entre renda e serviços cardíacos, caindo pela metade durante a pandemia (1,06 para 0,47); na região Sudeste é possível destacar o aumento do *odds ratio* da interação entre renda e serviços de câncer (0,96 para 1,37) e a queda o *odds ratio* na interação entre renda e instituições de ensino (de 1,13 para 0,79).

A região Sul, porém, destaca-se na interação entre renda e obstetrícia, com uma queda no valor do *odds ratio* (de 1,80 para 0,61) e a queda interação entre renda e hospitais privados (de 1,87 para 1,24). No restante dos *odds ratios* estimados há relativa estabilidade em relação ao resultado nacional.

No caso das interações de serviços de imagem há uma variabilidade de magnitude mais pronunciada – ver Figura 8, Tabela 10. Destaca-se que as interações entre diagnósticos de doenças cardíacas, cancerígenas e digestivas apresentam aumento substancial após o início da Pandemia. As demais interações, com diagnósticos de gravidez, doenças respiratórias e neurológicas, seguem uma tendência diferente: diminuem após o início da pandemia em termos de *odds ratio*. Essas tendências, porém, seguem caminhos diversos em uma análise regional (Figura 11, Tabela 13). Particularmente, as interações de serviços de imagem com diagnósticos digestivos e de câncer seguem a mesma tendência de aumento na maioria das regiões, com a exceção da região Sul.

Houve, porém, um conjunto de casos que foram julgados como bem-comportados e mais intuitivos para interpretação, conforme este trabalho destacou, além de alguns coeficientes que se julgou interessante a apresentação. Dois casos

de valores expressivos foram os efeitos estimados das interações de serviços de câncer e obstetrícias com os seus respectivos diagnósticos: nesses casos, os coeficientes são consideravelmente altos. No caso do diagnóstico de câncer interagido com o seu respectivo serviço, há tendência ao aumento durante pandemia do COVID-19 – ver Figura 8, Tabela 11. Observa-se que o coeficiente possui magnitude e direção não muito diferente de que alguns valores estimados em Ho (2006)⁶¹. Mas na análise regional observou-se *odds ratios* sensivelmente altos, o que seja fruto de problemas de convergência discutidos neste trabalho. Este resultado deve limitar as conclusões sobre esses achados por hora – estes valores são apresentados em Apêndice apenas (Figura A.11 do Apêndice A).

Quanto aos coeficientes que este trabalho julgou interessante destacar, os coeficientes associados à distância percorrida, a interação entre respiradores e o diagnóstico de doenças respiratórias, além da interação entre o diagnóstico de doenças cardíacas e os serviços cardíacos, ocorreram algumas mudanças na análise nacional em relação ao impacto da pandemia do COVID-19. O *odds ratio* relacionado à distância percorrida cai (comparando o antes e durante a pandemia), aumenta levemente na interação entre respiradores e diagnóstico de doenças respiratórias (mas é uma mudança muito pequena) e a interação entre serviços cardíacos e doenças cardíacas aumenta sensivelmente e adquire direção intuitiva.

A variável de distância percorrida possui uma direção esperada e significativa para um conjunto expressivo de resultados (ver Figura 9, Tabela 14 e Tabela A.3, com o número de valores significantes). O que se verificou com a variável de distância percorrida é que pacientes se tornaram (aparentemente) menos propensos a viajar para hospitais mais distantes após o início da pandemia COVID-19, mas é possível que outras características não controladas estejam afetando esta mudança, como uma mera mudança do perfil do paciente e não uma mudança de preferências – fator que pode estar associado a todos os demais resultados deste trabalho, considerando que não se realiza uma análise causal em função do evento da pandemia do COVID-19.

Em maior detalhe, na investigação regional destes valores, foi possível observar que a tendência foi diversa nacionalmente. As regiões Sudeste, Nordeste e

⁶¹ Por exemplo, o coeficiente de interação entre serviços de obstetrícia (“labor services”) de Ho (2006) atinge o valor de 3,544, o que resulta em um *odds ratio* de aproximadamente 33,6.

Sul tiveram uma queda mais expressiva no *odds ratio*. Na região Centro-Oeste o valor se manteve praticamente igual e na região Norte o coeficiente aumentou. A apuração mais detalhada dessa causa, porém, foge do escopo deste trabalho.

No caso da associação do diagnóstico de doenças respiratórias, verificou-se valores próximos a 1 de *odds ratio*, isto é, que alteram pouco a chance da escolha hospitalar. A mudança pouco pronunciada nos valores de *odds ratio* não foi capaz de assegurar se pacientes com diagnósticos respiratórios se tornaram mais propensos a escolher hospitais com mais respiradores durante o período da pandemia. Esse valor também é confirmado, em sua substancialidade e tendência, ao se observar o comportamento regional (Figura 12).

No caso das interações de coeficientes de serviços cardíacos e doenças cardíacas, encontrou-se interações que resultaram em *odds ratios* maiores do que 1 em todas as regiões, revelando um coeficiente de resultado intuitivo. Apenas em uma região (região centro-oeste) o *odds ratio* se manteve relativamente estável antes da pandemia e durante a pandemia: 3,02 (média de 5 anos, antes da pandemia) e 2,99 (durante a pandemia). Nas demais regiões, o *odds ratio* aumentou ou dobrou durante a pandemia.

Assim, tais resultados se demonstraram consistentes e robustos em alguns casos, sendo capazes de descrever tendências durante a pandemia do COVID-19, além de idiosincrasias regionais. As questões em aberto podem contribuir para estudos mais aprofundados sobre estimativa de demanda hospitalar no Brasil – incluindo aqui a questão da variabilidade regional.

REFERÊNCIAS

ALLISON, Paul D. Convergence failures in logistic regression. *SAS Global Forum*, v. 360, p. 1-11, 2008. Disponível em: http://www.people.vcu.edu/~dbandyop/BIOS625/Convergence_Logistic.pdf. Acesso em: 10 ago. 2021.

BRASIL. Presidência da República. Casa Civil. Lei nº 8080, de 19 de setembro de 1990. Dispõe sobre as condições para a promoção, proteção e recuperação da saúde, a organização e o funcionamento dos serviços correspondentes e dá outras providências. Disponível em http://www.planalto.gov.br/ccivil_03/LEIS/L8080.htm. Acesso em: 5 jun. 2021.

_____. Ministério da Saúde. Departamento de informática do SUS – DATASUS. Sobre o DATASUS. 2021a. Disponível em <https://datasus.saude.gov.br/sobre-o-datasus/>. Acesso em: 4 jun. 2021.

_____. Ministério da Saúde. Departamento de Informática do SUS – DATASUS. Institucional. 2021b. Disponível em <http://www2.datasus.gov.br/SIHD/institucional>. Acesso em: 4 de jun. 2021.

_____. Ministério da Saúde. Divisão de Disseminação de Informações em Saúde. 2021c. Disponível em <ftp://ftp.datasus.gov.br/dissemin/publicos/>. Acesso em 4 jun. 2021.

_____. Ministério da Saúde. Departamento de Informática do SUS – DATASUS. Origem dos Dados. 2010a. Disponível em <http://tabnet.datasus.gov.br/cgi/sih/rxdescr.htm>. Acesso em 4jun. 2021.

_____. Ministério da Saúde. Departamento de Informática do SUS – DATASUS. Tabela CID-10. 2010b. Disponível em <http://tabnet.datasus.gov.br/cgi/sih/mxcid10lm.htm>. Acesso em: 4 jun. 2021.

_____. Ministério da Saúde. Secretária de Atenção à Saúde, and Departamento de Regulação, Avaliação e Controle. Manual técnico do Sistema de Informação Hospitalar. 2004. Disponível em https://bvsms.saude.gov.br/bvs/publicacoes/07_0066_M.pdf Acesso em: 22 nov. 2021

_____. Ministério da Economia. Programa de Disseminação de Estatísticas do Trabalho. 2021d. Disponível: <http://acesso.mte.gov.br/portal-pdet/home/>. Acesso em 3 jun. 2021.

_____. IBGE. CENSO 2010. Disponível em < <https://censo2010.ibge.gov.br/>>. 2011. Acesso em: 20 set. 2021.

_____. IBGE. CENSO 2010. Malha de Setores Censitários. 2021e. Disponível em <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/26565-malhas-de-setores-censitarios-divisoes-intramunicipais.html?=&t=downloads>>. Acesso em: 20 set. 2021.

BRESLOW, Norman. Covariance analysis of censored survival data. *Biometrics*, p.89-99, 1974.

CAPPS, Cory; DRANOVE, David; SATTERTHWAITE, Mark. Competition and market power in option demand markets. *RAND Journal of Economics*, p. 737-763, 2003.

CAPPS, Cory; DRANOVE, David; ZABINSKI, Z. The Long, Slow Decline of ElzingaHogarty and What Comes After. *CPI Antitrust Chronicle*. Disponível em <https://www.bateswhite.com/assets/htmldocuments/07.2017.CPI-Capps-DranoveZabinski.pdf>. Acesso em: 10 set. de 2021.

CEP ABERTO. Disponível em <https://cepaberto.com/>. Acesso em: 3 jun. 2021.

CRUPPÉ, W.; GERAEDTS, M. How do patients choose a hospital for elective surgery? *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*. P. 951-957, 2011.

COATE, Malcolm B.; FISCHER, Jeffrey H. A practical guide to the hypothetical monopolist test for market definition. *Journal of Competition Law and Economics*. Vol. 4, n. 4, p. 1031-1063, 2008.

DUBEUX, Luciana Santos; FREESE, Eduardo; FELISBERTO, Eronildo. Acesso a hospitais regionais de urgência e emergência: abordagem aos usuários para avaliação do itinerário e dos obstáculos aos serviços de saúde. *Physis: Revista de Saúde Coletiva*. [S. l.]: FapUNIFESP (SciELO), jun. 2013.

ELZINGA, Kenneth G.; HOGARTY, Thomas F. The problem of geographic market delineation in antimerger suits. *The Antitrust Bulletin*. vol.18. n.1 p. 45-81, 1973.

FIGUEIREDO, Juliana Oliveira; PRADO, Nilia Maria de Brito Lima; MEDINA, Maria Guadalupe; PAIM, Jairnilson Silva. Gastos público e privado com saúde no Brasil e países selecionados. *Saúde em Debate*, v. 42, p. 37-47, 2018.

FIRTH, David. Bias reduction of maximum likelihood estimates. *Biometrika*. vol. 80, n.1, p. 27-38, 1993.

HARTUNG, Joachim; KNAPP, Guido; SINHA, Bimal K. *Statistical Meta-Analysis with Applications*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2008. (Wiley Series in Probability and Statistics). Disponível em: <http://doi.wiley.com/10.1002/9780470386347>. Acesso em: 13 out. 2021.

HO, Katherine. The welfare effects of restricted hospital choice in the US medical care market. *Journal of Applied Econometrics* v. 21, n. 7, p. 1039-1079, 2006.

Ho, Kate. PAKES, Ariel. Hospital Choices, Hospital Prices, and Financial Incentives to Physicians. *American Economic Review*, v. 104, n. 12, p 3841-84, 2014.

HO, Katherine; LEE, Robin S. Insurer competition in health care markets. *Econometrica*, v. 85, n. 2, p. 379-417, 2017.

MCFADDEN, Daniel. *Conditional logit analysis of qualitative choice behavior*. 1973.

MADDALA, Gangadharrao S. *Limited-dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press, 1986.

MEGA, Helena. *Por que o SUS de SP quer saber a distância entre pacientes e hospitais?* Tilt Uol. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2019/11/26/sistema-de-dados-do-sus-mostra-distancia-dos-hospitais-para-pacientes.htm>. Acesso em: 20 set.2021.

MEDICINANET. Doenças por vírus, de localização não especificada. Disponível em: https://www.medicinanet.com.br/cid10/1199/b34_doencas_por_virus_de_localizacao_nao_especificada.htm. Acesso em: 2 ago. 2020.

REIS, Cristiano Sathler Dos; NORONHA, Kenya. WAJNMAN, Simone. Envelhecimento populacional e gastos com internação do SUS: uma análise realizada para o Brasil entre 2000 e 2010. *Revista Brasileira de Estudos de População*, n. 33, p. 591-612, 2016.

RODRÍGUEZ, Germán. Lecture Notes. New Jersey: Princeton University, 2020. Disponível em: <https://data.princeton.edu/wws509/notes/c6s3>. Acesso em: 2 ago. 2020.

SANTOS, Lenir. Região de saúde e suas redes de atenção: modelo organizativo-sistêmico do SUS. *Ciência & Saúde Coletiva*, v. 22, p 1281-1289, 2017.

SALDANHA, Raphael de Freitas; BASTOS, Ronaldo Rocha; BARCELLOS, Christovam. Microdatasus: pacote para download e pré-processamento de microdados do Departamento de Informática do SUS (DATASUS). *Cad. Saúde Pública*, Rio de Janeiro , v. 35, n. 9, e00032419, 2019. Available Disponível em <<http://ref.scielo.org/dhcg3y>> Acesso em: 3 ago. 2020.

SMALL, Kenneth A.; ROSEN, Harvey S. Applied welfare economics with discrete choice models. *Econometrica: Journal of the Econometric Society*, p. 105-130, 1981.

STRYJEWSKI, Lisa. *40 years of boxplots*. 2010. Disponível em: <http://had.co.nz/stat645/project-03/boxplots.pdf>. Acesso em: 11 set. 2011.

SKINNER, Jonathan. Causes and Consequences of Regional Variations in Health Care. In: PAULY, Mark V.; MCGUIRE, Thomas G.; BARROS, Pedro Pita (Orgs.). *Handbook of Health Economics*. [s.l.]: Elsevier, 2011, v. 2, p. 45–93. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/B9780444535924000025>. Acesso em: 1 out. 2021.

SILVA, Everton Nunes da; SOARES, Fernando Ramalho Gameleira; FRIO, Gustavo Saraiva; *et al.* *Fluxo de internação por COVID-19 nas regiões de saúde do Brasil*. [s.l.: s.n.], 2021. Disponível em: <https://preprints.scielo.org/index.php/scielo/preprint/view/1849/version/1962> . Acesso em: 13 out. 2021.

TOWN, Robert; VISTNES, Gregory. Hospital competition in HMO networks. *Journal of health economics*, v. 20, n. 5, p. 733-753, 2001.

THERNEAU, Terry M.; LUMLEY, Thomas. Package 'survival'. *R Top Doc*, v. 128, n. 10, p. 28-33, 2015. Disponível em: <http://brieger.esalq.usp.br/CRAN/web/packages/survival/survival.pdf>. Acesso em: 10 ago.2021.

TUKEY, John Wilder. *Exploratory data analysis*. Reading, Mass: Addison-Wesley Pub. Co, 1977.

XAVIER, D.R.; MATOS, V.P.; MAGALHÃES M.A.F.M.; BELLIDO, J.G.; VELASCO, W.D.; VIACAVA, F. Polos e fluxos de deslocamento de pacientes para internação hospitalar e procedimentos selecionados no Sistema Único de Saúde. In: Noronha JC, Lima LD, Chorny AH, Dal Poz MR, Gadelha P (orgs.). *Brasil Saúde Amanhã: dimensões para o planejamento da atenção à saúde*. Rio de Janeiro: Editora Fiocruz, 2017, p. 113-149. Disponível em: <http://books.scielo.org/id/rs684/pdf/noronha-9786557080900-05.pdf>. Acesso em: 1 out.2021.

APÊNDICE A – DETALHES DAS VARIÁVEIS UTILIZADAS: TODOS RESULTADOS E COEFICIENTES ESTIMADOS

A.1 Análise completa

A Tabela A.2 e as Figuras A.1-A.2 complementam os resultados do corpo do texto, destacando eventuais variáveis adicionais estimadas. Trata-se das variáveis destacadas na análise do corpo, além de outras estimativas que podem ter a sua investigação à parte em trabalho futuros.

Tabela A. 1- Resultados completos – análise nacional

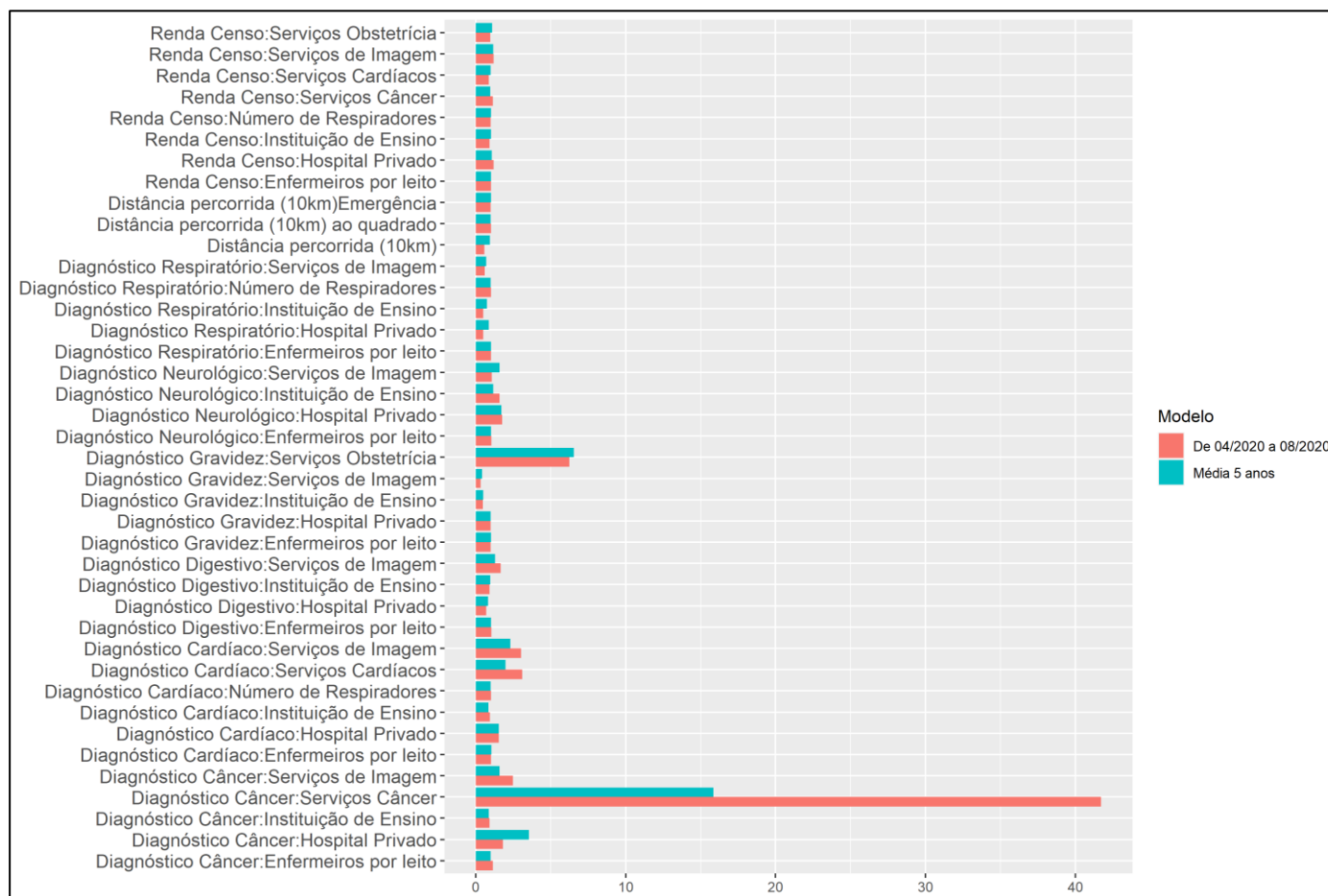
Variáveis	Período	Coefficiente	EP	OR
Diagnóstico Câncer:Enfermeiros por leito	De 04/2020 a 08/2020	0,1311	0,0015	1,1400
Diagnóstico Câncer:Enfermeiros por leito	Média 5 anos	-0,0162	0,0006	0,9839
Diagnóstico Câncer:Hospital Privado	De 04/2020 a 08/2020	0,5947	0,0122	1,8125
Diagnóstico Câncer:Hospital Privado	Média 5 anos	1,2627	0,0055	3,5350
Diagnóstico Câncer:Instituição de Ensino	De 04/2020 a 08/2020	-0,0766	0,0157	0,9263
Diagnóstico Câncer:Instituição de Ensino	Média 5 anos	-0,1588	0,0061	0,8531
Diagnóstico Câncer:Serviços Câncer	De 04/2020 a 08/2020	3,7311	0,0184	41,7250
Diagnóstico Câncer:Serviços Câncer	Média 5 anos	2,7627	0,0074	15,8423
Diagnóstico Câncer:Serviços de Imagem	De 04/2020 a 08/2020	0,9023	0,0165	2,4653
Diagnóstico Câncer:Serviços de Imagem	Média 5 anos	0,4613	0,0070	1,5862
Diagnóstico Cardíaco:Enfermeiros por leito	De 04/2020 a 08/2020	0,0231	0,0011	1,0234
Diagnóstico Cardíaco:Enfermeiros por leito	Média 5 anos	0,0328	0,0004	1,0334
Diagnóstico Cardíaco:Hospital Privado	De 04/2020 a 08/2020	0,4339	0,0135	1,5433
Diagnóstico Cardíaco:Hospital Privado	Média 5 anos	0,4281	0,0056	1,5344
Diagnóstico Cardíaco:Instituição de Ensino	De 04/2020 a 08/2020	-0,0751	0,0147	0,9276
Diagnóstico Cardíaco:Instituição de Ensino	Média 5 anos	-0,1622	0,0059	0,8503
Diagnóstico Cardíaco:Número de Respiradores	De 04/2020 a 08/2020	0,0070	0,0001	1,0070
Diagnóstico Cardíaco:Número de Respiradores	Média 5 anos	-0,0010	0,0001	0,9990
Diagnóstico Cardíaco:Serviços Cardíacos	De 04/2020 a 08/2020	1,1311	0,0190	3,0991
Diagnóstico Cardíaco:Serviços Cardíacos	Média 5 anos	0,6777	0,0072	1,9694
Diagnóstico Cardíaco:Serviços de Imagem	De 04/2020 a 08/2020	1,1016	0,0171	3,0089
Diagnóstico Cardíaco:Serviços de Imagem	Média 5 anos	0,8331	0,0066	2,3004
Diagnóstico Digestivo:Enfermeiros por leito	De 04/2020 a 08/2020	0,0303	0,0008	1,0307
Diagnóstico Digestivo:Enfermeiros por leito	Média 5 anos	0,0192	0,0003	1,0194
Diagnóstico Digestivo:Hospital Privado	De 04/2020 a 08/2020	-0,3648	0,0104	0,6944

Variáveis	Período	Coefficiente	EP	OR
Diagnóstico Digestivo:Hospital Privado	Média 5 anos	-0,2037	0,0041	0,8157
Diagnóstico Digestivo:Instituição de Ensino	De 04/2020 a 08/2020	-0,1005	0,0101	0,9044
Diagnóstico Digestivo:Instituição de Ensino	Média 5 anos	-0,0257	0,0037	0,9746
Diagnóstico Digestivo:Serviços de Imagem	De 04/2020 a 08/2020	0,5106	0,0131	1,6663
Diagnóstico Digestivo:Serviços de Imagem	Média 5 anos	0,2460	0,0048	1,2790
Diagnóstico Gravidez:Enfermeiros por leito	De 04/2020 a 08/2020	-0,0065	0,0006	0,9935
Diagnóstico Gravidez:Enfermeiros por leito	Média 5 anos	0,0032	0,0003	1,0032
Diagnóstico Gravidez:Hospital Privado	De 04/2020 a 08/2020	-0,0086	0,0077	0,9914
Diagnóstico Gravidez:Hospital Privado	Média 5 anos	-0,0204	0,0036	0,9798
Diagnóstico Gravidez:Instituição de Ensino	De 04/2020 a 08/2020	-0,7604	0,0075	0,4675
Diagnóstico Gravidez:Instituição de Ensino	Média 5 anos	-0,6894	0,0035	0,5019
Diagnóstico Gravidez:Serviços de Imagem	De 04/2020 a 08/2020	-1,1403	0,0106	0,3197
Diagnóstico Gravidez:Serviços de Imagem	Média 5 anos	-0,8966	0,0052	0,4080
Diagnóstico Gravidez:Serviços Obstetrícia	De 04/2020 a 08/2020	1,8293	0,0096	6,2294
Diagnóstico Gravidez:Serviços Obstetrícia	Média 5 anos	1,8768	0,0043	6,5323
Diagnóstico Neurológico:Enfermeiros por leito	De 04/2020 a 08/2020	0,0432	0,0014	1,0442
Diagnóstico Neurológico:Enfermeiros por leito	Média 5 anos	0,0064	0,0006	1,0064
Diagnóstico Neurológico:Hospital Privado	De 04/2020 a 08/2020	0,5600	0,0202	1,7507
Diagnóstico Neurológico:Hospital Privado	Média 5 anos	0,5381	0,0079	1,7128
Diagnóstico Neurológico:Instituição de Ensino	De 04/2020 a 08/2020	0,4652	0,0237	1,5923
Diagnóstico Neurológico:Instituição de Ensino	Média 5 anos	0,1511	0,0093	1,1631
Diagnóstico Neurológico:Serviços de Imagem	De 04/2020 a 08/2020	0,0695	0,0268	1,0720
Diagnóstico Neurológico:Serviços de Imagem	Média 5 anos	0,4649	0,0102	1,5919
Diagnóstico Respiratório:Enfermeiros por leito	De 04/2020 a 08/2020	0,0130	0,0007	1,0131
Diagnóstico Respiratório:Enfermeiros por leito	Média 5 anos	0,0163	0,0003	1,0165
Diagnóstico Respiratório:Hospital Privado	De 04/2020 a 08/2020	-0,7141	0,0087	0,4897
Diagnóstico Respiratório:Hospital Privado	Média 5 anos	-0,1453	0,0043	0,8648
Diagnóstico Respiratório:Instituição de Ensino	De 04/2020 a 08/2020	-0,7102	0,0080	0,4916
Diagnóstico Respiratório:Instituição de Ensino	Média 5 anos	-0,2884	0,0042	0,7495
Diagnóstico Respiratório:Número de Respiradores	De 04/2020 a 08/2020	0,0047	0,0001	1,0047
Diagnóstico Respiratório:Número de Respiradores	Média 5 anos	-0,0042	0,0001	0,9958
Diagnóstico Respiratório:Serviços de Imagem	De 04/2020 a 08/2020	-0,5230	0,0105	0,5927
Diagnóstico Respiratório:Serviços de Imagem	Média 5 anos	-0,3817	0,0057	0,6827
Distância percorrida (10km)	De 04/2020 a 08/2020	-0,5625	0,0008	0,5698
Distância percorrida (10km)	Média 5 anos	-0,0688	0,0000	0,9335
Distância percorrida (10km) ao quadrado	De 04/2020 a 08/2020	0,0012	0,0000	1,0012
Distância percorrida (10km) ao quadrado	Média 5 anos	0,0002	0,0000	1,0002
Distância percorrida (10km)Emergência	De 04/2020 a 08/2020	0,0005	0,0002	1,0005
Distância percorrida (10km)Emergência	Média 5 anos	0,0012	0,0001	1,0012
Renda Censo:Enfermeiros por leito	De 04/2020 a 08/2020	0,0112	0,0001	1,0112
Renda Censo:Enfermeiros por leito	Média 5 anos	0,0014	0,0001	1,0014
Renda Censo:Hospital Privado	De 04/2020 a 08/2020	0,1805	0,0020	1,1978
Renda Censo:Hospital Privado	Média 5 anos	0,0691	0,0010	1,0715
Renda Censo:Instituição de Ensino	De 04/2020 a 08/2020	-0,0798	0,0020	0,9233
Renda Censo:Instituição de Ensino	Média 5 anos	0,0162	0,0010	1,0164
Renda Censo:Número de Respiradores	De 04/2020 a 08/2020	-0,0007	0,0000	0,9993

Variáveis	Período	Coeficiente	EP	OR
Renda Censo:Número de Respiradores	Média 5 anos	0,0012	0,0000	1,0012
Renda Censo:Serviços Câncer	De 04/2020 a 08/2020	0,1277	0,0026	1,1362
Renda Censo:Serviços Câncer	Média 5 anos	-0,0292	0,0012	0,9712
Renda Censo:Serviços Cardíacos	De 04/2020 a 08/2020	-0,1414	0,0023	0,8682
Renda Censo:Serviços Cardíacos	Média 5 anos	-0,0154	0,0011	0,9847
Renda Censo:Serviços de Imagem	De 04/2020 a 08/2020	0,1746	0,0026	1,1908
Renda Censo:Serviços de Imagem	Média 5 anos	0,1546	0,0013	1,1672
Renda Censo:Serviços Obstetrícia	De 04/2020 a 08/2020	-0,0389	0,0020	0,9618
Renda Censo:Serviços Obstetrícia	Média 5 anos	0,0826	0,0010	1,0861

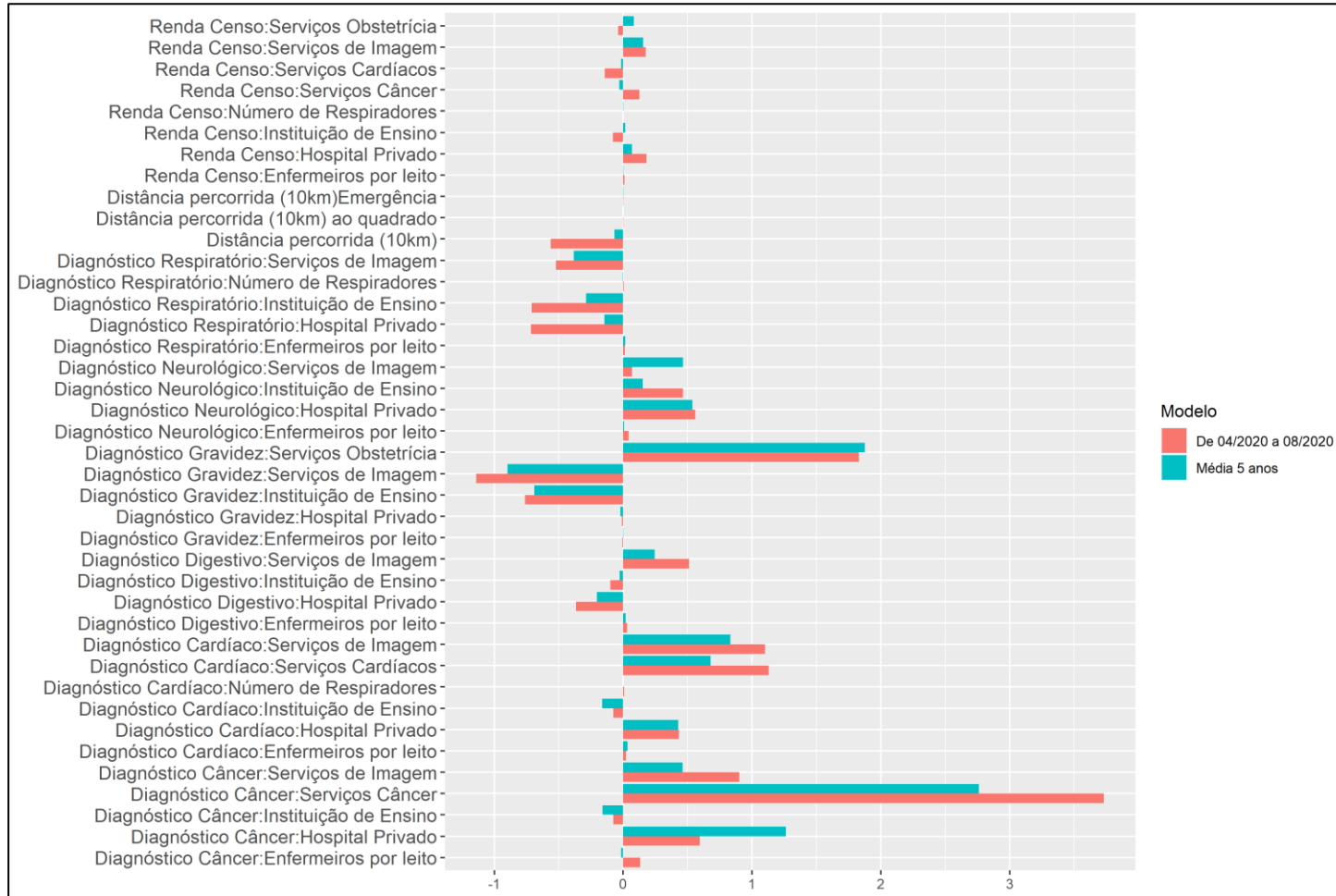
Fonte: Elaboração própria

Figura A.1 – Gráfico de Antes e Durante a Pandemia do COVID-19 – Todos os valores estimados (odds ratio)



Fonte: Elaboração própria.

Figura A.2 – Gráfico de Antes e Durante a Pandemia do COVID-19 – Todos os coeficientes estimados

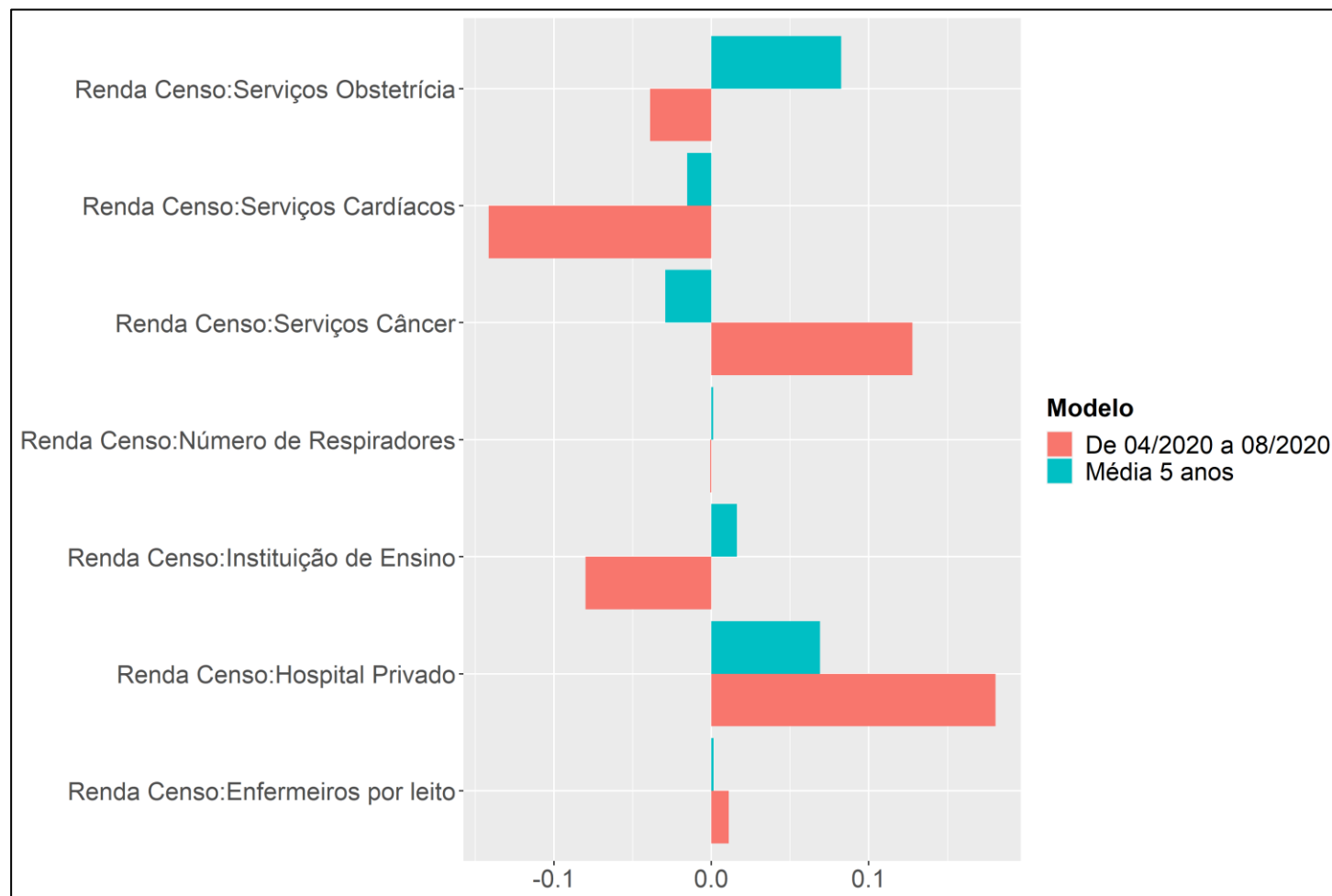


Fonte: Elaboração própria.

A.2 Análise agregada dos coeficientes

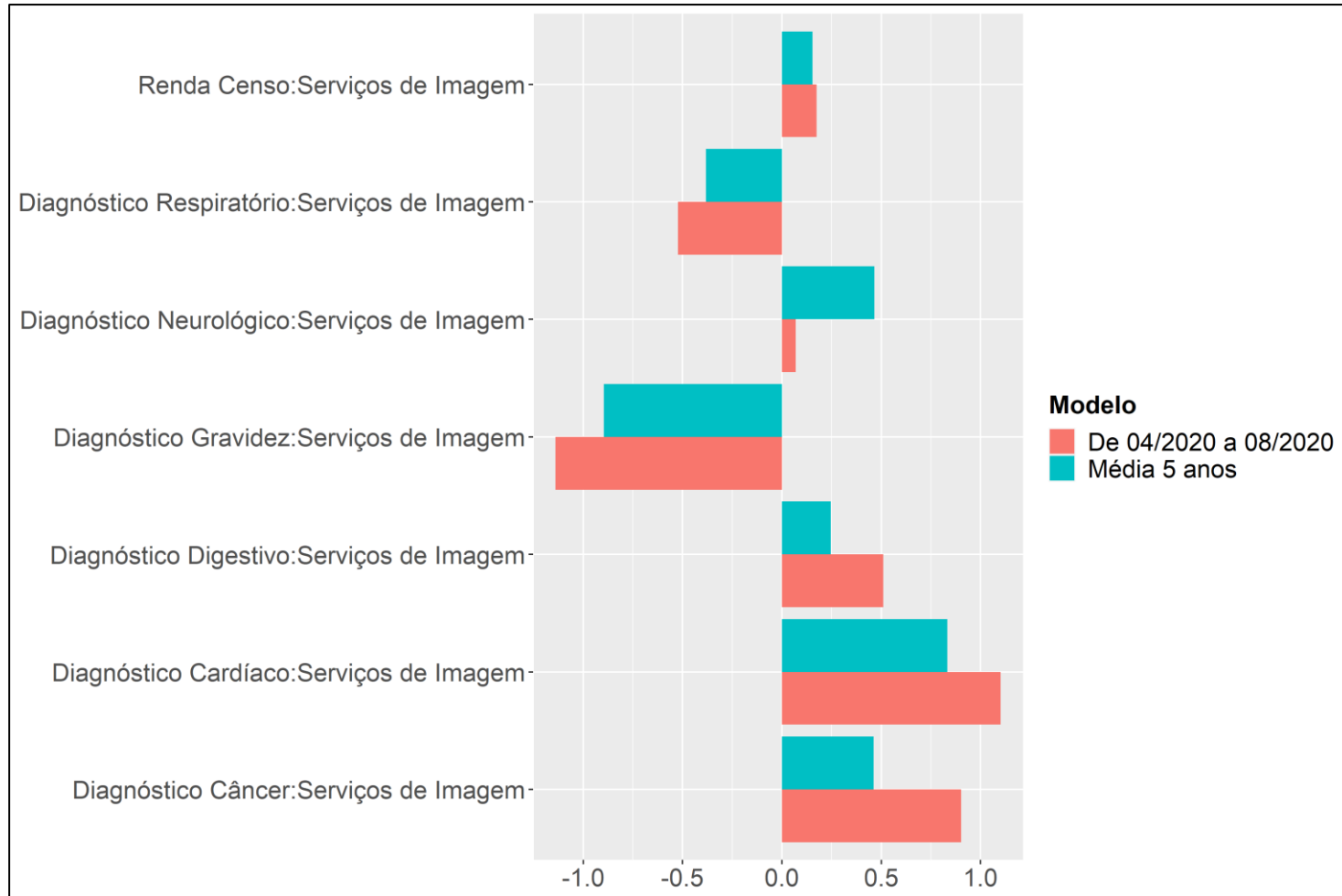
Este anexo contém os resultados dos coeficientes estimados para as regressões apresentadas no âmbito deste trabalho, mas em formato de gráfico para auxiliar visualizações que potencialmente não estavam claras pelas tabelas do corpo do texto.

As Figuras A.3-A.6 estão na ordem da apresentação dos resultados no corpo do texto.

Figura A.3 – Gráfico de Antes e Durante a Pandemia do COVID-19 – Coeficientes de interação com Renda

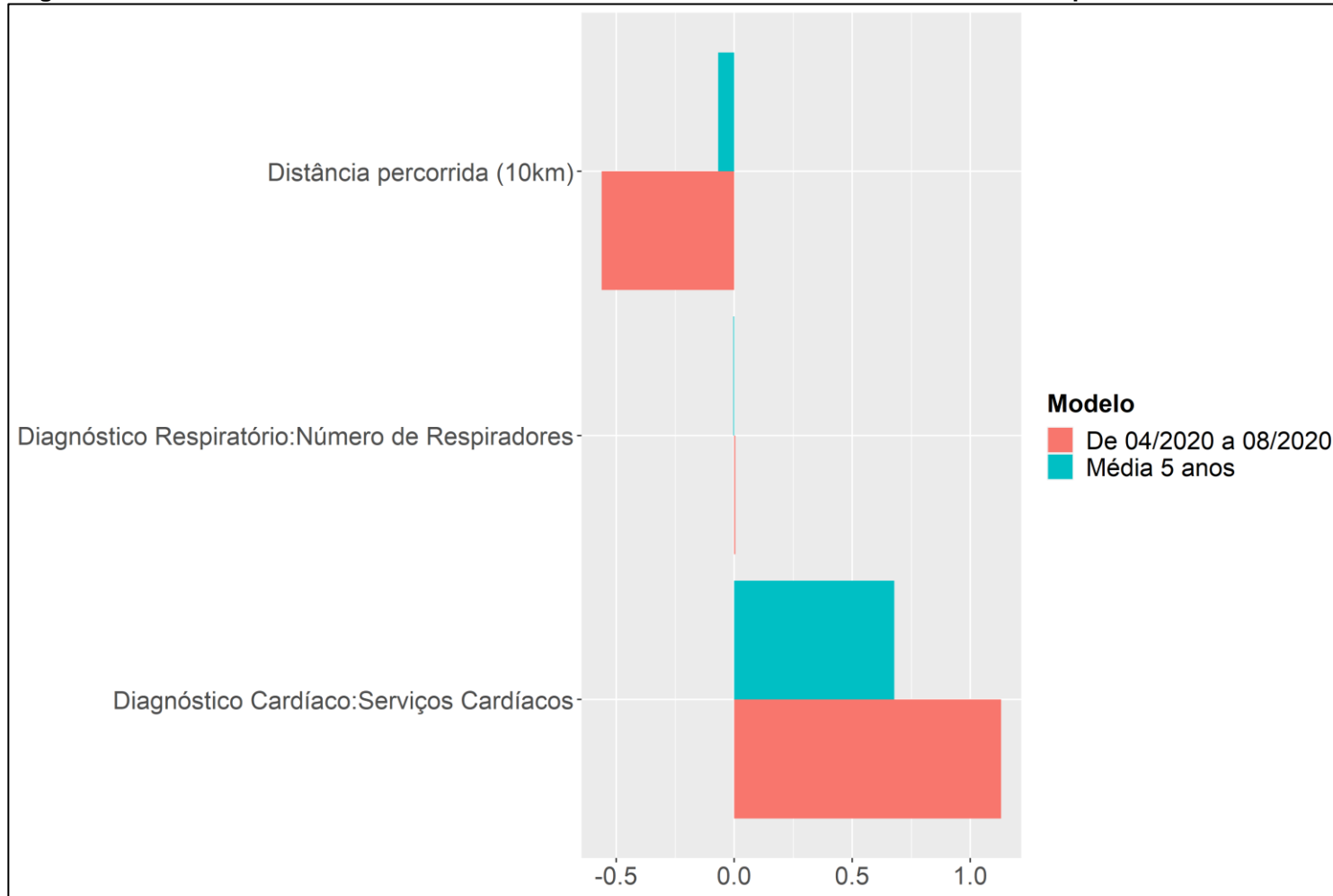
Fonte: Elaboração própria.

Figura A.4 – Gráfico de Antes e Durante a Pandemia do COVID-19 – Coeficientes de interação com Serviços de Imagem



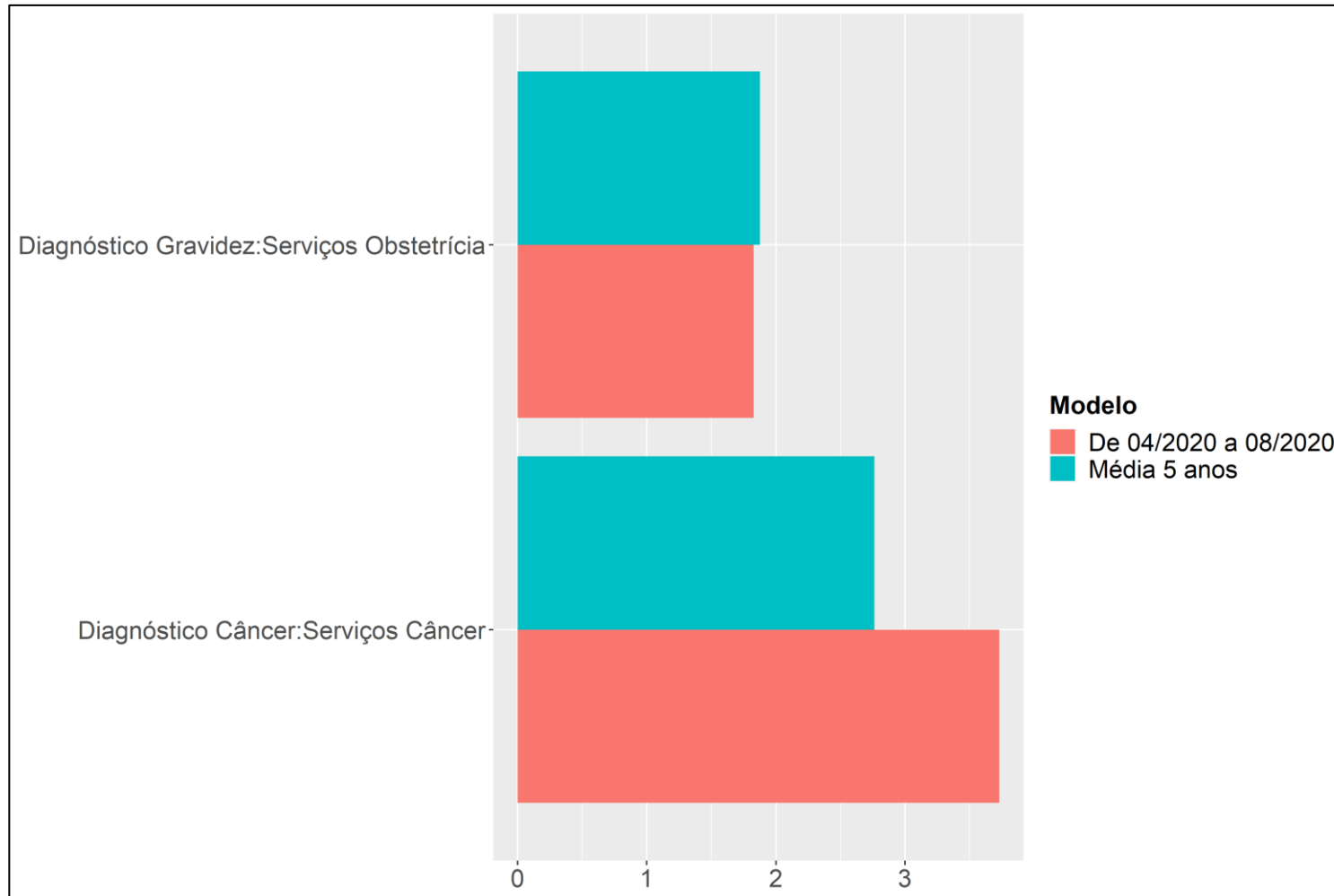
Fonte: Elaboração própria.

Figura A.5 – Gráfico de Antes e Durante a Pandemia do COVID-19 – Coeficientes com o comportamento consistente



Fonte: Elaboração própria.

Figura A.6 – Gráfico de Antes e Durante a Pandemia do COVID-19 – Coeficientes adicionais



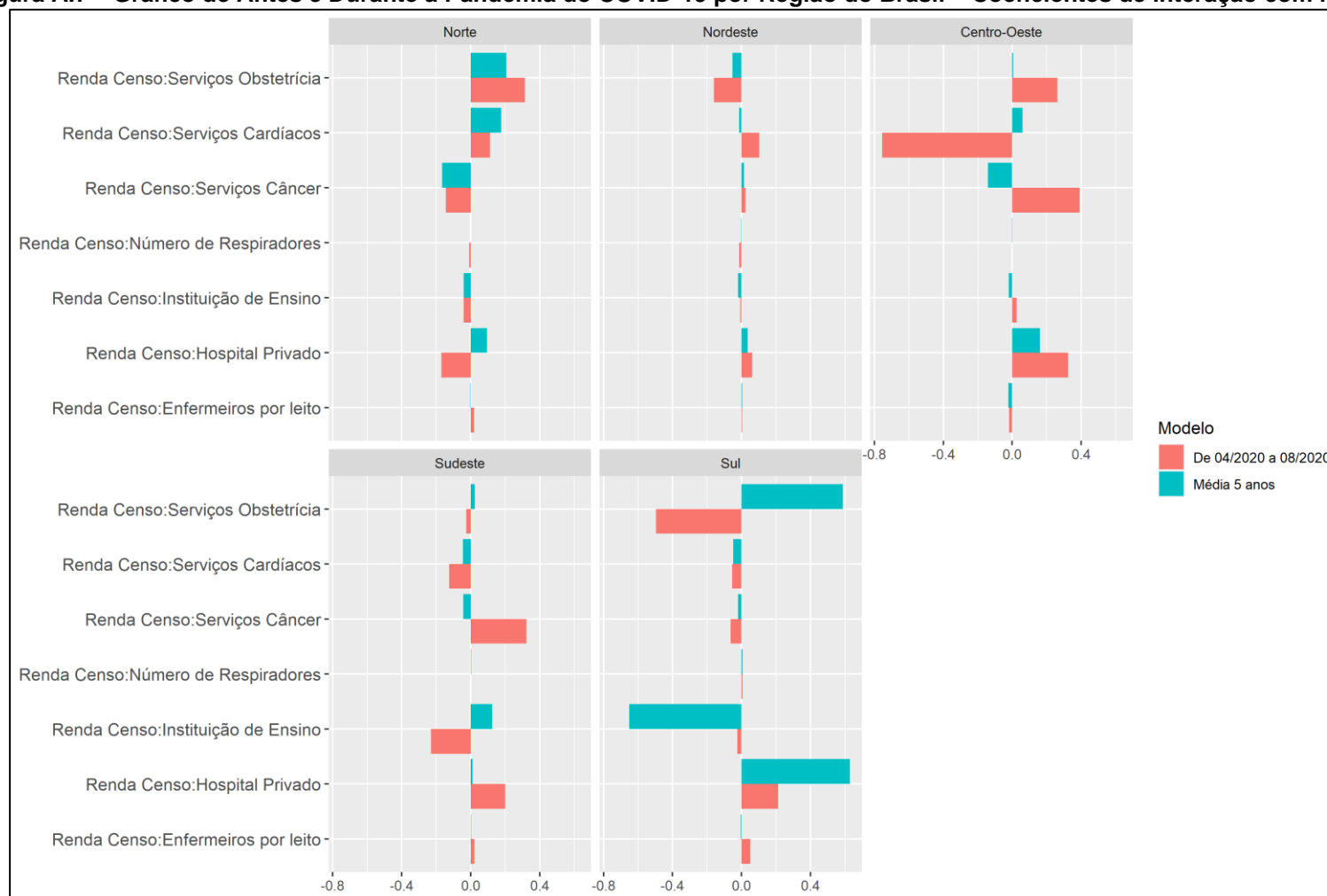
Fonte: Elaboração própria.

A.3 Análise regional

Este anexo contém os resultados dos coeficientes estimados para as regressões apresentadas no âmbito deste trabalho, mas em formato de gráfico para auxiliar visualizações que potencialmente não estavam claras pelas tabelas do corpo do texto. Neste caso, destaca-se as análises regionais.

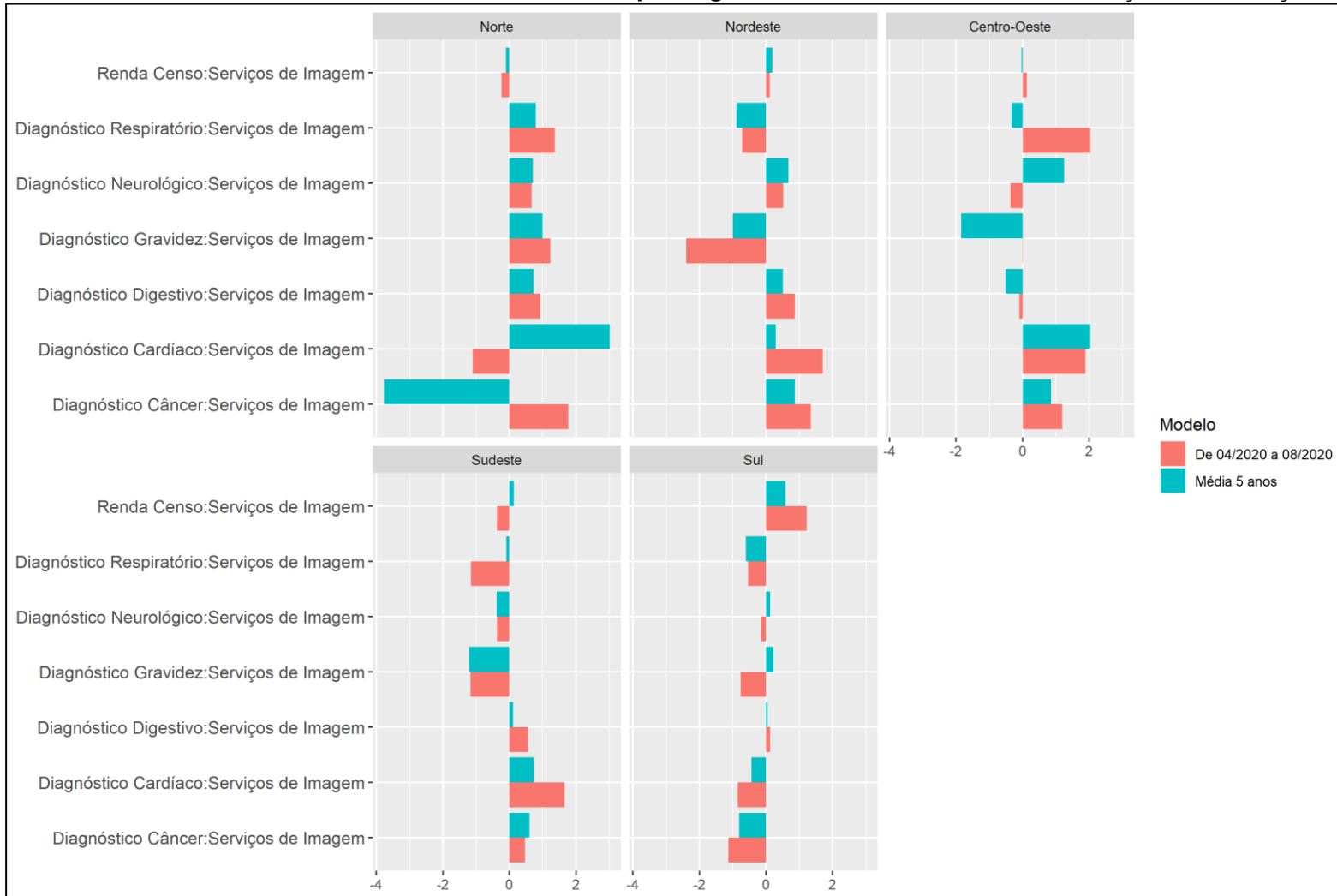
As Figuras A.7-A.9 estão na ordem da apresentação dos resultados no corpo do texto. As Figuras A.10-A.11 apresentam os resultados que não foram apresentados no corpo do texto, em virtude de incertezas a respeito da convergência de fato de alguns coeficientes que compõem os valores e como esses valores afetam as estimativas.

Figura A.7 – Gráfico de Antes e Durante a Pandemia do COVID-19 por Região do Brasil – Coeficientes de interação com Renda



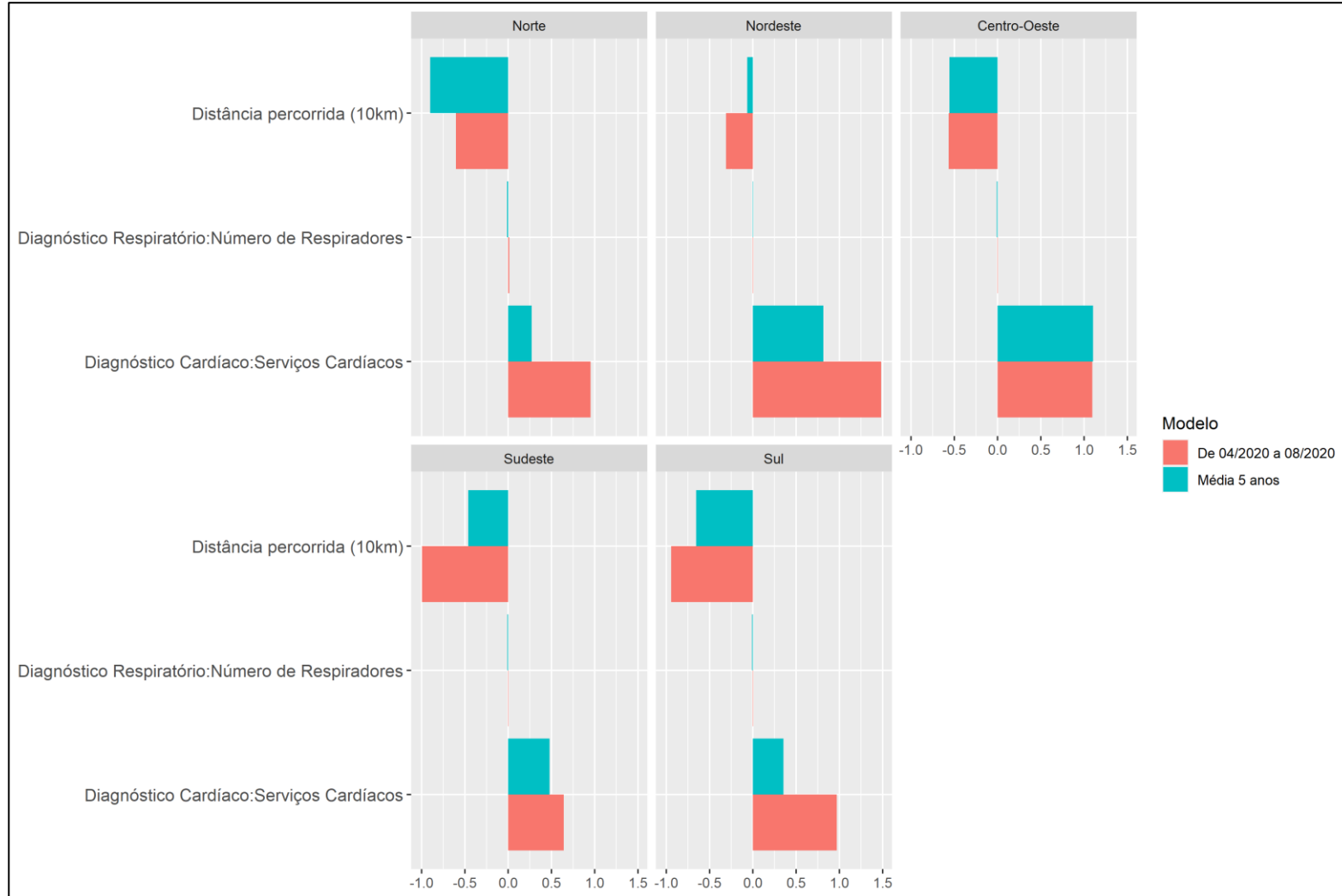
Fonte: Elaboração própria.

Figura A.8 – Gráfico de Antes e Durante a Pandemia do COVID-19 por Região do Brasil – Coeficientes de interação com Serviços de Imagem



Fonte: Elaboração própria.

Figura A.9 – Gráfico de Antes e Durante a Pandemia do COVID-19 por Região do Brasil – Coeficientes com o comportamento consistente



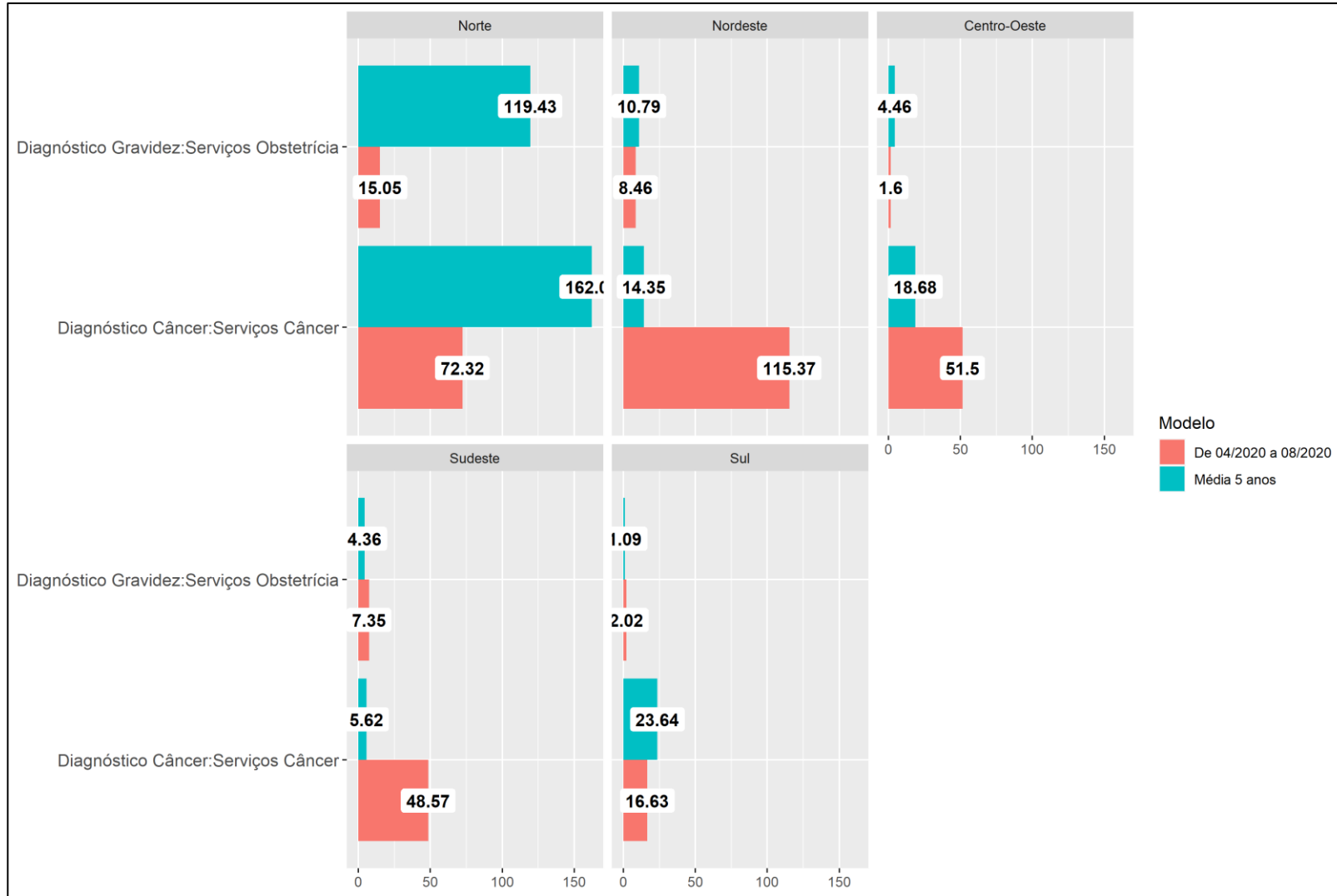
Fonte: Elaboração própria.

Figura A. 10 – Gráfico de Antes e Durante a Pandemia do COVID-19 por Região do Brasil – Coeficientes Adicionais



Fonte: Elaboração própria.

Figura A. 11– Gráfico de Antes e Depois da Pandemia do COVID-19 por Região do Brasil – Odds Ratio Adicionais



Fonte: Elaboração própria.

A.4 Significância estatística dos coeficientes estimados

A Tabela A.1 abaixo descreve a significância dos coeficientes estimados neste trabalho, separando o período que antecede a pandemia do COVID-19 (“Média 5 anos”) e o período de 2020 analisado neste trabalho (“De 04/2020 a 08/2020”). Como se pode observar as variáveis de distância são as mais recorrentemente significativas e as variáveis associadas a estruturas hospitalares ou diagnóstico de pacientes possuem uma proporção menor, o que potencialmente se explica em razão da heterogeneidade de oferta dessa estrutura na base de dados analisada.

Tabela A. 2 – Coeficientes Significantes estimados

Coeficientes	Modelo	Resultados Significantes	Total de Coeficientes	Percentual de Resultados Significantes*
Diagnóstico Câncer:Enfermeiros por leito	Média 5 anos - p-valor	164	384	42,7%
Diagnóstico Câncer:Enfermeiros por leito	De 04/2020 a 08/2020 - p-valor	107	384	27,9%
Diagnóstico Câncer:Hospital Privado	Média 5 anos - p-valor	122	384	31,8%
Diagnóstico Câncer:Hospital Privado	De 04/2020 a 08/2020 - p-valor	96	384	25,0%
Diagnóstico Câncer:Instituição de Ensino	Média 5 anos - p-valor	94	384	24,5%
Diagnóstico Câncer:Instituição de Ensino	De 04/2020 a 08/2020 - p-valor	86	384	22,4%
Diagnóstico Câncer:Serviços Câncer	Média 5 anos - p-valor	117	384	30,5%
Diagnóstico Câncer:Serviços Câncer	De 04/2020 a 08/2020 - p-valor	90	384	23,4%
Diagnóstico Câncer:Serviços de Imagem	Média 5 anos - p-valor	137	384	35,7%
Diagnóstico Câncer:Serviços de Imagem	De 04/2020 a 08/2020 - p-valor	96	384	25,0%
Diagnóstico Cardíaco:Enfermeiros por leito	Média 5 anos - p-valor	233	384	60,7%
Diagnóstico Cardíaco:Enfermeiros por leito	De 04/2020 a 08/2020 - p-valor	162	384	42,2%
Diagnóstico Cardíaco:Hospital Privado	Média 5 anos - p-valor	163	384	42,4%
Diagnóstico Cardíaco:Hospital Privado	De 04/2020 a 08/2020 - p-valor	107	384	27,9%
Diagnóstico Cardíaco:Instituição de Ensino	Média 5 anos - p-valor	98	384	25,5%
Diagnóstico Cardíaco:Instituição de Ensino	De 04/2020 a 08/2020 - p-valor	90	384	23,4%
Diagnóstico Cardíaco:Número de Respiradores	Média 5 anos - p-valor	212	384	55,2%
Diagnóstico Cardíaco:Número de Respiradores	De 04/2020 a 08/2020 - p-valor	164	384	42,7%
Diagnóstico Cardíaco:Serviços Cardíacos	Média 5 anos - p-valor	182	384	47,4%
Diagnóstico Cardíaco:Serviços Cardíacos	De 04/2020 a 08/2020 - p-valor	121	384	31,5%
Diagnóstico Cardíaco:Serviços de Imagem	Média 5 anos - p-valor	192	384	50,0%
Diagnóstico Cardíaco:Serviços de Imagem	De 04/2020 a 08/2020 - p-valor	136	384	35,4%
Diagnóstico Digestivo:Enfermeiros por leito	Média 5 anos - p-valor	268	384	69,8%
Diagnóstico Digestivo:Enfermeiros por leito	De 04/2020 a 08/2020 - p-valor	190	384	49,5%
Diagnóstico Digestivo:Hospital Privado	Média 5 anos - p-valor	194	384	50,5%
Diagnóstico Digestivo:Hospital Privado	De 04/2020 a 08/2020 - p-valor	124	384	32,3%
Diagnóstico Digestivo:Instituição de Ensino	Média 5 anos - p-valor	109	384	28,4%
Diagnóstico Digestivo:Instituição de Ensino	De 04/2020 a 08/2020 - p-valor	99	384	25,8%

Coeficientes	Modelo	Resultados Significantes	Total de Coeficientes	Percentual de Resultados Significantes*
Diagnóstico Digestivo:Serviços de Imagem	Média 5 anos - p-valor	203	384	52,9%
Diagnóstico Digestivo:Serviços de Imagem	De 04/2020 a 08/2020 - p-valor	147	384	38,3%
Diagnóstico Gravidez:Enfermeiros por leito	Média 5 anos - p-valor	282	384	73,4%
Diagnóstico Gravidez:Enfermeiros por leito	De 04/2020 a 08/2020 - p-valor	235	384	61,2%
Diagnóstico Gravidez:Hospital Privado	Média 5 anos - p-valor	171	384	44,5%
Diagnóstico Gravidez:Hospital Privado	De 04/2020 a 08/2020 - p-valor	147	384	38,3%
Diagnóstico Gravidez:Instituição de Ensino	Média 5 anos - p-valor	104	384	27,1%
Diagnóstico Gravidez:Instituição de Ensino	De 04/2020 a 08/2020 - p-valor	87	384	22,7%
Diagnóstico Gravidez:Serviços de Imagem	Média 5 anos - p-valor	204	384	53,1%
Diagnóstico Gravidez:Serviços de Imagem	De 04/2020 a 08/2020 - p-valor	165	384	43,0%
Diagnóstico Gravidez:Serviços Obstetrícia	Média 5 anos - p-valor	229	384	59,6%
Diagnóstico Gravidez:Serviços Obstetrícia	De 04/2020 a 08/2020 - p-valor	181	384	47,1%
Diagnóstico Neurológico:Enfermeiros por leito	Média 5 anos - p-valor	204	384	53,1%
Diagnóstico Neurológico:Enfermeiros por leito	De 04/2020 a 08/2020 - p-valor	114	384	29,7%
Diagnóstico Neurológico:Hospital Privado	Média 5 anos - p-valor	142	384	37,0%
Diagnóstico Neurológico:Hospital Privado	De 04/2020 a 08/2020 - p-valor	73	384	19,0%
Diagnóstico Neurológico:Instituição de Ensino	Média 5 anos - p-valor	92	384	24,0%
Diagnóstico Neurológico:Instituição de Ensino	De 04/2020 a 08/2020 - p-valor	63	384	16,4%
Diagnóstico Neurológico:Serviços de Imagem	Média 5 anos - p-valor	159	384	41,4%
Diagnóstico Neurológico:Serviços de Imagem	De 04/2020 a 08/2020 - p-valor	105	384	27,3%
Diagnóstico Respiratório:Enfermeiros por leito	Média 5 anos - p-valor	278	384	72,4%
Diagnóstico Respiratório:Enfermeiros por leito	De 04/2020 a 08/2020 - p-valor	207	384	53,9%
Diagnóstico Respiratório:Hospital Privado	Média 5 anos - p-valor	182	384	47,4%
Diagnóstico Respiratório:Hospital Privado	De 04/2020 a 08/2020 - p-valor	158	384	41,1%
Diagnóstico Respiratório:Instituição de Ensino	De 04/2020 a 08/2020 - p-valor	110	384	28,6%
Diagnóstico Respiratório:Instituição de Ensino	Média 5 anos - p-valor	106	384	27,6%
Diagnóstico Respiratório:Número de Respiradores	Média 5 anos - p-valor	260	384	67,7%
Diagnóstico Respiratório:Número de Respiradores	De 04/2020 a 08/2020 - p-valor	211	384	54,9%

Coeficientes	Modelo	Resultados Significantes	Total de Coeficientes	Percentual de Resultados Significantes*
Diagnóstico Respiratório:Serviços de Imagem	Média 5 anos - p-valor	191	384	49,7%
Diagnóstico Respiratório:Serviços de Imagem	De 04/2020 a 08/2020 - p-valor	178	384	46,4%
Distância percorrida (10km)	Média 5 anos - p-valor	380	384	99,0%
Distância percorrida (10km)	De 04/2020 a 08/2020 - p-valor	378	384	98,4%
Distância percorrida (10km) ao quadrado	Média 5 anos - p-valor	361	384	94,0%
Distância percorrida (10km) ao quadrado	De 04/2020 a 08/2020 - p-valor	339	384	88,3%
Distância percorrida (10km) Emergência	Média 5 anos - p-valor	271	384	70,6%
Distância percorrida (10km) Emergência	De 04/2020 a 08/2020 - p-valor	116	384	30,2%
Renda Censo:Enfermeiros por leito	Média 5 anos - p-valor	275	384	71,6%
Renda Censo:Enfermeiros por leito	De 04/2020 a 08/2020 - p-valor	212	384	55,2%
Renda Censo:Hospital Privado	Média 5 anos - p-valor	235	384	61,2%
Renda Censo:Hospital Privado	De 04/2020 a 08/2020 - p-valor	152	384	39,6%
Renda Censo:Instituição de Ensino	Média 5 anos - p-valor	112	384	29,2%
Renda Censo:Instituição de Ensino	De 04/2020 a 08/2020 - p-valor	99	384	25,8%
Renda Censo:Número de Respiradores	Média 5 anos - p-valor	227	384	59,1%
Renda Censo:Número de Respiradores	De 04/2020 a 08/2020 - p-valor	182	384	47,4%
Renda Censo:Serviços Câncer	Média 5 anos - p-valor	125	384	32,6%
Renda Censo:Serviços Câncer	De 04/2020 a 08/2020 - p-valor	96	384	25,0%
Renda Censo:Serviços Cardíacos	Média 5 anos - p-valor	239	384	62,2%
Renda Censo:Serviços Cardíacos	De 04/2020 a 08/2020 - p-valor	183	384	47,7%
Renda Censo:Serviços de Imagem	Média 5 anos - p-valor	231	384	60,2%
Renda Censo:Serviços de Imagem	De 04/2020 a 08/2020 - p-valor	185	384	48,2%
Renda Censo:Serviços Obstetrícia	Média 5 anos - p-valor	217	384	56,5%
Renda Censo:Serviços Obstetrícia	De 04/2020 a 08/2020 - p-valor	161	384	41,9%

Fonte: Elaboração própria.

APÊNDICE B – ROTINAS

As rotinas para a construção dos dados foram efetuadas sobretudo com o pacote *microdatasus* construído por Saldanha, Bastos e Barcelos (2019) e uma sequência de pelo menos dez rotinas para finalização da base de dados⁶². Trata-se, assim, de um conjunto de informações que tomariam um espaço considerável deste documento e que exigem uma organização de implementação. Logo, por economia de espaço, destaca-se apenas alguns exemplos de rotinas poderão ser solicitadas mediante contato com o autor. Embora as rotinas permitam que se entenda o exercício, não é possível reproduzi-las sem um conjunto de passos para a formação da base de dados.

As demais rotinas de implementação dos resultados seguem contempladas neste Anexo.

B.1 Rotina da regressão em *loop*

```
library(dplyr)
library(tidyr)
library(mlogit)
library(survival)

rm(list = ls()); gc()

# regressão das 450 regiões de Saúde #

#caminho do arquivo#

file=""

#inserção do arquivo
df_reg = readRDS(file)

#seleção dos meses da análise #

df_reg = df_reg %>%
  filter(MES>=4 & MES<=8)

#vetor de regiões para o loop#

REGIAO = unique(df_reg$REGIAO)

#criação data.frame vazio

df_model1=data.frame()
df_model1$REGIAO=numeric()
```

⁶² Essa é a expectativa atual com a redução e ajustes na programação, retirando trechos que podem ser otimizados.


```

df_model1$MODELO_TUDO=list()
df_model1$MODELO_PRE=list()
df_model1$MODELO_TUDO_ERRO=character()
df_model1$MODELO_PRE_ERRO=character()
df_model1$MODELO_TUDO_warning=character()
df_model1$MODELO_PRE_warning=character()
df_model1$ESCOLHAS_TUDO = list()
df_model1$ESCOLHAS_PRE = list()

#caminho dos arquivos de acompanhamento

dest_acompanhamento = ""

#loop da regressão das 450 regiões

for (i in v) {

  if(i==50){

    saveRDS(df_model1, dest_acompanhamento)

  }

  if(i==100){

    saveRDS(df_model1, dest_acompanhamento)

  }

  if(i==200){

    saveRDS(df_model1, dest_acompanhamento)

  }

  if(i==300){

    saveRDS(df_model1, dest_acompanhamento)

  }

  if(i==400){

    saveRDS(df_model1, dest_acompanhamento)

  }

  print(i)

#separação da base#

df=df_reg[df_reg$REGIAO==REGIAO[i],]
df_pre = df[df$Tempo=="Antes",]
df = df[df$Tempo=="Depois",]

##### mode depois - construção das variáveis #####

df$LONG_PAC = as.numeric(df$LONG_PAC)
df$LAT_PAC = as.numeric(df$LAT_PAC)
df$LONG_HOSP = as.numeric(df$LONG_HOSP)
df$LAT_HOSP = as.numeric(df$LAT_HOSP)

```

```

# separação da estrutura da base

df_hosp = df %>% select(CNES,
                      CARD, IMG,
                      OBS, CANC, ENF_LT,
                      FOR_PROFIT, ENSINO, RESP,
                      LAT_HOSP, LONG_HOSP, MES)

#hospitais com características completas

df_hosp = df_hosp[complete.cases(df_hosp),]

df_ind = df %>% select(CNES,
                      LAT_PAC, LONG_PAC, MES,
                      RENDA, DIAG, ID,
                      EMERG, CEP) %>%
distinct(.)

#pacientes com características completas

df_ind = df_ind[complete.cases(df_ind),]

#expansão da base

df=mlogit::mlogit.data(df_ind, choice = "CNES",
                      shape = "wide",
                      alt.levels = levels(df$CNES))

df = as.data.frame(df)

df$MES = as.numeric(as.character(df$MES))
df$alt = as.character(df$alt)

#estruturação das características com indivíduos

df = df %>% inner_join(df_hosp, by=c("alt"="CNES", "MES"))

#criação das variáveis adicionais.

df$dist = gmt::geodist(df$LONG_PAC,df$LAT_PAC,
                      df$LONG_HOSP,df$LAT_HOSP, units = "km")

df$RENDA = df$RENDA/1000
df$dist = df$dist/10
df$dist_sq=df$dist^2

df = fastDummies::dummy_cols(df, select_columns = "DIAG")

##### modelo pre - construção das variáveis - análogo #####

df_pre$LONG_PAC = as.numeric(df_pre$LONG_PAC)
df_pre$LAT_PAC = as.numeric(df_pre$LAT_PAC)
df_pre$LONG_HOSP = as.numeric(df_pre$LONG_HOSP)
df_pre$LAT_HOSP = as.numeric(df_pre$LAT_HOSP)

df_hosp = df_pre %>% select(CNES,
                          CARD, IMG,
                          OBS, CANC, ENF_LT,
                          FOR_PROFIT, ENSINO, RESP,

```

```

LAT_HOSP, LONG_HOSP, MES)

df_hosp = df_hosp[complete.cases(df_hosp),]

df_ind = df_pre %>% select(CNES,
                          LAT_PAC, LONG_PAC, MES,
                          RENDA, DIAG, ID,
                          EMERG, CEP) %>%
distinct(.)

df_ind = df_ind[complete.cases(df_ind),]

df_pre=mlogit::mlogit.data(df_ind, choice = "CNES",
                          shape = "wide",
                          alt.levels = levels(df_pre$CNES))

df_pre = as.data.frame(df_pre)

df_pre$MES = as.numeric(as.character(df_pre$MES))
df_pre$alt = as.character(df_pre$alt)

df_pre = df_pre %>% inner_join(df_hosp, by=c("alt"="CNES", "MES"))

df_pre$dist = gmt::geodist(df_pre$LONG_PAC,df_pre$LAT_PAC,
                          df_pre$LONG_HOSP,df_pre$LAT_HOSP, units = "km")

df_pre$RENDA = df_pre$RENDA/1000
df_pre$dist = df_pre$dist/10
df_pre$dist_sq=df_pre$dist^2

df_pre = fastDummies::dummy_cols(df_pre, select_columns = "DIAG")

#modelo logit apos

model1 = tryCatch({clogit(CNES ~ dist + dist_sq + dist:EMERG +
                          ENSINO:RENDA +
                          ENSINO:DIAG_GRAVIDEZ + ENSINO:DIAG_DIGESTIVO +
                          ENSINO:DIAG_CANCER +
                          ENSINO:DIAG_NEUROLOGICO +
                          ENSINO:DIAG_CARD +
                          ENSINO:DIAG_RESPIRATORIO +
                          ENF_LT:RENDA +
                          ENF_LT:DIAG_GRAVIDEZ + ENF_LT:DIAG_DIGESTIVO +
                          ENF_LT:DIAG_CANCER +
                          ENF_LT:DIAG_NEUROLOGICO +
                          ENF_LT:DIAG_CARD +
                          ENF_LT:DIAG_RESPIRATORIO +
                          IMG:RENDA +
                          IMG:DIAG_GRAVIDEZ + IMG:DIAG_DIGESTIVO +
                          IMG:DIAG_CANCER +
                          IMG:DIAG_NEUROLOGICO +
                          IMG:DIAG_CARD +
                          IMG:DIAG_RESPIRATORIO +
                          FOR_PROFIT:RENDA +
                          FOR_PROFIT:DIAG_GRAVIDEZ + FOR_PROFIT:DIAG_DIGESTIVO +
                          FOR_PROFIT:DIAG_CANCER +
                          FOR_PROFIT:DIAG_NEUROLOGICO +
                          FOR_PROFIT:DIAG_CARD +
                          FOR_PROFIT:DIAG_RESPIRATORIO+
                          CARD:RENDA +
                          CARD:DIAG_CARD +

```

```

CANC:RENDA +
CANC:DIAG_CANCER +
OBS:RENDA +
OBS:DIAG_GRAVIDEZ +
RESP:RENDA +
RESP:DIAG_RESPIRATORIO +
RESP:DIAG_CARD +
as.factor(alt) +
strata(ID) , data = df,
method = "approximate")), error=function(e){return(c("ERRO"))})

```

```
df_model=data.frame(REGIAO=REGIAO[i])
```

```
#modelo logit antes
```

```

model2 = tryCatch({clogit(CNES ~ dist + dist_sq + dist:EMERG +
  ENSINO:RENDA +
  ENSINO:DIAG_GRAVIDEZ + ENSINO:DIAG_DIGESTIVO +
  ENSINO:DIAG_CANCER +
  ENSINO:DIAG_NEUROLOGICO +
  ENSINO:DIAG_CARD +
  ENSINO:DIAG_RESPIRATORIO +
  ENF_LT:RENDA +
  ENF_LT:DIAG_GRAVIDEZ + ENF_LT:DIAG_DIGESTIVO +
  ENF_LT:DIAG_CANCER +
  ENF_LT:DIAG_NEUROLOGICO +
  ENF_LT:DIAG_CARD +
  ENF_LT:DIAG_RESPIRATORIO +
  IMG:RENDA +
  IMG:DIAG_GRAVIDEZ + IMG:DIAG_DIGESTIVO +
  IMG:DIAG_CANCER +
  IMG:DIAG_NEUROLOGICO +
  IMG:DIAG_CARD +
  IMG:DIAG_RESPIRATORIO +
  FOR_PROFIT:RENDA +
  FOR_PROFIT:DIAG_GRAVIDEZ + FOR_PROFIT:DIAG_DIGESTIVO +
  FOR_PROFIT:DIAG_CANCER +
  FOR_PROFIT:DIAG_NEUROLOGICO +
  FOR_PROFIT:DIAG_CARD +
  FOR_PROFIT:DIAG_RESPIRATORIO+
  CARD:RENDA +
  CARD:DIAG_CARD +
  CANC:RENDA +
  CANC:DIAG_CANCER +
  OBS:RENDA +
  OBS:DIAG_GRAVIDEZ +
  RESP:RENDA +
  RESP:DIAG_RESPIRATORIO +
  RESP:DIAG_CARD +
  as.factor(alt) +
  strata(ID) , data = df_pre,
  method = "approximate")), error=function(e){return(c("ERRO"))})

```

```
df_model$MODELO_TUDO_ERRO=character(1)
```

```
if(class(model1)[1]=="character") {
```

```
  df_model$MODELO_TUDO_ERRO="ERRO"
```

```
}
```

```
df_model$MODELO_PRE_ERRO=character(1)
```

```

if(class(model2)[1]=="character") {

  df_model$MODELO_PRE_ERRO="ERRO"

}

#evitando quebra do loop com a não convergência

if(class(model1)[1]=="character" || class(model2)[1]=="character") {

  df_model$MODELO_TUDO = list(1)
  df_model$MODELO_PRE = list(1)
  df_model$ESCOLHAS_TUDO = list(1)
  df_model$ESCOLHAS_PRE = list(1)

  df_model1 = rbind(df_model1, df_model)

  next

}

#características adicionais para a construção de uma análise de bem-estar

escolhas1=df[complete.cases(df),] %>% mutate(n=1) %>% group_by(ID, MES) %>% summarize(Escolhas=sum(n))
escolhas2=df_pre[complete.cases(df_pre),] %>% mutate(n=1) %>% group_by(ID, MES) %>% summarize(Escolhas
=sum(n))

#inserção de características adicionais.

df_model$ESCOLHAS_TUDO = list(1)
df_model$ESCOLHAS_PRE = list(1)
df_model$MODELO_TUDO = list(model1)
df_model$MODELO_PRE = list(model2)
df_model$ESCOLHAS_TUDO = list(escolhas1)
df_model$ESCOLHAS_PRE = list(escolhas2)

names(df_model)

df_model1 = rbind(df_model1, df_model)

}

#salvando o resultado do modelo

destino = ""

saveRDS(df_model1, destino)

```

B.2 Extração dos coeficientes e valores relevantes

```

rm(list=ls()); gc()

library(dplyr)
library(tidyr)
library(stringr)
library(survival)

df_reg = readRDS("C:/Users/joao.morais/Desktop/demand_sus/REG/rds/regressao_mun_nova_logica_regiao2_a
te_tudo_c_warning_alt3_22.rds")
names(df_reg)

REGIAO = unique(df_reg$REGIAO)

leitura1 = function(i, df) {

  gc()

  print(i)

  mensagem_modelo=df_reg$MODELO_TUDO[[i]]
  mensagem_modelo2=df_reg$MODELO_PRE[[i]]

  if(class(mensagem_modelo)[1]!="clogit") {

    return(NULL)

  }

  if(class(mensagem_modelo2)[1]!="clogit") {

    return(NULL)

    # next

  }

  teste = data.frame(df_reg$MODELO_TUDO[[i]]$coefficients)
  teste$Variaveis = row.names(teste)
  names(teste)[1] = "Coeficientes"

  df_spread = teste %>%
    filter(str_detect(Variaveis,"as\\.factor")==F) %>%
    spread(Variaveis, Coeficientes)

  df_nomes = data.frame(REGIAO = as.character(df_reg$REGIAO[i]))

  df_estruturado1 = cbind(df_nomes, df_spread)

  df_estruturado1$Modelo = "De 04/2020 a 08/2020 - coeficientes"

  ###

  teste = data.frame(summary(df_reg$MODELO_TUDO[[i]]$coefficients[,5])
  teste$Variaveis = row.names(teste)
  names(teste)[1] = "P_Valor"

  df_spread = teste %>%
    filter(str_detect(Variaveis,"as\\.factor")==F) %>%
    spread(Variaveis, P_Valor)

```

```

df_nomes = data.frame(REGIAO = as.character(df_reg$REGIAO[i]))

df_estruturado2 = cbind(df_nomes, df_spread)

df_estruturado2$Modelo = "De 04/2020 a 08/2020 - p-valor"

###

teste = data.frame(summary(df_reg$MODELO_TUDO[[i]])$coefficients[,3])
teste$Variaveis = row.names(teste)
names(teste)[1] = "SE"

df_spread = teste %>%
  filter(str_detect(Variaveis,"as\\.factor")==F) %>%
  spread(Variaveis, SE)

df_nomes = data.frame(REGIAO = as.character(df_reg$REGIAO[i]))

df_estruturado3 = cbind(df_nomes, df_spread)

df_estruturado3$Modelo = "De 04/2020 a 08/2020 - SE"

df_estruturado=bind_rows(df_estruturado1,
                        df_estruturado2,
                        df_estruturado3)

###

warning1=data.frame(warning=rep(df_reg$MODELO_TUDO_warning[i],3))
obs=data.frame(observacoes=rep(length(df_reg$MODELO_TUDO[[i]]$residuals),3))
r2 = 1-(df_reg$MODELO_TUDO[[i]]$loglik[2]/df_reg$MODELO_TUDO[[i]]$loglik[1])
r2 = data.frame(pseudo_r2=rep(r2, 3))

df_estruturado = cbind(df_estruturado, warning1)
df_estruturado = cbind(df_estruturado, obs)
df_estruturado = cbind(df_estruturado, r2)

###

teste = data.frame(df_reg$MODELO_PRE[[i]]$coefficients)
teste$Variaveis = row.names(teste)
names(teste)[1] = "Coeficientes"

df_spread = teste %>%
  filter(str_detect(Variaveis,"as\\.factor")==F) %>%
  spread(Variaveis, Coeficientes)

df_nomes = data.frame(REGIAO = as.character(df_reg$REGIAO[i]))

df_estruturado1 = cbind(df_nomes, df_spread)

df_estruturado1$Modelo = "Média 5 anos - coeficientes"

###

teste = data.frame(summary(df_reg$MODELO_PRE[[i]])$coefficients[,5])
teste$Variaveis = row.names(teste)
names(teste)[1] = "P_Valor"

df_spread = teste %>%
  filter(str_detect(Variaveis,"as\\.factor")==F) %>%
  spread(Variaveis, P_Valor)

df_nomes = data.frame(REGIAO = as.character(df_reg$REGIAO[i]))

df_estruturado2 = cbind(df_nomes, df_spread)

```

```

df_estruturado2$Modelo = "Média 5 anos - p-valor"

# warning2=data.frame(warning=rep(df_reg$MODELO_PRE_warning[i],2))
#
###

teste = data.frame(summary(df_reg$MODELO_PRE[[i]])$coefficients[,3])
teste$Variaveis = row.names(teste)
names(teste)[1] = "SE"

df_spread = teste %>%
  filter(str_detect(Variaveis,"as\\.factor")==F) %>%
  spread(Variaveis, SE)

df_nomes = data.frame(REGIAO = as.character(df_reg$REGIAO[i]))

df_estruturado3 = cbind(df_nomes, df_spread)

df_estruturado3$Modelo = "Média 5 anos - SE"

df_estruturado2=bind_rows(df_estruturado1,
  df_estruturado2,
  df_estruturado3)

###

warning1=data.frame(warning=rep(df_reg$MODELO_PRE_warning[i],3))
obs=data.frame(observacoes=rep(length(df_reg$MODELO_PRE[[i]]$residuals),3))
r2 = 1-(df_reg$MODELO_PRE[[i]]$loglik[2]/df_reg$MODELO_PRE[[i]]$loglik[1])
r2 = data.frame(pseudo_r2=rep(r2, 3))

df_estruturado2 = cbind(df_estruturado2, warning1)
df_estruturado2 = cbind(df_estruturado2, obs)
df_estruturado2 = cbind(df_estruturado2, r2)

###

df_estruturado = plyr::rbind.fill(df_estruturado, df_estruturado2)

return(df_estruturado)
}

lista = lapply(1:dim(df_reg)[1], leitura1, df_reg)

df_final = do.call("rbind", lista)

df_final = df_final %>% distinct(.)

saveRDS(df_final, "")

```


B.3 Construção dos resultados

```

library(tidyverse)
library(ggplot2)
library(systemfonts)
library(ggrepel)

rm(list=ls()); gc()

df= readRDS("C:/Users/joao.morais/Desktop/demand_sus/REG/rds/coeficientes_extraidos_mun_antes_e_depois
_nova_logica_regiao3_n22.rds")

tipos_alerta=as.character(tipos_alerta)

#adequação dos nomes

nomes = names(df)[2:41]

nomes = data.frame(nomes)

nomes = nomes %>% mutate(Nomes_Final1=case_when(str_detect(nomes, "DIAG_CANCER:") ~ "Diagnóstico
Câncer:",
                                str_detect(nomes, "DIAG_RESPIRATORIO:") ~ "Diagnóstico Respiratório:",
                                str_detect(nomes, "DIAG_NEUROLOGICO:") ~ "Diagnóstico Neurológico:",
                                str_detect(nomes, "DIAG_GRAVIDEZ:") ~ "Diagnóstico Gravidez:",
                                str_detect(nomes, "DIAG_DIGESTIVO:") ~ "Diagnóstico Digestivo:",
                                str_detect(nomes, "DIAG_CARD:") ~ "Diagnóstico Cardíaco:",
                                str_detect(nomes, ":DIAG_CANCER") ~ "Diagnóstico Câncer:",
                                str_detect(nomes, ":DIAG_RESPIRATORIO") ~ "Diagnóstico Respiratório:",
                                str_detect(nomes, ":DIAG_NEUROLOGICO") ~ "Diagnóstico Neurológico:",
                                str_detect(nomes, ":DIAG_GRAVIDEZ") ~ "Diagnóstico Gravidez:",
                                str_detect(nomes, ":DIAG_DIGESTIVO") ~ "Diagnóstico Digestivo:",
                                str_detect(nomes, ":DIAG_CARD") ~ "Diagnóstico Cardíaco:",
                                str_detect(nomes, "RENDA") ~ "Renda Censo:",
                                str_detect(nomes, "dist") ~ "Distância percorrida (10km)",
                                str_detect(nomes, "_sq") ~ "Distância percorrida (10km) ao quadrado"),
                                Nomes_Final1=case_when(str_detect(nomes, "_sq") ~ "Distância percorrida (10km) ao quadrado"
                                ,
                                T ~ Nomes_Final1),
                                Nomes_Final2=case_when(str_detect(nomes, "ENF_LT") ~ "Enfermeiros por leito",
                                str_detect(nomes, "IMG") ~ "Serviços de Imagem",
                                str_detect(nomes, ":EMERG") ~ "Emergência",
                                str_detect(nomes, "PROFIT") ~ "Hospital Privado",
                                str_detect(nomes, ":CANC") ~ "Serviços Câncer",
                                str_detect(nomes, ":CARD") ~ "Serviços Cardíacos",
                                str_detect(nomes, ":OBS") ~ "Serviços Obstetria",
                                str_detect(nomes, ":RESP") ~ "Número de Respiradores",
                                str_detect(nomes, "ENSINO") ~ "Instituição de Ensino"),
                                Nomes_Final2 = case_when(is.na(Nomes_Final2) ~ "",
                                T ~ Nomes_Final2),
                                Nomes_Final = paste0(Nomes_Final1,Nomes_Final2)) %>%
                                select(nomes, Nomes_Final)

## excluindo os coeficiente insignificantes ##

#separação dos erros padrões#

df_ggplot1_SE = df %>%
gather(Coeficientes, Valores, -c(REGIAO, Modelo, warning)) %>%
mutate(Modelo2=case_when(str_detect(Modelo,"De 04/2020")==T ~ "De 04/2020 a 08/2020",
                                str_detect(Modelo,"Média")==T ~ "Média 5 anos"),
                                cutoff=case_when(str_detect(Modelo,"p-valor")==T & Valores<=0.05 ~ 1,
                                T ~ 0)) %>%
group_by(REGIAO, Modelo2, Coeficientes) %>%

```

```

mutate(cutoff=max(cutoff, na.rm = T)) %>%
filter(str_detect(Modelo, "SE")) %>%
# filter(cutoff==1) %>%
ungroup() %>%
inner_join(nomes, by=c("Coeficientes"="nomes")) %>%
select(-Coeficientes) %>%
rename(Coeficientes = Nomes_Final,
       Valores_SE = Valores) %>%
filter(str_detect(Modelo, "SE")) %>%
select(-Modelo) %>%
rename(Modelo=Modelo2) %>%
select(Coeficientes, Valores_SE, Modelo, REGIAO) %>%
distinct(.)

#separação dos coeficientes

df_ggplot1_coef = df %>%
gather(Coeficientes, Valores, -c(REGIAO, Modelo, warning)) %>%
mutate(Modelo2=case_when(str_detect(Modelo, "De 04/2020")==T ~ "De 04/2020 a 08/2020",
                        str_detect(Modelo, "Média")==T ~ "Média 5 anos",
                        cutoff=case_when(str_detect(Modelo, "p-valor")==T & Valores<=0.05 ~ 1,
                                         T ~ 0)) %>%
group_by(REGIAO, Modelo2, Coeficientes) %>%
mutate(cutoff=max(cutoff, na.rm = T)) %>%
filter(str_detect(Modelo, "coeficientes")) %>%
ungroup() %>% inner_join(nomes, by=c("Coeficientes"="nomes")) %>%
select(-Coeficientes) %>%
group_by(Nomes_Final, Modelo) %>%
mutate(q1=quantile(Valores, na.rm = T, probs = seq(0, 1, 0.25))[2],
       q2=quantile(Valores, na.rm = T, probs = seq(0, 1, 0.25))[4],
       cutoffq1=q1-(q2-q1)*1.5,
       cutoffq2=q2+(q2-q1)*1.5) %>%
filter(Valores>cutoffq1 & Valores<cutoffq2) %>%
filter(cutoff==1) %>%
rename(Coeficientes = Nomes_Final,
       Valores_coef = Valores) %>%
filter(str_detect(Modelo, "coeficientes")) %>%
select(-Modelo) %>%
rename(Modelo=Modelo2) %>%
select(Coeficientes, Valores_coef, Modelo, REGIAO) %>%
distinct(.)

#coeficientes mais EP#

df_conj_sign = inner_join(df_ggplot1_SE, df_ggplot1_coef, by=c("Coeficientes", "Modelo", "REGIAO"))

#média pelo inverso da variância#

res_sign=df_conj_sign %>% mutate(numerador = Valores_coef/Valores_SE^2,
                               denominador = 1/Valores_SE^2,
                               numerador=case_when(is.nan(numerador) ~ NA_real_,
                                                    is.infinite(numerador) ~ NA_real_,
                                                    T ~ numerador),
                               denominador=case_when(is.infinite(denominador) ~ NA_real_,
                                                      T ~ denominador)) %>%
group_by(Coeficientes, Modelo) %>%
summarize(numerador=sum(numerador, na.rm=T),
          denominador=sum(denominador, na.rm=T)) %>%
mutate(Coef = numerador/denominador,
       Coef_odds = exp(Coef))

```

APÊNDICE C – ANÁLISE ANTES E DEPOIS DE BEM-ESTAR

O cálculo do excedente do consumidor i , conforme Small e Rosen (1981), pode ser obtido realizando o logaritmo da soma dos exponenciais da equação (A.1):

$$EC_i = \ln \sum_h \exp(\delta_h + x_h v_{il} \beta) \quad (A.1)$$

Assumindo uma utilidade marginal constante da renda (α), é possível identificar uma variação do excedente do consumidor (no caso, do paciente):

$$\Delta EC = \frac{EC_{Depois} - EC_{Antes}}{\alpha} \quad (A.2)$$

Ou seja, considerando que α em média não se altera em uma análise *antes e depois*, sobretudo em uma mesma região, é possível identificar a existência de uma variação positiva ou negativa do excedente do consumidor.

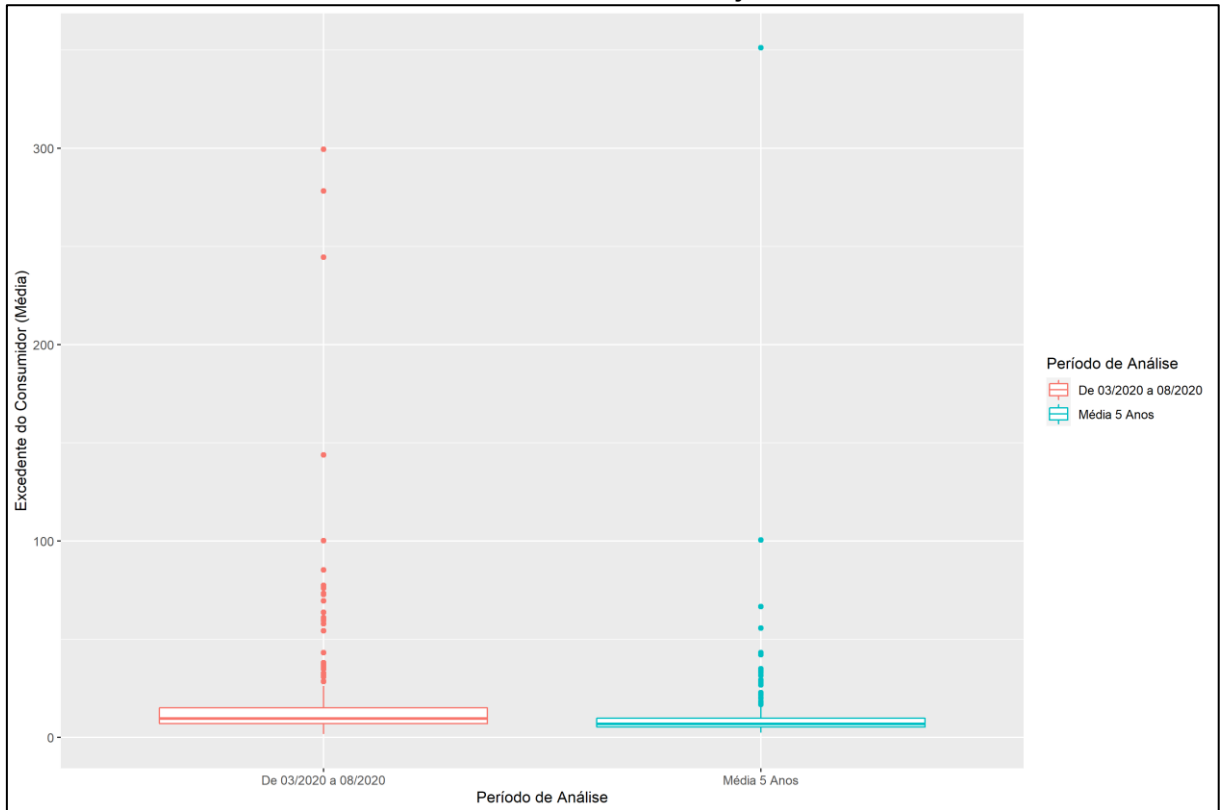
C.1 Resultados preliminares

Esta subseção identifica os efeitos antes e depois da pandemia analisando os dados pela métrica *logsum* ou a expressão identificada em A.2. Trata-se, portanto, dos resultados relacionados à investigação da variação de bem-estar.

Primeiro, é importante ressaltar que houve uma queda de pacientes no período da pandemia, conforme identificado na Figura 1 apresentado indiretamente na seção de dados deste trabalho, através do diagnóstico de doenças respiratórias. Em uma análise de bem-estar preliminar que considere o quantitativo de atendimento a pacientes internados, há aparentemente uma perda de bem-estar durante a pandemia.

Contudo, é possível realizar uma análise identificando efeitos de bem-estar na média — esses resultados são identificados na análise a seguir. A Figura A.12 abaixo identifica uma dispersão maior do Excedente do Consumidor após o início da pandemia — provavelmente substanciando a possibilidade de que os pacientes internados durante a pandemia *são mais informados* sobre as características hospitalares ou *necessitam* de mais atendimento prontamente, derivando maior utilidade do que antes da pandemia. Assim, na média, embora não seja possível no momento estabelecer uma causa clara, pacientes durante a pandemia derivaram um maior Excedente do Consumidor.

Figura A.12 – Análise com pareando pacientes – Análise antes e depois com os mesmos tipos de Pacientes com o mês julho



Fonte: Elaboração própria. Nota: a análise compreende somente o mês de julho.

Este resultado, naturalmente, como a seção destaca, é preliminar.

APÊNDICE D – DETALHAMENTO SOBRE A BASE SIH-RD

A Tabela A.1 demonstra as variáveis que são passíveis de serem extraídas para internações em hospitais conveniados do SUS. Considerando as variáveis que detalham diagnóstico, CEP, idade e sexo do paciente, há variáveis suficientes para reproduzir a estimação uma aproximação da estimação de demanda aplicada em Ho (2006) ou Ho e Lee (2017).

Tabela A.3 – Variáveis Disponíveis para acesso com a base SIH-RD

Nome da Variável	Descrição	Tipo de Variável
UF_ZI	Município de residência	texto
ANO_CMPT	Ano de processamento da AIH, formato aaaa	inteiro
MES_CMPT	Mês de processamento da AIH, formato mm	inteiro
ESPEC	Especialidade do leito	fator
CGC_HOSP	CNPJ do estabelecimento	texto
N_AIH	Número da AIH	texto
IDENT	Identificação do tipo da AIH	fator
CEP	CEP do paciente	texto
MUNIC_RES	Município de residência do paciente	inteiro
NASC	Data de nascimento do paciente	data
SEXO	Sexo do paciente	fator
UTI_MES_IN	variável desconhecida	inteiro
UTI_MES_AN	variável desconhecida	inteiro
UTI_MES_AL	variável desconhecida	inteiro
UTI_MES_TO	variável desconhecida	inteiro
MARCA_UTI	Indica qual o tipo de UTI utilizada pelo paciente	fator
UTI_INT_IN	variável desconhecida	inteiro
UTI_INT_AN	variável desconhecida	inteiro
UTI_INT_AL	variável desconhecida	inteiro
UTI_INT_TO	Quantidade de diárias em unidade intermediária	inteiro
DIAR_ACOM	Quantidade de diárias de acompanhante	inteiro
QT_DIARIAS	Quantidade de diárias	inteiro
PROC_SOLIC	Procedimento solicitado	texto
PROC_REA	Procedimento realizado	texto
VAL_SH	Valor de serviços hospitalares	numérico
VAL_SP	Valor de serviços profissionais	numérico
VAL_SADT	variável desconhecida	numérico
VAL_RN	variável desconhecida	numérico
VAL_ACOMP	variável desconhecida	numérico
VAL_ORTP	variável desconhecida	numérico
VAL_SANGUE	variável desconhecida	numérico
VAL_SADTSR	variável desconhecida	numérico
VAL_TRANSP	variável desconhecida	numérico
VAL_OBSANG	variável desconhecida	numérico
VAL_PED1AC	variável desconhecida	numérico
VAL_TOT	Valor total da AIH	numérico
VAL_UTI	Valor de UTI	numérico
US_TOT	Valor total, em dólar	numérico
DT_INTER	Data de internação, no formato aaaammdd	data
DT_SAIDA	Data de saída, no formato aaaammdd	data
DIAG_PRINC	Código de diagnóstico principal (CID10)	texto
DIAG_SECUN	Código de diagnóstico secundário (CID10)	texto
COBRANCA	Motivo de saída/permanência	fator
NAT_JUR	Natureza jurídica do estabelecimento, conforme a Comissão Nacional de Classificação - CONCLA	fator

Nome da Variável	Descrição	Tipo de Variável
GESTAO	Indica o tipo de gestão do hospital	fator
RUBRICA	variável desconhecida	numérico
IND_VDRL	Indica exame VDRL	fator
MUNIC_MOV	Município do estabelecimento	inteiro
COD_IDADE	Unidade de medida da idade	fator
IDADE	Idade	inteiro
DIAS_PERM	Dias de permanência	inteiro
MORTE	Indica óbito	fator
NACIONAL	Nacionalidade do paciente	fator
NUM_PROC	variável desconhecida	numérico
CAR_INT	Caráter da internação	fator
TOT_PT_SP	variável desconhecida	numérico
CPF_AUT	variável desconhecida	numérico
HOMONIMO	Indicador se o paciente da AIH é homônimo de outra AIH	fator
NUM_FILHOS	Número de filhos do paciente	inteiro
INSTRU	Grau de instrução do paciente	fator
CONTRACEP1	Tipo de contraceptivo utilizado	fator
CONTRACEP2	Tipo de contraceptivo utilizado	fator
GESTRISCO	Indicador se é gestante de risco	fator
INSC_PN	Número da gestante no pré-natal	texto
SEQ_AIH5	Sequencial de longa permanência (AIH tipo 5)	fator
CBOR	Ocupação do paciente, segundo a Classificação Brasileira de Ocupações - CBO	texto
CNAER	Código de acidente de trabalho	texto
VINCPREV	Vínculo com a previdência	fator
GESTOR_COD	Motivo de autorização da AIH pelo Gestor	fator
GESTOR_TP	Tipo de gestor	inteiro
GESTOR_CPF	Número do CPF do gestor	inteiro
GESTOR_DT	Data de autorização do gestor, formato aaaammdd	data
CNES	Código CNES do hospital	texto
CNPJ_MANT	CNPJ da mantenedora	texto
INFEHOSP	Status de infecção hospitalar	fator
CID_ASSO	CID causa	texto
CID_MORTE	CID da morte	texto
COMPLEX	Complexidade	fator
FINANC	Tipo de financiamento	fator
FAEC_TP	Subtipo de financiamento FAEC	fator
REGCT	Regra contratual	fator
RACA_COR	Raça/cor do paciente	fator
ETNIA	Etnia do paciente, se raça cor for indígena	fator
SEQUENCIA	Sequencial da AIH na remessa	texto
REMESSA	Número da remessa	texto
AUD_JUST	Justificativa do auditor para aceitação da AIH sem o número do Cartão Nacional de Saúde	texto
SIS_JUST	Justificativa do estabelecimento para aceitação da AIH sem o número do Cartão Nacional de Saúde	texto
VAL_SH_FED	Valor do complemento federal de serviços hospitalares. Está incluído no valor total da AIH	numérico
VAL_SP_FED	Valor do complemento federal de serviços federais. Está incluído no valor total da AIH	numérico
VAL_SH_GES	Valor do complemento do gestor (estadual ou municipal) de serviços hospitalares. Está incluído no valor total da AIH	numérico
VAL_SP_GES	Valor do complemento do gestor (estadual ou municipal) de serviços profissionais. Está incluído no valor total da AIH	numérico
VAL_UCI	Valor de UCI	numérico
MARCA_UCI	Tipo de UCI utilizada pelo paciente	fator
DIAGSEC1	Diagnóstico secundário 1	texto
DIAGSEC2	Diagnóstico secundário 2	texto
DIAGSEC3	Diagnóstico secundário 3	texto
DIAGSEC4	Diagnóstico secundário 4	texto
DIAGSEC5	Diagnóstico secundário 5	texto
DIAGSEC6	Diagnóstico secundário 6	texto

Nome da Variável	Descrição	Tipo de Variável
DIAGSEC7	Diagnóstico secundário 7	texto
DIAGSEC8	Diagnóstico secundário 8	texto
DIAGSEC9	Diagnóstico secundário 9	texto
TPDISEC1	Tipo de diagnóstico secundário 1	fator
TPDISEC2	Tipo de diagnóstico secundário 2	fator
TPDISEC3	Tipo de diagnóstico secundário 3	fator
TPDISEC4	Tipo de diagnóstico secundário 4	fator
TPDISEC5	Tipo de diagnóstico secundário 5	fator
TPDISEC6	Tipo de diagnóstico secundário 6	fator
TPDISEC7	Tipo de diagnóstico secundário 7	fator
TPDISEC8	Tipo de diagnóstico secundário 8	fator
TPDISEC9	Tipo de diagnóstico secundário 9	fator

Fonte: Wiki das convenções da base SIH-RD (github.com/rfsaldanha).⁶³

⁶³ As informações da wiki se encontram disponíveis no site do autor do pacote *microdatasus* (Software R), Raphael Freitas Saldanha, hospedado pela plataforma github. Disponível em <https://github.com/rfsaldanha/microdatasus/wiki/Conven%C3%A7%C3%B5es-SIH-RD>. Acesso em: 31 de mai. 2021.

APÊNDICE E – ANÁLISE DE FORTALEZA

Este trabalho em uma frase preliminar investigou estimações de demanda apenas para a cidade de Fortaleza-CE, com o objetivo de construir uma estimação que demonstrasse coeficientes suficientemente significativos e identificasse direções similares às encontradas por Ho (2006) e Ho e Lee (2017) em suas estimações de primeiro estágio. A análise foi apresentada no âmbito da qualificação deste trabalho.

A análise foi implementada com absolutamente todas as características descritas ao longo do corpo deste trabalho, consideradas as algumas exceções listadas abaixo.

A análise foi implementada com intuito de verificar potenciais problemas e adequações que seriam necessárias à análise. É possível discutir sobre a magnitude e direção de alguns coeficientes estimados⁶⁴, mas se tratou de uma análise preliminar para a investigação implementada neste trabalho.

E.1 Exceções da Análise de Fortaleza

Nessa análise preliminar não foi concebida a construção de algumas variáveis, como as interações de respiradores e diagnóstico de doenças respiratórias. No início dessa análise, que foi iniciada na primeira metade do ano de 2020, este trabalho não havia ainda constatado como deveria considerar o papel do COVID-19 na presente análise.

Outras exceções seguem à frente. As variáveis que compõe informações salariais foram obtidas através de dados da RAIS,⁶⁵ como uma *proxy* dos salários dos CEPs de residência dos pacientes — hipotetiza-se que os salários de empresas localizadas nas regiões de residência dos pacientes é uma *proxy* suficientemente adequada para o salário. Com essa informação, cria-se a variável “Salário CEP6” ou “Salário”, que será utilizada neste trabalho.

Os dados de distância entre pacientes e hospitais foram obtidos através da captura latitude e longitudes associadas com os CEPs brasileiros através da API “CEP

⁶⁴ Algumas interações adquiriram valores bem altos ainda que com a direção esperada – interações de serviços de câncer e o diagnóstico de câncer.

⁶⁵ Em virtude da defasagem de dados do último Censo, optou-se por utilizar a aproximação da RAIS para obter uma aproximação potencialmente mais crível dos salários da região.

Aberto”, desde pacientes a hospitais. Com os dados de latitude e longitude, computou-se a distância geodésica⁶⁶ entre os hospitais e pacientes com o fim de obter a variável descrita em Ho (2006). Com esse detalhamento, foi possível efetuar a construção das variáveis “Distância (10 km)”, onde uma unidade da variável representa 10 km, e a variável “Distância (10 km) ao quadrado” — uma versão elevada ao quadrado da mesma variável.

A construção da base de hospitais consistiu na verificação de CNPJs com um número positivo de leitos pela base CNES-ST e que se encontravam na base SIH-RD para a cidade de Fortaleza. Foram considerados, assim, 33 hospitais no município de Fortaleza para a estimação de demanda.

A Tabela A.3 à frente sintetiza os coeficientes e o fato de que os coeficientes aparentemente estavam corretos seguiam uma direção intuitiva.

⁶⁶ Uma derivação da fórmula pode ser obtida através de Jekeli (2012), com a demonstração da menor distância entre dois pontos em um elipsóide.

Tabela A. 4 – Resultados da Regressão Logística com os dados obtidos para a Cidade de Fortaleza

Interações	Variáveis	(1)	(2)	(3)	(4)	(5)
	Distância (10 km)	-0,521*** (0,02)	-0,953*** (0,019)	-1,215*** (0,046)	-1,329*** (0,047)	-1,234*** (0,049)
	Distância (10 km) ao quadrado	0,0069*** (0,0009)	0,0069*** (0,0009)	0,0076*** (0,001)	0,0069*** (0,001)	0,0074*** (0,001)
	Distância (10 km)*Emergência	0,145*** (0,022)	0,523*** (0,022)	0,31*** (0,05)	0,44*** (0,051)	0,339*** (0,052)
Enfermo	Salário	0,551*** (0,01)	0,639*** (0,011)	0,073*** (0,021)	0,081*** (0,023)	0,091*** (0,026)
	Gravidez	0,014 (0,033)	-0,156*** (0,033)	-0,133*** (0,039)	-0,502*** (0,042)	-1,011*** (0,054)
	Digestivos	0,394*** (0,039)	0,429*** (0,04)	-0,012 (0,046)	-0,14*** (0,048)	-0,121** (0,048)
	Câncer	-0,411*** (0,048)	-0,586*** (0,047)	-1,069*** (0,053)	-0,459*** (0,059)	-1,69*** (0,087)
	Neurológico	0,151** (0,066)	0,002 (0,066)	-0,13* (0,072)	0,617*** (0,088)	0,412*** (0,088)
	Cardíacos	0,552*** (0,048)	0,897*** (0,05)	0,609** (0,057)	0,641*** (0,062)	0,685*** (0,092)
	Respiratório	0,001 (0,044)	-0,04 (0,044)	-0,222*** (0,048)	0,015 (0,054)	0,006 (0,053)
	Numero de enfermeiros por leito	Salário		-0,038*** (0,001)	0,01*** (0,002)	0,009*** (0,003)
Gravidez			-0,082*** (0,006)	-0,667*** (0,021)	-0,6*** (0,021)	-0,319*** (0,027)
Digestivos			-0,02*** (0,005)	0,017* (0,01)	0,035*** (0,009)	0,03*** (0,008)
Câncer			0,005 (0,004)	0,097*** (0,006)	0,055*** (0,006)	-0,015** (0,006)
Neurológico			-0,084*** (0,013)	-0,739*** (0,042)	-0,79*** (0,041)	-0,601*** (0,039)
Cardíacos			0,05*** (0,003)	0,129** (0,006)	0,118*** (0,006)	0,156*** (0,007)
Respiratório			-0,007 (0,004)	0,063*** (0,007)	0,036*** (0,007)	0,025*** (0,006)
Serviços de Imagem		Salário			-0,032 (0,034)	-0,024 (0,035)
	Gravidez			-1,149*** (0,077)	-1,373*** (0,078)	-0,488** (0,098)
	Digestivos			0,242** (0,067)	0,157** (0,067)	0,16** (0,066)
	Câncer			0,46*** (0,08)	1,244*** (0,09)	-0,686** (0,111)
	Neurológico			1,198*** (0,093)	1,646*** (0,107)	1,48*** (0,107)
	Cardíacos			-2,158*** (0,133)	-2,132*** (0,134)	-3,457** (0,136)
	Respiratório			-2,31*** (0,135)	-2,163*** (0,136)	-2,181*** (0,136)
	Hospitais Privados	Salário				0,026 (0,028)
Gravidez					-1,381*** (0,069)	0,406*** (0,08)
Digestivos					-0,513*** (0,067)	-0,505*** (0,066)
Câncer					1,909*** (0,062)	1,256*** (0,085)
Neurológico					2,183*** (0,085)	1,881*** (0,081)
Cardíacos					0,114 (0,071)	3,14*** (0,148)
Respiratório					0,658*** (0,056)	0,696*** (0,057)
Serviços Cardíacos		Salário				
	Cardíacos					5,774*** (0,162)
Serviços para Câncer	Salário					-0,038 (0,042)
	Câncer					6,028*** (0,162)
Serviços de Obstetrícia	Salário					0,033 (0,037)
	Gravidez					4,342*** (0,089)
Observações		919644	919644	919644	919644	919644
Efeitos fixos de Hospitais		Não	Não	Sim	Sim	Sim
Pseudo-R ²		0,043	0,053	0,187	0,203	0,268
log-verossimilhança		-93202,1651	-92283,56409	-79211,74349	-77636,95759	-71268,08762

Nota: Os asteriscos representam os graus de significância dos coeficientes *p<0.1; **p<0.05; ***p<0.01.