

UNIVERSIDADE DE LISBOA  
FACULDADE DE CIÊNCIAS  
DEPARTAMENTO DE INFORMÁTICA



# **ESTUDO DO IMPACTO DA PANDEMIA NAS ESTRUTURAS RESIDENCIAIS PARA IDOSOS DE PORTUGAL EM 2020**

Pedro Sávio Arruda Pires

**Mestrado em Bioinformática e Biologia Computacional**

Dissertação orientada por:  
Prof. Ana Nunes  
e Prof. Ganna Rozhnova

# Dedicatória e agradecimentos

Dedico este trabalho aos meus queridos pais Roberto César Lopes Pires e a Dra. Cleide Maria Furtado Arruda Pires, advogado e médica especialista em Endocrinologia. Ambos são graduados pela Universidade Federal do Ceará (UFC) e a última é também mestre pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Eles sempre me apoiaram incondicionalmente e serão sempre os meus maiores modelos de virtude e inspiração por toda a minha caminhada. Faço um especial agradecimento às minhas orientadoras Prof. Ana Nunes e Prof. Ganna Rozhnova, que não mediram esforços para cooperar e unir forças para o alcance do propósito deste trabalho. Faço igualmente um grande agradecimento aos prezados Prof. Manuel Carmo Gomes, Dr. André Peralta, Dra. Eugénia Fernandes e Dr. Pedro Casaca, sendo o primeiro professor da Faculdade de Ciências da Universidade de Lisboa, que colaborou imenso com o nosso projeto, e os seguintes da Direção de Serviços de Informação e Análise (DSIA) pertencente à Direção-Geral da Saúde (DGS), pelo excelente trabalho em parceria e colaboração no fornecimento dos dados e mais informações relevantes, vitais para a viabilização deste trabalho. Sem o apoio de cada um desses indivíduos, este trabalho não teria sido possível de ser implementado.

# Resumo

O nosso projeto baseia-se na análise dos registos de casos e mortes por COVID-19 em utentes ocorridos em ERPI entre 15 de abril e 27 de dezembro de 2020. Com base nestes dados, determinou-se Risco Relativo (RR) de infeção e de morte dos utentes em ERPIs relativamente à população idosa em geral para as cinco regiões de saúde (com a exceção da região Alentejo devido a problemas de qualidade nos seus registos de casos e óbitos), e para os dois períodos de 2020 que correspondem à primeira e à segunda vaga daquele ano. Verificámos que em todas as regiões o RR de infeção foi reduzido na segunda vaga, o que demonstra a capacidade de adaptação das ERPIs em geral no sentido de melhorar a proteção dos utentes. Esta análise mostra também que em Lisboa e Vale do Tejo (LVT) surge como um outlier de excesso de RR e de menor redução de RR na segunda vaga. Isto motivou um estudo mais detalhado em que LVT foi comparada com o conjunto das outras regiões (OUTRAS). Fizemos então uma análise por tamanho das ERPIs, depois de verificar que a distribuição das ERPIs por tamanhos é aproximadamente a mesma em LVT e em OUTRAS. Os indicadores analisados em função do tamanho foram o odds ratio (OR) de um lar registar um ou mais casos entre os seus utentes e o tamanho de cada surto registado (TS) medido como a percentagem dos utentes infetados. A métrica TS comporta-se em todas as regiões de acordo com o esperado e o reportado na literatura, diminuindo com o tamanho da ERPI. Quanto ao OR, em OUTRAS ele comporta-se como esperado, aumentando significativamente com o tamanho da ERPI, mas em LVT o OR mínimo ocorre nas ERPIs médias, e as ERPIs pequenas surgem com um nível de risco muito elevado. Assim, o comportamento de outlier de LVT resulta em parte de um excesso de risco de infeção nas ERPIs pequenas. A análise do ratio funcionário utente mostra também que as ERPIs pequenas de LVT são outliers com valores muito pequenos desse ratio, relativamente às outras classes de ERPIs.

**Palavras Chave:** Covid-19, Risco Relativo, Estruturas Residenciais de Pessoas Idosas, Estudo Observacional de Coorte, Aprendizados durante a pandemia.

# Abstract

Our project is based on the analysis of records of COVID-19 cases and deaths among residents of nursing homes that occurred between April 15th and December 27th, 2020. Based on these data, the relative risk (RR) of infection and death of residents in nursing homes was determined relative to the general elderly population for the five health regions (excluding the Alentejo region due to problems with the quality of its case and death records), and for the two periods of 2020 corresponding to the first and second waves of that year. We found that in all regions the RR of infection was reduced in the second wave, which demonstrates the ability of nursing homes in general to adapt in order to better protect residents. This analysis also shows that Lisbon and Vale do Tejo (LVT) appears as an outlier with a high RR and a lower reduction in RR in the second wave. This motivated a more detailed study in which LVT was compared to the remaining regions (OUTRAS). We then carried out analysis by size of nursing homes after verifying that the distribution of nursing homes by size is approximately the same in LVT and in OUTRAS. The indicators analyzed based on size were the odds ratio (OR) of a nursing home having one or more cases among its residents and the size of each outbreak recorded (TS) measured as the percentage of infected residents. The TS metric behaves in all regions as expected and reported in the literature, decreasing with the size of the nursing home. As for the OR, in OUTRAS it behaves as expected, significantly increasing with the size of the nursing home, but in LVT the minimum OR occurs in medium-sized nursing homes, and small nursing homes have a very high risk level. Thus, LVT's outlier behavior is partially due to an excess risk of infection in small nursing homes. The analysis of the staff-to-resident ratio also shows that small nursing homes in LVT are outliers with very low values of this ratio relative to other classes of nursing homes.

**Keywords:** Covid-19, Relative Risk, Residential Structures for Elderly People, Observational Cohort Study, Lessons from the Pandemic.

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Contextualização	1
1.2	Objetivos	3
1.3	Descrição dos dados	3
1.4	Definições em uso	6
1.4.1	Caso de infecção por SARS-CoV-2	6
1.4.2	Óbitos por COVID-19	6
1.4.3	Estabelecimento Residencial para idosos ou ERPI	6
1.4.4	Funcionários de ERPI	7
<b>2</b>	<b>Materiais e métodos</b>	<b>8</b>
2.1	Pré-processamento e análise dos dados	8
2.1.1	Integração e Limpeza dos Dados	8
2.1.2	Transformação dos Dados	10
2.1.3	Agregação de regiões e estratificação do número de ERPIS e de utentes por tamanhos das ERPIS	13
2.2	Cálculo do RR	15
2.3	Análise do OR do risco de um lar apresentar pelo menos um caso de COVID-19	16
2.4	Análise do tamanho médio do surto	18
<b>3</b>	<b>Resultados</b>	<b>19</b>
3.1	População total do estudo em 2020	19
3.2	Análise dos surtos ocorridos em ERPI no ano de 2020	23
3.2.1	Análise do RR em ERPIS por região, por período e por tamanho	23
3.2.2	Caracterização dos surtos em ERPIS por percentagem de casos	27
3.2.3	Comparação do OR e da proporção do tamanho do surto em ERPI por ARS e pelas suas dimensões	31
<b>4</b>	<b>Conclusões e discussão</b>	<b>36</b>
	<b>Bibliografia</b>	<b>39</b>



# Lista de Tabelas

1.1	Representação das variáveis de interesse inerentes aos dados do registo epidemiológico da COVID-19 em ERPIs, presentes em cada folha representante das ARS nos 35 ficheiros de dados.	4
1.2	Variáveis dos dados referentes ao número de indivíduos residentes em cada região de saúde ou ARS estratificado por grupos etários.	4
1.3	Variáveis de interesse nos dados oriundos do ficheiro com a contagem de casos por COVID-19 em território nacional.	5
1.4	Variáveis dos dados obtidos no endereço eletrónico da Carta Social com o uso de Web Scraping.	5
1.5	Representação tabular com três variáveis de interesse nos dados geográficos em conjunto com a região de saúde correspondente.	6
1.6	Capa de cada ficheiro de dados com o número total de ERPIs, utentes e funcionários estratificados por cada ARS	6
2.1	Representação do cálculo para estimação da incidência com a seleção de sete variáveis, incluindo a exibição dos decréscimos da prevalência entre utentes e funcionários, para obtenção do número de casos nestes dois grupos.	12
3.1	Incidência calculada da COVID-19 em utentes dos lares para os meses de abril a dezembro de 2020 em cada ARS.	21
3.2	Número de utentes e ERPIs em cada classe de tamanhos em ARSLVT e nas outras três regiões.	25
3.3	Resultados da Regressão Logística para o risco de um ou mais casos nos lares e seu respetivo OR em relação à cada classe de referência.	34
3.4	Resultados da Regressão Linear para o tamanho médio do surto em função dos tamanhos dos lares como única variável preditora.	34
3.5	Média do rácio entre utentes e funcionários, dos casos em funcionários, da proporção de casos por funcionários e os seus respetivos desvios padrões.	35



# Lista de Figuras

3.1	Distribuição dos utentes e ERPIS em números na sua totalidade por cada ARS.	19
3.2	Distribuição dos funcionários e ERPIS em números na sua totalidade por cada ARS.	20
3.3	Distribuição dos utentes e ERPIS nos dados com reporte de surtos.	20
3.4	Distribuição dos funcionários e ERPIS nos dados com reporte de surtos.	21
3.5	Comparação da dimensão média das ERPIS reportados com surtos (em vermelho) e do total de ERPIS existentes (em azul) no ano de 2020.	22
3.6	Curvas de Incidência da COVID-19 no ano de 2020 para os utentes das ERPIS (azul) em comparação com todos os idosos de Portugal continental (vermelho).	22
3.7	RR para a incidência em ERPIS nas regiões de saúde no ano de 2020.	24
3.8	RR para a incidência nas regiões de saúde (com exceção da ARSA) até agosto e após agosto.	24
3.9	RR em função do tamanho em ARSLVT e nas outras três regiões.	25
3.10	RR para os óbitos nas regiões de saúde.	26
3.11	RR para os óbitos nas regiões de saúde (com exceção da ARSA) no primeiro e no segundo período.	27
3.12	RR para os óbitos em função do tamanho na ARSLVT e nas outras três regiões.	28
3.13	Proporções de utentes (vermelho) e casos (azul) distribuídos por intervalos do número de utentes em ERPI na ARSLVT dentro dos dois períodos do ano de 2020.	28
3.14	Proporções de utentes (vermelho) e casos (azul) distribuídos por intervalos do número de utentes em ERPI nas regiões ARSALG, ARSC e ARSN dentro dos dois períodos do ano de 2020.	29
3.15	Frequência relativa das ERPIS da região ARSLVT distribuída pela proporção do número de casos em utentes pelo número total destes indivíduos nos dois períodos do ano.	30
3.16	Frequência relativa das ERPIS nas regiões ARSALG, ARSC e ARSN distribuída pela proporção do número de casos de utentes pelo número total destes indivíduos nos dois períodos do ano.	31
3.17	Distribuição de ERPI por número de utentes na região ARSLVT.	32
3.18	Distribuição de ERPI por número de utentes nas regiões ARSALG, ARSC e ARSN.	32
3.19	Distribuição por tamanhos de todas as ERPIS existentes em cada classe regional (ARSLVT e o conjunto das outras três regiões)	33



# Capítulo 1

## Introdução

### 1.1 Contextualização

Com o desenrolar da pandemia de COVID-19 em Portugal, as políticas de saúde pública e de tomada de decisão por parte das autoridades sanitárias visaram a proteção dos grupos de risco, em especial o de idosos residentes em instalações do tipo ERPI (Estrutura Residencial para Pessoas Idosas).

O grupo dos idosos tem sido ao longo do tempo bastante vulnerável à infeção pelo SARS-CoV-2, em parte devido ao contato múltiplo entre os utentes e funcionários nos ambientes do tipo ERPI. Acresce que, tal como reportado em [1], entre os motivos que contribuem para um maior risco de prognóstico negativo (hospitalização e morte) na terceira idade estão a frequência de comorbidades pré-existentes ou os impactos na imunidade com o processo de envelhecimento.

Como observa [2], o contágio inicial da doença em Portugal, no ano de 2020, se expressou mais fortemente em territórios com uma menor densidade populacional e com mais idosos. Os focos ocorridos em ERPIs tiveram, conforme os autores, uma parcela bastante significativa das novas contaminações a partir do primeiro dia do mês de abril daquele ano. Como resposta ao alto contágio descrito nestas instalações, que se traduziram numa ameaça para os residentes ali presentes, algumas medidas não farmacológicas para mitigação do contágio foram adotadas pelas ERPIs.

Estas medidas, implementadas desde o início da pandemia, agiram em concordância com a Orientação de n.º 009/2020 da Direção-Geral da Saúde (DGS) datada de 11 de março de 2020 e acessível em [3]. Nesse âmbito as medidas gerais definidas lidaram com as visitas aos utentes, higiene e limpeza do espaço e dos resíduos, concentração de pessoas e distanciamento social. No documento também foi feita a apresentação das diretrizes sobre como tratar a admissão de novos utentes, casos suspeitos ou confirmados em utentes, bem como hospitalizações e óbitos ocorridos nas instituições.

No que diz respeito aos casos suspeitos nas instalações, o documento detalhou os sintomas

a serem considerados, determinou o isolamento destes utentes e o uso de testes laboratoriais para despiste. Em situações de casos confirmados nos utentes, a Autoridade de Saúde Local foi orientada a fazer o reporte imediato da situação para a Autoridade Regional de Saúde (ARS) e igualmente para a Direção Executiva do Agrupamento de Centros de Saúde (ACeS), Direção Técnica do Estabelecimento, a Direção Distrital da Segurança Social, Direção do Estabelecimento e o Presidente da Câmara. Outra atribuição dada a Autoridade de Saúde Local foi a de definir e acompanhar estratégias para a realização de testes e de isolamento de espaços para mitigar os riscos de transmissão para outros idosos institucionalizados.

Outro ponto tratado no documento diz respeito aos óbitos ocorridos em instituições com casos confirmados da doença ou que tenham tido utentes/funcionários com sintomas compatíveis com a doença. Nestas situações a orientação dada foi a de sempre considerar o óbito como caso suspeito de infeção até que se provasse o oposto via teste laboratorial.

Conforme descreveu [4], a medida que a pandemia seguia o seu rumo, algumas ERPIs adotaram o uso de testes rápidos de antigénio para uma maior agilidade na identificação dos casos positivos e rápido isolamento dos mesmos dentro das próprias instalações.

No entanto, como relatou [5], existiram níveis de risco diferentes conforme as características específicas de cada lar. Na Espanha, notou-se no estudo retrospectivo de [6], por uma análise de oito *clusters*, agrupados com base nas características extraídas de uma amostra de 167 instalações, que em quatro destes grupos a alta mortalidade foi resultante dos fatores: maior proporção de residentes com comorbidades crónicas ou doenças avançadas, baixa capacidade de resposta em relação às medidas preventivas necessárias face ao surgimento da pandemia, tamanhos maiores das instituições e localização em zonas rurais.

O comportamento do alto contágio em lares seguido de mortes foi igualmente reportado em mais estudos: nos Estados Unidos, conforme observa [7], até 21 de maio de 2020, foram registadas pelo menos 35.000 mortes ocasionadas pela COVID-19 nestas instalações, o que representou 42% de todos os óbitos ocorridos pela doença até àquela data em 38 estados americanos. Na supracitada Espanha, conforme outro estudo [8], entre os dias de 20 de março e 1 de junho de 2020, foi observado o número de 282 mortes numa amostra de 630 pacientes (47.5%) oriundos de lares em Madrid com quadro clínico compatível com a COVID-19. Uma percentagem de 44% das mortes causadas pela doença na França, como descrito por [9], ocorreram igualmente em lares de idosos até novembro de 2020.

O cenário mudou a partir do início de 2021, com a implementação do programa nacional da vacinação indicada para a imunização contra o SARS-CoV-2, que teve princípio no dia 27/12/2020 [10]. Nesse contexto, foi priorizada a população envelhecida, portanto, foram vacinados a maioria dos utentes e funcionários em ERPIs de idosos visando reduzir a mortalidade ocasionada pela doença, sendo a vacinação muito protetora da hospitalização e morte como demonstrou [5].

É a hipótese central deste presente capítulo a observação de um maior risco de incidência e de

mortes nas instalações do tipo ERPI em território nacional em comparação com casos ocorridos fora destes locais, a exemplo do que foi reportado em outros países conforme os estudos [2], [4], [5], [6], [7], [8], [9], previamente citados. Neste âmbito, será feito no presente estudo a análise dos riscos inerentes a estes ambientes com uma base nos dados de infeções e mortes ocorridos na população idosa dentro e fora destas instalações.

## 1.2 Objetivos

O presente estudo tem por objetivo maior o de dar uma contribuição no entendimento da dinâmica dos riscos inerentes aos lares de idosos no ano de 2020. São os seus objetivos específicos:

1. Obtenção da incidência e dos óbitos ocorridos nas instalações do tipo ERPI para a pandemia de COVID-19 no ano de 2020.
2. Comparação da incidência e mortalidade em utentes de ERPI com a população idosa em geral através do Risco Relativo (RR) por regiões de saúde ou Autoridade Regional de Saúde (ARS) e as suas agregações. Também procurou-se avaliar se existiram diferenças entre as regiões de saúde e tamanhos dos lares, através do Odds Ratio (OR), no que diz respeito ao risco de um lar apresentar um ou mais casos de COVID-19.
3. Procurar identificar se houve diferenças significativas no valor do RR, OR e as suas agregações, bem como avaliar quais fatores podem ter levado a essa diferenciação.

## 1.3 Descrição dos dados

O trabalho desenvolvido contou com a análise de quatro fontes de dados descritivas do contexto da pandemia da COVID-19 em todo o território em Portugal continental. Podem ser detalhadas como sendo:

1 - Um conjunto de 35 ficheiros de dados epidemiológicos correspondentes a uma lista de 35 datas de registos intercalados com intervalo de aproximadamente uma semana cada, em formato “xlsx”, no qual cada ficheiro contém dados de surtos ativos em ERPIs descritos em cada uma destas datas que compreendem ao total um período aproximado de 35 semanas. Estes registos foram feitos para os lares das cinco regiões do continente, no contexto da COVID-19, com variáveis descritivas dos casos, hospitalizações e mortes decorrentes dos surtos em cada ERPI no ano de 2020, conforme a tabela 1.1. Estes dados foram partilhados no âmbito de colaboração académica pela Direção de Serviços de Informação e Análise (DSIA), pertencente à DGS.

Conforme indicado na tabela 1.1, cada ficheiro possui na sua maioria cinco folhas respetivas à cada uma das cinco ARS continentais e, com exceção de um único ficheiro, que continha uma folha adicional sem identificação, estas foram integradas num mesmo modelo de dados. Os

Variável	Descrição
Data do Registo	Data de registo dos casos no lar.
ARS	ARS na qual o lar se localiza.
Distrito	Distrito no qual o lar se localiza.
Nome/Identificação	Nomes dos lares com casos confirmados ou suspeitos de COVID-19.
Localização (Concelho)	Concelho ao qual o lar está localizado.
Ocupação	Número de utentes residentes no lar na altura da data registada.
Casos Confirmados	Número de casos confirmados para COVID-19 em utentes na altura da data registada (prevalência).
Casos Pendentes	Número de utentes que aguardam o resultado do teste.
Óbitos	Número cumulativo dos óbitos ocorridos em cada lar dentro do intervalo correspondente aos meses de abril e dezembro.
Número de Funcionários	Número de funcionários que trabalham no lar.
Casos em Funcionários	Número de casos confirmados para COVID-19 em funcionários dos lares na altura da data registada (prevalência).

Tabela 1.1: Representação das variáveis de interesse inerentes aos dados do registo epidemiológico da COVID-19 em ERPIs, presentes em cada folha representante das ARS nos 35 ficheiros de dados.

dados das folhas estavam uniformemente estruturados para todos os ficheiros, com a exceção de uma única variável de nome Distrito presente em algumas folhas e noutras não. Para além das variáveis descritas, cumpre observar que os ficheiros possuíam nas suas capas ou folha inicial, o número total de utentes e de ERPIs contabilizados para o ano de 2020.

2 - Ficheiro de dados com o número total dos residentes em cada uma das ARS pelo grupo etário. Dados igualmente partilhados pela DGS e originados do Instituto Nacional de Estatística, para a população estimada de residentes em 2020. Descritos na tabela 1.2.

Variável	Descrição
ARS	Representa as regiões nacionais de saúde.
Grupo Etário	Classes de idade da população de cada ARS por intervalos de 5 anos até o limiar de 80 anos e, a partir desta, até o limiar de 120 anos.
População	População correspondente a cada ARS e o seu grupo etário.

Tabela 1.2: Variáveis dos dados referentes ao número de indivíduos residentes em cada região de saúde ou ARS estratificado por grupos etários.

3 - Ficheiros de dados originados da DGS e respetivos às infeções e óbitos ocorridos em território nacional, para o ano de 2020, com as suas variáveis de interesse descritas na tabela 1.3.

Variável	Descrição
Sexo	Sexo do utente registado como caso.
ARS	Região de Saúde a qual o caso foi registado .
Concelho	Município em que o caso foi registado.
ID	Identificador Numérico do caso.
Data de Notificação	Descreve quando o caso foi notificado.
Idade	Idade do utente na altura em que o caso foi registado.
Grupo Etário	Classe de idades a qual o caso pertence.
Data de Início dos Sintomas	Descreve quando o utente teve os primeiros sintomas.
Data do Resultado Final	Descreve quando o caso foi confirmado.
Data de Nascimento	Descreve a data de nascimento do falecido.
Data do Óbito	Descreve a data do falecimento do caso.

Tabela 1.3: Variáveis de interesse nos dados oriundos do ficheiro com a contagem de casos por COVID-19 em território nacional.

4 - Ficheiro de dados produzido com o recurso de extração *Web Scrapping* para obtenção da lista total dos dados de cada ERPI em todos os distritos de Portugal continental disponibilizados no sítio “web” da Carta Social [11] com os registos respetivos a cada lar sendo descritos pelas variáveis: Morada, Concelho, Capacidade Máxima, Nome da ERPI e Código Postal (tabela 1.4).

Variável	Descrição
Morada	Localização física de cada ERPI descrita pelo nome da rua, código postal e freguesia.
Concelho	Concelho ao qual a ERPI está localizada.
Capacidade Máxima	Número máximo de residentes em ERPI.
Nome da ERPI	Nome da Instalação.

Tabela 1.4: Variáveis dos dados obtidos no endereço eletrônico da Carta Social com o uso de Web Scraping.

5 - Ficheiro de dados igualmente facultado pela DGS com a relação de concelhos e distritos de Portugal continental com a sua respetiva região de saúde (tabela 1.5).

6 - Dados obtidos na capa do ficheiro relativo à tabela 1.1, em que constam números totais e estratificados por regiões do número de ERPIs, utentes e estimado de funcionários (tabela 1.6).

Variável	Descrição
Concelho	Todos os concelhos de Portugal continental.
Região	Região Administrativa (NUTS II).
ARS	Região de Saúde.

Tabela 1.5: Representação tabular com três variáveis de interesse nos dados geográficos em conjunto com a região de saúde correspondente.

Variável	Descrição
Data de Registo	Data de quando os dados foram coletados no formato dd/mm/aaaa.
ARS	Autoridade Regional de Saúde.
ERPI	Contabilização do número de ERPIs existentes na região.
Nº Utentes	Contabilização do número de utentes das ERPIs da região.
Nº estimado de Funcionários	Contabilização do número de funcionários a trabalharem em ERPIs da região.

Tabela 1.6: Capa de cada ficheiro de dados com o número total de ERPIs, utentes e funcionários estratificados por cada ARS

## 1.4 Definições em uso

### 1.4.1 Caso de infeção por SARS-CoV-2

Um caso de SARS-CoV-2 corresponde a uma infeção por SARS-CoV-2 confirmada laboratorialmente conforme a plataforma do Sistema Nacional de Vigilância Epidemiológica (SINAVE). Conforme a definição de caso, tanto a Reação em Cadeia da Polimerase (PCR) quanto o Teste Rápido de Antígeno (RAT) podem ser usados para fins de diagnóstico da doença. Essa definição é adotada tanto para os casos em utentes quanto na população geral.

### 1.4.2 Óbitos por COVID-19

Na população geral, os óbitos por COVID-19 correspondem aos que existem com o diagnóstico clínico ou laboratorial compatível com a doença. Na população em ERPIs, as mortes em utentes correspondem aos casos confirmados ou suspeitos da doença em instalações com surtos. Neste âmbito, todas as mortes que ocorreram em ERPIs nestas situações foram tratadas como suspeitas de COVID-19, embora nem todas pudessem corresponder como sendo decorrentes da doença.

### 1.4.3 Estabelecimento Residencial para idosos ou ERPI

Na legislação vigente, conforme [12], uma instalação deste tipo: “destina-se à habitação de pessoas com 65 ou mais anos que, por razões familiares, dependência, isolamento, solidão ou

insegurança, não podem permanecer na sua residência”. Conforme a mesma portaria aponta, exceções em relação à idade são permitidas desde que sejam justificadas. Ainda, o referido texto fornece algumas diretrizes para o funcionamento destas instalações, como o rácio entre os funcionários e utentes, serviços prestados, entre outras características.

#### **1.4.4 Funcionários de ERPI**

Tratam-se de pessoal que compõe o corpo técnico dos lares como cuidadores, enfermeiros, técnicos em geriatria, encarregados do serviço doméstico e cozinheiros.

## Capítulo 2

# Materiais e métodos

### 2.1 Pré-processamento e análise dos dados

Em todas as etapas, que serão posteriormente descritas, fez-se a utilização da aplicação RStudio e da linguagem de programação R. Foram codificadas as etapas do pré-processamento dos dados relativos aos casos em ERPIs no ano de 2020 com o uso das seguintes bibliotecas: tidyverse, readxl, tibble, openxlsx, tidyr, purr, plyr, stringi, stringdist, fuzzyjoin, zoo, emmeans. Para realização dos cálculos que serão posteriormente descritos e das representações gráficas, fez-se o uso das bibliotecas: scales, ggplot2, SciViews, forestplot, ggpubr, AICcmodavg e emmeans.

Para início do pré-processamento dos dados no ano de 2020, fez-se a unificação dos 35 ficheiros de dados num único ficheiro de dados. Após a junção de todos os dados num único conjunto, obteve-se um novo ficheiro constituído por 7351 linhas.

#### 2.1.1 Integração e Limpeza dos Dados

Na fase seguinte, realizou-se a remoção de strings inseridas em algumas colunas quantitativas, o que poderia gerar valores vazios e conseqüente perda de informação na conversão destes dados para o tipo numérico. As strings constituíam-se de anotações e observações qualitativas feitas aos valores preenchidos. A localização destes dados foi feita com a sua identificação automática dentro das variáveis quantitativas no dito ficheiro de dados e, após a exportação para o formato "xlsx", foi feita a sua remoção manual com o Microsoft Excel. Em seguida, foram seleccionadas apenas as linhas sem valores vazios em algumas das colunas de interesse, que constituíram um total de **7064** linhas. Conforme a tabela 1.1, são as variáveis indispensáveis para as linhas:

- **Nome/Identificação;**
- **Datas:** representam o dia em que o registo do número de casos foi efetuado, sendo ao todo 35 datas de registo extraídas do conjunto dos 35 ficheiros de dados;

- **ARS;**
- **Casos confirmados em utentes:** aqui representados como sendo a prevalência da doença nos utentes, com importância central para a posterior extração da incidência através do cálculo desta variável em razão dos dias decorridos visando a constituição do risco relativo e igualmente para o cálculo do número total de infetados visando a obtenção do tamanho do surto ocorrido em utentes por cada lar;
- **Casos confirmados em funcionários:** tal como os casos confirmados em utentes, essa variável representa-se como sendo a prevalência da doença nos funcionários e é também indispensável para a posterior extração do número total de infetados para o cálculo do tamanho do surto entre os indivíduos deste grupo por cada lar;
- **Óbitos:** trata-se do registo cumulativo das mortes ocorridas em cada ERPI. De importância central para efetuar-se o objetivo da geração do Risco Relativo para este indicador;

Uma vez concluída a unificação e seleção das linhas, procedeu-se a utilização das funções e técnicas de *Record Linkage* em conjunto com as correções manuais necessárias para a normalização das variáveis e das entidades respectivas as ERPIs. Para esta etapa, foram enumeradas as seguintes tarefas abaixo.

1. Utilização da função *stringdist* em conjunto da métrica de semelhança de Jaro-Winkler para aproximação de *strings*, com um parâmetro de valor 0.09 dentro de um intervalo entre 0.0 (nenhuma similaridade) e 1 (correspondência exata) com o objetivo de atribuir um identificador numérico.
2. Utilização da função *group\_indices* da biblioteca *dplyr*, que gerou novos identificadores numéricos tendo como argumentos a recém-criada variável identificadora do tipo numérico e a variável qualitativa do Concelho de cada entidade lar.
3. Ordenação das linhas por meio da função *arrange* da supracitada biblioteca *dplyr* com os argumentos definidos pelas variáveis Concelho e ARS.
4. Exportação deste novo conjunto de dados para o formato *xlsx* sendo feita em seguida a correção manual com o auxílio da aplicação *Microsoft Excel*.

Na primeira tarefa o *output* esperado foi uma variável identificadora do tipo numérico para cada entidade representativa de uma ERPI aproximada pela variável respetiva ao seu nome de identificação. No segunda, justificou-se uma nova geração da dita variável para uma melhor especificidade e redução no número de erros, pois foi atribuído novo identificador para cada entidade não só pela similaridade dos seus nomes de registo, como também pelos concelhos em

que estavam localizados. Na terceira tarefa, o objetivo foi o de permitir ordenar os dados pela municipalidade e região de cada lar. Com a ordenação concluída, prosseguiu-se com a exportação para um ficheiro do tipo "xlsx", como descreveu a quarta tarefa. Este último passo foi objetivado pela verificação de possíveis falhas no processo de geração do identificador para cada entidade. Com o ficheiro exportado, procurou-se normalizar os nomes existentes para cada entidade lar num único nome e igualmente modificar os números dos diferentes identificadores de um mesmo lar com problemas na aproximação para um único identificador numérico.

Foram igualmente corrigidos outras variáveis qualitativas das ERPIs respectivas à sua localização física que continham valores vazios e variações de nomes. De igual modo, foram alterados os valores da variável de ocupação de cada lar, para o seu valor inicial do registo da primeira data de cada grupo de ERPIs.

Em seguida, foram removidas as linhas com os registos de prevalência repetidos durante várias datas. Para estes casos, considerou-se a premissa de serem erros de imputação e/ou reportes repetidos. Portanto, para cada grupo representativo dos registos de um lar nessas condições, manteve-se apenas a primeira entrada de cada sequência repetida. Foram igualmente removidas as linhas que continham o valor zero ou valores vazios em todos os registos do número de funcionários. Finalmente, foram excluídos do presente estudo as ERPIs sem registos de casos nos utentes, ou seja, com casos somente nos cuidadores ou funcionários, mas sem registo nos utentes. Após a integração, foi feita ainda uma filtragem dos registos de ERPIs com rácio de funcionários por utentes, maior do que dois, já que para estes casos haveria um número muito maior de funcionários do que de utentes o que foi assumido como sendo erros de imputação nos registos.

Para o número de óbitos, conforme já antes mencionado, registou-se um valor cumulativo. Esse valor acumulava os números de óbitos entre a primeira e a última data de registo em cada lar. Cinco ERPIs que não tiveram o registo crescente na variável dos óbitos, somados a mais nove destas instalações que registaram um número de óbitos maior do que o número de casos durante o processamento do tamanho do surto nos utentes (total de casos registados em cada lar) foram descartados por terem erros de imputação. Portanto, para obtenção da variável respetiva aos óbitos tomou-se o último valor cumulativo de cada lar.

### 2.1.2 Transformação dos Dados

Na próxima etapa, o objetivo centrou-se em transformar os dados com o cálculo dos valores existentes no campo da prevalência contabilizada das ERPIs na altura da data de cada registo, para a obtenção dos valores do tamanho do surto e da incidência diária. Para este efeito, em primeiro lugar foi classificada a ordem do surto de cada lar. Em razão de que um mesmo lar poderia ter mais de um surto em épocas distintas do ano, foi calculado e comparado o número de dias decorridos entre um registo da prevalência com o seu registo posterior, dentro de cada grupo

de linhas representativas de um lar. Aquando a diferença de dias entre dois registos corresponde a um número menor ou igual a 28 dias, classificou-se como sendo o mesmo surto. Quando o número de dias entre os registos era maior que este limiar, foi classificada esta diferença como um surto posterior e assim em diante. Cumpre observar que o limiar de até 28 dias foi escolhido de forma empírica com base na distribuição observada nos dados, que indicaram que a maior parte dos acréscimos ou novos casos registados ocorreu com menos de um mês. Portanto, este intervalo foi definido para maior precisão na identificação de diferentes surtos, o que resultou no armazenamento qualitativo numa variável que passou a ser utilizada para a identificação de cada surto.

A nova variável, representativa do número do surto, passou a ser tratada como o identificador numérico dos seus respectivos lares. Portanto, as linhas pertencentes a mais de um surto num mesmo lar foram consideradas outras entidades para todas as operações, com a exceção do cálculo relativo ao número de óbitos. Para a obtenção da incidência no grupo dos utentes e dos funcionários, foram contabilizados os registos dos surtos até o seu último máximo (valor no qual houve um último incremento na prevalência em relação ao registo anterior) nas variáveis representativas dos casos ou prevalência nos dois grupos. Para alcance da obtenção do tamanho do surto e, posteriormente, da incidência temporal em utentes, foi feito o cálculo desses indicadores separadamente. Em relação aos funcionários, fez-se apenas o cálculo da obtenção do tamanho do surto. Portanto, foram feitos três procedimentos distintos no mesmo conjunto de dados, o que resultou em três tabelas distintas: as duas primeiras em relação aos números inerentes ao tamanho do surto para os dois grupos presentes nas ERPIs e, em seguida, a incidência temporal obtida somente para o grupo dos utentes. Considerou-se o último máximo na variável de casos, ou seja, o valor após o último incremento na prevalência, correspondendo os posteriores decrementos já à extinção do surto no lar em questão. Para obtenção do número de casos e posteriormente o tamanho do surto tanto em utentes como em funcionários, repetiu-se o mesmo procedimento. Em seguida, fez-se para cada um dos dois conjuntos de dados a soma do valor do último máximo com os decréscimos (se existirem e até o último máximo de cada variável). Com isso, obteve-se o número total de casos ocorridos.

Conforme a tabela 2.1 indica, no exemplo de um lar na ARSALG (ARS Algarve) com um único surto registado, houve um incremento de 6 para 14 pessoas infetadas na primeira data de 8 de outubro para a segunda de 18 de outubro. Foi assumido o decréscimo na data seguinte de 29 de outubro e novamente no dia 3 de novembro como a recuperação de utentes da doença. Do dia 03 ao dia 12 de novembro não houve nenhum aumento ou diminuição no número de utentes infetados e por essa razão foi sendo assumida a não existência de novos casos ou recuperações entre as duas datas. No entanto, no dia 19 de novembro, a prevalência do número de utentes infetados volta a subir de 9 para 10, o que indica que uma pessoa foi infetada porquanto, o surto ainda estava ativo.

Na última linha, correspondente ao dia 24 de novembro, foi contabilizado um decréscimo

Data	N.º de Utentes	Casos (Utentes)	Decréscimos (Utentes)	N.º de STAFF	Casos em STAFF	Decréscimos (STAFF)
08/10/2020	36	6	0	26	7	0
18/10/2020	36	14	0	26	11	0
29/10/2020	36	13	1	26	11	0
03/11/2020	36	9	4	26	5	6
12/11/2020	36	9	0	26	3	2
19/11/2020	36	10	0	26	1	2
24/11/2020	36	8	2	26	0	1

Tabela 2.1: Representação do cálculo para estimação da incidência com a seleção de sete variáveis, incluindo a exibição dos decréscimos da prevalência entre utentes e funcionários, para obtenção do número de casos nestes dois grupos.

em relação ao registo na data anterior. Isso indica que, por razão de não constar mais nenhum outro aumento no registo após a dita data, o fim do surto registou-se na data anterior de 19 de novembro. Portanto, para o surto no lar em questão, foram consideradas somente as linhas com as datas entre 8 de outubro e 19 de novembro, sendo descartado o registo com a última data. A operação resultou no número total de 15 pessoas infetadas, que representa o último máximo (10) em conjunto com os decréscimos dos registos anteriores (5). Na análise respetiva aos casos registados em funcionários, observou-se que o único incremento foi o de 7 para 11 casos logo na segunda data. Não foram considerados os casos e decréscimos nas datas dos registos seguintes, dado que não registaram-se novos casos entre funcionários, o que totalizou unicamente no registo de 11 funcionários infetados para o dito surto.

Foi igualmente possível atribuir a cada registo inicial de um surto o seu período respetivo do ano de 2020, sendo feita a classificação em dois períodos daquele ano correspondentes aos meses anteriores e posteriores ao início da segunda vaga: abril a agosto e setembro a dezembro.

Em seguida, para obtenção da incidência temporal em utentes, foi feito separadamente para cada grupo de ERPIs e os seus respetivos surtos a operação de preenchimento sequenciado dos dias entre a primeira data e a data do último máximo dos registos da prevalência no grupo dos utentes. Para a obtenção destes números para o grupo dos utentes, foi feita a divisão do número de casos pelo número de dias decorridos para cada grupo de lar e o seu respetivo surto. A dita operação teve como objetivo a distribuição da contribuição diária do número de casos em utentes, calculada por cada lar com os seus respetivos surtos. No exemplo da tabela anterior (2.1) o cálculo entre o dia 8 de outubro (início do surto) e 19 de novembro (data do último máximo) confere o número de 42 dias decorridos para o surto em questão. A seleção do último máximo com os decréscimos anteriores e a sua divisão pelo número de dias contados teve como objetivo a distribuição da incidência aproximada da doença nos utentes para os dias correspondentes a cada surto nas ERPIs.

Uma vez distribuída a incidência dos utentes por dias, foi possível efetuar a agregação dos

valores para outros formatos temporais. Optou-se primeiro pela agregação no formato de meses, em que as datas existentes para todos os surtos foram transformadas com a função *month* da biblioteca *lubridate* e sumarizadas com o uso da função *summarise* da biblioteca *dplyr* para uma variável representativa de cada mês e em conjunto com os valores da incidência (arredondados com a função *round*) no grupo dos utentes agregados pelo novo formato.

Criou-se, portanto, duas relações: a primeira com os tamanhos dos surtos ocorridos nas ERPIs para ser utilizada nas análises dos números de casos nas classes dos utentes e funcionários em função do tamanho das ERPIs nos dois períodos do ano; e a segunda com a incidência temporal para ser usada nas análises do RR entre as regiões.

### 2.1.3 Agregação de regiões e estratificação do número de ERPIs e de utentes por tamanhos das ERPIs

Para todas as análises que serão tratadas posteriormente, fez-se, além da divisão por regiões e períodos do ano, a agregação de regiões (como adiante se verá o motivo) e a estratificação do número de ERPIs e de utentes por classes de tamanhos. Foi mantida a região correspondente à Autoridade Regional de Saúde Lisboa e Vale-do-Tejo (ARSLVT) para ser comparada com as três regiões correspondentes às Autoridades Regionais de saúde do Algarve (ARSALG), Centro (ARSC) e Norte (ARSN), com a exceção da Autoridade Regional de Saúde do Alentejo (ARSA), que não entrou para este tipo de análise dado o baixo número de casos observados ao longo de 2020. A agregação consistiu em categorizar todos as ERPIs das três regiões correspondentes numa única classe regional para posteriormente servir na sumarização das variáveis representativas do número de utentes, números de casos e de óbitos em utentes, assim como o número de funcionários e o número de casos em funcionários.

Além da agregação, fez-se igualmente a estratificação em cada classe de tamanhos das instalações (divididas entre pequenas, médias e grandes). Vale destacar que para este tipo de análise não encontrou-se na literatura um *cut-off* padrão para categorização das instalações pelo número de utentes, com os intervalos a variarem de país para país e também conforme a questão alvo de cada investigação. Isso pode ser exemplificado num estudo feito nos Estados Unidos conforme [7] no qual os autores utilizaram a contagem de até 50 utentes como classificação para as instalações pequenas, 50 a 150 para as médias e mais de 150 para as grandes. Os mesmos números para categorias de tamanhos foram adotados para a estratificação presente num estudo que avaliou a associação entre as características dos lares de idosos nos Estados Unidos com a prevalência da doença nos seus funcionários ([13]).

Já num outro estudo feito na Austrália, conforme [14], foram tomados por categorias os números de até 50 utentes para lares pequenos, 51 até 100 para lares médios e maiores que 101 para grandes. Outros estudos adotaram números menores para este fim, como é o caso de [15] que avaliou a associação do excesso de risco de mortes nos lares de idosos do Reino Unido com as

características destes lares e que tiveram por estratificação de tamanho os lares pequenos ( $\leq 23$ ), médios (24-40) e grandes ( $> 40$ ). Outro caso pode ser exemplificado num estudo canadense em [16], que fez a estratificação das instalações estudadas por classes de intervalos de 10 utentes, com a primeira categoria sendo das instalações com menos de 20 utentes ( $< 20$ ) e a última com uma contagem maior ou igual a 90 utentes ( $\geq 90$ ). Foi igualmente encontrado um estudo que fez a estratificação em somente duas categorias, como é o caso de [17] que dividiu o tamanho dos lares pelo número de utentes em pequenos-médios ( $\leq 70$ ) e grandes ( $> 70$ ).

Portanto, no presente estudo, o critério adotado para escolha do *cut-off* para estratificação foi feito com base na distribuição das ERPIs em três categorias, de modo que nenhuma destas categorias ficasse com um número muito baixo de ERPIs em sua composição.

Para isso, foram utilizados como métrica os intervalos de: 1 a 25 utentes para ERPIs pequenas, 25 a 60 utentes para médios e mais de 60 utentes para grandes. Também foi feita a estimação do número de ERPIs e de utentes distribuídos pelas categorias de tamanhos pequeno, médio e grande presentes no ano de 2020, a contar inclusive os números respetivos as ERPIs que não tiveram ou registaram surtos. Para esta ultima operação, fez-se primeiramente para os dados obtidos no endereço eletrônico da Carta Social (conforme a tabela 1.4) a mesma estratificação por tamanhos respetivos aos números de utentes em cada uma das regiões já mencionadas. Cumpre ainda observar que para categorizar os dados obtidos da Carta Social em regiões de saúde, fez-se o uso do ficheiro “linkage”(vide a tabela 1.5), igualmente fornecido pela DGS, para integração com o uso da função *left-join* destes dois conjuntos através das colunas respetivas ao concelho e ao distrito, a resultar num novo conjunto de dados com a coluna da região de saúde correspondente.

Após o processo de categorização das classes de tamanhos e regiões no conjunto total de ERPIs, o próximo passo foi o de obter a proporção de instalações pequenas, médias e grandes dentro do universo de cada classe regional. O mesmo foi feito com os números de utentes dentro das ditas classes. Uma vez obtidas as ditas proporções, o seguinte passo foi o de multiplicar esta proporção pelo número total de ERPIs e utentes conforme os números indicados na capa dos ficheiros de dados descritos na tabela 1.1. Com o uso das funções *round* e *ceiling* presentes na base da linguagem de programação R, foi possível converter os números resultantes da operação com casas decimais em números inteiros para posterior conferência do seu somatório pelas três classes de tamanho, com o número total de utentes correspondente em cada região. Após a conferência dos valores obtidos, usou-se o mesmo tipo de agregação previamente descrita, para a sumarização destes dados pelas duas classes regionais que correspondem à ARSLVT e as regiões de saúde ARSALG, ARSC e ARSN.

## 2.2 Cálculo do RR

Para fins de comparabilidade entre os números de casos obtidos nas instalações do tipo ERPI com os números da população idosa em geral, foi feito o cálculo do RR para o *cut-off* de 75 anos. A escolha deste limiar foi feita tendo em conta a ausência da observação de diferenças relevantes para a mesma análise com outros tipos de *cut-off* a partir dos 65 anos (idade na qual os idosos geralmente estão aptos a ingressar em ERPIS). Porém, dado que a população nestas instalações tende a ter uma média superior de idade, como observado no gráfico de distribuição de idades no relatório da Carta Social em [18], optou-se por escolher o limiar de idade mais próximo da média reportada no gráfico (aproximadamente 80 anos) que fosse possível de fazer os cálculos. Devido ao *cut-off* de 80 ou mais anos resultar num número de casos em ERPIS superior ao reportado na população idosa sem incluir as ditas instalações, o limiar de 75 anos ficou escolhido para este fim.

Portanto, para calcular corretamente o RR foi necessário recorrer ao conjunto de dados de todos os casos em território nacional e filtrar somente os casos de maiores de 75 anos por cada ARS do continente. Em seguida, fez-se a agregação do número de casos, oriundos desses dados, para os formatos temporais de meses e os dois períodos do ano de 2020. Uma vez agregadas e transformadas, as colunas foram integradas ao já existente ficheiro de dados “relativo às ERPIS. A mesma tarefa foi realizada igualmente pela agregação das regiões em duas classes regionais, como consta na subsecção anterior.

Finalmente, para efetuar o cálculo do RR, foi antes calculado o risco da incidência nas ERPIS ( $RL$ ) a dividindo o número de casos das ERPIS ( $I_1$ ) pelo número total de utentes ( $N_1$ ). Para que o mesmo princípio fosse aplicado ao risco da incidência na população idosa em geral ( $RG$ ) fez-se, em primeiro lugar, a subtração do número de casos em ERPI do número de casos no conjunto total dos idosos. Com isso, foi possível obter a incidência fora das ERPIS ( $I_2$ ). De modo homólogo, efetuou-se a subtração do número de residentes em ERPI do número total de pessoas com mais de 75 anos em cada ARS, que resultou no número de idosos fora das ERPIS ( $N_2$ ) em cada ARS elencada. Em seguida, fez-se a divisão do  $I_2$  por  $N_2$  obtendo-se o  $RG$ . No final obtemos:

$$RL = \frac{I_1}{N_1} \quad (2.1)$$

$$RG = \frac{I_2}{N_2} \quad (2.2)$$

$$RR = \frac{RL}{RG} \quad (2.3)$$

O cálculo do RR, conforme as equações 2.1, 2.2 e 2.3, foi feito em função do risco de incidência para as ERPIs de uma ARS, a dividir pelo risco de incidência para a população idosa em geral na dita ARS. Por fim, foi realizada a estimação do IC (Intervalo de Confiança) a 95% para cada valor obtido de cada RR. Para isso, foram aplicadas as fórmulas demonstradas nas equações 2.4 que refere-se ao erro padrão do logaritmo ( $e(\ln RR)$ ), 2.5 ao limite inferior ( $Lim_{inf}$ ) do intervalo e 2.6 ao limite superior ( $Lim_{sup}$ ) da estimação intervalar 19.

$$e(\ln RR) = \sqrt{\frac{1}{I_1} - \frac{1}{N_1} + \frac{1}{I_2} - \frac{1}{N_2}} \quad (2.4)$$

$$Lim_{inf} = e^{\ln RR - 1.96 e(\ln RR)} \quad (2.5)$$

$$Lim_{sup} = e^{\ln RR + 1.96 e(\ln RR)} \quad (2.6)$$

Uma vez obtido o número total dos óbitos ocorrido em cada lar, como sendo o último valor cumulativo, foi agregado este número para cada ARS ou classes regionais (para as análises com regiões agregadas). Prosseguiu-se, então, ao cálculo do RR com o mesmo princípio do método aplicado para o ficheiro de dados da incidência. Foi realizada a filtragem dos óbitos em maiores de 75 anos no conjunto de dados respeitante aos óbitos por COVID-19 em território continental e feita a contagem e agregação do número de óbitos por classes de regiões numa nova variável e, posteriormente, a integração desta em conjunto com a variável de identificação das regiões de saúde correspondentes, para todo o ano de 2020 ao novo ficheiro de dados das ERPIs. O mesmo procedimento foi repetido para a estratificação em dois períodos do ano. Por fim, aplicaram-se as fórmulas definidas nas equações 2.1 a 2.6.

### 2.3 Análise do OR do risco de um lar apresentar pelo menos um caso de COVID-19

Para comparar o risco de um lar apresentar um ou mais casos em 2020, primeiro foi feito o uso dos dados estratificados que foram originados da Carta Social, conforme previamente indicado nas tabelas 3.2 e 1.4. Procurou-se, então, obter a estimação do OR de um lar ter pelo menos um caso da doença, com o uso de regressão logística pela função glm (*Generalized Linear*

*Model*) em conjunto com as funções *emmeans* e *pairs* da biblioteca *emmeans* que serviram para exponenciação dos coeficientes e a sua transformação no dito OR.

Para as 2526 instalações do tipo ERPI, fez-se a remoção dos registos pertencentes à região ARSA, o que resultou no total de 2276 ERPIs. O próximo passo foi o de fazer a contagem, numa coluna, das ERPIs que tiveram surtos em relação ao número de total destas instalações. Para isso, foi contabilizado numa coluna, a partir do conjunto de dados respetivos ao tamanho do surto, o número de ERPIs que reportaram casos dentro das classes de tamanho indicadas na tabela 3.2 e posteriormente fez-se a integração desta coluna no conjunto anterior. Portanto, ficou constituído o ficheiro de dados a ser trabalhado com duas colunas: uma com o número total de ERPIs e a outra com o número de ERPIs que reportaram surtos.

Deu-se o seguimento no processo com a criação de uma nova variável do tipo binária para a classificação das ERPIs que tiveram casos e igualmente para os que não tiveram nenhum caso. A dita variável foi definida como sendo dependente enquanto as variáveis independentes foram as mesmas elencadas no processo acima respetivas ao tamanho das ERPIs e as classes de regiões. Para a seleção de um modelo que melhor se ajustasse aos dados, foram analisados quatro modelos construídos *a priori* e provenientes das possibilidades de combinações entre as variáveis independentes. Para este efeito, foi ajustado para o primeiro modelo somente a região como variável independente, dado que cada região manteve a sua autonomia e diferenciação em termos de recursos para o combate à pandemia, o que gerou diferentes tipos de engajamentos e resultados nos impactos, para o segundo modelo a variável respetiva as categorias dos tamanhos das ERPIs foi escolhida como o único preditor, dado que a dimensão destes poderia influenciar nos riscos de ocorrência de surtos;

Nos dois últimos modelos considerou-se, respetivamente, a combinação das duas variáveis (terceiro modelo) e a interação das duas variáveis ou o efeito que ambas exercem de forma mutualmente condicional na variável dependente (quarto modelo).

Para a seleção do melhor modelo, foi feito o uso da métrica *Akaike Information Criterion* (AIC) na qual o quanto menor for a pontuação da métrica para um modelo, melhor será a sua capacidade de descrever os dados. Para cada modelo, o AIC foi calculado da seguinte forma [20]:

$$AIC = -2 \ln \hat{L} + 2K \quad (2.7)$$

Onde na equação 2.7, a variável  $\hat{L}$  representa a estimação da máxima-verosimilhança num modelo e a  $K$  representa o número de parâmetros nele definidos. O cálculo foi feito com o auxílio da função *aictab* da supracitada biblioteca *AICcmodavg*. Foram também calculados os valores do peso do Akaike para cada modelo. Estes valores representam em termos de proporção o poder preditivo de cada modelo sendo obtidos com o uso da equação 2.8 conforme [21]:

$$W_k = \frac{\exp\left(\frac{-1}{2}\Delta_k\right)}{\sum_{r=1}^R \exp\left(\frac{-1}{2}\Delta_r\right)} \quad (2.8)$$

Onde no numerador a variável  $\Delta_k$  representa a subtração entre a pontuação do AIC no modelo vigente e a do melhor modelo (com o menor AIC pontuado). O dito numerador representa o valor da estimativa da verosimilhança relativa para o mesmo  $\Delta_k$ . Já o denominador representa o somatório destes mesmos valores para todos os modelos em questão.

Com a escolha do melhor modelo, deu-se prosseguimento a regressão logística ajustada às variáveis selecionadas para obtenção do OR.

## 2.4 Análise do tamanho médio do surto

Para obtenção do tamanho médio do surto (média da proporção de casos nos utentes de uma ERPI) fez-se uma regressão linear na qual a variável resposta do tamanho do surto (contagem de casos / contagem total de utentes em cada ERPI) teve as suas observações, assim como quatro modelos resultantes das combinações entre as variáveis preditoras do tamanho do lar e região de saúde. Foram construídos quatro modelos, o primeiro com somente a variável região como variável preditora, o segundo com somente a variável do tamanho do lar, o terceiro com a combinação das duas variáveis e o quarto com o efeito da interação entre ambas. Para seleção do melhor modelo, fez-se o uso da já tratada métrica AIC com as equações 2.7 e 2.8 da subsecção anterior. O conjunto de dados utilizados neste processo são respetivos apenas às ERPIs que reportaram surtos.

## Capítulo 3

# Resultados

### 3.1 População total do estudo em 2020

Verificou-se, em princípio, nos dados reportados de 2020, um número total de 99.234 utentes distribuídos por 2526 instalações do tipo ERPI nas cinco ARS, tal como se observa na figura 3.1:

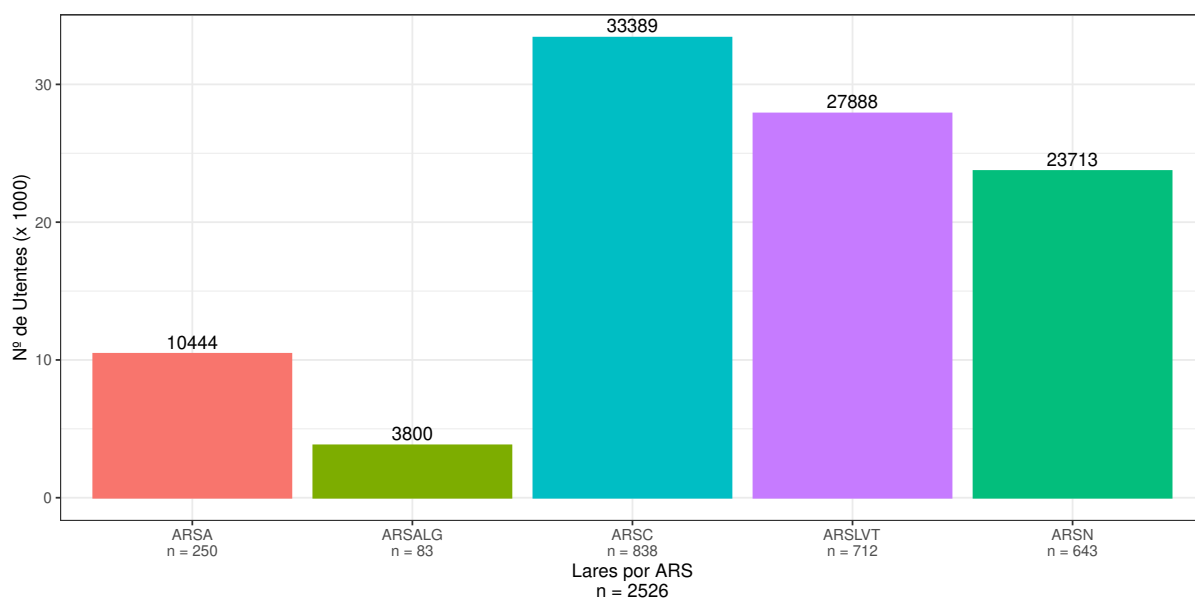


Figura 3.1: Distribuição dos utentes e ERPIs em números na sua totalidade por cada ARS.

De igual modo, foi observado um número total de 60991 funcionários destas ERPIs distribuídos nas cinco regiões conforme a figura 3.2.

Os números da dita figura 3.2 e da 3.1 correspondem a população total de utentes e funcioná-

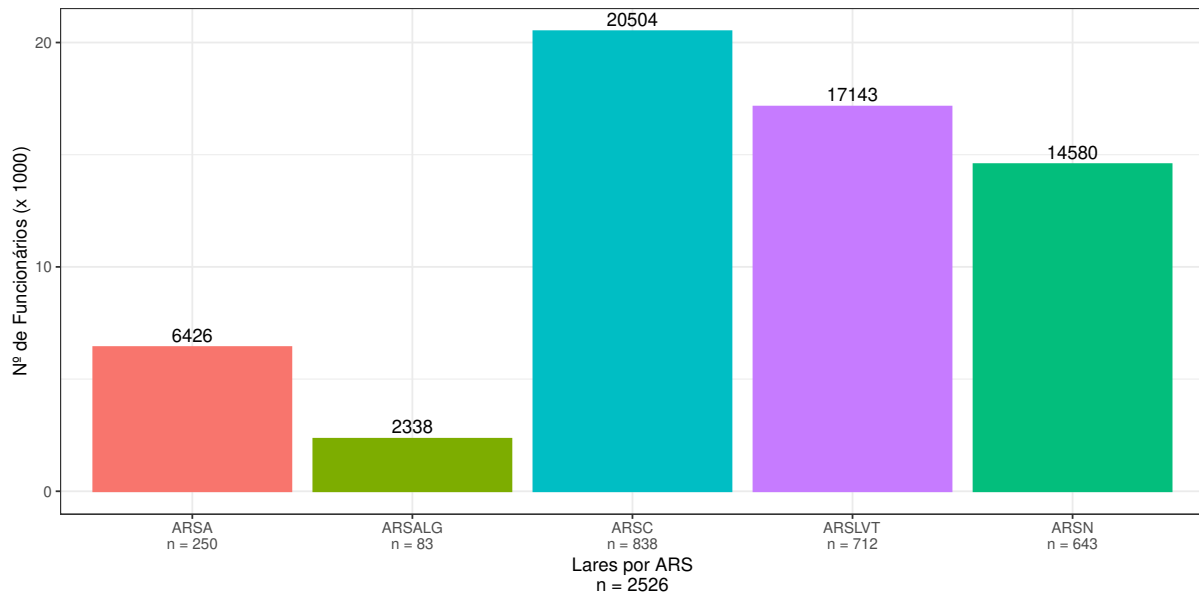


Figura 3.2: Distribuição dos funcionários e ERPIs em números na sua totalidade por cada ARS.

rios com as ERPIs instalados por todo o continente e tiveram como origem a folha da capa dos ficheiros de dados de surtos em ERPIs de idosos (ver dados e materiais). O processo de seleção dos dados das ERPIs que reportaram casos resultou nas seguintes composições de utentes e de funcionários nas ERPIs, como constam nas figuras 3.3 e 3.4.

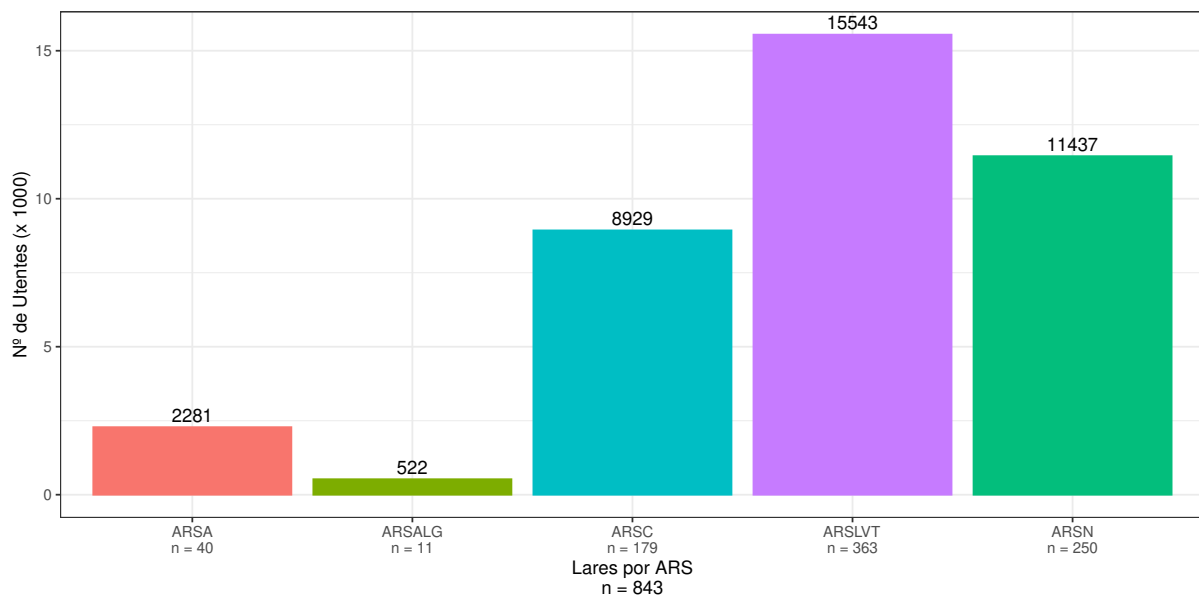


Figura 3.3: Distribuição dos utentes e ERPIs nos dados com reporte de surtos.

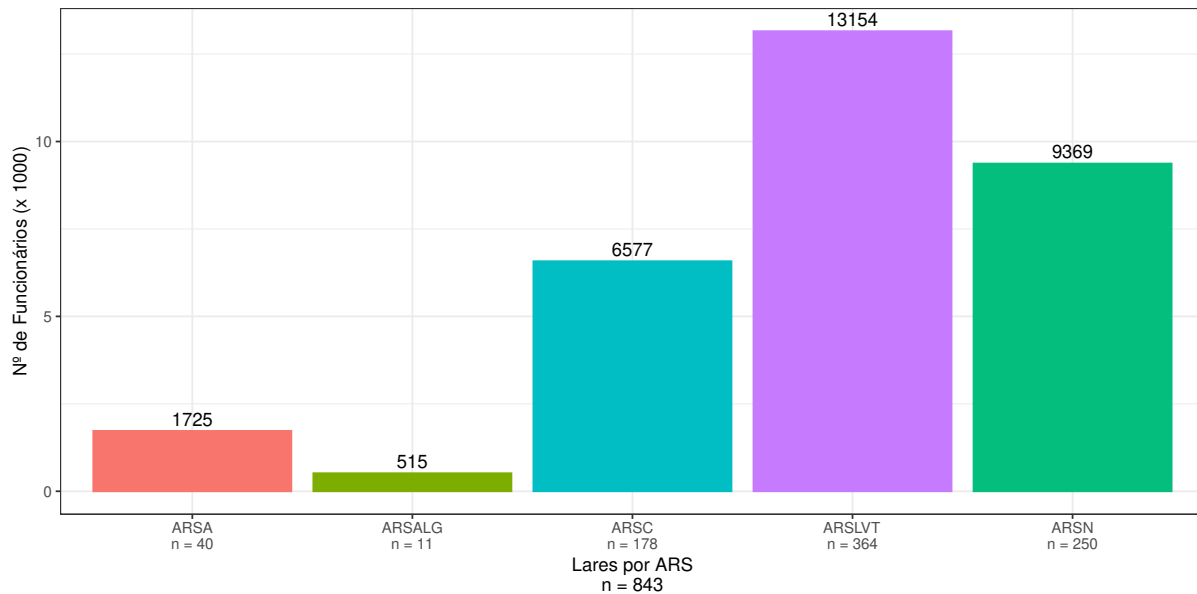


Figura 3.4: Distribuição dos funcionários e ERPIs nos dados com reporte de surtos.

A figura 3.5 mostra a comparação, em cada região, da dimensão média das ERPIs que reportaram surtos com a dimensão média do conjunto de todas as ERPIs.

As ERPIs de tamanho grande contribuíram mais do que proporcionalmente para a carga de infecção, o que evidencia que, no conjunto dos dados, o tamanho das instituições influencia o risco de infecção dos seus utentes.

A tabela 3.1 mostra a incidência em ERPIs por região e por mês desde o início dos registos até ao fim de 2020:

Regiões	Abr.	Mai.	Jun.	Jul.	Ago.	Set.	Out.	Nov.	Dez.
ARSA	6	3	61	19	0	45	129	149	393
ARSALG	4	14	0	0	20	1	47	30	12
ARSC	649	62	15	2	0	16	285	623	1377
ARSLVT	319	363	283	320	271	363	777	1880	1416
ARSN	1301	396	118	22	34	178	485	1037	1052
<b>Total</b>	<b>2279</b>	<b>838</b>	<b>477</b>	<b>363</b>	<b>325</b>	<b>603</b>	<b>1723</b>	<b>3719</b>	<b>4250</b>

Tabela 3.1: Incidência calculada da COVID-19 em utentes dos lares para os meses de abril a dezembro de 2020 em cada ARS.

Na figura 3.6 estão representadas, no mesmo período, a incidência em ERPIs no conjunto das regiões continentais e a incidência na população idosa não institucionalizada.

O gráfico demonstra um agravamento na curva epidémica a partir do fim de agosto, com o início da segunda vaga em território nacional. A incidência mensal na população idosa não institucionalizada passou de 756 casos em agosto para 1779 em setembro, um aumento de cerca

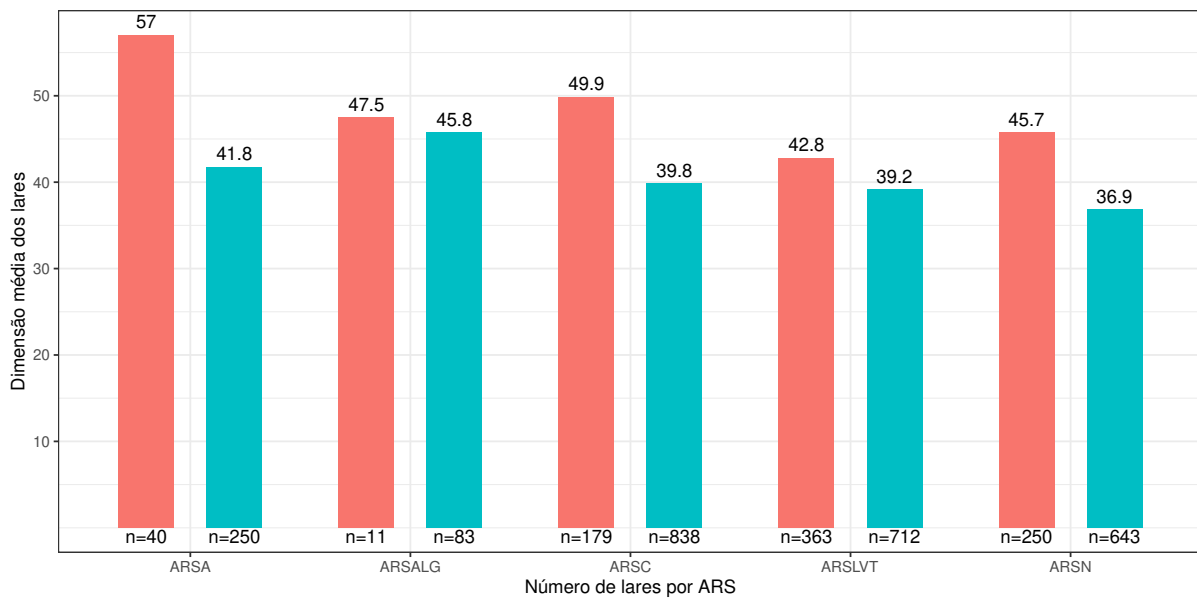


Figura 3.5: Comparação da dimensão média das ERPIs reportados com surtos (em vermelho) e do total de ERPIs existentes (em azul) no ano de 2020.

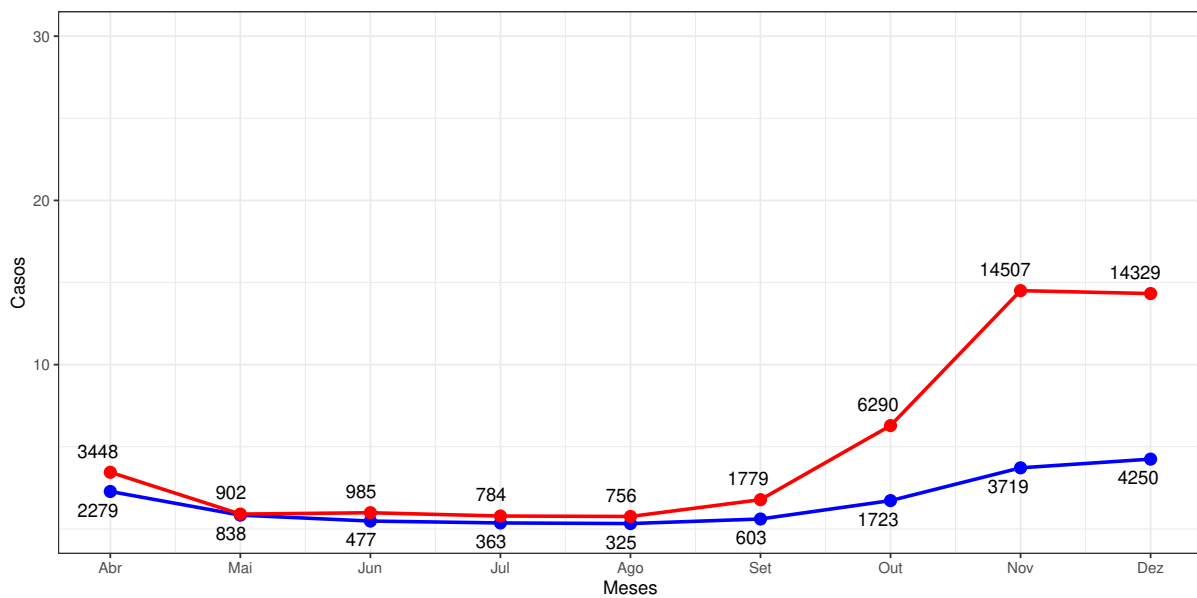


Figura 3.6: Curvas de Incidência da COVID-19 no ano de 2020 para os utentes das ERPIs (azul) em comparação com todos os idosos de Portugal continental (vermelho).

de 135%. Para os casos em instalações ERPI, foi observado um aumento de 325 casos em agosto para 603 casos em setembro, um incremento de cerca de 85.5% entre os dois meses. O número de casos seguiu em alta até o mês de novembro, altura que a segunda vaga atingiu o pico. Apesar de o primeiro período ter tido uma grande contribuição das ERPIs, o que aproximou as duas curvas de incidência referentes as duas populações, o segundo período em que foram registados mais casos também contou com um maior distanciamento entre as duas curvas. Portanto, os casos ocorridos no segundo período tiveram uma menor contribuição dos casos ocorridos em ERPIs, revelando uma maior eficácia neste período na proteção aos utentes destas.

## 3.2 Análise dos surtos ocorridos em ERPI no ano de 2020

### 3.2.1 Análise do RR em ERPIs por região, por período e por tamanho

#### RR da incidência

Nesta secção serão abordados os padrões observados nas cinco regiões de saúde relativamente ao RR dos casos e óbitos nelas registados. Neste contexto, para o ano de 2020, obteve-se o RR com o seu respetivo IC (Intervalo de Confiança) para a incidência da COVID-19 em utentes de instalações do tipo ERPI em comparação com a incidência da doença na população de maiores de 75 anos de cada região.

Conforme a figura 3.7, é possível observar o RR por todo o ano de 2020 nas cinco regiões.

A região com maior número de casos em ERPI foi a de Lisboa e Vale do Tejo, com 5992 infeções ocorridas naqueles ambientes e com um RR estimado em 10.5 (IC 95%: 10.17–10.83) sendo o maior de todos os observados. A ARSN, segunda região com o maior número de casos absolutos, teve um RR de 4.07 (IC 95%: 3.95–4.2). A região com terceiro maior RR foi a ARSC (RR: 3.92; IC 95%: 3.75–4.1). A ARSA teve um RR de 4.05 (IC 95%: 3.7–4.43) e a ARSALG, a menor em número total de casos em ERPI, teve um RR de 3.93 (IC 95%: 3.23–4.79). Cumpre observar que a ARSA, devido ao baixo número de casos no primeiro período, foi excluída das análises seguintes.

Para fins de especificidade do RR nas regiões de saúde por períodos pré e pós-segunda vaga, fez-se uma divisória em dois períodos, com o primeiro período a corresponder de abril até agosto e o segundo de setembro até dezembro, como descreve a figura 3.8.

A estimação do RR nas regiões de saúde ARSC e ARSN resultou num número bastante acrescido nos cinco primeiros meses da pandemia. Entretanto, o RR sofreu uma acentuada queda nos meses seguintes. A ARSALG teve um RR estimado em 14.39 (IC 95%: 8.97–23.1) no primeiro período e 3.01 (IC 95%: 2.39–3.79) no segundo, além de uma margem de erro acrescida naquela primeira vaga. O valor do RR na região ARSLVT também revelou uma redução de 13.34 (IC 95%: 12.46–14.3) para 9.77 (IC 95%: 9.42–10.13) após a segunda vaga, porém não

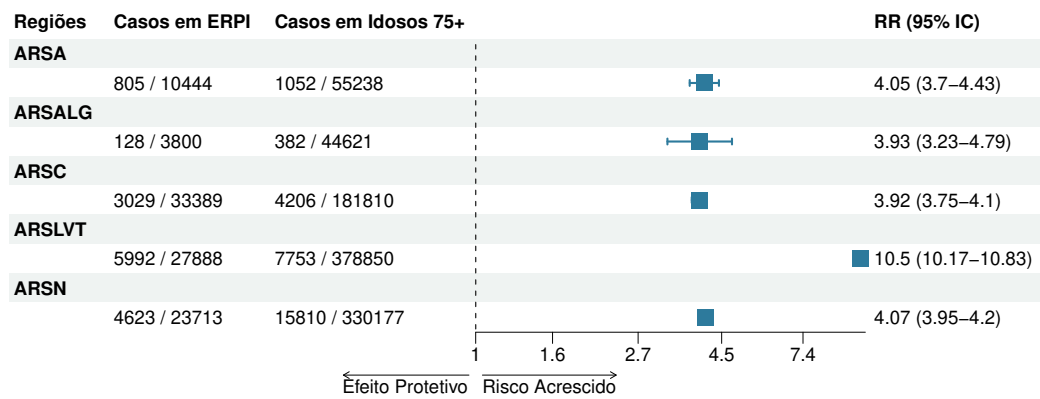


Figura 3.7: RR para a incidência em ERPIs nas regiões de saúde no ano de 2020.

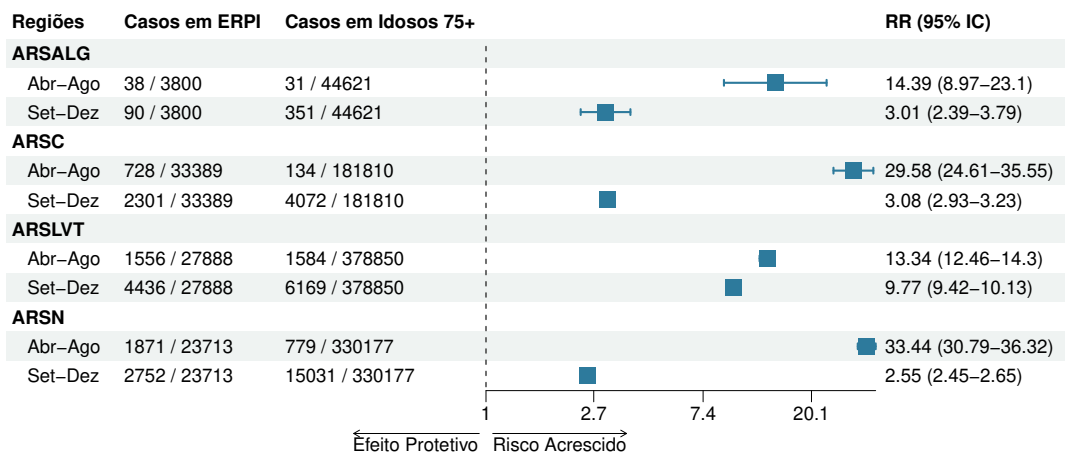


Figura 3.8: RR para a incidência nas regiões de saúde (com exceção da ARSA) até agosto e após agosto.

tão expressiva quanto as regiões anteriores. Dado o padrão observado nas regiões ARSALG, ARSC e ARSN, porém não na região ARSLVT, faremos em secções posteriores e a seguir nesta mesma secção uma análise regional comparando a região ARSLVT por um lado, com o conjunto das regiões ARSN, ARSC e ARSALG, por outro.

Em relação à estratificação dos números de utentes nas ERPIs por tamanhos das instalações, a dita operação resultou nos seguintes números conforme a tabela 3.2:

Regiões de Saúde	Tamanhos	Lares em 2020	Utentes em 2020
ARSALG, ARSC e ARSN	1-25 utentes	508	8593
	25-60 utentes	829	33000
	60+ utentes	227	19309
ARSLVT	1-25 utentes	256	4462
	25-60 utentes	342	14000
	60+ utentes	114	9426

Tabela 3.2: Número de utentes e ERPIs em cada classe de tamanhos em ARSLVT e nas outras três regiões.

Neste âmbito, após aplicado o cálculo do RR para os números resultantes, observou-se na figura 3.9 que a classe correspondente a agregação das três regiões ARSALG, ARSC e ARSN também teve um padrão diferente do observado na ARSLVT.

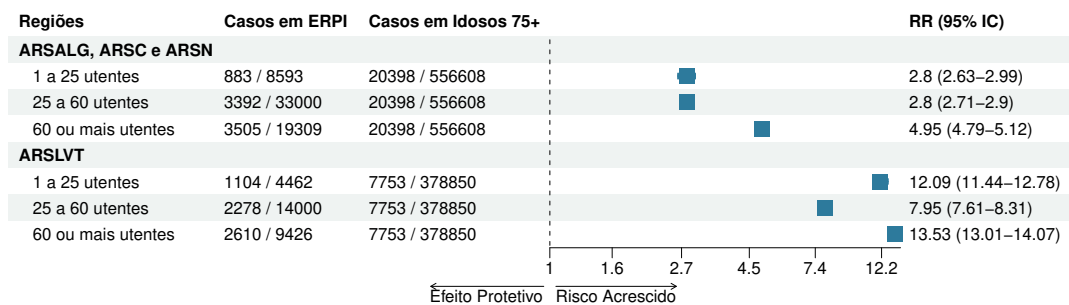


Figura 3.9: RR em função do tamanho em ARSLVT e nas outras três regiões.

Para a dita região ARSLVT, registou-se um risco muito maior nas ERPIs pequenas (RR: 12.08; IC 95%: 11.42–12.77) em comparação com as ERPIs médias (RR: 7.95; IC 95%: 7.61–8.3) e comparável aos de tamanho grande (RR: 13.53; IC 95%: 13.01–14.07). Observa-se, portanto,

que o valor nas ERPIs pequenas desta região é discrepante do observado nas ERPIs de mesmo tamanho nas demais regiões.

### RR dos óbitos

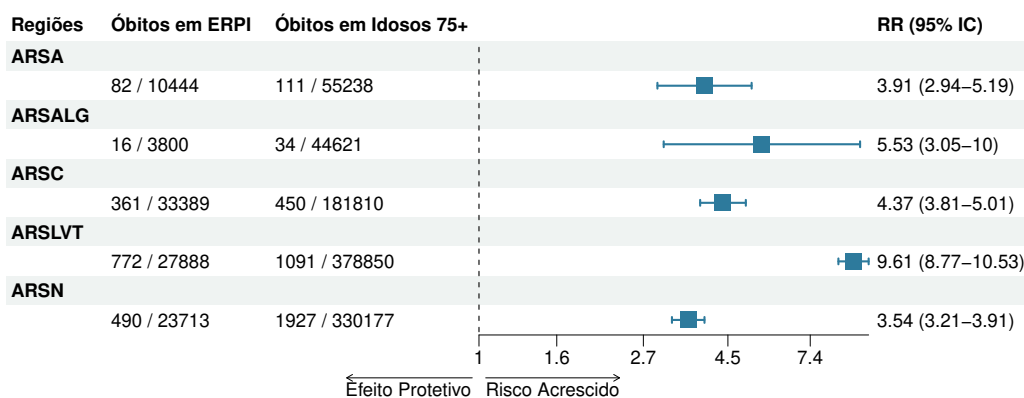


Figura 3.10: RR para os óbitos nas regiões de saúde.

Conforme ilustra a figura 3.10, o RR observado para cada uma das cinco ARS continentais, em todo o ano de 2020, teve em todos os casos um valor acima de um com um intervalo a 95% de confiança acima deste limiar. A ARSA teve um valor de 3.91 (IC 95%: 2.94–5.19) para todo aquele ano. O valor registado para a ARSALG foi de 5.53 (IC 95%: 3.05–10), sendo o segundo de maior valor entre as cinco regiões observadas e o primeiro em amplitude na estimação intervalar. A ARSC teve um valor de 4.37 (IC 95%: 3.81–5.01). Para o RR da ARSLVT, observou-se o valor máximo estimado que foi de 9.61 (IC 95%: 8.77–10.53). Finalmente, na ARSN, foi observado o valor de 3.54 (IC 95%: 3.21–3.91).

Foi também possível avaliar, conforme a figura 3.11, o RR da mortalidade ocorrida as regiões de saúde pelos dois períodos. O RR reportado nesta figura mostra que uma redução do risco de mortalidade ocorreu de forma mais acentuada nas regiões ARSC e ARSN. Conforme ficou reportado anteriormente na figura 3.8 estas duas regiões foram as que tiveram justamente um maior risco de infecção, seguido de uma queda mais acentuada do primeiro para o segundo período. A região de saúde ARSLVT teve uma redução muito menor do primeiro para o segundo período,

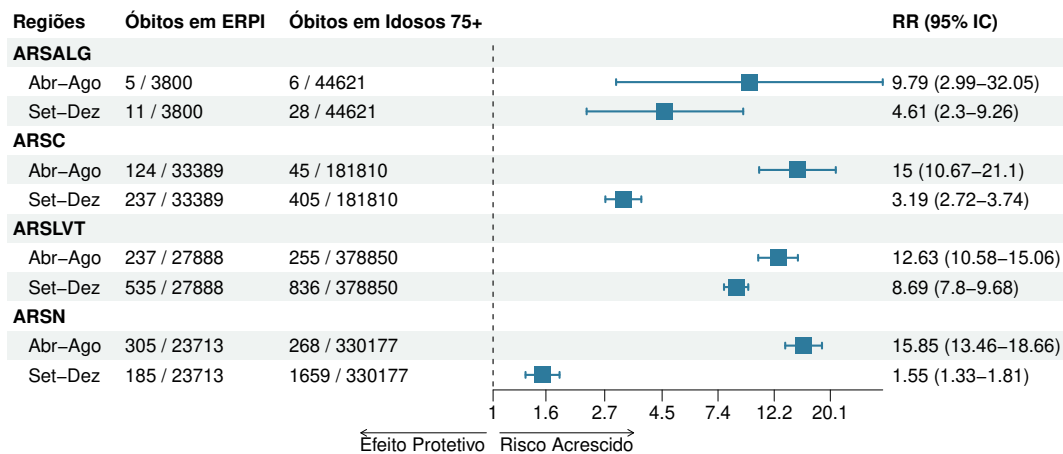


Figura 3.11: RR para os óbitos nas regiões de saúde (com exceção da ARSA) no primeiro e no segundo período.

o que também condiz com o que foi anteriormente reportado para o risco de infecções. A região ARSALG também teve uma redução do risco registrada no segundo período, porém, dado o pequeno número de mortes ocorridos dentro e fora das ERPIs na região, o intervalo de confiança a 95% demonstrou ser inconclusiva a diminuição deste risco. Em termos gerais, estes resultados para o RR de óbito concordam com os da Figura 3.8 para o RR de infecção.

Também foi visto, conforme a figura 3.12, que o RR dos óbitos estratificado por tamanho das ERPIs não acompanhou o padrão visto anteriormente nos casos para a ARSLVT na figura 3.9. A comparação dos dados relativos a ARSLVT nas ditas figuras sugere que possa ter havido sobrenotificação de casos nas ERPIs pequenas desta região.

### 3.2.2 Caracterização dos surtos em ERPIs por percentagem de casos

Como mostra a figura 3.13, foi observada na região ARSLVT uma proporção de 17.8% das infecções ocorridas em utentes nas ERPIs pequenas no primeiro período de 2020 e um aumento para 18.7% no segundo. Nas duas classes seguintes, observou-se que a proporção de infecções nas ERPIs intermédias teve um aumento no segundo período (42.6%) em comparação com o primeiro período (26.3%). Na classe respetiva, as ERPIs grandes, observou-se uma redução no segundo período (38.8%) em comparação com o primeiro (55.9%). Cumpre observar que nesta última classe a proporção de ocupação supera consideravelmente a dos infetados, o que indica

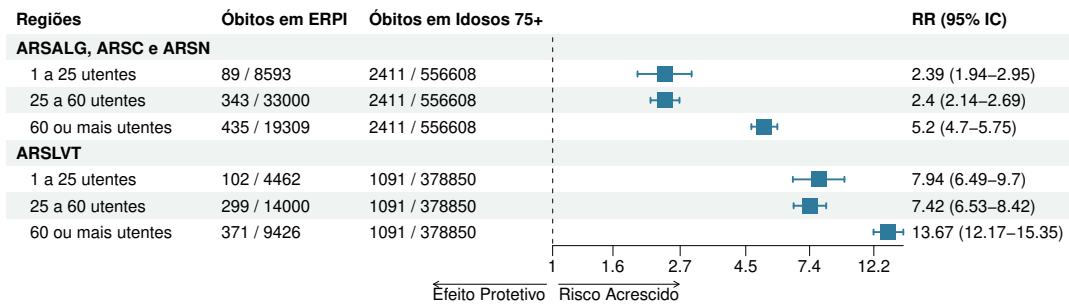


Figura 3.12: RR para os óbitos em função do tamanho na ARSLVT e nas outras três regiões.

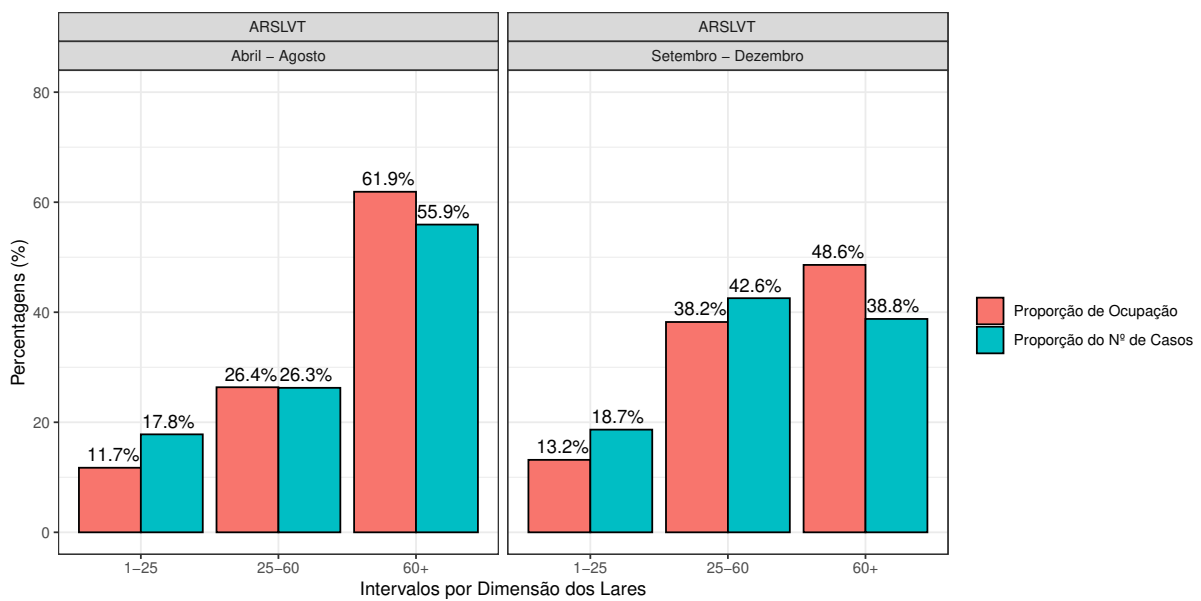


Figura 3.13: Proporções de utentes (vermelho) e casos (azul) distribuídos por intervalos do número de utentes em ERPI na ARSLVT dentro dos dois períodos do ano de 2020.

que as ERPIs de maior tamanho tiveram menos infeções por utentes.

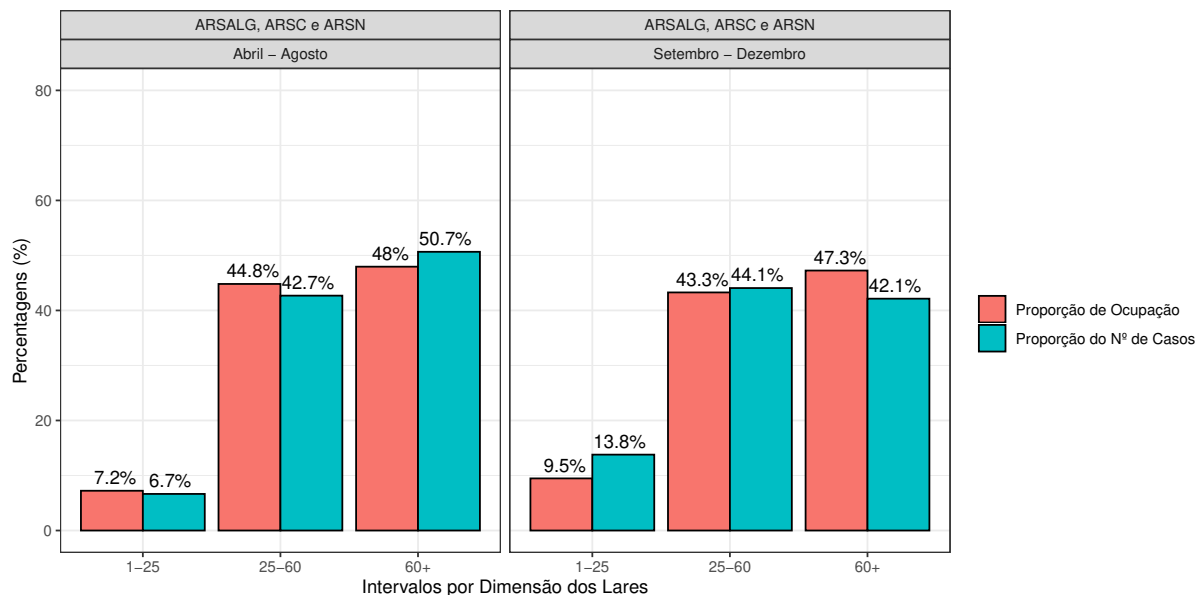


Figura 3.14: Proporções de utentes (vermelho) e casos (azul) distribuídos por intervalos do número de utentes em ERPI nas regiões ARSALG, ARSC e ARSN dentro dos dois períodos do ano de 2020.

Conforme a figura 3.14, a situação nas regiões ARSALG, ARSC e ARSN demonstrou que as ERPIs pequenas representaram nos dois períodos, respetivamente, apenas 6.7% e 13.8% das novas infeções, com as proporções de ocupação em 7.2% e 9.5% relativamente aos conjuntos do primeiro e do segundo período. Nas ERPIs médias observou-se uma maior proporção de residentes infetados, com ambos os períodos a registarem respetivamente as proporções de 42.7% e 44.1% das infeções. A proporção da ocupação dentro destas ERPIs permaneceu igualmente com pouca diferença nos dois períodos (44.8% e 43.3%). Nas ERPIs grandes a proporção de ocupação no segundo período sofreu uma sutil diferença quando comparada ao primeiro (48% e 47.3%). No entanto, a proporção de infetados sofreu uma diminuição de 8.6 pontos percentuais do primeiro (50.7%) para o segundo (42.1%) o que indica que esta classe teve menos infeções no segundo período em comparação com o primeiro, apesar de terem ambos os períodos um número aproximado de pessoas.

É intuitivo, reportado na literatura ([7]), que a probabilidade de que haja pelo menos uma infeção é maior para ERPIs grandes do que pequenas, tendo as primeiras mais oportunidades de contatos com o exterior. É igualmente intuitivo e conhecido ([7] e [22]) que sendo mais frequente as ERPIs de maior dimensão terem sucesso a confinar a uma percentagem relativamente pequena de utentes um surto que nelas ocorra. A figura 3.14 mostra que estes dois efeitos se equilibram, de maneira que a contribuição para a prevalência é proporcional à população de cada classe de tamanho.

Pelo contrário, na ARSLVT, as ERPIs pequenas têm mais casos reportados do que os que lhe corresponderiam em percentagem de utentes, o que indica que o efeito protetor do tamanho ou não existe, ou não é suficiente para compensar a dificuldade de contenção de surtos.

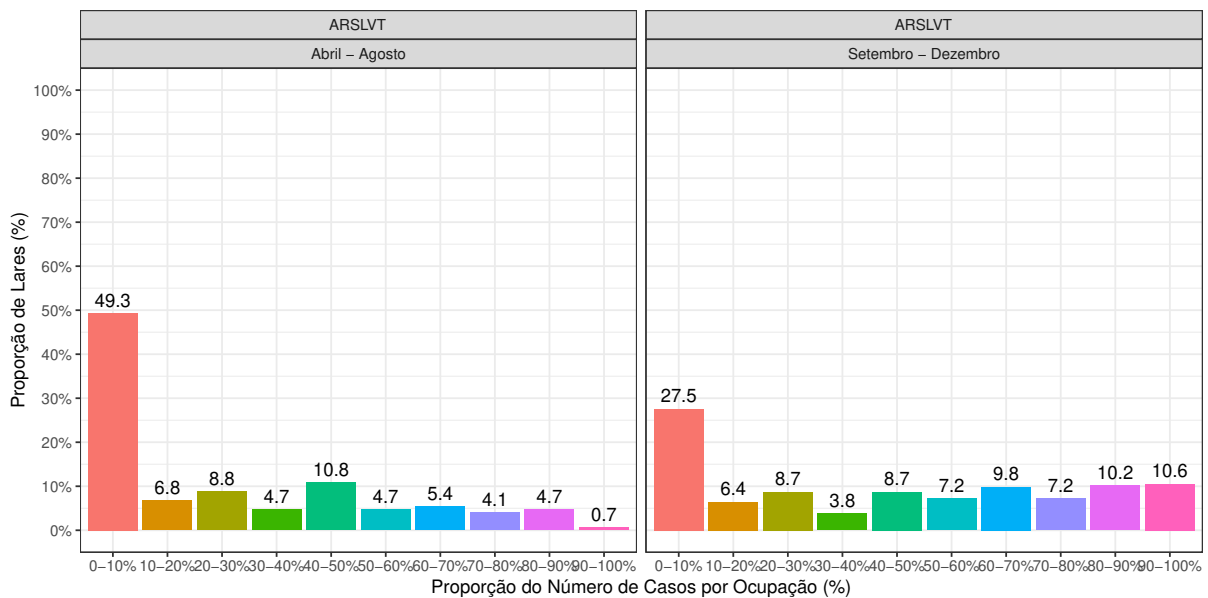


Figura 3.15: Frequência relativa das ERPIs da região ARSLVT distribuída pela proporção do número de casos em utentes pelo número total destes indivíduos nos dois períodos do ano.

Outro fator analisado, conforme a figura 3.15, diz respeito a distribuição da frequência relativa das ERPIs em intervalos da proporção de casos por utentes. Foi observado que no primeiro período a região ARSLVT teve quase a metade dos seus casos em ERPIs ocorridos em surtos sob controlo com até 10% dos utentes (49.3%). Porém, o segundo período destacou-se por uma drástica redução dos casos em ERPIs com a situação controlada, sendo agora 27.5% do total de infeções nos surtos.

Na figura 3.16, é possível observar que no primeiro período as regiões de saúde ARSALG, ARSC e ARSN tiveram, juntas, uma percentagem de 47.6% das ERPIs com os casos a limitarem-se em até 10% da ocupação. No segundo período, observou-se, assim como a região ARSLVT, uma queda na capacidade de contenção dos surtos destas regiões de saúde para 29%. Neste resultados vemos que as diferenças entre regiões não são significativas, e que, em geral, os surtos ocorridos a partir de agosto afetam uma maior percentagem de utentes. Não obstante, o risco relativo de infeção diminui bastante em relação à população idosa não institucionalizada.

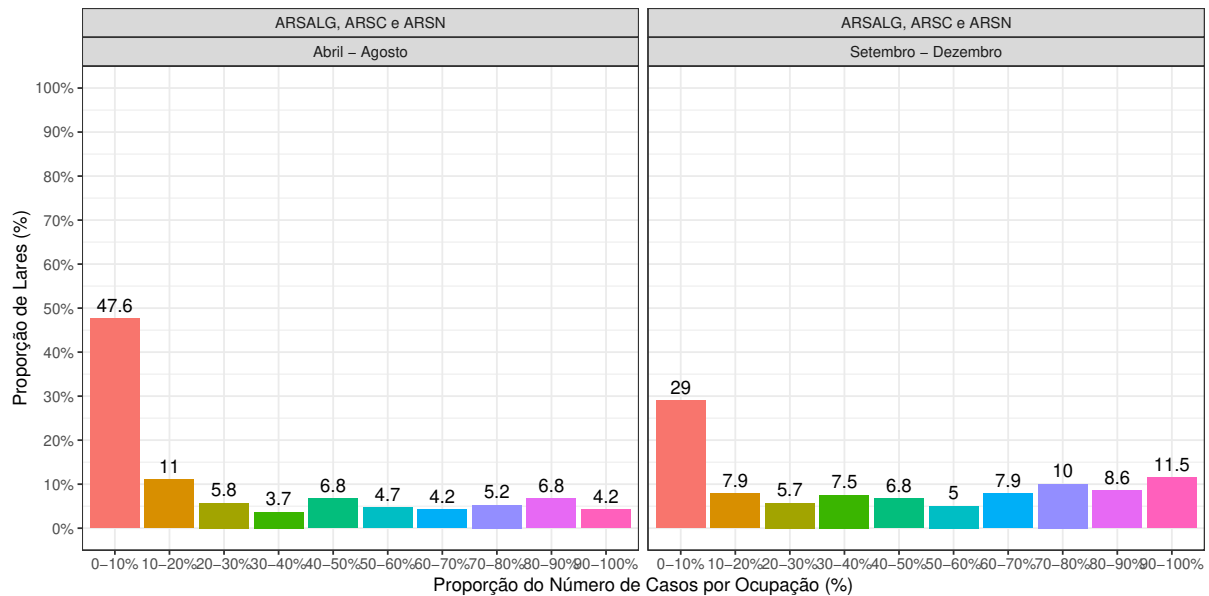


Figura 3.16: Frequência relativa das ERPIs nas regiões ARSALG, ARSC e ARSN distribuída pela proporção do número de casos de utentes pelo número total destes indivíduos nos dois períodos do ano.

### 3.2.3 Comparação do OR e da proporção do tamanho do surto em ERPI por ARS e pelas suas dimensões

A ocupação das ERPIs foi definida como o número de utentes em cada ERPI e fez-se uma análise por tamanho e por região. Tal como definido na secção 2.1.3, foram consideradas ERPIs pequenas (respetivamente médias, grandes) as que têm número de utentes inferior a 25 (respetivamente entre 25 e 60, superior a 60). Cumpre observar que o universo das ERPIs a serem analisadas nesta secção foi obtido do conjunto daqueles que apresentaram surtos (vide a figura 3.3).

Como consta na figura 3.17, a região ARSLVT apresentou uma frequência acrescida de ERPIs pequenas no primeiro período. Para este intervalo foi observada uma frequência relativa de 38.5% do número total de instalações. A frequência das ERPIs médias foi menor, com 31.8% do total de ERPIs no dito período, enquanto as ERPIs grandes representaram uma percentagem ainda menor de 29.5%. Já no segundo período, observou-se que a percentagem de ERPIs menores diminuiu para 36.2%, dos intermédios aumentou para 41.9% e dos maiores diminuiu para 21.9%.

Cumpre observar na figura 3.18 que no primeiro período as regiões ARSALG, ARSC e ARSN, tiveram uma frequência relativa de infeções nas ERPIs menores aproximada a das ERPIs maiores (22.5% e 22%) e maior de ERPIs intermédias (55.5%). O segundo período teve uma distribuição com padrão similar, já que as ERPIs menores igualmente representaram uma frequência relativa aproximada a dos maiores (24.4% e 24.7%) e nos intermédios a de 50.9%.

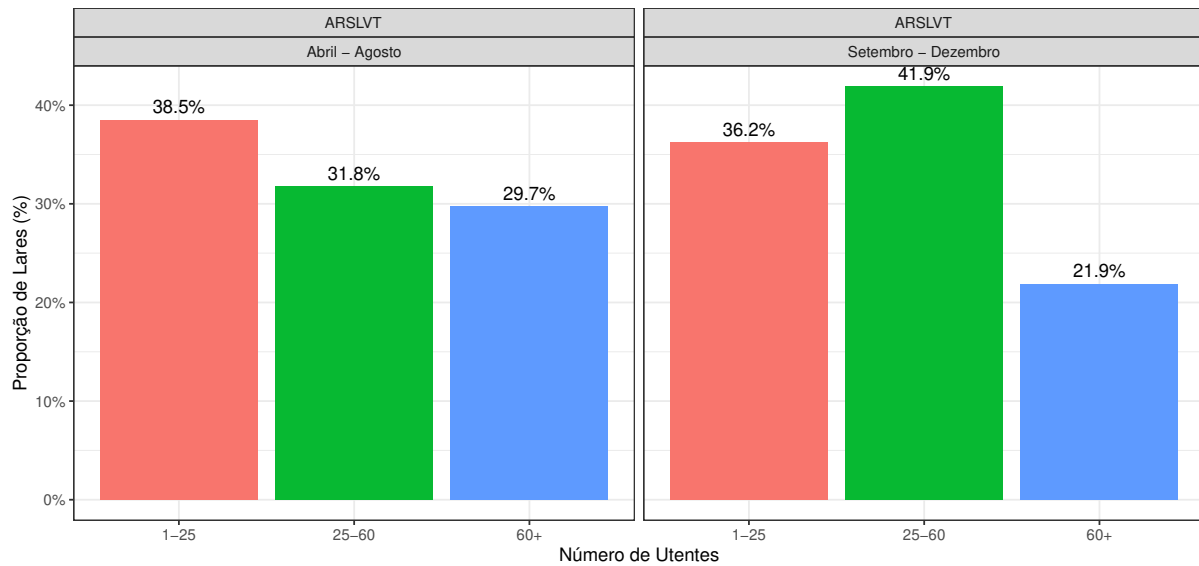


Figura 3.17: Distribuição de ERPI por número de utentes na região ARSLVT.

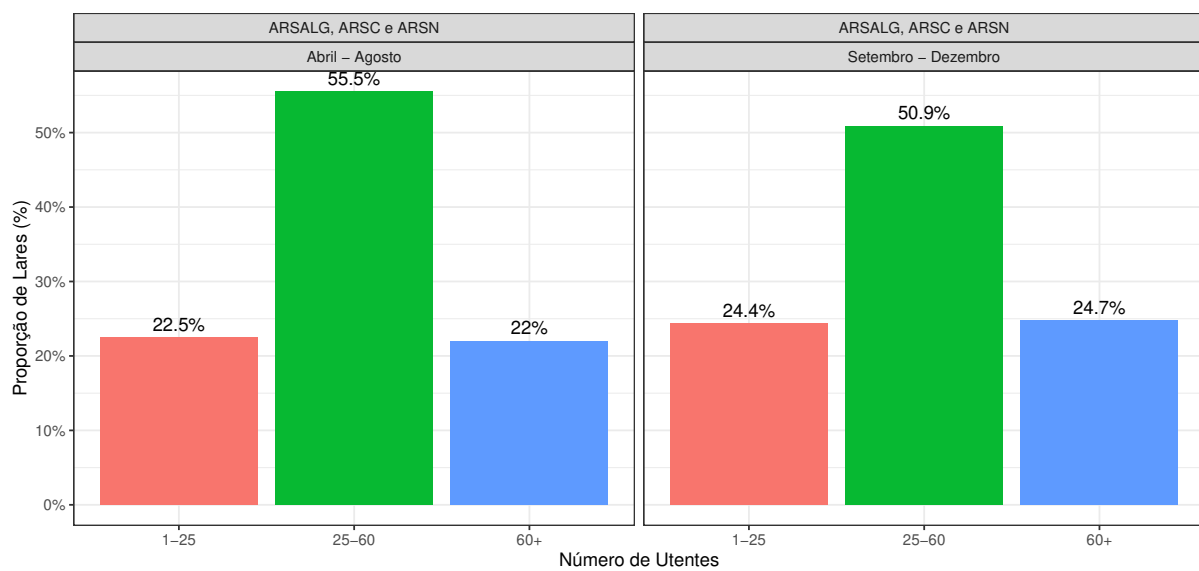


Figura 3.18: Distribuição de ERPI por número de utentes nas regiões ARSALG, ARSC e ARSN.

Neste aspeto, como outros que veremos a seguir, a região de saúde ARSLVT apresenta um padrão diferente. A fração das ERPIs de pequena dimensão no conjunto das que reportam casos é muito mais elevada, sobretudo até agosto.

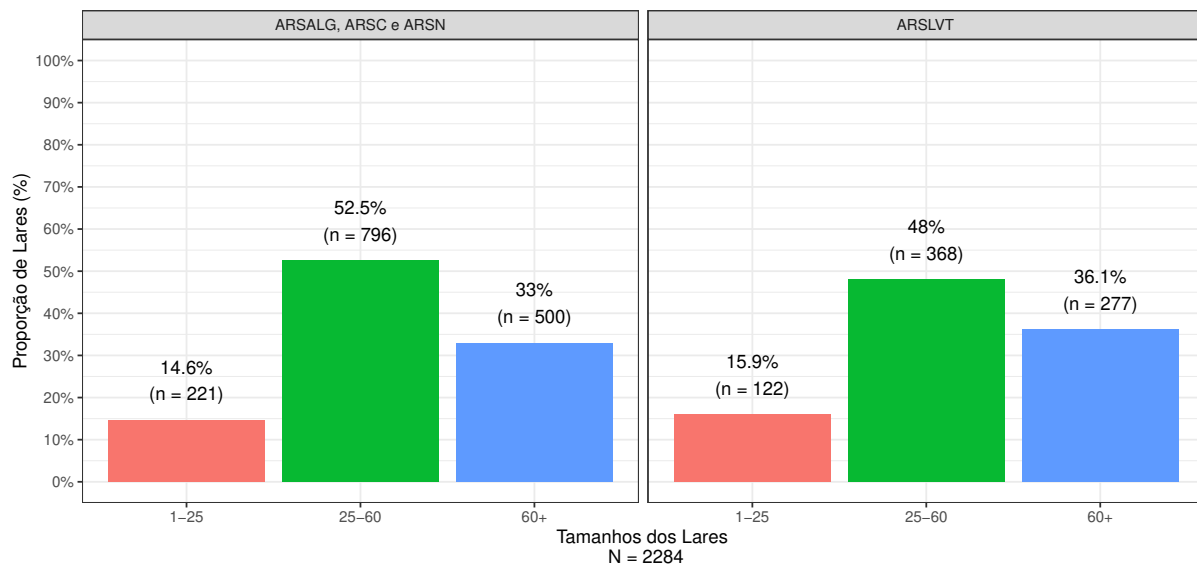


Figura 3.19: Distribuição por tamanhos de todas as ERPIs existentes em cada classe regional (ARSLVT e o conjunto das outras três regiões)

Conforme mostra a figura 3.19, após ser observada no universo de todas as instalações do tipo ERPI conforme a Carta Social (N = 2284) uma distribuição similar das instalações por categorias tamanhos entre a região ARSLVT e as demais, constatou-se a não existência de grandes diferenças na distribuição das ERPIs por tamanhos entre as duas classes regionais. Portanto, quando feita a análise em função do tamanho, essa análise não reflete diferenças estruturais no conjunto do tecido de todas as ERPIs a nível regional, mas reflete de facto o impacto por tamanhos na maneira como a doença progrediu nestas instalações.

Para a análise da regressão logística em função do risco de um lar reportar um ou mais casos da doença, fez-se o uso da métrica AIC, na qual o quarto modelo, com a interação entre as duas variáveis independentes região e tamanho, foi selecionado com a proporção de 0.99 no peso do Akaike (vide a definição desta métrica na secção Materiais e Métodos).

Conforme a tabela 3.3, ao contrário das demais regiões, em que o risco aumenta com o tamanho, foi observado que a região ARSLVT teve um maior risco de apresentar um ou mais casos de COVID-19 nas ERPIs com dimensão limitada a 25 utentes do que nas ERPIs de tamanho médio.

As ERPIs de tamanho grande tiveram um OR de 1.7 (IC 95%: 0.86- 3.36) em comparação com as ERPIs de tamanho pequeno. Contudo, dado que o dito intervalo continha o valor

Regiões de Saúde	Tamanho dos Lares	Lares com Surto	Lares sem Surto	Odds do Risco	OR (IC 95%)	P-valor
ARSALG, ARSC e ARSN	1-25 utentes	106 (111 surtos)	402	0,26	NA	
	25-60 utentes	235 (248 surtos)	594	0,4	1.5 (1.03-2.2)	0.029
	60+ utentes	99 (111 surtos)	128	0,77	2.93 (1.79-4.8)	0.001
ARSLVT	1-25 utentes	146 (153 surtos)	110	1,33	NA	
	25-60 utentes	138 (158 surtos)	204	0,68	0.51 (0.32-0.82)	0.001
	60+ utentes	79 (102 surtos)	35	2,26	1.7 (0.86-3.36)	0.228

Tabela 3.3: Resultados da Regressão Logística para o risco de um ou mais casos nos lares e seu respetivo OR em relação à cada classe de referência.

1, constatou-se ser inconclusiva a associação de risco entre as duas classes, o que também é evidenciado pelo p-valor de 0.228.

Nas regiões ARSALG, ARSC e ARSN, o OR manteve-se dentro do padrão observado conforme reporta (17): com um crescente risco ordenado pelas ERPIs pequenas, médias e grandes. O OR das ERPIs médias nestas regiões foi de 1.5 (IC 95%: 1.03-2.2) em comparação com a referência (ERPIs pequenas). As ERPIs grandes registaram um OR de 2.93 (IC 95%: 1.79-4.8) em comparação com a dita categoria de referência. Isso demonstra que nas outras regiões o OR comportou-se de forma esperada tendo em conta a literatura. Enquanto na ARSLVT, o comportamento observado foi diferente e paradoxal pelo odds desproporcionalmente elevado nas ERPIs pequenas que tiveram um risco comparável às grandes de apresentarem um ou mais casos.

Relativamente à regressão linear para o tamanho do surto, foi selecionado o modelo que fez uso apenas da variável tamanhos devido a sua menor pontuação face aos três outros modelos porquanto, uma proporção do peso do Akaike de 67%. Para validar a dita seleção, foi feita igualmente uma análise da significância das variáveis nos modelos lineares em que se fez o uso da variável regional e as suas combinações com a dita variável de tamanhos, e constatou-se que o este preditor não se mostrou significativo para explicar a distribuição dos dados (P-valor > 0.05). Portanto, o efeito observado é esperado para todas as regiões.

Tamanho dos Lares	Tamanho do Surto (%)	Efeito Marginal	P-Valor
1-25 utentes		44,52	NA
25-60 utentes		35,01	-11,04 0.001
60+ utentes		33,04	-13,37 0.001

Tabela 3.4: Resultados da Regressão Linear para o tamanho médio do surto em função dos tamanhos dos lares como única variável preditora.

Observa-se conforme a tabela 3.4 que, após o emprego da regressão linear com a variável

resposta do tamanho do surto em função do tamanho das ERPIs como variável preditora, o valor médio do tamanho do surto para o conjunto total das ERPIs diminuiu das ERPIs menores para as maiores.

Ainda para os dados referentes a quantidade de utentes e funcionários reportados, visualiza-se para cada região e tamanho do lar na tabela 3.5 os seguintes números médios do rácio de utentes por STAFF, número de casos em funcionários e proporção de casos em funcionários com os seus respetivos desvios padrões. Nota-se que nas ERPIs da região ARSLVT a média de rácio entre utentes e STAFF foi de 1.86, um valor bem acima dos observados em outros tamanhos e mesmo no conjunto das outras regiões agregadas.

Caracterização dos lares		Rácio de Utentes/STAFF		Casos em STAFF		Proporção de Casos em STAFF	
Regiões de Saúde	Tamanho dos Lares	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
ARSALG, ARSC e ARSN	1-25 utentes	1,5	0,88	4,2	4,17	0,31	0,29
	25-60 utentes	1,42	0,63	6,38	7,25	0,2	0,2
	60+ utentes	1,5	0,6	12,36	10,58	0,2	0,17
ARSLVT	1-25 utentes	1,86	2,07	3,2	3,25	0,29	0,28
	25-60 utentes	1,29	0,47	6,79	5,93	0,21	0,19
	60+ utentes	1,32	0,74	13,76	12,08	0,2	0,16

Tabela 3.5: Média do rácio entre utentes e funcionários, dos casos em funcionários, da proporção de casos por funcionários e os seus respetivos desvios padrões.

Também neste aspeto de subdotação de recursos humanos as ERPIs de tamanho pequeno da ARSLVT aparecem singularizadas.

## Capítulo 4

# Conclusões e discussão

O presente trabalho teve foco na análise dos registros de casos e mortes por COVID-19 em utentes ocorridos em ERPI entre as datas de 15 de abril e 27 de dezembro de 2020. Com base nos dados analisados, determinou-se o RR relativo à infecção e morte dos utentes em ERPIs comparativamente à população idosa em geral para as cinco regiões de saúde (figura 3.7), e para os dois períodos de 2020 que correspondem à primeira e à segunda vaga daquele ano (figura 3.8).

Conforme a análise das figuras mencionadas, verificou-se que em todas as regiões o RR de casos e mortes foi reduzido na segunda vaga, o que demonstra a capacidade de adaptação das ERPIs em geral no sentido de melhorar a proteção dos utentes. Esta constatação pode ser feita com base na figura 3.6, onde foi observada uma incidência em ERPIs muito próxima à incidência na população em geral nos meses de abril a agosto (primeiro período). Esta tendência no primeiro período corresponde ao que foi citado em [2] em relação aos lares de idosos como foco inicial da doença nos concelhos do interior, centro e norte, sendo justamente estas duas últimas regiões as que foram registadas na figura 3.8 como tendo um maior acréscimo de RR em comparação com a população geral naquele mesmo período inicial.

A partir do mês de setembro (segundo período) foi observado um agravamento da curva epidémica na população idosa, com os casos aumentando de forma bastante acentuada até a estabilização em dezembro. O aumento também foi registado nos lares de idosos, conforme se observou na mesma figura 3.6 porém com a diferença de que o número de casos fora dos lares foi muito maior do que dentro destas instalações, demonstrando um melhor preparo e capacidade de lidar com os surtos da doença naquele período. Isso também ficou evidenciado nas figuras 3.8 e 3.11 em que o número de casos e mortes sofreram uma redução sobretudo mais acentuada nas regiões de saúde ARSC e ARSN.

Estes resultados, que demonstraram uma clara limitação do excesso de risco de casos e mortes em ERPIs a partir do segundo período em comparação com a população idosa em geral, também foram condizentes com o que relatou [23]. Pois, como ficou reportado pelos autores do dito estudo, a proporção de óbitos ocorridos em lares de idosos apresentou uma redução

substancial na segunda vaga em relação à primeira em 8 dos 11 países analisados. Os autores ainda atribuíram como uma das possíveis causas desta redução as medidas não farmacológicas somadas a um melhor aprendizado com a primeira vaga. Portanto, condizente com aquilo que foi reportado na já referida figura 3.11.

A mesma constatação de uma possível melhoria nas medidas tomadas pelos lares somadas a alguma imunidade adquirida pelos residentes após a primeira vaga foi feita por [24], no qual os autores analisaram o risco de mortalidade nos lares de idosos do Reino Unido, nas duas vagas da pandemia em 2020, comparado com o risco em idosos nas suas casas privadas (em ambiente externo a estes lares). Mais uma vez, os resultados demonstraram uma redução expressiva do risco na segunda vaga comparada à primeira. Com os autores a concluírem que o excesso de risco ocorreu sobretudo na primeira vaga, não tendo sido observado o mesmo impacto na segunda.

Vale destacar que ambos os estudos [23] e [24] referem as diferenças nas características demográficas da população mais atingida como um dos possíveis fatores explicativos para esta redução. Já [24] avalia que outra explicação baseada nas mudanças demográficas poderia ser o elevado número de mortes na primeira vaga a resultarem nos residentes de lares de idosos que sobreviveram a doença com alguma imunidade e menor número de comorbidades, tornando-os mais aptos a sobreviverem nestes ambientes durante a segunda vaga.

Outro estudo ([25]) avaliou a proporção de infecções ocorridas em idosos institucionalizados em relação a todas as infecções ocorridas em território nacional. Os resultados do dito estudo revelaram que a proporção de infecções nos idosos foi alta nos meses iniciais da epidemia em Portugal, de 3 de março a 15 de maio de 2020, representando o período de maior risco para os utentes. Já no segundo período, o número de casos em utentes foi maior, porém a proporção face ao total de positivos foi menor, o que indica que apesar do aumento de casos em ERPIs, a sua proporção em relação a todos os casos foi menor do que no início do ano, evidenciando um menor risco no último período. Esses resultados, apesar de ter seguido uma metodologia diferente, põe em evidência uma tendência similar a apresentada neste presente estudo.

Porém, em relação à análise do comportamento do risco nas regiões conforme as figuras supracitadas, verificou-se que na região ARSLVT a queda do RR de infecção foi muito menor do que nas outras regiões observadas, em especial devido à contribuição dos casos em ERPIs pequenas.

Isto motivou uma análise por tamanho das ERPIs, após verificar que a distribuição das ERPIs por tamanhos é aproximadamente a mesma na ARSLVT e nas demais (figura 3.19). Os indicadores analisados em função do tamanho foram o odds ratio (OR) de um lar registar um ou mais casos entre os seus utentes e o tamanho de cada surto registado (TS) medido como a percentagem dos utentes infetados.

Conforme indicado na tabela 3.4, a métrica TS comportou-se em todas as regiões conforme o esperado e o reportado na literatura ([7] e [22]), diminuindo com o tamanho da ERPI. Quanto ao OR, reportado na tabela 3.3, nas demais regiões ele comportou-se como esperado ([7]), au-

mentando com o tamanho da ERPI, mas na região ARSLVT, onde o OR mínimo ocorre nas ERPIs de tamanho intermediário. Assim, o comportamento de *outlier* da dita região resultou em grande parte de um excesso de risco de infecção nas ERPIs pequenas. A análise do rácio de utentes por funcionários, conforme demonstra a tabela 3.5, constata que as ERPIs de menor tamanho na ARSLVT são também *outliers* com valores muito grandes desse rácio.

Tomados no seu conjunto, os resultados mostram que houve, em geral, uma melhor proteção dos utentes de ERPIs após a primeira vaga. Mostram também que o desempenho da ARSLVT foi pior do que a das demais regiões, mantendo-se o RR de óbito muito elevado pelo menos até ao fim de 2020.

A análise dos dados em função do tamanho das ERPIs mostra ainda que as ERPIs pequenas de ARSLVT são também *outliers* tanto no número muito elevado de casos que reportam como no rácio também muito elevado de utentes por funcionários. Dado que os dados relativos a óbitos não seguem os dados de casos, ou a letalidade é menor nas ERPIs pequenas e com falta de pessoal, ou nesses casos há irregularidades nos registos de casos que se traduzem em sobre-notificação.

O presente estudo contou com as seguintes limitações: O registo de casos nas instalações do tipo ERPI foi feito com base na prevalência da doença nos utentes num determinado dia da semana. Dado que foi necessário realizar operações de soma e divisão para extração da incidência e do número de casos ocorridos nas instalações, estes dois valores devem ser vistos como aproximados.

Algumas ERPIs tiveram de ser filtradas por possuírem registos inconsistentes. Por esse motivo o universo das instalações que tiveram casos não está completo.

Dado que algumas ERPIs registaram casos com valores repetidos ao longo de várias semanas, é assumido que alguns valores nas listas de determinadas ERPIs tiveram erros ou problemas de imputação. Considerou-se para estas situações apenas a primeira data de registo e removidas as demais linhas.

Os dados obtidos para o universo total de lares no continente via o sítio “web” da Carta Social se referem ao ano de 2022 e não o ano de 2020 que respetiva ao ficheiro de dados dos casos em ERPIs. Fez-se necessário aplicar a distribuição das instalações e o seu número de utentes pelas categorias de regiões e tamanhos, obtida da lista de todas as ERPIs de 2022, para estimar o número de instalações e de utentes dentro das categorias mencionadas tendo em conta o número agregado de ERPIs e utentes nas suas regiões em 2020 (presente no ficheiro 1.6).

# Bibliografia

- [1] Dadras, O, SeyedAlinaghi, S. A, Karimi, A, Shamsabadi, A, Qaderi, K, Ramezani, M, Mirghaderi, S. P, Mahdiabadi, S, Vahedi, F, Saeidi, S, Shojaei, A, Mehrtak, M, Azar, S. A, Mehraeen, E, & Voltarelli, F. A. (2022) COVID-19 mortality and its predictors in the elderly: A systematic review. *Health Science Reports* **5**.
- [2] Marques da Costa, E & Marques da Costa, N. (2020) A pandemia covid-19 em portugal continental: uma análise geográfica da evolução verificada nos meses de março e abril. *Hygeia-Revista Brasileira de Geografia Médica e da Saúde* pp. 72–79.
- [3] Solidário, Rotass – Rede Operacional e Técnica da Apoio ao Setor. (2022) Atualização Orientação 009 de DGS | ROTASS. Acesso em 25.09.2022.
- [4] Sá, M. M, Pires, A. L, Cabral, D, Leite, T, Mendonça, C, & Neves-Reina, L. (2021) Nursing homes and covid-19, an avoidable fatality: Results of a multidisciplinary approach. *Acta Médica Portuguesa* **34**, 239–239.
- [5] Suetens, C, Kinross, P, Berciano, P. G, Nebreda, V. A, Hassan, E, Calba, C, Fernandes, E, Peralta-Santos, A, Casaca, P, Shodu, N, Dequeker, S, Kontopidou, F, Pappa, L, Kacelnik, O, Børseth, A. W, O'Connor, L, Garvey, P, Liausedienė, R, Valinteliene, R, Ernst, C, Mossong, J, Štefkovičová, M, Prostináková, Z, Danielsen, A. C, Mougkou, A, Lamb, F, Cenciarelli, O, Monnet, D. L, & Plachouras, D. (2021) Increasing risk of breakthrough COVID-19 in outbreaks with high attack rates in European long-term care facilities, July to October 2021. *Eurosurveillance* **26**, 2101070.
- [6] Suñer, C, Ouchi, D, Mas, M. À, Lopez Alarcon, R, Massot Mesquida, M, Prat, N, Bonet-Simó, J. M, Expósito Izquierdo, M, Garcia Sánchez, I, Rodoreda Noguerola, S, et al. (2021) A retrospective cohort study of risk factors for mortality among nursing homes exposed to covid-19 in spain. *Nature Aging* **1**, 579–584.
- [7] Abrams, H. R, Loomer, L, Gandhi, A, & Grabowski, D. C. (2020) Characteristics of us nursing homes with covid-19 cases. *Journal of the American Geriatrics Society* **68**, 1653–1656.

- [8] Bielza, R, Sanz, J, Zambrana, F, Arias, E, Malmierca, E, Portillo, L, Thuissard, I. J, Lung, A, Neira, M, Moral, M, et al. (2021) Clinical characteristics, frailty, and mortality of residents with covid-19 in nursing homes of a region of madrid. *Journal of the American Medical Directors Association* **22**, 245–252.
- [9] Tarteret, P, Strazzulla, A, Rouyer, M, Gore, C, Bardin, G, Noel, C, Benguerdi, Z.-E, Berthaud, J, Hommel, M, Aufaure, S, et al. (2021) Clinical features and medical care factors associated with mortality in french nursing homes during the covid-19 outbreak. *International Journal of Infectious Diseases* **104**, 125–131.
- [10] Direção-Geral da Saúde. (2021) Plano de Vacinação da COVID-19. Acesso em 05.12.2021.
- [11] Ministério do Trabalho, Solidariedade e Segurança Social (MTSSS). (2022) Início - Carta Social. Acesso em 24.11.2022.
- [12] Social, Ministério da Solidariedade e da Segurança. (2012) Portaria n.º 67/2012, de 21 de março. Acesso em 24.11.2022.
- [13] Ryskina, K. L, Yun, H, Wang, H, Chen, A. T, & Jung, H. Y. (2021) Characteristics of Nursing Homes by COVID-19 Cases Among Staff: March to August 2020. *Journal of the American Medical Directors Association* **22**, 960.
- [14] Ibrahim, J. E, Li, Y, Brown, C, McKee, G, Eren, H, & Pham, T. (2021) Risk Stratification of Nursing Homes to Plan COVID-19 Responses: A Case Study of Victoria, Australia. *Disaster Medicine and Public Health Preparedness* p. 1.
- [15] Morciano, M, Stokes, J, Kontopantelis, E, Hall, I, & Turner, A. J. (2021) Excess mortality for care home residents during the first 23 weeks of the COVID-19 pandemic in England: a national cohort study. *BMC Medicine* **19**.
- [16] Burton, J. K, McMinn, M, Vaughan, J. E, Fleuriot, J, & Guthrie, B. (2021) Care-home outbreaks of COVID-19 in Scotland March to May 2020: National linked data cohort analysis. *Age and Ageing* **50**, 1482.
- [17] Cazzoletti, L, Zanolin, M. E, Tocco Tussardi, I, Alemayohu, M. A, Zanetel, E, Visentin, D, Fabbri, L, Giordani, M, Ruscitti, G, Benetollo, P. P, Tardivo, S, & Torri, E. (2021) Risk Factors Associated with Nursing Home COVID-19 Outbreaks: A Retrospective Cohort Study. *International Journal of Environmental Research and Public Health* **18**.
- [18] GEP - Gabinete de Estratégia e Planeamento. (2021) Carta social - rede de serviços e equipamentos - relatório 2020.

- 
- [19] Woodward, M. (2013) *Epidemiology: Study Design and Data Analysis, Third Edition*, Chapman & Hall/CRC Texts in Statistical Science. (Taylor & Francis).
- [20] Symonds, M. R. E & Moussalli, A. (2011) A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion.
- [21] Wagenmakers, E. J & Farrell, S. (2004) AIC model selection using Akaike weights. *Psychonomic Bulletin Review* 2004 11:1 **11**, 192–196.
- [22] Kruse, F. M, Mah, J. C, Metsemakers, S. J, Andrew, M. K, Sinha, S. K, & Jeurissen, P. P. (2021) Relationship between the Ownership Status of Nursing Homes and Their Outcomes During the COVID-19 Pandemic: A Rapid Literature Review. *Journal of Long-Term Care* **2021**, 207–220.
- [23] Ioannidis, J. P, Axfors, C, & Contopoulos-Ioannidis, D. G. (2021) Second versus first wave of COVID-19 deaths: Shifts in age distribution and in nursing home fatalities. *Environmental Research* **195**, 110856.
- [24] Schultze, A, Nightingale, E, Evans, D, Hulme, W, Rosello, A, Bates, C, Cockburn, J, MacKenna, B, Curtis, H. J, Morton, C. E, Croker, R, Bacon, S, McDonald, H. I, Rentsch, C. T, Bhaskaran, K, Mathur, R, Tomlinson, L. A, Williamson, E. J, Forbes, H, Tazare, J, Grint, D, Walker, A. J, Inglesby, P, DeVito, N. J, Mehrkar, A, Hickman, G, Davy, S, Ward, T, Fisher, L, Green, A. C, Wing, K, Wong, A. Y, McManus, R, Parry, J, Hester, F, Harper, S, Evans, S. J, Douglas, I. J, Smeeth, L, Eggo, R. M, Goldacre, B, & Leon, D. A. (2022) Mortality among Care Home Residents in England during the first and second waves of the COVID-19 pandemic: an observational study of 4.3 million adults over the age of 65. *The Lancet Regional Health - Europe* **14**, 100295.
- [25] De Mello-Sampayo, F. (2022) Spatial and Temporal Analysis of COVID-19 in the Elderly Living in Residential Care Homes in Portugal. *International Journal of Environmental Research and Public Health* **19**.