



Universidade do Estado do Rio de Janeiro

Centro Biomédico

Instituto de Medicina Social

Cleber Vinicius Brito dos Santos

Como o desmatamento influencia a ocorrência de leishmaniose visceral?

Uma abordagem contrafactual

Rio de Janeiro

2020

Cleber Vinicius Brito Dos Santos

**Como o desmatamento influencia a ocorrência de leishmaniose visceral? Uma
abordagem contrafactual**

Dissertação a ser apresentada, como requisito parcial
para obtenção do título de Mestre, ao Programa de
Pós-Graduação em Saúde Coletiva, da Universidade
do Estado do Rio de Janeiro. Área de concentração:
Epidemiologia.

Orientador: Prof. Dr. Cláudio José Struchiner

Coorientadores: Prof. Dr. Guilherme Loureiro Werneck

Dr^a. Anaiá da Paixão Sevá

Rio de Janeiro

2020

CATALOGAÇÃO NA FONTE
UERJ/REDE SIRIUS/CB/C

S237 Santos, Cleber Vinicius Brito dos
Como o desmatamento influencia a ocorrência de leishmaniose visceral? Uma abordagem contrafactual / Cleber Vinicius Brito dos Santos – 2020.
86 f.

Orientador: Cláudio José Struchiner
Coorientadores: Guilherme Loureiro Werneck
Anaiá da Paixão Sevá

Dissertação (Mestrado) – Universidade do Estado do Rio de Janeiro, Instituto de Medicina Social.

1. Leishmaniose visceral – Teses. 2. Doenças transmitidas por vetores – Teses. 3. Doenças negligenciadas – Teses. 4. Epidemiologia – Teses. I. Struchiner, Cláudio José. II. Werneck, Guilherme Loureiro. III. Sevá, Anaiá da Paixão. IV. Universidade do Estado do Rio de Janeiro. Instituto de Medicina Social. V. Título.

CDU 616993.161

Bibliotecária: Joice Soltosky Cunha – CRB 7 5946

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação, desde que citada a fonte.

Assinatura

Data

Cleber Vinicius Brito dos Santos

Como o desmatamento influencia a ocorrência de leishmaniose visceral? Uma abordagem contrafactual

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Saúde Coletiva, da Universidade do Estado do Rio de Janeiro. Área de concentração: Epidemiologia.

Aprovada em 05 de março de 2020.

Orientador: Prof. Dr. Cláudio José Struchiner

Instituto de Medicina Social - UERJ

Coorientadores: Prof. Dr. Guilherme Loureiro Werneck

Instituto de Medicina Social - UERJ

Dr^a. Anaiá da Paixão Sevá

Universidade Estadual de Santa Cruz - UESC

Banca Examinadora:

Prof. Dr. José Ueleres Braga

Instituto de Medicina Social - UERJ

Prof. Dr. Michael Eduardo Reichenheim

Instituto de Medicina Social - UERJ

Prof. Dr. Luiz Max Fagundes de Carvalho

Fundação Getúlio Vargas

Rio de Janeiro

2020

DEDICATÓRIA

Dedico esta dissertação à Célio José Alves dos Santos (in memoriam) e Isaura Maria de Brito (in memoriam).

AGRADECIMENTOS

A realização deste trabalho só foi possível com o incentivo, apoio e carinho de muitas pessoas que já estavam presentes em minha vida, e outras que cruzaram o meu caminho durante a minha estadia no Rio de Janeiro. Sendo assim, gostaria de prestar meus sinceros agradecimentos a todos que me ajudaram ao longo dessa conquista.

Primeiramente gostaria de agradecer a minha família pelo constante apoio moral, intelectual e, mais importante, pelo amor e carinho.

A minha esposa Bianca, pelo carinho e suporte durante os momentos difíceis.

A minha amada filha Linda, por me dar forças para levantar da cama todos os dias.

Meus sinceros agradecimentos ao Cláudio e Gui, pelos valiosos ensinamentos e por serem exemplos de humildade, competência e serenidade.

A Anaiá por, desde o início, ter aceitado me coorientar e sempre ter encontrado um tempinho para fazer sugestões e discutir assuntos relacionados ao trabalho.

Aos amigos que fiz no Rio de Janeiro, aos companheiros de Nepidi, Nebin e PIEVF em especial Ana Paula, Ana Sara, Anna Beatriz, Bárbara, Beatriz, Camilla, João, Luciane, Taísa e Vanessa, obrigado por terem tornado meus dias mais alegres.

Aos professores e amigos do Instituto de Medicina Social pela partilha de conhecimento que foram essenciais para a construção desse trabalho, especialmente Evandro Coutinho, Ponce de León, Uelers Braga e Michael Reichenheim.

A toda equipe de funcionários da UERJ e do IMS por, apesar de todas adversidades, manterem esse lugar na ativa, principalmente as meninas da secretaria (Aline, Eliete e Silvia), Artur, Aleksandra e Elir.

Ao Programa de Pós-Graduação em Saúde Coletiva por ter me aceitado como aluno.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de mestrado.

He uses statistics as a drunken man uses lamp-posts – for support rather than illumination.

Andrew Lang

RESUMO

SANTOS, Cleber Vinicius Brito dos. **Como o desmatamento influencia a ocorrência de leishmaniose visceral? Uma abordagem contrafactual**. 2020. 86 f. Dissertação (Mestrado em Saúde Coletiva) – Instituto de Medicina Social, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2020.

A leishmaniose visceral (LV) é uma doença sistêmica de transmissão vetorial. No Brasil, ela é causada pelo protozoário *Leishmania infantum* e é transmitida por flebotômíneos do gênero *Lutzomyia*, tendo os cães como a principal fonte de infecção em áreas urbanas. Historicamente a LV era conhecida como uma doença endêmica rural, porém desde a década de 1980, ela atingiu proporções endêmicas e epidêmicas em grandes cidades brasileiras. Desde então vários fatores são considerados como norteadores da expansão da LV, como variáveis socioeconômicas, climáticas e ambientais. Mais especificamente sobre o efeito do desmatamento e perturbações antrópicas no ambiente, a maioria do que se encontra na literatura tem natureza qualitativa ou utiliza de abordagens reducionistas, sem considerar a complexidade da dinâmica de uma doença infecciosa de transmissão vetorial e de caráter zoonóticos. Na presente dissertação, investigamos o efeito do desmatamento na ocorrência de *Lutzomyia longipalpis* (*Lu. longipalpis*), leishmaniose visceral canina (LVC) e leishmaniose visceral humana (LVH), tomando como exemplo o estado de São Paulo. Para isso, utilizamos uma abordagem contrafactual para estimar os efeitos (geral, direto e indireto) do desmatamento na ocorrência do *Lu. longipalpis*/LVC/LVH. Isso foi feito em dois passos, primeiro estimamos os parâmetros por meio de um algoritmo de Metropolis-Hastings duplo e, por fim, estimamos os efeitos causais através de um amostrador de Gibbs, por meio do pacote autognet no R. Vimos que municípios desmatados apresentam 2.63, 2.07 e 3.18 maiores chances de apresentar o vetor, LVC e LVH, respectivamente quando comparados com os municípios que não apresentaram desmatamento. Foi observada também uma forte influência da presença do vetor, LVC e LVH dos municípios vizinhos na ocorrência dos mesmos em municípios previamente livres dos desfechos (6.67, 4.26 e 4.27). Já sob mudanças hipotéticas de prevalência do desmatamento de 50% para 0% no estado, são esperadas quedas na prevalência do vetor, LVC e LVH de 11%, 6.67% e 29.87% respectivamente. O desmatamento influi na ocorrência do vetor, doença em cães e humanos por duas principais vias, (i) alterando o funcionamento do ecossistema e estrutura da comunidade, permitindo a reprodução e colonização do vetor; e (ii) promovendo uma aproximação entre todos os componentes do ciclo da LV. De tal modo, para correto controle da LV e doenças infecciosas como um todo, é imprescindível um desenvolvimento ecologicamente correto com soluções viáveis para as compensações entre a agricultura, urbanização e conservação. urbanização e conservação.

Palavras-chave: Inferência causal. Doenças transmitidas por vetores. Doenças tropicais negligenciadas. Doenças emergentes e reemergentes. Inferência bayesiana. Desfechos potenciais. Saúde Única.

ABSTRACT

SANTOS, Cleber Vinicius Brito dos. **How deforestation influence the occurrence of Visceral Leishmaniasis? A counterfactual approach**. 2020. 86 f. Dissertação (Mestrado em Saúde Coletiva) –Instituto de Medicina Social, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2020.

Visceral leishmaniasis (VL) is a systemic vector-borne disease. In Brazil, it caused by the protozoan *Leishmania infantum* and is transmitted by sandflies of the genus *Lutzomyia*, with dogs as the principal source of infection in urban areas. Historically, VL was known as a rural endemic disease, since the 80's it has become endemic and epidemic in large Brazilian cities. Since then, many factors were hypothesised as driving VL expansion, as socioeconomic, climatic and environmental variables. More specifically, concerning deforestation and human-made actions in the environment, most studies tend to be qualitative in nature or use traditional reductionist approaches, ignoring the complexities that are inherent of vector-borne zoonotic infectious diseases. The present study aimed to investigate the effect of deforestation in the occurrence of *Lutzomyia longipalpis*, canine visceral leishmaniasis (CVL), and human visceral leishmaniasis (HVL), taking as a motivating example the São Paulo state (Brazil). To this end, we chose a counterfactual approach to estimate the effects (overall, direct and indirect) of deforestation in the occurrence of vector/CVL/HVL. We did it in two steps; first, we estimated the parameters through a double Metropolis-Hastings algorithm and, finally, we estimated the causal effects through a Gibbs sampler, using the autognet package in R. We observe that deforested cities show 2.63, 2.07, and 3.18 higher odds of vector/CVL/HVL occurrence, respectively, when compared to non-deforested municipalities. We also see a significant influence of vector, CVL, and HVL presence in the neighbours in its appearance in previous naive cities, 6.67, 4.26, 4.27 respectively. Lastly, under hypothetical changes in deforestation's prevalence from 50% to 0% in the whole state, is expected a decrease in its prevalence of the vector, LVC and LVH of 11%, 6.67% and 29.87% respectively. Deforestation in the occurrence of infectious diseases and, more specifically, VL importance, is two-folded: (i) changing's the ecosystem equilibrium and community structure, allowing its vector to reproduce and colonise; (ii) promoting a close contact through the VL cycle components. In such a way, for correct control of VL and infectious disease as a whole, it is essential an eco-friendly development with viable solutions for trade-offs between agriculture, urbanization and conservation.

Keywords: Causal inference. Vector-borne diseases. Neglected tropical diseases. Emerging infectious diseases. Bayesian inference. Potential outcomes. One health

LISTA DE ILUSTRAÇÕES

Figura 1-	Ciclo da leishmaniose visceral.....	18
Figura 2-	Grafo acíclico direcionado hipotético.....	31
Figura 3-	Grafo em cadeia hipotético.....	35
Figura 4-	Diagrama Causal completo do estudo.....	39
Tabela 1	Distribuição dos dados.....	45
Tabela 2	Estimativas dos parâmetros para o <i>Lu. Longipalpis</i> , LVC e LVH.....	46
Figura 5-	Estimativas dos efeitos causais sob diferentes prevalências de desmatamento na rede.....	48

LISTA DE ABREVIATURAS E SIGLAS

CG	Grafo de cadeias – <i>Chain Graph</i>
CVE/SES-SP	Centro de Vigilância Epidemiológica da Secretaria de Estado de Saúde de São Paulo
DAG	Grafo Acíclico Direcionado – <i>Directed Acyclic Graph</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
LC	Leishmaniose cutânea
<i>Lu. longipalpis</i>	<i>Lutzomyia longipalpis</i>
LV	Leishmaniose visceral
LVC	Leishmaniose visceral canina
LVH	Leishmaniose visceral humana
MapBiomass	Projeto de Mapeamento Anual da Cobertura e Uso do Solo no Brasil
MCMC	Monte Carlo em cadeias de Markov – <i>Markov Chain Monte Carlo</i>
PIB	Produto Interno Bruto
SEADE	Fundação Sistema Estadual de Análise de Dados do Estado de São Paulo

LISTA DE SÍMBOLOS

	Independente de
$<$	Menor que
$>$	Maior que
	Para todos
	É um elemento de
$ $	Dado
\propto	Proporcional a
iid	Independentes e identicamente distribuídos
G	Grafo

SUMÁRIO

	INTRODUÇÃO.....	12
1	OBJETIVOS.....	15
2	REFERENCIAL TEÓRICO.....	16
2.1	Leishmania e leishmanioses.....	16
2.2	Leishmaniose visceral.....	16
2.3	Vetor.....	17
2.4	Ciclo biológico.....	18
2.5	Doença em humanos.....	19
2.6	Doença em cães.....	19
2.7	Fatores Climáticos.....	20
2.8	Modificações no ambiente.....	20
2.9	Situação epidemiológica.....	21
2.10	Inferência causal.....	22
2.10.1	Modelo de respostas potenciais.....	23
2.10.2	Efeitos causais na presença de interferência.....	26
2.11	Inferência Bayesiana.....	27
2.11.1	Teorema de Bayes.....	27
2.11.2	Distribuição a priori.....	28
2.11.3	Monte Carlo em cadeias de Markov (MCMC) e amostragem de Gibbs.....	29
2.12	Grafos acíclicos direcionados.....	30
2.13	Grafos de cadeia.....	34
3	METODOLOGIA.....	36
3.1	Caracterização da unidade de análise.....	36
3.2	Dados de desfecho.....	36
3.3	Dados de exposição.....	37
3.4	Covariáveis.....	38
3.4.1	Gasoduto.....	40
3.4.2	Produto Interno Bruto Municipal.....	40
3.4.3	População Humana por Município.....	40

3.5	Notações.....	41
3.6	Assunções.....	42
3.7	Estimação dos efeitos causais.....	44
3.8	Especificações.....	44
4	RESULTADOS.....	45
	DISCUSSÃO E CONCLUSÕES.....	49
	REFERÊNCIAS	52
	APÊNDICE A- Descrição das notações, assunções e método de estimação.....	64
	APÊNDICE B - Descrição das análises de dados de uso do solo e covariáveis..	70
	APÊNDICE C- Descrição do script dos DAGs e variáveis para ajuste.....	73
	APÊNDICE D - script bayes auto-g-computation.....	77

INTRODUÇÃO

As leishmanioses são um grupo de doenças de transmissão vetorial causadas por parasitos do gênero *Leishmania* (Família Trypanosomatidae, ordem Kinetoplastida) que é composto por mais de 20 espécies patogênicas para humanos (ASHFORD, 2000; DESJEUX, 2004).

As leishmanioses apresentam duas formas clínicas principais que variam de acordo com a espécie de *Leishmania*: A leishmaniose cutânea (LC) que causa lesões ulcerativas na pele, mas raramente leva a óbito (PACE, 2014); e a leishmaniose visceral (LV), que é a forma mais severa e letal se não tratada (DESJEUX, 2004; ORGANIZAÇÃO MUNDIAL DA SAUDE, 2010).

A LV é endêmica em mais de 60 países, sendo que mais de 95% dos casos ocorrem em dez países (Bangladesh, Brasil, China, Etiópia, Índia, Quênia, Nepal, Somália, Sudão e Sudão do Sul) (ALVAR et al., 2012; ORGANIZAÇÃO MUNDIAL DA SAUDE, 2018).

No Brasil a LV é causada pelo *Leishmania infantum* (sin. *Leishmania chagasi*), que é transmitido para humanos e outros mamíferos principalmente pela picada do flebótomo *Lutzomyia longipalpis* (ORGANIZAÇÃO MUNDIAL DA SAUDE, 2010). Mamíferos de diversas ordens podem ser reservatórios da *L. infantum*, incluindo roedores, marsupiais, edentados, primatas e canídeos (DESJEUX, 2004; READY, 2010). O cão doméstico (*Canis familiaris*) é considerado o mais importante reservatório em áreas urbanas (QUINNELL E COURTENAY, 2009).

As doenças de transmissão vetorial dependem de complexas interações entre agente infeccioso, vetor, hospedeiro e reservatórios, ocorrendo em um ambiente propício (REISEN, 2010). Assim, a presença de *L. infantum*, seu vetor e reservatórios são favorecidos por algumas condições ambientais e climáticas, tais como climas quentes e áreas de mata (HLAVACOVA et al., 2013; MOŠKOVSKIJ E DUHANINA, 1971). De fato, os componentes do ciclo da LV eram inicialmente encontrados apenas em ambientes naturais e não perturbados por ações humanas (GRIMALDI E TESH, 1993). Nas últimas décadas, no entanto, a expansão das sociedades nos levou a uma maior aproximação com ecossistemas naturais, onde o ciclo de LV já estava presente, aumentando sua exposição a flebotomíneos infectados (DESJEUX, 2001; DUJARDIN, 2006).

Apesar de ser historicamente conhecida como doença endêmica rural, a LV atingiu proporções endêmicas e epidêmicas em grandes cidades brasileiras desde a década de 1980,

representando grave problema de saúde pública (DRUMOND E COSTA, 2011; WERNECK, 2008; ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2010).

O estado de São Paulo é um exemplo desse fenômeno de expansão da LV, associado à expansão geográfica do vetor. Em 1998, o *Lu. longipalpis* havia sido observado em dois municípios, já em 2018 se fez presente em 194 dos 645 municípios do estado (CASANOVA et al. 2015; SECRETARIA DE ESTADO DE SAÚDE DE SÃO PAULO, 2018). Em 1998 e em 1999 foram identificados os primeiros casos caninos e humanos, respectivamente (CAMARGO-NEVES E KATZ, 1999).

A expansão da LV em São Paulo ocorreu ao longo de um eixo que se estendia do Noroeste para o sudeste do estado. O fenômeno foi inicialmente descrito por Antonialli et al. (2007), no Mato Grosso do Sul, e por Cardim et al. (2013), em São Paulo, onde foi observado que a entrada da doença coincidiu com a construção do gasoduto Bolívia-Brasil e se expandiu por toda a sua extensão. A construção, que ocorreu entre 1997 e 1999, fomentou a migração de um grande número de trabalhadores de regiões endêmicas, como os municípios limítrofes com a Bolívia, por exemplo Corumbá-MS (ANTONIALLI et al., 2007; CARDIM et al. 2013). Vários fatores são citados como norteadores da expansão da LV, como variáveis socioeconômicas, climáticas e ambientais, em particular os desequilíbrios ecológicos antropogênicos.

Surtos de doenças infecciosas apresentam dimensões espaciais e temporais (SATTENSPIEL E LLOYD, 2009). As heterogeneidades espaciais podem surgir devido a diferenças nas distribuições de vetores, hospedeiros e outros fatores de risco (WERNECK, 2008) e devem ser levadas em consideração na avaliação do potencial de disseminação de doenças infecciosas.

Apesar dos esforços para modelar fatores associados à LV no Brasil (ARAÚJO et al., 2013; KARAGIANIS-VOULES et al., 2013) e no mundo (GÁLVEZ et al., 2011; GOMEZ-BARROSO et al., 2015), poucos estudos avaliaram a distribuição da LV integrando dados do vetor e os principais hospedeiros (humanos e cães), levando em consideração sua dispersão e fatores ambientais em território brasileiro (ASSUNÇÃO ET AL. 2001, CARDIM ET AL. 2016 E SEVÁ ET AL 2017).

Mais especificamente, no que concerne o efeito do desmatamento e ações antrópicas no ambiente, a maioria dos estudos apresentam algumas limitações, tendendo a ser de natureza qualitativa ou focados puramente em associações estatísticas, sem elaboração/apresentação de modelo teórico sobre o fenômeno de interesse (WILCOX E GUBLER; 2005).

Baseando-se nas informações previamente expostas, o presente trabalho tem por propósito: (i) quantificar o efeito do desmatamento na ocorrência de *Lu. longipalpis*, leishmaniose visceral canina (LVC) e humana (LVH), nos municípios do estado de São Paulo, e (ii) investigar a mudança esperada no risco de presença de vetor/LVC/LVH em mudanças hipotéticas da prevalência do desmatamento no estado de São Paulo.

Um melhor entendimento dos fatores que subjazem a ocorrência da LV serve não somente para contribuir para a compreensão da dinâmica da infecção, como no desenvolvimento de medidas de controle mais efetivas no estado, mas também para informar o risco da ocorrência da doença em outras áreas com condições socioeconômicas, ambientais e climáticas semelhantes.

Para melhor entendimento dos processos analíticos utilizados ao longo do trabalho, essa dissertação foi estruturada em tópicos. Dessa forma, o tópico inicial apresenta o referencial teórico, inicialmente com os subtópicos 2.1 a 2.9 versando detalhes sobre a LV, bem como sua situação no mundo, no Brasil e em São Paulo, além de alguns fatores que considerados associados à sua ocorrência e que serviram de auxílio para construção do modelo teórico. Mais adiante, entre os subtópicos 2.10 a 2.13, foram reunidas algumas fundamentações teóricas básicas sobre metodologia empregada, mais especificamente inferência causal e inferência bayesiana. O tópico três descrever a metodologia empregada no trabalho. O tópico 4 resume os resultados obtidos das análises. Por conseguinte, são discutidos os achados e expostas as principais conclusões. Por último, mas não menos importante, o manuscrito dispõe na seção de apêndices, os scripts utilizados para o desenvolvimento das análises, assim como detalhes dos métodos e resultados.

A leitura sequencial do trabalho é opcional. Aos leitores cujo foco não se restrinja a aspectos metodológicos e estatísticos, recomenda-se pular a leitura dos subtópicos 2.10-2.13 e tópico 3 e retomar a leitura a partir do tópico 4.

1 OBJETIVOS

No presente estudo estamos interessados no efeito do desmatamento (perda de cobertura vegetal) na ocorrência de *Lu. longipalpis*, leishmaniose visceral canina (LVC) e humana (LVH), no estado de São Paulo. Mais especificamente, nós investigamos:

(1) O efeito do desmatamento na ocorrência do vetor/LVC/LVH no estado de São Paulo entre 2000 e 2018;

(2) A mudança esperada no risco de ocorrência de vetor/LVC/LVH de um município, correspondente a uma alteração em seu nível de desmatamento, sob uma distribuição hipotética (contrafactual) do nível de desmatamento de todos os outros municípios do estado (efeito direto);

(3) A mudança esperada no risco de ocorrência de vetor/LVC/LVH de um município, correspondente a uma alteração no nível de desmatamento dos demais municípios da rede (estado de São Paulo), mantendo o nível de desmatamento do município constante (efeito indireto);

(4) A mudança esperada de ocorrência de vetor/LVC/LVH na rede (estado) correspondente a uma alteração hipotética na distribuição do status de desmatamento no estado (e.g., 100% do estado passou por desmatamento em comparação com 0%) (efeito geral).

2 REFERENCIAL TEÓRICO

2.1 A *Leishmania* e as leishmanioses

As leishmanioses são um grupo de doenças parasitárias causadas por protozoários do gênero *Leishmania* e são transmitidos entre hospedeiros mamíferos por flebótomos do gênero *Phlebotomous* e *Lutzomyia* (BATES et al., 2015). As leishmanioses envolvem primariamente ciclos de transmissão zoonóticos envolvendo uma ou mais espécies de reservatórios mamíferos, com exceção de duas espécies (*L. donovani* e *L. tropica*) que são transmitidas por meio de ciclos antroponóticos (ALVAR et al., 2004).

Distintas espécies de *Leishmania* causam diferentes manifestações clínicas, que variam em severidade de lesões cutâneas leves até a doença visceral que pode levar a morte. Em humanos, existem duas formas principais da doença: a cutânea que, dependendo da espécie, pode resultar em cutânea difusa (*L. amazonensis*), recidivans (*L. tropica*) ou mucocutânea (*L. braziliensis*), e a forma visceral (*L. donovani* e *L. infantum*). Sendo que a última, se causada por *L. donovani*, pode eventualmente evoluir para a leishmaniose dérmica pós kala-azar ou PKDL (*Post Kala-azar Dermal Leishmaniasis*) (ALVAR et al., 2004).

O parasito foi descoberto no fim do século XIX, por vários pesquisadores, entretanto foi Sir Ronald Ross (1903) que nomeou o gênero e espécie de *Leishmania donovani*, em homenagem aos achados do patologista escocês William Leishman (1903) e o médico irlandês Charles Donovan (1903). Alguns anos depois, o bacteriologista francês Charles Nicolle (1908) descreveu a *Leishmania infantum*, outra espécie responsável por LV e no mesmo ano, juntamente com Charles Comte, encontrou o parasito em cães na Tunísia (1908). Para mais detalhes sobre a clínica e história das leishmanioses, ver Burza et al. (2018) e Steverding (2017), respectivamente.

2.2 Leishmaniose Visceral

A leishmaniose visceral, também chamada de *kala-azar*, é causada por espécies do complexo *Leishmania donovani* (*L. donovani* [= *Leishmania archibaldi*] no sul da Ásia e leste da África e *Leishmania infantum* [= *Leishmania chagasi*] nas América Central e do Sul, Mediterrâneo, Ásia central e partes da China) (LUKES et al, 2007).

No ambiente silvestre, mamíferos de diversas ordens podem ser reservatórios da LV, incluindo roedores, marsupiais, edentados, primatas e canídeos. Podendo destacar no Brasil a raposas (*Dusicyon vetulus* e *Cerdocyon thous*) e marsupiais (*Didelphis albiventris*) (DESJEUX, 2004; READY, 2010).

Já em áreas urbanas, o cão doméstico (*Canis familiaris*) é considerado, por seu intenso contato com o homem, além da alta infectividade e prevalência, o principal reservatório da LV (ALVAR et al., 2004). A enzootia canina tem precedido a ocorrência de casos humanos e a infecção canina tem sido mais prevalente que no homem (MORENO E ALVAR, 2002; OLIVEIRA et al, 2001). Além disso, estudos apontam que havendo presença do vetor e alta prevalência de cães infectados, maior é a chance de infecção de humanos (MORENO E ALVAR, 2002; QUINNELL E COURTENAY, 2009; MINISTÉRIO DA SAÚDE, 2014).

2.3 Vetor

Os principais vetores da LV são pertencentes à ordem Diptera, família Psychodidae, subfamília Phlebotominae (LAINSON E SHAW, 1987). Cerca de 900 espécies de flebotomíneos já foram identificados em todo o mundo, sendo que nas Américas, 60 espécies foram reconhecidas como vetor ou vetor em potencial da *Leishmania* (PETERS E KILLICK-KENDRICK; 1987). No Brasil, duas espécies foram confirmadas como envolvidas na transmissão da LV, *Lutzomyia longipalpis* (MINISTÉRIO DA SAÚDE, 2014) e *Lutzomyia cruzi* (GALATI, 2018), porém a última ocorre somente no Mato Grosso do Sul.

O *Lu. longipalpis* é amplamente distribuído no Brasil e é encontrado em diversos ecossistemas, variando de ambientes secos como a Caatinga a florestas úmidas como a Amazônia. Apesar da sua distribuição geográfica, uma característica dessa espécie é a capacidade de se adaptar ao ambiente peridomiciliar, particularmente em áreas rurais e áreas periféricas de centros urbanos, em qualquer período do ano (LAINSON E SHAW 1987; GALATI, 2003).

Os flebotomíneos são conhecidos popularmente no Brasil como mosquito palha, birigui, tatuquira e etc. São insetos pequenos, que medem entre 1 e 3 mm de comprimento e possuem o corpo revestido por pelos de coloração castanha (ou cor de palha) (GALATI, 2018). Eles são facilmente reconhecidos pelo seu comportamento de voar em pequenos saltos e pousar com asas entreabertas. Na fase adulta eles se adaptam a vários ambientes, porém em

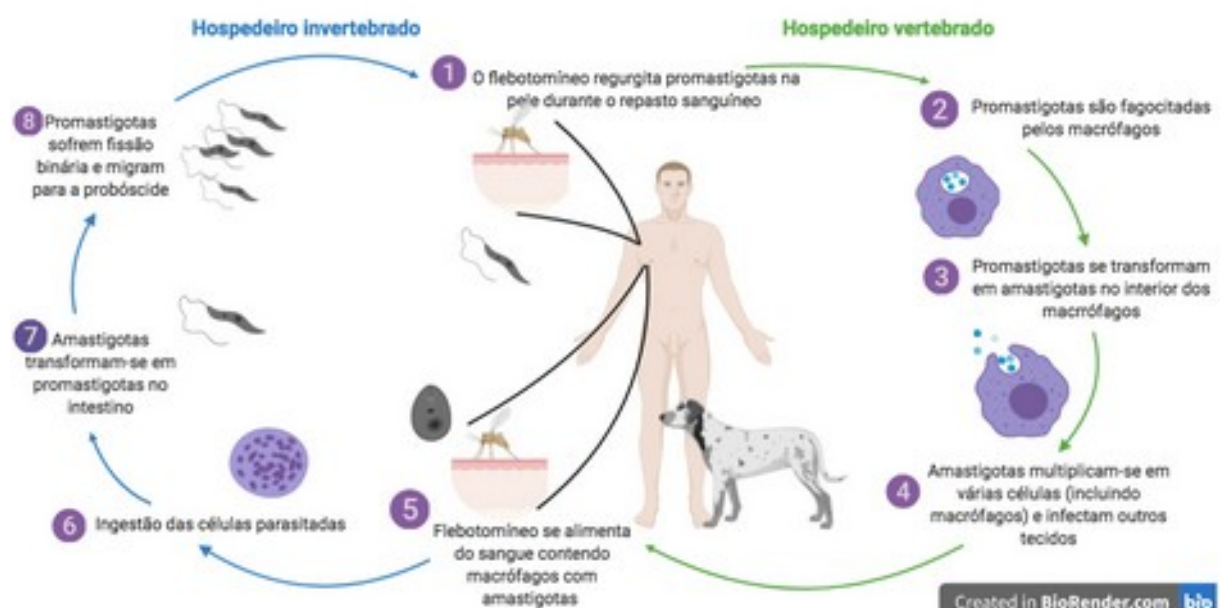
sua fase larvária, desenvolvem-se em ambiente terrestre úmido, com baixa incidência de luz, e rico em matéria orgânica em decomposição, pois a usam para se alimentar (RANGEL, 2005; GALATI, 2003; GALATI, 2018).

Ambos, machos e fêmeas, se alimentam de carboidratos. Eles normalmente deixam seus abrigos para se alimentar ao entardecer ou à noite. Apenas as fêmeas são hematófagas, uma condição necessária para que seus ovários amadureçam. As fêmeas são infectadas quando picam reservatórios infectados e podem, posteriormente, transmitir formas infecciosas da doença picando seres humanos e outros mamíferos (BRAZIL E BRAZIL, 2003; GALATI, 2018).

2.4 Ciclo biológico

A *Leishmania* apresenta duas fases de desenvolvimento: amastigota, dentro dos macrófagos do hospedeiro vertebrado, e promastigota, fase infectante no trato digestivo do flebotomo (DESJEAUX, 2004; READY, 2010). Flebotomos fêmeas adquirem o macrófago infectado com a forma amastigota de *Leishmania* ao se alimentarem do sangue do hospedeiro. No intestino do vetor, os amastigotas se transformam em promastigota, se reproduzem e migram para o aparelho bucal para infectarem o hospedeiro no próximo repasto sanguíneo (Figura 1) (DESJEAUX, 2004; READY, 2010; REITHINGER, 2007).

Figura 1- Ciclo da leishmaniose visceral.



Fonte: O autor.

2.5 Doença em humanos

A infecção pela *L. infantum* é caracterizada por um amplo espectro clínico, podendo variar desde manifestações ausentes (assintomática), discretas, moderadas e graves (MINISTÉRIO DA SAÚDE, 2014). A LV é comumente caracterizada por início com febre, palidez muco-cutânea, emagrecimento, hepatoesplenomegalia, podendo evoluir para edema de membros inferiores e anasarca, além de hemorragia e icterícia (BURZA et al. 2018, ORGANIZAÇÃO MUNDIAL DA SAUDE, 2018). Quando não tratada é quase sempre fatal, com uma letalidade estimada de 95%, causando entre 20.000-40.000 mortes por ano em todo o mundo, podendo esse número ser ainda maior devido a subnotificação e má-classificação (COLLIN et al., 2006; ALVAR et al., 2012; ORGANIZAÇÃO MUNDIAL DA SAUDE, 2018). A LV atinge principalmente crianças, adultos jovens e indivíduos imunodeprimidos (e.g. coinfeção por HIV), sendo este último o grupo com maior mortalidade, juntamente com pessoas com desnutrição e com dificuldades no acesso a tratamento (e.g., conflitos armados, fome, migrações) (ALVAR et al., 2008; ALVARENGA et al., 2010; DESJEUX, 2001; REITHINGER et al., 2007). Até mesmo em pacientes com acesso ao tratamento, a letalidade é cerca de 10 % (ORGANIZAÇÃO MUNDIAL DA SAUDE, 2008).

2.6 Doença em cães

A infecção por *L. infantum* em cães, também tem um amplo espectro clínico que varia de assintomático a grave. Apresenta evolução lenta e início insidioso, evoluindo para sinais semelhantes aos desenvolvidos em humanos (e.g., perda de peso, hepatoesplenomegalia e linfadenopatia), além de alterações cutâneas (e.g., prurido, alopecia, pelagem seca e úlceras crostosas), onicogribose, paresia dos membros, caquexia, inanição e óbito (ALVAR et al., 2004). Uma revisão sobre o aspecto clínico da doença nos caninos pode ser vista em Solano-Gallego e colaboradores (2009).

A carga da doença em cães é difícil de quantificar, pois a prevalência é muito variada e os testes diagnóstico apresentam diferentes sensibilidade (60-100%) e especificidade (77-96%), dificultando a comparação ou sumarização (SILVA et al., 2013). Uma revisão de inquéritos sorológicos na Europa encontrou, numa amostra de 500.000 cães, uma

soroprevalência de 23% (FRANCO et al., 2011), mas estudos sugerem que a verdadeira prevalência pode ser ainda maior (BANETH et al., 2008), com o uso de métodos moleculares (80%) (LACHAUD et al., 2002).

2.7 Fatores Climáticos

Fatores abióticos influenciam o ciclo de transmissão da leishmaniose. Dentre esses fatores, as condições climáticas têm papel capital porque os vetores necessitam de condições específicas para seu desenvolvimento e sobrevivência (MUNSTERMANN, 2019; HLAVACOVA et al., 2013).

Temperaturas quentes são necessárias para adequado desenvolvimento e funcionamento do metabolismo do vetor, além de influenciar no desenvolvimento da *Leishmania* dentro do vetor (HLAVACOVA et al., 2013). Assim, as variações climáticas podem afetar a competência vetorial, desenvolvimento da *Leishmania* e a abundância do vetor, assim como alterar a distribuição e sobrevida de hospedeiros (DESJEUX, 2001; ORGANIZAÇÃO MUNDIAL DA SAUDE, 2018). Além disso, outros fatores climáticos, como a redução da precipitação podem intensificar as vulnerabilidades socioeconômicas, devido à secas e perdas de colheitas, que podem aumentar a migração de infectados de áreas rurais endêmicas para áreas urbanas/periurbanas livres da doença, podendo criar assim novos focos de transmissão (CHAVES E PASCUAL, 2006; RODRIGUEZ-MORALES et al., 2010; ROY et al., 2016; DESJEUX, 2001).

2.8 Modificações no ambiente

Originalmente, a *Leishmania*, seus vetores e reservatórios eram presentes em áreas florestais sem perturbação antrópica (MOŠKOVSKIJ E DUHANINA, 1971). No entanto, a expansão humana levou a uma maior aproximação de ecossistemas naturais, onde o ciclo de LV já estava presente, aumentando a exposição humana a flebotomíneos infectados (DESJEUX, 2001; DUJARDIN, 2006). Essas mudanças no ambiente resultaram em alterações na ecologia do parasita, vetor e reservatório, favorecendo a ocorrência da leishmaniose em áreas urbanas (SHAW, 2007). No século passado, acreditava-se que desmatamento poderia levar a redução da incidência de leishmaniose, por meio do

desaparecimento dos hospedeiros silvestres e redução da população de vetores (ESTERRE et al., 1986). Porém, o processo de desmatamento em conjunto a adaptabilidade dos parasitas e flebotomíneos, especialmente o *Lu. longipalpis*, fez com que a doença continuasse a ser um problema de saúde pública (WALSH et al., 1993).

A extensa mudança que humanos promovem ambiente vem ocorrendo de forma mais rápida e intensa em países em desenvolvimento. Em muitos desses países a leishmaniose está presente, levando à um aumento da sua incidência (DESJEUX, 2001; LAMBIN et al., 2010). Além disso, novas áreas de transmissão surgiram em áreas periurbanas onde a vegetação oferece abrigo para vetores e reservatórios (DUJARDIN 2006; SILVA et al., 2011). Ao mesmo tempo espécies de vetores conseguiram se adaptar ao ambiente urbano e gradualmente deixaram de depender de regiões de mata, permitindo a transmissão da doença de áreas urbanas (SALOMÓN et al., 2015).

Autores já associaram a dispersão do vetor com destruição ou alterações de ambientes naturais (MARZOCHI E MARZOCHI, 1997; ASHFORD, 2000; SHAW, 2007; WALSH et al., 1993). Por exemplo, estudos no Mato Grosso do Sul (ANTONIALLI et al., 2007) e São Paulo (CARDIM et al., 2013; SEVÁ et al., 2017) evidenciaram o papel da construção do gasoduto Bolívia-Brasil, com a expansão da LV. A associação entre o gasoduto e a ocorrência de LV é que sua construção fomentou a migração e movimentação de trabalhadores provindos de áreas endêmicas (ANTONIALLI et al., 2007).

2.9 Situação epidemiológica

A LV é endêmica em mais de 60 países, estimam-se que ocorram 50.000-90.000 novos casos por ano, em todo o mundo (ALVAR et al., 2012). Em 2017, 95% dos casos ocorreram em apenas dez países (Bangladesh, Brasil, China, Etiópia, Índia, Quênia, Nepal, Somália, Sudão e Sudão do Sul) (ORGANIZAÇÃO MUNDIAL DA SAUDE, 2018).

Na América do Sul a doença já foi descrita em mais de 12 países, sendo que 97% dos casos ocorrem no Brasil (ORGANIZAÇÃO PANAMERICANA DA SAÚDE, 2019). O primeiro registro data 1913, quando Migone, em Assunção (Paraguai), descreveu o primeiro caso em necropsia de um paciente oriundo de Corumbá-MS (Brasil) (ALENCAR et al., 1991), sendo novamente relatado em 1934, num estudo de febre amarela, em pacientes no Norte e Nordeste do Brasil (PENNA et al., 1934). A doença se manteve restrita nessas regiões

do país até os anos 90, quando se dispersou e, até 2018, foram relatados casos autóctones em todas as regiões do país (MINISTÉRIO DA SAÚDE, 2019).

Mais de 60 mil casos foram relatados no Brasil no período de 2001 a 2017, resultando em mais de 4.000 óbitos (Brasil, 2019). Na década de 90, 90% dos casos do país ocorreram no Nordeste, já em 2007 a participação da região nos casos totais foi reduzida para 50% (MINISTÉRIO DA SAÚDE, 2014). Em 2018, a transmissão autóctone foi relatada em mais de 4.000 municípios e 24 unidades federativas, com 3466 novos casos e incidência de 5,05 casos/100 mil habitantes (MINISTÉRIO DA SAÚDE, 2019).

No estado de São Paulo, até a década de 90, eram observados somente casos alóctones (SECRETARIA DE ESTADO DE SAÚDE DE SÃO PAULO, 2018). Até que em 1997, na região Oeste do estado, mais precisamente em Araçatuba, foi encontrado o *Lu. longipalpis*, já em 1998 e em 1999 foram relatados o primeiro caso canino e humano autóctone, respectivamente (CAMARGO-NEVES E KATZ; 1999). Até 2013 o vetor foi encontrado 25,3% (163/645) dos municípios do estado, foram confirmados 2204 casos e 192 mortes (CASANOVA et al. 2015; SECRETARIA DE ESTADO DE SAÚDE DE SÃO PAULO, 2018).

2.10 Inferência Causal

Historicamente diversas ciências, como filosofia, passando pelos tratados de Aristóteles, ciências sociais, com Blalock (1971), e medicina (e.g., postulados de Koch e critérios de Hill), se interessam por causalidade, enquanto a estatística somente mais recentemente parece ter se interessado por causalidade (LUIZ E STRUCHINER, 2002).

Holland (1986) cita que, ao se falar de causalidade, o primeiro obstáculo está na diversidade de questões que surgem. Os filósofos têm maior interesse em discussões sobre o significado de causa. Já em áreas como sociologia e medicina, há o interesse na identificação de causas para um dado efeito.

Na epidemiologia essa busca fica claramente evidente. Miettinen e Karp (2012 p.35), por exemplo, comentam que “a pesquisa epidemiológica é, quase exclusivamente, interessada em etiologias de doenças” (“Epidemiological research is, almost exclusively, concerned with etiology of illness”). Nesse contexto, a etiologia se refere a origens causais de doença, mais especificamente, o que causou o início ou a progressão da patogênese. Podemos citar como primeiros exemplos da busca por causalidade na

epidemiologia os postulados de Koch (ou Henle-Koch) (1890) e os “critérios”¹ de Sir Bradford Hill (1965). Porém, ambos apresentam limitações, que acabam por impedir a sua generalização.

Os postulados de Koch², apesar de terem sido por décadas a principal referência para estabelecer causalidade de um agente em uma doença, possuem algumas limitações como (i) assumem somente um fator causal e não leva em consideração múltiplos agentes causais, (ii) o terceiro postulado exige condições de controle, que nem sempre são possíveis, e (iii) são aplicáveis somente para alguns casos de doenças infecciosas (COCHRAN ET AL., 2000; EVANS, 1976).

Já Hill, que foi o primeiro a tentar separar formalmente associação de causalidade, apesar do extenso uso dos critérios (HILL, 1965; CALVERT, 1994; COLDITZ, 1998), comenta que com exceção da temporalidade, nenhum dos outros fatores³ por ele descritos pode ser considerado condição *sine qua non* para julgar se uma associação é causal ou não (HILL, 1965).

De forma diferente, a abordagem utilizada pelos estatísticos tem um cunho mais objetivo e quantitativo. Apesar da constante negação, vide a frase ‘correlação não implica causalidade’ (PEARSON, 1892), o objetivo da maioria, senão toda, inferência estatística é inferir relações causais. Quando um pesquisador constrói um modelo de regressão, ele está interessado em saber se uma variável independente, X , está associada a uma variável dependente, Y . Almejando saber, não se os eventos simplesmente coocorrem, mas se ao alterar valores de X , é possível alterar os de Y .

A abordagem utilizada na estatística para viabilizar a inferência causal, baseia-se na formulação de um modelo construído sob uma ótica contrafactual⁴, comumente chamado de modelo de respostas potenciais ou, como alcunhado por Holland (1986), o Modelo de Rubin.

¹ Também chamados de pontos de vista (do inglês *viewpoints*) de Hill, dado que o autor nunca utilizou a palavra critério de causalidade ao longo da publicação.

² (i). O parasita ocorre em todos os casos da doença em questão e sob circunstâncias que podem explicar as alterações patológicas e o curso clínico da doença.
(ii) Não ocorre em nenhuma outra doença como um parasita fortuito e não patogênico.
(iii) Depois de ser devidamente isolado do corpo e cultivado repetidamente em pura cultura, pode induzir a doença novamente.

³ (I) Força da associação; (II) Consistência; (III) Especificidade; (IV) Temporalidade; (V) Gradiente biológico; (VI) Plausibilidade; (VII) Coerência; (VIII) Evidência experimental e (IX) Analogia.

⁴ O que significa que o modelo é baseado em valores hipotéticos, que não são totalmente observados.

2.10.1 Modelo de respostas potenciais:

A ideia de conceitualizar causalidade em termos de estados contrafactuais remonta pelo menos a Hume (1748). Mais formalmente, na literatura estatística, notações para abordagem contrafactuais foram dadas por Neyman (1923) em sua tese de doutorado, no contexto de experimentos randomizados na agricultura. Essa abordagem foi posteriormente desenvolvida por Rubin (1974) e ampliada para estudos observacionais. Mais tarde a mesma abordagem foi estendida para múltiplas exposições e exposições que variam no tempo por Robins (1986) e relacionada com representações gráficas (grafos acíclicos direcionados, ie., DAG's) por Spirtes et al. (1993) e Pearl (1995).

A lógica contrafactual envolve teóricas viagens no tempo para remover um fator causal e demonstrar que, na sua ausência, o efeito (i.e., doença) não ocorre. Na epidemiologia, o ensaio clínico randomizado cruzado (*cross-over RCTs*) emula o método contrafactual, permitindo a comparabilidade das unidades experimentais. Elas são selecionadas de forma que os grupos (controle e tratamento) se assemelhem em todas as condições, exceto na exposição ao suposto fator causal.

Porém, estudos de natureza experimental nem sempre podem ser realizados, seja por dificuldades práticas (e.g., custos), ou éticas (e.g., estudar a relação entre cigarro e câncer). Dessa maneira, podemos imaginar um estudo observacional como um experimental imperfeito, desde que certos pressupostos sejam assumidos.

Digamos que temos, n indivíduos em uma população N , onde $n \in N$, $N = \{i, j, \dots, n\}$. Um conjunto de agentes causais t com valores 0 (tratamento) e 1 (controle), $t = (\text{pertence}) \{0, 1\}$. Por último, um desfecho, Y , dicotômico, $Y = (\text{pertence}) \{0, 1\}$.

O efeito causal do tratamento⁵, sobre o indivíduo i , relativo ao controle, é calculado pela diferença entre $Y^{t=1}(i)$ e $Y^{t=0}(i)$, isto é:

$$Y^{t=1}(i) - Y^{t=0}(i)$$

Porém, é fácil perceber que, para o mesmo indivíduo i , somente é possível observar uma das respostas, a chamada factual. Essa limitação foi chamada por Holland (1986) como o problema fundamental da inferência causal, porém no mesmo trabalho ele assinala algumas soluções para o problema. Uma delas é, ao invés de se estimar o efeito causal a nível

⁵ Ao longo da dissertação utilizaremos o termo tratamento, porém não nos referindo exclusivamente a ensaios clínicos, somente tentamos manter a nomenclatura mais comumente utilizada na literatura. O termo tratamento pode ser cambiado por causa ou exposição em qualquer parte do texto.

individual, estimar o efeito a nível populacional ou subpopulacional, onde o efeito causal médio é a diferença entre a esperança de $Y^{t=1}$ e a esperança de $Y^{t=0}$. Ou seja:

$$E(Y^{t=1}) - E(Y^{t=0}).$$

Mas a utilidade das soluções para o problema fundamental da inferência causal, destacadas por Holland (1986), depende de diferentes conjuntos de suposições que serão resumidas a seguir.

Digamos que temos um estudo observacional⁶ onde o interesse é estimar o efeito causal médio de um tratamento T sobre um desfecho Y em n indivíduos, sob C fatores de confusão⁷. Para identificação do efeito causal, são necessárias algumas suposições⁸:

(1) Consistência, também chamado de versão única do tratamento (*single version of treatment*), ou seja, o tratamento deve ser bem definido e específico, não havendo diferentes formas ou versões do tratamento. Mais especificamente, assumimos um indivíduo i com o tratamento observado T igual a t , tem o desfecho Y igual ao desfecho potencial Y^t . Ou seja:

$$Y_i(T) = Y_{t_i} = Y_i, \text{ se } T_i = t$$

(2) Ignorabilidade (a.k.a., permutabilidade condicional), dadas as covariáveis, C , a atribuição do tratamento é independente dos desfechos potenciais. Também chamada de ‘não ausência de confundimento’ (*no unmeasured confounding*):

$$Y_i \perp T_i \vee C_i$$

(3) Positividade, simplesmente diz que todos indivíduos têm probabilidade não nula de receber o tratamento, ou seja:

$$P(T=t|C=c) > 0 \text{ para todos } t \text{ e } c$$

(4) Estabilidade (ou não interferência), ou seja, o desfecho de um indivíduo, é independente do tratamento, ou desfecho, de outro indivíduo:

$$Y_i \perp T_j$$

⁶ Para simplificação, utilizaremos tratamento e desfecho binário

⁷ Em epidemiologia, fator de confusão é uma variável relacionada ao desfecho e a exposição

⁸ Para mais detalhes sobre o modelo de respostas potenciais e suas suposições, nós recomendamos Hernan e Robins (2006, 2020).

Porém, em algumas situações esses pressupostos não podem assumidos ou mesmo testados. Como comentado em Halloran e Struchiner (1991, 1995), um aspecto fundamental das doenças infecciosas é a transmissão do agente infeccioso de um indivíduo para outro, de forma que a ocorrência da doença em um indivíduo depende de quem mais está infectado na população ao qual o indivíduo pertence. Esse fenômeno, chamado por Ross (1916) de acontecimento dependentes, tem consequência direta na aplicação do modelo de Rubin, graças a violação da assunção de estabilidade/não interferência. Porém alguns estudos, como será destacado a seguir, estão avançando de forma a permitir o relaxamento dessa assunção.

2.10.2 Identificação de efeitos causais presença de interferência

Como exposto no tópico anterior, tradicionalmente, os métodos estatísticos para inferir efeitos causais assumem dados *iid*⁹ e não interferência entre as unidades de observação (COX 1958, RUBIN 1974).

Somente décadas depois, alguns estudos permitiram considerar situações onde a não interferência não é necessária, por meio da assunção de interferência parcial, onde a interferência é observada somente em subgrupos ou blocos de indivíduos. Essa assunção é apropriada para grupos bem definidos, separados no espaço ou no tempo (GRAHAM, 2008; HUDGENS E HALLORAN, 2008; SOBEL, 2006; TCHETGEN TCHETGEN E VANDERWEELE, 2012).

Porém em contextos, como doenças infecciosas, a dependência entre as unidades impossibilita a definição *clusters* não sobrepostos. Apenas recentemente Tchetgen Tchetgen et al. (2017) propuseram um algoritmo, nomeado *auto-g-computation*, para avaliar os efeitos causais (total, direto e indireto) considerando as interações das unidades como uma rede e que pudesse ser representada na forma de um grafo de cadeias.

Entretanto, essa abordagem sofre de algumas limitações quanto a convergência e eficiência dos estimadores (BESAG, 1977; GUYON E KUSH, 1992; FULCHER 2019). Para contornar essas limitações, Fulcher (2019) desenvolveu uma extensão da abordagem de Tchetgen Tchetgen e colaboradores (2017) em um *framework* bayesiano para estimação e inferência nomeado *Bayes auto-g-computation algorithm*.

Os detalhes sobre o algoritmo desenvolvido por Fulcher (2019) e que foi utilizado por nós, será descrito no próximo tópico, na seção de métodos e no Apêndice A. A seguir

⁹ Independentes e identicamente distribuídos (*independent and identically distributed*)

dissertaremos um pouco sobre a inferência bayesiana que foi utilizada para estimação no algoritmo.

2.11 Inferência Bayesiana

Existem, no ramo da inferência estatística, duas visões principais de como se aprender os valores dos parâmetros em um sistema estocástico (i.e., sistemas que não são totalmente previsíveis): a inferência frequentista (também chamada de clássica) e a inferência bayesiana.

Discussões filosóficas acerca das duas abordagens vão além do escopo dessa dissertação. Porém, ao longo dessa seção, de forma introdutória, será brevemente comentado as principais diferenças entre as duas vertentes.

Ambas as formas veem os dados como realizações de sistemas estocásticos que contém um ou mais processos aleatórios. Porém, na estatística clássica, as quantidades usadas para descrever esses processos aleatórios (e.g., parâmetros) são considerados fixos e constantes desconhecidas, enquanto na estatística bayesiana, os parâmetros são vistos como realizações não observadas de processos aleatórios. Na estatística clássica, a incerteza é descrita em termos de frequência relativa (daí o nome frequentista) de um evento contido em um infinito número de ensaios hipotéticos baseados em um banco de dados (i.e., amostra). Na estatística bayesiana, a incerteza é avaliada usando a distribuição posterior do parâmetro, que é a distribuição de probabilidade condicional de todas quantidades desconhecidas (e.g., parâmetros), dado os dados (amostra), o modelo (verossimilhança) e o nosso conhecimento prévio (*priori*) sobre essas quantidades antes de conduzir a análise (BOLSTAD E CURRAN, 2016; BERNARDO E SMITH, 2000).

Em outras palavras, na estatística frequentista, a probabilidade é relativa à frequência de uma quantidade (e.g., parâmetro) dos dados observados. Em contraste, na estatística bayesiana, a probabilidade é usada para expressar a incerteza sobre a provável magnitude de um parâmetro, onde nenhuma replicação hipotética dos dados é necessária.

A estatística bayesiana vem se mostrando uma poderosa ferramenta com um vasto campo de aplicações, como genética (BEAUMONT, 2004), linguística (FRANK, GOODMAN; 2012), evolução (GEISLER E DIEHL, 2002), ecologia (PARENT E RIVOT, 2012), epidemiologia (LAWSON, 2013) e até nos trabalhos do fictício detetive Sherlock

Holmes (KADANE; 2009). Historicamente, métodos bayesianos foram aplicados por Alan Turing na decodificação dos códigos alemães na segunda grande guerra (GOOD, 1979).

2.11.1 Teorema de Bayes

O reverendo Thomas Bayes (1702-1761) formulou o teorema que hoje leva seu nome, descrito no artigo *An Essay Towards Solving a Problem in the Doctrine of Chances*, publicado após a sua morte, por Richard Price. Ele demonstrou como a probabilidade inversa pode ser usada para calcular a probabilidade de eventos antecedentes a partir da ocorrência do evento seguinte. Esse método foi adotado por Laplace (1814) e outros cientistas no século 19 e a partir do século 20 principalmente por de Finetti (1971), Jeffreys (1961) e Savage (1954), que desenvolveram métodos completos de inferência estatística baseada no teorema de Bayes.

O teorema de Bayes, também chamado de regra de Bayes descreve a relação entre duas probabilidades condicionais $p(A|B)$ e $p(B|A)$, onde “|” lê-se como “dado”:

$$p(A|B) = \frac{p(B \vee A) p(A)}{p(B)} \quad (1)$$

Essa equação é utilizada para derivar a distribuição de probabilidade dos parâmetros θ , dado os dados x , ou seja, a distribuição posterior $p(\theta|x)$:

$$p(\theta \vee x) = \frac{p(x \vee \theta) p(\theta)}{p(x)} \quad (2)$$

Onde a distribuição posterior $p(\theta|x)$ é proporcional ao produto da função de verossimilhança $p(x|\theta)$ e a distribuição a priori do parâmetro $p(\theta)$. Para fazer esse produto uma função de distribuição de probabilidade, com integral igual a 1, uma constante normalizadora $p(x)$ é utilizada como denominador. Ignorando o denominador, que é somente uma constante normalizadora e não envolve os parâmetros θ , a regra de Bayes pode ser resumida como

$$\text{Distribuição posterior} \propto \text{Verossimilhança} \times \text{Distribuição a priori} \quad (3)$$

Onde \propto lê-se como “proporcional à”. Desse modo, a inferência Bayesiana usa as leis de probabilidade para combinar informação sobre o parâmetro θ contido nos dados x , como quantificado na função de verossimilhança $p(x|\theta)$.

2.11.2 Distribuição a priori

Na análise Bayesiana, a distribuição a priori é utilizada para representar o conhecimento prévio sobre os parâmetros desconhecidos, antes de avaliar-se os dados.

A escolha da distribuição a priori para um parâmetro pode ser feita de diversas formas:

- a) Podemos considerar, por exemplo, quando o objeto de estudo é uma doença, alguma medida de sua ocorrência na população (e.g., prevalência);
- b) Podemos definir uma priori baseada na incorporação do conhecimento de especialistas;
- c) Quando dados de ocorrência ou consenso de especialistas não ocorrem e queremos ser conservadores, podemos considerar distribuições a priori não informativas;

Para outras formas ou maior detalhamento, recomendamos Box e Tiao (1973).

2.11.3 Monte Carlo em cadeias de Markov (MCMC) e amostragem de Gibbs

Para se obter amostras da distribuição posterior é necessário resolver integrais múltiplas complicadas e que, na quase totalidade das vezes, exige o uso de métodos numéricos ou de aproximações de integrais, principalmente quando a dimensão do vetor de parâmetros é grande. Surge daí a necessidade do uso de poderosos métodos computacionais, como os métodos de Monte Carlo em cadeias de Markov (MCMC).

MCMC é um conjunto de técnicas para simular amostras da distribuição posterior $p(\theta|x)$ dado o modelo, a verossimilhança $p(x|\theta)$ e os dados x , usando sequências dependentes de variáveis aleatórias. MCMC foi desenvolvido pelos físicos Metropolis et al. (1953), e depois generalizado por Hastings (1970) logo um dos principais algoritmos de MCMC é conhecido por Metropolis-Hastings. Existem várias formas de MCMC atualmente para mais detalhes ver Bernardo e Smith (2000), Gamerman e Lopes (2006) e Gelfand e Smith (1990).

Um dos métodos mais amplamente utilizados, e que foi utilizado na nossa pesquisa, é a amostragem de Gibbs (do inglês *Gibbs sampling*) (GEMAN, GEMAN; 1984; GELFAND, SMITH; 1990). Uma analogia para fácil entendimento foi descrita por Kéry (2010), partindo da ideia que temos um grande problema para resolver (i.e., uma amostra da distribuição a posteriori de n parâmetros), ao invés de tentar resolvê-lo de uma vez só, é mais eficiente dividi-lo em subunidades (k subunidades) e solucionar (i.e., amostrar) cada subunidade k de cada vez.

Por exemplo, temos nossos dados x e nosso vetor de parâmetros desconhecidos θ , consistindo em k subcomponentes $\theta = (\theta_1, \theta_2, \dots, \theta_k)$, portanto, queremos estimar k parâmetros.

1. Escolhemos valores iniciais: $\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_k^{(0)}$

2. Geramos $\theta_1^{(1)}$ a partir de $p(\theta_1 \vee \theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_k^{(0)}, x)$

Amostramos $\theta_2^{(1)}$ a partir de $p(\theta_2 \vee \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)}, x)$

...

Amostramos $\theta_k^{(1)}$ a partir de $p(\theta_2 \vee \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)}, x)$

3. Repetimos o passo 2 várias vezes (e.g., 100 vezes, 1000 vezes) e eventualmente obtemos amostras de $p(\theta|x)$

O passo 2 é chamado de atualização (*update*) ou iteração da amostragem de Gibbs e após a convergência, resulta em uma amostra contendo k valores da probabilidade conjunta da distribuição posterior $p(\theta|x)$. A sequência de amostras aleatórias para cada um dos k parâmetros resultantes do passo 3 forma uma cadeia da Markov (BEST E RICHARDSON, 2005).

Após a geração de amostras da distribuição posterior, utilizamos essas amostras para obter estimadores de Monte Carlo nos permitindo calcular medidas sumárias a posteriori, como a média, desvio-padrão e intervalos de credibilidade.

Conforme mencionado no início da sessão, visamos somente uma breve introdução sobre a estatística Bayesiana e destacar algumas das vantagens de sua utilização. Para maior detalhamento da inferência bayesiana, é recomendada a leitura dos livros-texto de Kruschke (2014), Bernardo e Smith (2000), Kéry (2010), Gamerman e Lopes (2006) e Gelfand e Smith (1990).

2.12 Grafos Acíclicos Direcionados

Grafos acíclicos direcionados (*Directed Acyclic Graphs* ou, simplesmente, DAGs), que compõem a teoria dos diagramas causais, são uma ferramenta que está cada vez mais em uso na epidemiologia. A teoria dos diagramas causais pode ser utilizada em diversas questões na pesquisa epidemiológica, como a identificação de parâmetros causais e aspectos relacionados à validade externa e a síntese dos resultados (PEARL 1995; 2000; PEARL E ROBINS, 1995).

Esta ferramenta também tem sido utilizada para demonstrar em quais situações o uso de critérios e técnicas tradicionais para controlar confundimento e viés de seleção podem levar a conclusões inadequadas (HERNÁN et al., 2002; 2004; ROBINS, 2001). Além disso, os DAGs facilitam o entendimento de algumas controvérsias presentes na epidemiologia, como o paradoxo de Simpson e o viés de Berkson (PEARL, 2014; ROTHMAN et al., 2008).

Os DAGs, também chamados de redes bayesianas, são uma forma gráfica de explicitar as relações entre variáveis em estudo de forma qualitativa, que muitas vezes são utilizados para explicitar relações causais. A partir da assunção de nosso modelo teórico estar correto, o DAG tem uma dupla função: (1) nos ajuda a observar se a premissa de ignorabilidade (*no unmeasured confounding*) é correta; além de ser essencial na (2) identificação de quais variáveis devem ser incluídas no ajuste para evitar estimativas enviesadas.

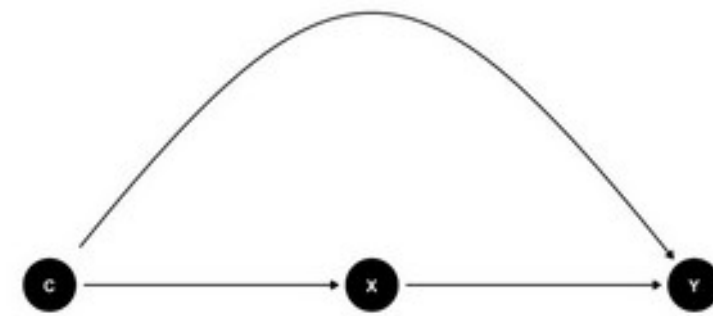
Um DAG é composto de dois elementos: nó e aresta. Onde cada nó representa uma variável, as arestas formalizam as relações entre as variáveis. Já a ausência de aresta, a assunção mais forte em um DAG, representa a total assunção de não efeito entre as variáveis.

Além disso, os DAGs são acíclicos, o que significa que uma variável não pode ser conectada a ela mesma. Um DAG causal, também chamado de diagrama causal, é um DAG onde as arestas representam efeitos causais e que inclui todas as causas comuns (i.e., variáveis de confusão) de qualquer par de variáveis presentes no grafo¹⁰. Usaremos o diagrama¹¹ na Figura 2 para fins de ilustração dos componentes de um DAG.

Figura 2 - Grafo acíclico direcionado hipotético

¹⁰ Ao longo do tópico, os termos DAGs e diagramas causais serão utilizados intercambiavelmente.

¹¹ O diagrama na figura 2, bem como todos os seguintes, foram gerados por meio do pacote `ggdag`(Barrett, 2018).



Fonte: O autor.

As três arestas representam, respectivamente, o efeito de C em X , C em Y e X em Y .

Para descrever a relação entre variáveis em um DAG, são usadas nomenclaturas de parentesco e ancestralidade. Filhos e pais se referem a relações causais diretas, como $X \rightarrow Y$, onde X é pai (ou causa) de Y e, equivalentemente, Y é filho (ou efeito) de X . Descendentes e ancestrais se referem a quaisquer variáveis ao longo de um diagrama, por exemplo no caminho causal $C \rightarrow X \rightarrow Y$, a variável C é ancestral de X e Y , por outro lado, X e Y são descendentes de C .

Em estudos empíricos, estamos interessados no efeito de uma variável dita independente¹² sobre uma variável que consideramos dependente¹³. Todas as variáveis restantes no diagrama são chamadas covariáveis.

As covariáveis podem ser classificadas como confundidoras, mediadoras e colisoras. Variável confundidora é aquela que é ancestral da exposição e do desfecho, como a variável C da figura 2. Mediadoras são variáveis que estão no caminho causal entre a exposição e o desfecho, por exemplo, a variável X media o efeito entre C e Y no caminho causal $C \rightarrow X \rightarrow Y$. Já uma variável colisor é aquela que tem duas setas convergentes, ou seja, é efeito da exposição e do desfecho, como Y em $C \rightarrow Y \leftarrow X$.

Um caminho em um DAG é qualquer sequência de arestas conectando uma variável a outra, independente da direção da seta. Observando o diagrama representado na Figura 2, podemos identificar vários caminhos como:

- (1) $C \rightarrow X$,
- (2) $X \rightarrow Y$,
- (3) $C \rightarrow Y$,
- (4) $Y \leftarrow X$,
- (5) $Y \leftarrow C$,

¹² Também chamado de exposição, tratamento ou causa

¹³ Também conhecido por efeito ou variável dependente

$$(6) C \rightarrow X \rightarrow Y,$$

$$(7) C \rightarrow Y \leftarrow X,$$

$$(8) X \leftarrow C \rightarrow Y,$$

$$(9) X \rightarrow Y \leftarrow C$$

Nos diagramas causais, os caminhos podem ser classificados em:

- i. Cadeias: são caminhos causais diretos, como por exemplo o caminho $C \rightarrow X \rightarrow Y$ da Figura 2, onde cada variável é causadora da próxima formando uma cadeia causal (C causa X e X causa Y).
- ii. Garfos: são caminhos onde duas variáveis possuem uma causa/antecessor comum, como em $X \leftarrow C \rightarrow Y$, onde ambos X e Y são causados por C .
- iii. Garfos invertidos (caminho com variáveis colisoras): são o oposto de garfos, onde duas arestas encontram um único nó, representando um efeito/descendente comum. Como no caminho $C \rightarrow Y \leftarrow X$, onde C e X causam uma variável Y .

Um caminho é considerado desbloqueado (ou aberto), quando a informação flui de uma ponta a outra do caminho (i.e., quando os caminhos são cadeias ou garfos), ou bloqueado, quando o fluxo de informação, ou dependência, é interrompido em algum ponto ao longo do caminho. Se todos os caminhos causais entre duas variáveis são bloqueados, então as variáveis são direccionalmente separadas ou d-separadas (do inglês *d-separated*), o que significa que elas são estatisticamente independentes.

Um caminho causal é naturalmente bloqueado quando há uma variável colisor no caminho (i.e., um caminho de garfo invertido). Na Figura 2, o caminho de C a X ($C \rightarrow Y \leftarrow X$) é bloqueado pela variável colisor Y . Um caminho desbloqueado pode ser bloqueado se ajustarmos/condicionarmos qualquer variável não colisor (PEARL 1995, 2000).

No nosso exemplo da Figura 2, digamos que desejamos estimar o efeito causal de X em Y , mas existem outros caminhos indiretos, também chamado de porta de trás ou porta dos fundos (do inglês *backdoor path*) onde há uma aresta apontando para X , ao invés de saindo dele. Isso causa uma situação de confundimento ou, aproximando da linguagem epidemiológica, evidencia um fator de confusão (do inglês *confounding*). Portanto, para estimar corretamente o efeito de X em Y devemos bloquear todos os caminhos, que não o de interesse, para estimar corretamente o fluxo causal. Na Figura 2, por exemplo, o conjunto mínimo para ajuste inclui somente a variável de confusão C .

Existe uma vasta literatura sobre diagramas causais, aqui citamos ao leitor, Judea Pearl, que desenvolveu muito da teoria dos diagramas causais, com diversas publicações que variam em dificuldade técnica, de introduções amigáveis como o seu livro mais recente *The Book of Why* (PEARL et al, 2018) até mais técnicos, como *Causal Inference in Statistics: A primer* (PEARL et al, 2016). Já para uma aproximação maior com epidemiologia, há uma excelente revisão em português escrita por Cortes e colegas (2016) e o capítulo sobre diagramas causais no livro-texto *Modern Epidemiology* (ROTHMAN et al 2008).

2.13 Grafos de cadeia

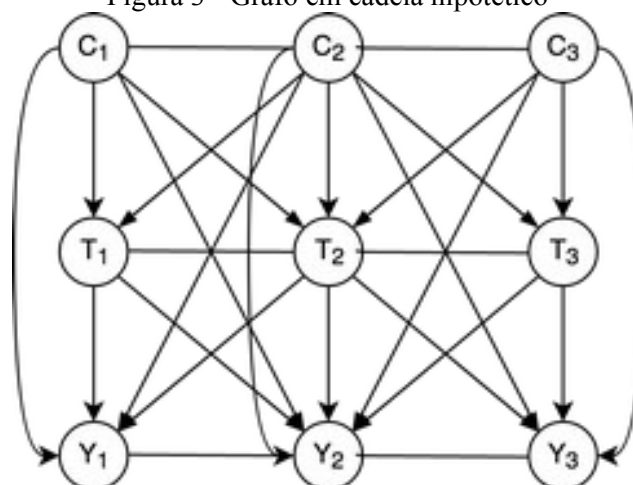
Grafos de cadeia (CG) são representações gráficas bastante semelhantes a um DAG, porém com diferentes aplicações e pertencentes a outra teoria, a teoria dos grafos. Eles são chamados de grafos mistos, pois contém arestas direcionadas (\rightarrow) e não direcionadas ($-$) (LAURITZEN, 1996). Ao contrário dos DAG's, a rede de variáveis (ou indivíduos) conectados pode ser cíclica, não assumindo ordem ou temporalidade entre as variáveis, porém satisfazendo a propriedade de Markov (MAATHUIS et al., 2019).

A propriedade de Markov é um processo estocástico, onde as distribuições de probabilidade dos dados de um indivíduo i , numa rede de k indivíduos, depende somente dos valores dos vizinhos de i , de forma que indivíduos não adjacentes a i são irrelevantes nas inferências feitas para i (KINDERMANN, 1980).

Utilizaremos a Figura 3 para ilustrar um CG no nosso contexto de redes. Chamemos o nosso grafo de G_ε , que apresenta três unidades (i.e., indivíduos), ou seja $G_\varepsilon = \{1, 2, 3\}$, e cada indivíduo contém três variáveis $\{C, T, Y\}$ e dois blocos de adjacências, $\varepsilon = \bar{i}$, onde ε corresponde ao conjunto de pares na rede (i.e., conjunto de indivíduos conectados).

O grafo possui conexões com arestas direcionadas e não direcionadas. Arestas direcionadas, como as que conectam C_1 a T_1 e T_2 , representam que as covariáveis do indivíduo (i.e., município) 1 possam ser antecedentes e causas diretas do tratamento na unidade 1 e também para o tratamento do vizinho, 2, ou seja $C_1 \circledast \{T_1, T_2\}$. As setas $C_1 \circledast \{Y_1, Y_2\}$ e $T_1 \circledast \{Y_1, Y_2\}$ estão presentes em G_ε pelas mesmas razões. Já as arestas não direcionadas, como em $C_1 - C_2$, representam uma associação contemporânea entre o par de variáveis, ou seja, não assumimos direção de causalidade e temporalidade.

Figura 3 - Grafo em cadeia hipotético



Fonte: O autor

Sob essa estrutura gráfica, foi observado por Shpitser et al., (2017) que os desfechos (Y) e covariáveis (C) podem ser vistos, como uma realização de um campo aleatório condicional de Markov e um campo aleatório de Markov (do inglês *Markov Random Field* - MRF), respectivamente, permitindo assim quantificar os efeitos causais sob um contexto de interferência entre unidades de observação.

Para detalhes sobre a teoria dos grafos, grafos de cadeia, modelos gráficos ler Maathuis et al. (2019); Kindermann (1980) e Lauritzen (1996).

3 METODOLOGIA

3.1 Caracterização da unidade de análise

O estado de São Paulo localiza-se no Sudeste do Brasil e é dividido em 645 municípios (IBGE, 2018). O estado é responsável por 33,5% do PIB nacional e possui 22% da população do país. Áreas urbanas e destinadas à agropecuária são cercadas por remanescentes de vegetação nativa de Mata Atlântica e Cerrado, que foram reduzidos de mais de 80% para somente 3% da área do estado nos últimos dois séculos (VICTOR et al., 2005).

A população de estudo consistiu em todos os municípios do estado de São Paulo que não apresentavam os desfechos (*Lu. longipalpis*/LVC/LVH) na linha de base (ano de 2000), resultando em $n=620$. Foi gerada uma rede com os municípios, onde a conexão era dada se dois municípios faziam fronteira. Para tal, foram utilizados o arquivo *shapefile* (.shp) do estado de São Paulo, disponível no site do IBGE e os pacotes *maptools* e *spdep* do R (BIVAND, 2014; BIVAND et al., 2015).

3.2 Dados de desfecho

Dados sobre casos caninos e humanos de LV e presença do vetor da LV foram obtidos para cada município e ano, no período de 2000-2018. As informações foram tratadas de forma binária (0=ausente, 1=presente).

No Brasil, a LV humana é uma doença de notificação compulsória. Os dados de casos confirmados humanos foram obtidos do website¹⁴ do Centro de Vigilância Epidemiológica da Secretaria de Estado de Saúde de São Paulo (CVE/SES-SP) (SECRETARIA DE ESTADO DE SAÚDE DE SÃO PAULO, 2018).

Desde o início das atividades do programa de vigilância e controle da leishmaniose visceral no estado de São Paulo, criado em 1999, todos os municípios devem notificar o primeiro caso canino confirmado (SECRETARIA DE ESTADO DE SAÚDE DE SÃO PAULO, 2018). Assim, os dados do presente estudo foram obtidos a partir de inquéritos sorológicos caninos realizados pelos municípios e confirmados pelo Instituto Adolfo Lutz e CVE/SES-SP.

¹⁴ <http://www.saude.sp.gov.br/cve-centro-de-vigilancia-epidemiologica-prof.-alexandre-vranjac/areas-de-vigilancia/doencas-de-transmissao-por-vetores-e-zoonoses/agrivos/leishmaniose-visceral/dados-estatisticos>

Já os dados de presença de *Lu. longipalpis* foram obtidos a partir de publicações anteriores e, principalmente, de coleções entomológicas realizadas pela CVE/SES-SP durante as atividades de vigilância. As ações de vigilância são feitas por meio de busca ativa, com coleções anuais ou semestrais com armadilhas de luz do tipo CDC em domicílios urbanos de todos os municípios considerados silenciosos (i.e., sem histórico do vetor) (CASANOVA et al., 2015; SECRETARIA DE ESTADO DE SAÚDE DE SÃO PAULO, 2018, SÃO PAULO, 2006).

3.3 Dados da exposição

Os dados referentes a cobertura vegetal foram obtidos do Projeto de Mapeamento Anual da Cobertura e Uso do Solo no Brasil (MAPBIOMAS, 2018). O MapBiomas é uma iniciativa colaborativa que envolve mais de vinte instituições e produz mapas anuais de cobertura e uso do solo do Brasil (mais detalhes sobre o [projeto](http://mapbiomas.org/)¹⁵ e [metodologia](https://storage.googleapis.com/mapbiomas/base-de-dados/metodologia/colecao-3_0/1-ATBD-Collection-3-version-1.pdf)¹⁶). Os mapas de alta resolução (resolução de 30m) são produzidos por meio de algoritmos de aprendizagem de máquina através da plataforma *Google Earth Engine* a partir de imagens dos satélites *Landsat*¹⁷ e disponíveis no *Google Cloud Storage*.

Os dados do MapBiomas são de acesso livre e originalmente disponíveis em formato *raster*¹⁸. Eles são classificados em mais de 20 categorias de uso do solo para o período de 1985 a 2018. Nós utilizamos os dados para o estado de São Paulo no período de 2000 a 2018 e calculamos a porcentagem de área de floresta por município e ano:

$$For_{mun}^{year} = \left(PF_{mun}^{year} / \sum P_{mun}^{year} \right) * 100 \quad (4)$$

¹⁵ <http://mapbiomas.org/>

¹⁶ https://storage.googleapis.com/mapbiomas/base-de-dados/metodologia/colecao-3_0/1-ATBD-Collection-3-version-1.pdf

¹⁷ De acordo com as análises de acurácia realizadas a acurácia geral dos dados gerados no MapBiomas é de 88.7%.

¹⁸ um raster consiste basicamente numa matriz de células (ou pixels) organizados em linhas e colunas, onde cada célula contém valores representando uma informação (no nosso caso, a classificação do uso do solo).

onde PF_{mun}^{year} representa a quantidade de *pixels*¹⁹ de floresta para cada município e ano, e P_{mun}^{year} representa o total de pixels para cada município. Calculamos a variação de área de floresta para o período:

$$\Delta_{mun}^{For} = For_{mun}^{18} - For_{mun}^{00} \quad (5)$$

e, por fim, classificamos os municípios com $\Delta_{mun}^{For} < 0$, como desmatados (1), e com $\Delta_{mun}^{For} \geq 0$, como não desmatados (0).

Nesse processo, foram utilizados os pacotes *raster* (HIJMANS et al., 2015), *tidyverse* (WICKHAM, 2017) e *sp* (PEBESMA E BIVAND, 2015) do *software R* (R CORE TEAM, 2013). O relatório com o script utilizado para as análises foi gerado no *R Markdown* (ALLAIRE et al., 2016) e pode ser visto no Apêndice B.

3.4 Covariáveis

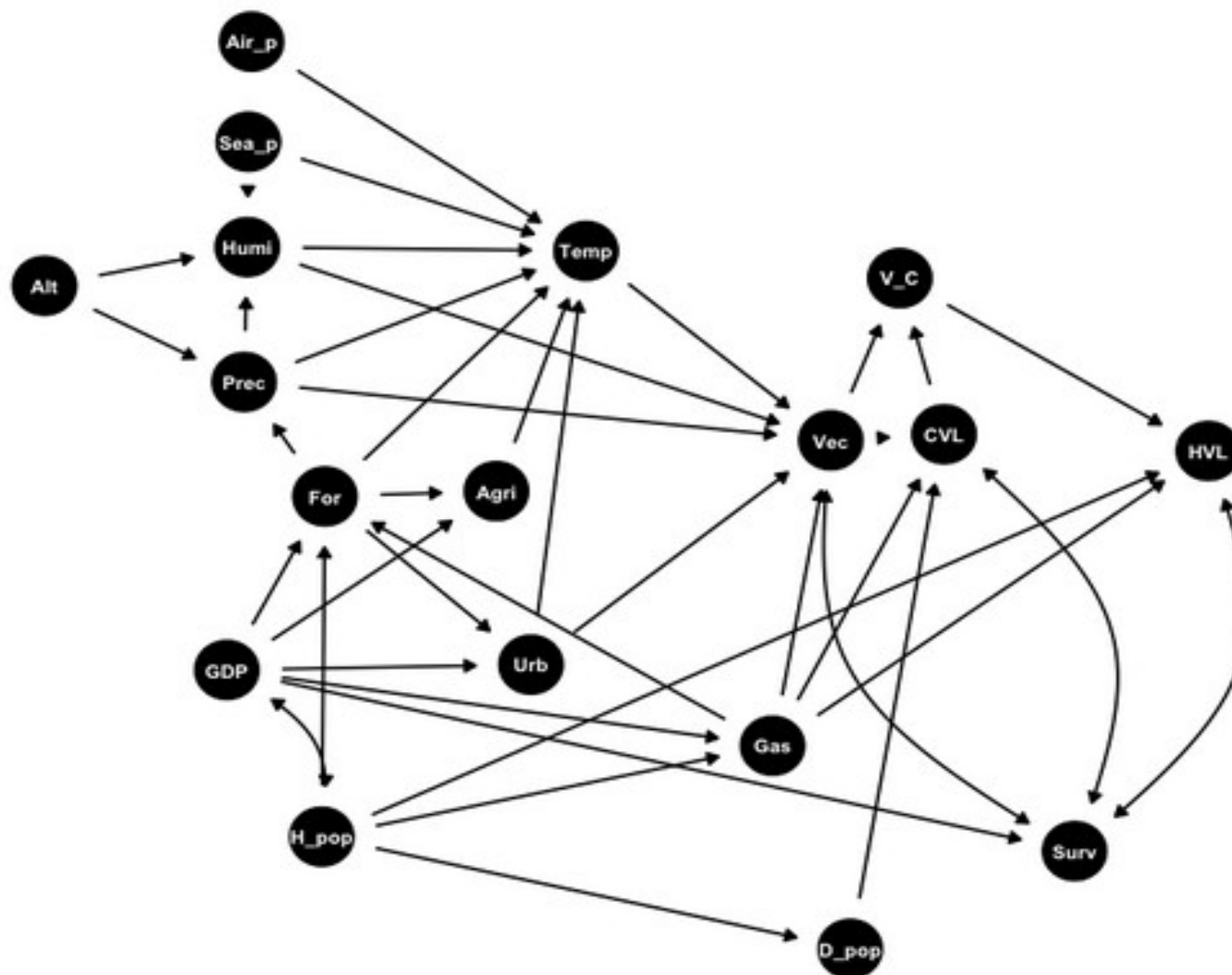
Após uma rigorosa revisão de literatura, as seguintes covariáveis foram consideradas como associadas à ocorrência de leishmaniose visceral: Produto Interno Bruto (*GDP*), ações de vigilância (*Surv*), presença do gasoduto Bolívia-Brasil (*Gas*), população canina (*D_pop*), população humana (*H_pop*), urbanização (*Urb*), área de agricultura (*Agri*), precipitação (*Prec*), umidade (*Hum*), altitude (*Alt*), pressão do ar (*Air_p*), proximidade do mar (*Sea_p*), temperatura (*Temp*).

A Figura 4 mostra o nosso gráfico acíclico direcionado demonstrando as possíveis relações causais entre as variáveis²⁰.

¹⁹ Um pixel é o menor componente de uma imagem digital. Cada pixel é uma amostra de uma imagem original, no nosso caso, um pixel representa a cobertura de uso do solo numa área de 30 metros

²⁰ Script para gerá-lo está apresentado no apêndice C

Figura 4 - Diagrama causal completo do estudo. Os nós representam variáveis, as arestas representam relações e as setas a direção da relação causal. Arestas bidirecionadas representam a relação entre o par de variáveis onde não inferimos temporalidade entre as duas.



Fonte: O autor

Considerando o nosso diagrama, as variáveis mínimas para ajuste foram:

- Presença do vetor (Vec): $\{GDP, Gas\}$;
- Leishmaniose visceral canina (CVL): $\{GDP, Gas, H_{pop}\}$ ou $\{GDP, Gas, D_{pop}\}$;
- Leishmaniose visceral humana (HVL): $\{GDP, Gas, H_{pop}\}$

Sendo assim, dentre as opções acima, as covariáveis utilizadas foram a presença do gasoduto Bolívia-Brasil, PIB municipal e população humana. A seguir as covariáveis serão descritas.

3.4.1 Gasoduto

Dados referentes ao gasoduto Bolívia-Brasil foram obtidos do Ministério dos Transportes (2010). Utilizamos o arquivo *shapefile* (.shp) e com a função calculadora de campo do *QGIS* versão 2.18 (QGIS DEVELOPMENT TEAM, 2016), classificamos os municípios como 1, se o gasoduto o atravessa, e 0, caso contrário.

3.4.2 Produto Interno Bruto municipal

Os dados de Produto Interno Bruto (PIB) foram provenientes da Fundação Sistema Estadual de Análise de Dados do Estado de São Paulo (Fundação Seade). A fundação é um órgão da Secretaria de Gestão e Planejamento do Governo do Estado de São Paulo, com objetivo de produzir e disseminar análises estatísticas socioeconômicas e demográficas. Todos os dados são de livre acesso e gratuitos. Nós utilizamos a média do PIB (100 milhões de reais/ano) para o período de 2000-2018 (Fundação SEADE, 2019).

3.4.3 População humana por município

Os dados sobre estimativa populacional também foram provenientes da Fundação Seade (2019). Coletamos dados sobre a população estimada e calculamos média (milhões de pessoas) para o período correspondente (2000-2018).

3.5 Notações

Temos n indivíduos (i.e., municípios) conectados em uma rede (i.e., estado de São Paulo), ε corresponde ao conjunto de pares de vizinhos, de forma que i e j estão diretamente conectados se, e somente se, compartilham fronteira, ou seja $(i, j) \in \varepsilon$. Já $N_i = \{j : (i, j) \in \varepsilon\}$ representa o conjunto de rede de vizinhos do indivíduo i , isto é, o conjunto de municípios com o qual i faz fronteira. Para cada $i = 1, \dots, n$, existem dados observados (Y_i, T_i, C_i) , onde Y_i representa o desfecho (presença do vetor, LVC, ou LVH), T_i corresponde ao tratamento binário (i.e., status de desmatamento) recebido pelo indivíduo i , e C_i o vetor de covariáveis. Definimos $Y = (Y_1, \dots, Y_n)$, $T = (T_1, \dots, T_n)$ e $C = (C_1, \dots, C_n)$ como a alocação de desfecho, tratamento e covariáveis, respectivamente. Denotamos $O = (Y, T, C)$ como sendo o conjunto de alocação do desfecho, tratamento e covariáveis através da rede.

Seja $t = (t_1, \dots, t_n)$ denotado como a alocação do tratamento realizado na rede e $t_{-j} = (t_1, \dots, t_n) \setminus t_j$ significando o subvetor $n-1$ com o j -ésimo valor deletado. Nós definimos $T(n)$ como o conjunto de vetores de possíveis alocações do tratamento de tamanho n . De forma que t toma um dos 2^n valores possíveis em $T(n)$.

O risco de presença *Lu. longipalpis*/LVC/LVH de um indivíduo sobre alocação hipotética do tratamento \mathbf{t} é dada por $\psi_i(t) = E(Y_i(t))$ para $i = 1, \dots, n$. Ou seja, esse é o risco do desfecho no município i sobre a alocação hipotética do tratamento \mathbf{t} a todo o estado de São Paulo. As próximas definições são baseadas nas definições análogas dadas por Halloran e Struchiner (1995) e Hudgens e Halloran (2008). Primeiro, temos o efeito causal direto médio do município i , ao mudar o seu status de tratamento de $t=0$ (sem desmatamento) para $t=1$ (desmatado), enquanto mantemos o status de desmatamento dos municípios restantes em t_{-i} :

$$DE_i(t_{-i}) = \psi_i(t_{-i}, t_i = 1) - \psi_i(t_{-i}, t_i = 0) \quad (6)$$

O efeito indireto, ou efeito *spillover*, do município i quando seu status de desmatamento é fixado em 0, quando o status de desmatamento dos outros municípios é cambiado para o oposto de i (i.e., t_{-i}), ou seja, quando o status de todo o estado, exceto i , passa de não exposto para exposto:

$$IE_i(t_{-i}) = \psi_i(t_{-i}, t_i = 0) - \psi_i(t_{-i} = 0, t_i = 0) \quad (7)$$

Seja a estratégia de alocação definida como $g_i(t_{-i}; \gamma)$, a densidade de probabilidade de t_{-i} , indexada por γ .

Já o efeito causal geral para o município i compara o risco dos desfechos sob duas estratégias de alocação γ_1 e γ_2 , onde

$$OE_i(\gamma_1, \gamma_2) = \psi_i(\gamma_1) - \psi_i(\gamma_2)$$

$$\sum_{t \in T(n)} \dots$$

3.6 Assunções

Para estimação dos efeitos causais, nós seguimos as três assunções de Tchetgen Tchetgen et al., (2017):

$$1. Y_i(T) = Y_i \text{ a.e.}^{21}$$

A primeira assunção é uma versão de rede da assunção de consistência e assume que o desfecho Y observado para o indivíduo i (i.e. Y_i) é igual ao seu desfecho potencial, $Y_i(T)$, sob a alocação de tratamento observada.

$$2. T_i \perp Y_i(t) \vee C_i \forall t \in T(n)$$

A assunção 2 diz que para todo i , o tratamento T é independente do desfecho, $Y_i(t)$, dadas as suas covariáveis, C_i , para todos níveis de t . Em outras palavras, dizemos que todas as informações relevantes (fatores de risco), usadas para gerar a alocação da exposição (i.e., status de desmatamento) estão contidas nas covariáveis da linha de base (ignorabilidade/permutabilidade condicional).

$$3. P(O = o) > 0 \text{ para todos os possíveis valores de } o$$

Por último temos a assunção de positividade a nível de rede, onde o conjunto de dados (tratamento, desfecho e covariáveis) tem probabilidade maior que zero. Sob as assunções 1-3, pode-se definir $\psi_i(t)$ como:

$$E(Y_i(t)) = \sum_c E(Y_i \vee T=t, C=c) f(c) \sum_{y,c} y_i f(y \vee T=t, C=c) f(c)$$

Então o desfecho potencial para o município i sob uma estratégia de alocação γ pode ser expresso como:

²¹ a.e.: do inglês *for almost everywhere* e do francês *p.p (presque partout)*, se referindo a quase todos valores possíveis.

$$\psi_i(\gamma) = \sum_{t \in T(n)} \sum_{y, c} y_i f(y \vee T=t, C=c) g(t; \gamma) f(c) \quad (10)$$

E, por fim, pode ser calculada a média para obter efeitos causais a nível de rede (i.e., estado). Como descrito em Tchetgen Tchetgen (2017), a equação 9 pode ser calculada através de um algoritmo de amostragem de Gibbs. Assumindo que as variáveis (\mathbf{Y} , \mathbf{T} , \mathbf{C}) podem ser representadas na forma de um gráfico de cadeia G_ε com ε adjacências. Dado o nosso grafo G_ε , as seguintes independências condicionais seguidas por propriedade Markov global associada aos gráficos de cadeia,

$$Y_i \perp \{Y_k, T_k, C_k\} \vee (T_i, C_i, O_i) \text{ para todo } i \text{ e } k \notin N_i, k \neq i \quad (11)$$

$$C_i \perp C_{-i} \{O_i \vee C_{-i} \cap O_i \text{ para todo } i \text{ e } k \notin N_i, k \neq i \quad (12)$$

A equação 11 significa que o status do desfecho (vetor/LVC/LVH) de um município é independente dos dados (exposição, covariáveis e desfecho) dos municípios com o qual ele não faz fronteira, condicionado aos dados dos municípios que ele faz fronteira. Da mesma maneira, a equação 12 afirma que as covariáveis de um município são independentes das covariáveis dos municípios que ele não faz fronteira, dado as covariáveis dos seus municípios contactantes. Essas suposições permitem formalizar a ideia que os contactantes de um município criam um cobertor de Markov²², deixando o município protegido da influência do restante da rede.

No nosso caso, a equação A11 pode ser resumido como:

$$Pr(Y_i=1 | T=t, C=c, Y_{-i}=y_{-i}) = Pr(y_i=1 | T_i=t_i, C_i=c_i, O_i) \quad (13)$$

Já a equação para as covariáveis (12) é definida como:

$$Pr(C_{i,p}=1 | C_{i,-p}=c_{i,-p}, C_{-i}=c_{-i}) = Pr(C_{i,p}=1 | C_{i,-p}=c_{i,-p}, \{C_j: j \in N_i\})$$

$$\propto \expit \left(K(c_{i,p}; \tau_p) + \sum_{p \neq s} P_{ps} c_{i,s} + v_p \sum_{j \in N_i} \frac{c_{j,p}}{w_i} \right)$$

²² O termo cobertor de Markov, ou *Markov blanket*, foi criado por Pearl (1988). O cobertor Markov para um nó (i.e., variável) em um modelo gráfico contém todas as variáveis que protegem o nó do resto da rede. Isso significa que o cobertor Markov de um nó é o único conhecimento necessário para prever o comportamento desse nó e de seus filhos.

(14)

Para uma função suave conhecida $K(c_{i,p}; \tau_p)$ com parâmetro desconhecido τ_p . Uma versão mais detalhada das notações, assunções e estimativas pode ser vista no Apêndice A.

3.7 Estimação dos efeitos causais

A estimação dos efeitos causais da rede se deu em dois passos, como descrito em Fulcher (2019) e detalhado no Apêndice A. No primeiro passo foram estimados os valores dos parâmetros (α e β) por meio de uma abordagem bayesiana chamada de algoritmo de Metropolis-Hastings duplo, que se trata de dois algoritmos de Metropolis-Hastings aninhados (Liang, 2010). Com os resultados da distribuição posterior do passo anterior, nós implementamos o amostrador de Gibbs chamado de *auto-g-computation* bayesiano, produzido por Tchetgen Tchetgen (2017), resultando na distribuição posterior dos efeitos causais geral, direto e indireto.

O procedimento do passo 1 foi implementado por meio da função `agcParam`, já a segunda parte foi feita a partir da função `agcEffect`, ambas funções presentes no pacote `autonet` (Fulcher, 2019).

3.8 Especificações

Para a estimação dos parâmetros (passo 1) nós utilizamos priori não informativas, 3 cadeias de Markov com 25.000 interações cada e *burn-in* de 5. Já para a segunda parte das análises, rodamos o algoritmo *auto-g-computation* bayesiano com prevalência do tratamento de 0%, 20%, 50%, 80% e 100%, ou seja, $\gamma = \{0, 0.2, 0.5, 0.8, 1\}$. O script da análise pode ser visto no Apêndice D.

4 RESULTADOS

A distribuição de todas as variáveis incluídas na análise é dada na Tabela 1. Como citado anteriormente na descrição da população (tópico 3.1), 25 indivíduos (municípios) foram retirados da análise porque já havia o desfecho no início do período de estudo.

Tabela 1 - Distribuição dos dados (n=620)

	Proporção (%)	n
Desmatamento	69.67	432
<i>Lu. Longipalpis</i>	31.29	194
LVC	15.80	98
LVH	15.80	98
Gasoduto	10.80	67
	média	Desvio padrão
PIB (100 milhões)	3225.3	28247.7
População humana (milhões)	62765	448180

Fonte: o autor

Os valores das estimativas dos parâmetros e intervalos de credibilidade a 95% estão dispostos na tabela 2. Como os valores estão ponderados pelo número de conexões, os valores dos parâmetros podem ser interpretados como o efeito médio das variáveis entre vizinhos.

Os municípios que apresentaram desmatamento têm, respectivamente, 2.63, 2.07 e 3.18 maiores chances de apresentar o vetor (*Lu. Longipalpis*), LVC e LVH, quando comparados com os municípios que não apresentaram desmatamento. Nenhuma das covariáveis foi significativa a nível individual. Já o efeito do desmatamento de municípios vizinhos só foi significativamente associado com o risco de LVH (OR=1.59 [IC 95%: 1.05 – 2.70]). Os municípios previamente livres dos desfechos (i.e., sem *Lu. Longipalpis*, LVC e LVH) e que tem vizinhos que apresentam os desfechos possuem maiores chances de passar a tê-los (respectivamente 6.67, 4.26 e 4.27). Entre as covariáveis dos vizinhos, somente a população humana (1.66, 1.27) foi associada a risco dos desfechos (LVC e LVH). Os *trace plots*, que servem para o diagnóstico do MCMC, e gráficos detalhados para cada um dos desfechos podem ser vistos no Apêndice D.

Tabela 2 - Estimativas dos parâmetros para o *Lu. Longipalpis*, LVC e LVH

Variável	Desfecho		
	Vetor	LVC	LVH
Média (Intervalo de Credibilidade a 95%)			
Intercepto	0.53 (0.21 - 1.45)	0.09 (0.03 – 0.35)	0.20 (0.11 – 0.56)
Desmatamento	2.63 (2.11 – 3.33) *	2.07 (1.61 – 2.74) *	3.18 (2.26 – 4.26) *
Gasoduto	2.17 (0.45 – 8.58)	1.04 (0.58 – 3.06)	1.55 (0.60 - 3.60)
PIB	0.86 (0.81 – 0.92)	0.92 (0.85 – 0.99)	0.54 (0.50– 0.58)
População humana	=====	1.01 (0.94 – 1.09)	0.89 (0.82 – 0.96)
Vetor vizinhos	6.67 (3.92 – 12.19) *	=====	=====
LVC vizinhos	=====	4.26 (1.68 – 7.71) *	=====
LVH vizinhos	=====	=====	4.27 (1.79 – 6.90) *
Desmatamento vizinhos	1.72 (0.98 – 2.67)	1.51 (0.93 – 2.14)	1.59 (1.05 – 2.70) *
Gasoduto vizinhos	1.64 (0.61 – 5.84)	1.32 (0.80 – 2.50)	1.13 (0.46 – 2.05)
PIB vizinhos	1.05 (0.91 – 1.18)	1.23 (1.01– 1.42)	0.63 (0.53– 0.71)
População humana vizinhos	=====	1.66 (1.37 – 1.92) *	1.27 (1.07-1.43) *

*Significativo a 95% (intervalo de credibilidade não inclui o 1)

Fonte: o autor

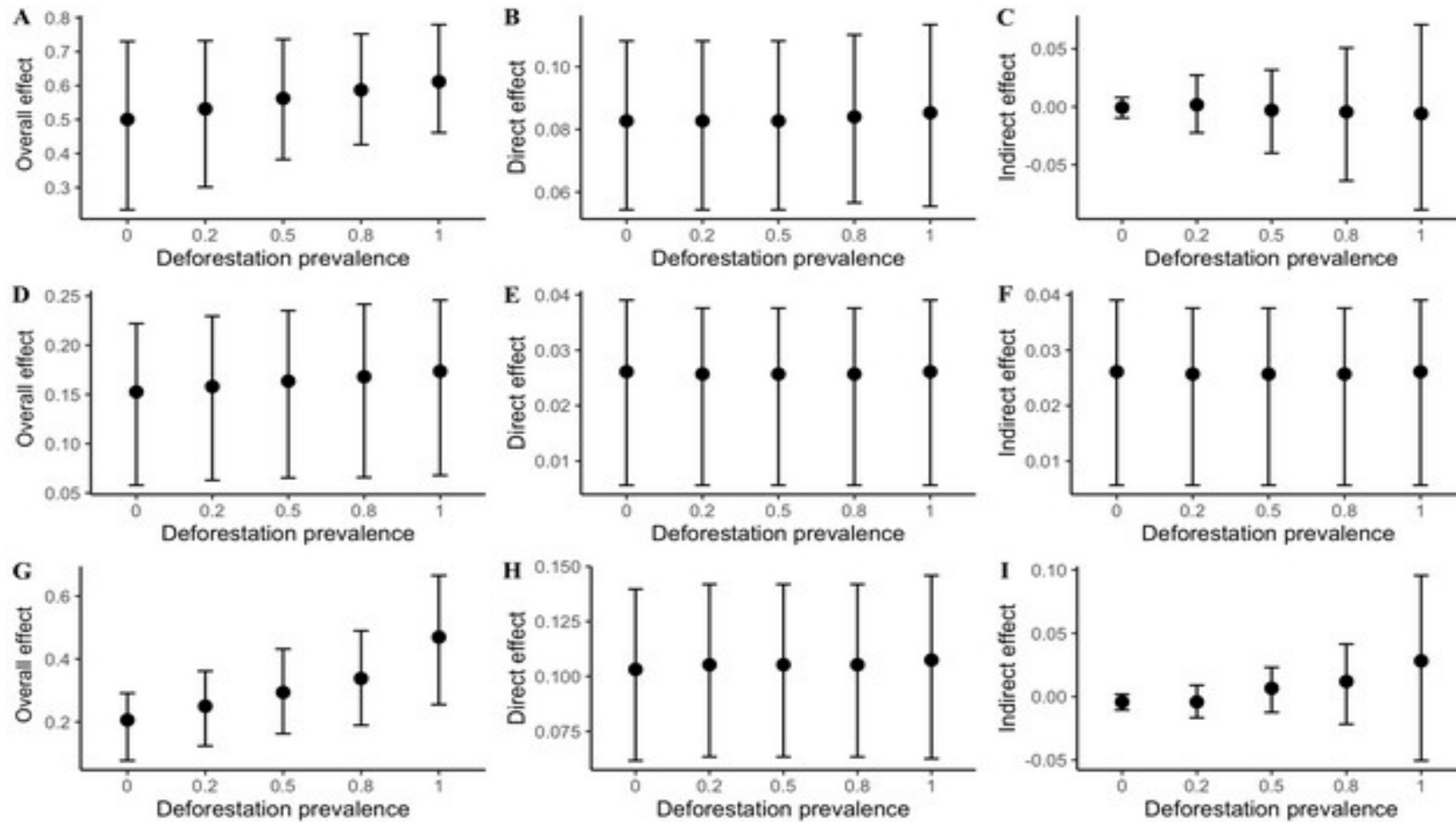
O efeito direto pode ser interpretado como a mudança esperada no risco da ocorrência dos desfechos (vetor/LVC/LVH) de um município sob uma mudança hipotética do seu status de exposição (não exposto→exposto), mantendo o nível de exposição (i.e., desmatamento) dos outros municípios em 0. O efeito indireto é a mudança esperada do risco de ocorrência dos desfechos de um município, correspondente a uma alteração no status de exposição dos demais municípios do estado (0→1), mantendo o nível de exposição do município constante. Já o efeito geral é a mudança esperada na ocorrência dos desfechos em todo o estado correspondente a uma mudança na prevalência da exposição no estado.

Para o vetor, os efeitos geral, direto e indireto estão expostos na Figura 5 (A-C), respectivamente. A prevalência de *Lu. longipalpis* diminui em 11% quando a prevalência de desmatamento na rede diminui de 50% (0.56) para 0% (0.5), já sob um aumento do desmatamento de 50% para 80%, a prevalência aumenta em 4.40% (Figura 5A). O efeito direto na prevalência do vetor é 0.08 com 50% de desmatamento na rede, esse efeito não mudou com variações na prevalência do tratamento nos vizinhos (Figura 5B). Já os efeitos indiretos não apresentaram evidência significativa de efeito (Figura 5C).

Os efeitos para a LVC estão presentes na Figura 5 (D-F). Estima-se uma redução de 6.67% do efeito geral com a diminuição da prevalência do desmatamento de 50% para 0% (Figura 5D). Os efeitos diretos apresentaram diferenças negligenciáveis entre as prevalências, bem como os efeitos indiretos.

Já para a LVH, apresentado na Figura 5 (G-I), a redução de prevalência de desmatamento de 50% (0.293) para 0% (0.20) resulta numa diminuição do efeito geral em 29,87% (FIGURA 5G). Enquanto o efeito direto não apresentou diferenças

Figura 5 - Estimativas dos efeitos causais na rede sob diferentes prevalências de desmatamento. Pontos representam as médias e linhas o intervalo de credibilidade 95%. (A-C) Representam, respectivamente, os efeitos geral, direto e indireto de diferentes prevalências do desmatamento na ocorrência do vetor na rede (i.e., estado). (D-F) Representam, os efeitos geral, direto e indireto de diferentes prevalências do desmatamento na ocorrência de LVC no estado. (G-I) representam, respectivamente os efeitos geral, direto e indireto de diferentes prevalências do desmatamento na ocorrência de LVH na rede.



Fonte: O autor, 2020.

DISCUSSÃO E CONCLUSÕES

Nosso estudo se trata da primeira tentativa de aplicar métodos de inferência causal para estimar o efeito do desmatamento na ocorrência de doenças infecciosas, mais especificamente a leishmaniose visceral.

Apesar da crescente aceitação de que a ocorrência, dinâmica e emergência de doenças infecciosas possa, ao menos em parte, ser relacionado ao desmatamento, há uma clara escassez de estudos para confirmar ou até mesmo refutar se, e como, mudanças na cobertura vegetal teriam papel chave no fenômeno (PLOWRIGHT et al., 2008). Vários fatores contribuíram para nosso pobre entendimento sobre a relação entre o meio ambiente e doenças infecciosas, sendo possível destacar três. O primeiro entrave é a inerente dificuldade de estudar ecossistemas naturais, graças aos altos custos exigidos para coleta de dados, por exemplo. Segundo, podemos citar a falta ou má qualidade de dados sobre prevalência dos componentes dos ciclos de doenças infecciosas, principalmente a falta de dados sobre população de vetores e/ou hospedeiros selváticos, além de problemas de diagnóstico e notificação de casos. Por último, podemos destacar como limitante, a falta de abordagens tradicionais de inferência causal que sejam aplicáveis a sistemas de tamanha complexidade (ALLEMBY 2005; PLOWRIGHT 2008).

Isso resultou no uso recorrente de métodos científicos reducionistas, como o teste de hipótese, que separam o problema de estudo em elementos e então focam nos elementos isoladamente, ignorando os complexos, e muitas vezes não lineares e não independentes, efeitos em cadeia das mudanças de um elemento sobre outro (HILBORN E MANGEL, 1997). Consequentemente, apesar do desmatamento ser frequentemente citado na literatura como associado à ocorrência, emergência e/ou reemergência de doenças infecciosas, principalmente as de transmissão vetorial, não há consenso sobre o seu papel na ocorrência da doença. Mais especificamente no tocante a leishmaniose, há uma confusão sobre os efeitos de áreas verdes e desmatamento, bem como suas consequências, (e.g., urbanização e expansão da agricultura e pecuária e assentamentos rurais). Alguns estudos comentam que há uma maior ocorrência da doença em localidades com mais vegetação densa (CERBINO-NETO, WERNECK E COSTA, 2009; WERNECK E MAGUIRE., 2007; WERNECK et al., 2002), outros estudos ficam no meio termo e não observaram associação entre desmatamento e LV (ARAÚJO, 2011); enquanto outros pesquisadores citam que áreas urbanizadas há uma maior ocorrência

do vetor e por consequência da doença em cães e humanos (CHAVES, COHEN E PASCUAL, 2008).

Nós identificamos o desmatamento como um fator de risco na ocorrência do vetor e da leishmaniose visceral em cães e humanos na nossa rede. O desmatamento tem um papel fundamental em duas fases do estabelecimento da doença.

Inicialmente, o desmatamento altera o funcionamento do ecossistema (VITOUSEK, 1990) e a estrutura da comunidade (CHAPIN ET AL., 2000). Isso se dá por uma cadeia de eventos, inicialmente por meio da fuga ou morte de espécies do topo da cadeia alimentar do ecossistema, além de predadores e competidores naturais do próprio vetor. Com isso, o vetor experimenta um processo chamado de liberação ecológica (*ecological release*), onde a espécie fica livre do controle natural de predadores e competição por alimento (DANGLES E MALMQVIST; 2004). A partir desse relaxamento de controle, a população se torna hábil em alocar recursos para aumento populacional, tornando-se eficientes colonizadores e rapidamente se dispersando ao longo do ambiente (MACK et al 2000). Uma extensão desse conceito de liberação ecológica já foi sugerida para explicar o fenômeno de invasão bem-sucedida de espécies como *Rattus* e *Aedes*, que são hospedeiras e vetores de algumas das mais importantes doenças que acometem humanos (WILCOX E GUBLER, 2005; PRENTER et al., 2004).

Além de afetar diretamente a ecologia do vetor, o desmatamento e suas consequências (e.g., urbanização e expansão agrícola e pecuária) promovem uma aproximação entre o vetor já adaptado no novo ambiente e os reservatórios (i.e., cães e humanos) (PATZ et al., 2000). Esse estreitamento da relação entre vetor e reservatórios é um fator chave para a ocorrência e emergência de doenças zoonóticas. O contato é uma característica chave da dinâmica de doenças; R_0^{23} está intimamente ligado à taxa de contato entre indivíduos suscetíveis e infectantes e à taxa de recuperação ou mortalidade de indivíduos infectados (LLOYD-SMITH et al., 2009).

O nosso estudo conta com algumas limitações. A primeira diz respeito a simplificação dos dados, principalmente o tratamento, que foram utilizados de forma binária (0/1) para melhor convergência e permitir a estimação dos efeitos. Isso nos impede de, por exemplo, observar se há efeito de dose-resposta do desmatamento sobre os desfechos (e.g.,

²³ R_0 , ou número básico de reprodução, é o número médio de casos secundários que surgem a partir de um caso primário, em uma população susceptível (MACDONALD, 1956). Para que um patógeno se espalhe e seja mantido dentro de uma população o valor de R_0 deve ser >1 (HALLORAN, LONGINI E STRUCHINER; 2010).

quanto maior o nível de desmatamento de um município, maior o risco dos despechos). Outra limitação é a nossa impossibilidade de inferir temporalidade na relação entre os indivíduos da rede, onde não podemos definir, por exemplo, ‘quem infectou quem’.

Finalmente, ressaltamos a necessidade de colocar o controle da LV e, de maneira mais ampla as doenças infecciosas, dentro de uma estrutura que engloba desenvolvimento ecologicamente correto e soluções viáveis para as compensações entre agricultura, urbanização e conservação. As áreas de floresta são relevantes não apenas para a conservação da biodiversidade, como já reconhecido em estudos de agroecossistemas, mas também para prevenir o surgimento e emergências de doenças infecciosas (CHAVES, COHEN E PASCUAL, 2008; VANDERMEER E PERFECTO, 2007). Em futuros trabalhos deve-se buscar abordagens integradas e multidisciplinares, como a Saúde Única (mais conhecida pelo termo em inglês, *One Health*). Abordagens de Saúde Única são definidas como ‘... o esforço colaborativo de várias disciplinas - trabalhando local, nacional e globalmente - para alcançar a saúde ideal para pessoas, animais e nosso meio ambiente ...’ (AMVA, 2008), que reconhecem que a saúde das pessoas está conectada à saúde das pessoas animais e meio ambiente.

REFERÊNCIAS

- ALENCAR J.E.; DIETZE, R. Leishmaniose visceral (Calazar). Em: **Doenças infecciosas e parasitárias**, Veronesi R. Rio de Janeiro: Guanabara Koogan; p.706-17, 1991.
- ALLAIRE, Joseph; et al. Rmarkdown: Dynamic Documents for R; 2016. **URL** <https://CRAN.R-project.org/package=rmarkdown>. **R package version**, v. 1, p. 506, 2016.
- ALLENBY, Brad. Technology at the global scale: Integrative cognitivism and Earth Systems Engineering Management. 2005.
- ALVAR, Jorge et al. The relationship between leishmaniasis and AIDS: the second 10 years. **Clinical microbiology reviews**, v. 21, n. 2, p. 334-359, 2008.
- ALVAR, Jorge et al. Canine leishmaniasis. **Advances in parasitology**, v. 57, n. 3, p. 1-88, 2004.
- ALVAR, Jorge et al. Leishmaniasis worldwide and global estimates of its incidence. **PloS one**, v. 7, n. 5, 2012.
- ALVARENGA, Daniel Gomes de et al. Leishmaniose visceral: estudo retrospectivo de fatores associados à letalidade. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 43, n. 2, p. 194-197, 2010.
- AMERICAN VETERINARY MEDICAL ASSOCIATION. One health: A new professional imperative. **One Health Initiative Task Force Final Report**, 2008.
- ANTONIALLI, Suely A. C. et al. Spatial analysis of American visceral leishmaniasis in Mato Grosso do Sul state, Central Brazil. **Journal of infection**, v. 54, n. 5, p. 509-514, 2007.
- ARAÚJO, Valdelaine E. M. et al. Relative risk of visceral leishmaniasis in Brazil: a spatial analysis in urban area. **PLoS neglected tropical diseases**, v. 7, n. 11, 2013.
- ARAÚJO, Valdelaine E. M. Análise da distribuição espaço-temporal da leishmaniose visceral e perfil clínico-epidemiológico dos casos e óbitos, Belo Horizonte Minas Gerais, 1994 a 2009. 2011
- ASHFORD, Richard. The leishmaniasis as emerging and reemerging zoonoses. **International journal for parasitology**, v. 30, n. 12-13, p. 1269-1281, 2000.
- ASSUNÇÃO, Renato M.; REIS, I.A.; OLIVEIRA; C.D.L. Diffusion and prediction of Leishmaniasis in a large metropolitan area in Brazil with a Bayesian space-time model. **Statistics in Medicine**, v. 20, n. 15, p. 2319-2335, 2001.
- BANETH, Gad et al. Canine leishmaniosis—new concepts and insights on an expanding zoonosis: part one. **Trends in parasitology**, v. 24, n. 7, p. 324-330, 2008.
- BARRETT, Malcolm. ggdag: Analyze and Create Elegant Directed Acyclic Graphs. **R package version 0.1.0**, 2018.

BATES, Paul A. et al. Recent advances in phlebotomine sand fly research related to leishmaniasis control. **Parasites & vectors**, v. 8, n. 1, p. 131, 2015.

BAYES, Thomas; PRICE, Richard. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. **Philosophical Transactions of the Royal Society**, v. 53, p. 370-418, 1763. <http://dx.doi.org/10.1098/rstl.1763.0053>

BEAUMONT, Mark A.; RANNALA, Bruce. The Bayesian revolution in genetics. **Nature Reviews Genetics**, v. 5, n. 4, p. 251-261, 2004.

BERNARDO, José M.; SMITH, Adrian FM. **Bayesian theory**. John Wiley & Sons, 2000.

BESAG, Julian. Efficiency of pseudolikelihood estimation for simple Gaussian fields. **Biometrika**, p. 616-618, 1977.

BEST, Nicky; RICHARDSON, S. **Introduction to Bayesian analysis using WinBUGS (Summer course)**, Londres: Imperial College of London, 2005.

BIVAND, Roger et al. Package ‘Spdep’: Spatial dependence: Weighting schemes, statistics and models. **Repository CRAN**, 2015.

BIVAND, Roger. maptools: Tools for reading and handling spatial objects. 2014. **R package version 0.8-39** <https://CRAN.R-project.org/package=maptools>.

BLALOCK, Hubert M. Causal models in the social sciences. Chicago: Aldine-Atherton, 1971.

BOLSTAD, William M.; CURRAN, James M. **Introduction to Bayesian statistics**. John Wiley & Sons, 2016.

BOX, George EP; TIAO, George C. **Bayesian inference in statistical analysis**. John Wiley & Sons, 2011.

BRAZIL, Reginaldo Peçanha; BRAZIL, Beatriz Gomes. Biologia de flebotomíneos neotrópicos. In: **Flebotomíneos no Brasil**. 2003. p. 257-274.

BURZA, Sakib; CROFT S.L.; BOELAERT M. **Seminar Leishmaniasis**. Lancet Infectious Diseases. 6736:1-20, 2018.

CALVERT, G. Dennis. A review of observational studies on the relationship between cholesterol and coronary heart disease. **Australian and New Zealand journal of medicine**, v. 24, n. 1, p. 89-91, 1994.

CAMARGO-NEVES, Vera L.; KATZ G. Leishmaniose visceral americana no estado de São Paulo. **Revista da Sociedade Brasileira de Medicina Tropical**. 32 (Supl.II): 63-4, 1999.

CARDIM, Marisa F. M. et al. Visceral leishmaniasis in the state of São Paulo, Brazil: spatial and space-time analysis. **Revista de saúde pública**, v. 50, p. 48, 2016.

- CARDIM, Marisa F. M. et al. Introdução e expansão da Leishmaniose visceral americana em humanos no estado de São Paulo, 1999-2011. **Revista de Saúde Pública**, v. 47, p. 691-700, 2013.
- CASANOVA, Claudio et al. Distribution of *Lutzomyia longipalpis* chemotype populations in São Paulo state, Brazil. **PLoS neglected tropical diseases**, v. 9, n. 3, 2015.
- CERBINO NETO, José; WERNECK, Guilherme Loureiro; COSTA, Carlos Henrique Nery. Factors associated with the incidence of urban visceral leishmaniasis: an ecological study in Teresina, Piauí State, Brazil. **Cadernos de Saúde Pública**, v. 25, p. 1543-1551, 2009.
- CHAPIN III, F. Stuart et al. Consequences of changing biodiversity. **Nature**, v. 405, n. 6783, p. 234-242, 2000.
- CHAVES, Luis et al. Social exclusion modifies climate and deforestation impacts on a vector-borne disease. **PLOS Neglected tropical diseases**, v. 2, n. 2, 2008.
- CHAVES, Luis.F.; PASCUAL, M. Climate cycles and forecasts of cutaneous leishmaniasis, a nonstationary vector-borne disease. **PLoS medicine**, v. 3, n. 8, p. e295, 2006.
- COCHRAN, Gregory M.; EWALD, Paul W.; COCHRAN, Kyle D. Infectious causation of disease: an evolutionary perspective. **Perspectives in biology and medicine**, v. 43, n. 3, p. 406-448, 2000.
- COLDITZ, Graham A. Relationship between estrogen levels, use of hormone replacement therapy, and breast cancer. **JNCI: Journal of the National Cancer Institute**, v. 90, n. 11, p. 814-823, 1998.
- COLLIN, Simon M. et al. Unseen Kala-azar deaths in South Sudan (1999–2002). **Tropical Medicine & International Health**, v. 11, n. 4, p. 509-512, 2006.
- CORTES, Taísa Rodrigues; FAERSTEIN, Eduardo; STRUCHINER, Claudio José. Utilização de diagramas causais em epidemiologia: um exemplo de aplicação em situação de confusão. **Cadernos de Saúde Pública**, v. 32, p. e00103115, 2016.
- COX, David Roxbee; COX, David R. **Planning of experiments**. New York: Wiley, 1958.
- DANGLES, Olivier; MALMQVIST, Björn. Species richness–decomposition relationships depend on species dominance. **Ecology letters**, v. 7, n. 5, p. 395-402, 2004.
- de FINETTI, Bruno. Theory of probability. John Wiles & Sons. Bristol. 1979.
- DESJEUX, Philip. Leishmaniasis: current situation and new perspectives. **Comparative immunology, microbiology and infectious diseases**, v. 27, n. 5, p. 305-318, 2004.
Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/15225981>. Acesso em: 10 out. 2018
- DESJEUX, Philip. The increase in risk factors for leishmaniasis worldwide. **Transactions of the royal society of tropical medicine and hygiene**, v. 95, n. 3, p. 239-243, 2001.

DONOVAN Charles. On the possibility of the occurrence of trypanosomiasis in India. **British Medical Journal**. 1903;2:79.

DRUMOND, Karina Oliveira; COSTA, Francisco Assis Lima. Forty years of visceral leishmaniasis in the State of Piauí: a review. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 53, n. 1, p. 3-11, 2011.

DUJARDIN, Jean-Claude. Risk factors in the spread of leishmaniasis: towards integrated monitoring?. **Trends in parasitology**, v. 22, n. 1, p. 4-6, 2006.

ESTERRE, P.; Chippaux, J.P.; Lefait, J.F.; Dedet, J.P. Evaluation of a cutaneous leishmaniasis control program in a forest village of French Guyana. **Bulletin of the World Health Organization**, v. 64, n. 4, p. 559-565, 1986.

EVANS, Alfred S. Causation and disease: the Henle-Koch postulates revisited. **The Yale journal of biology and medicine**, v. 49, n. 2, p. 175, 1976.

FRANCO, Ana O. et al. Predicting the distribution of canine leishmaniasis in western Europe based on environmental variables. **Parasitology**, v. 138, n. 14, p. 1878-1891, 2011.

FRANK, Michael C.; GOODMAN, Noah D. Predicting pragmatic reasoning in language games. **Science**, v. 336, n. 6084, p. 998-998, 2012.

FULCHER, Isabel Rose. **Statistical Inference for Causal Mechanisms: Mediation and Interference**. 2019. Tese (Doutorado em Bioestatística) - Departamento de Bioestatística, Universidade de Harvard.

Fundação Sistema Estadual de Análise de Dados (SEADE). Estatísticas de demografia [online] São Paulo; 2019. Disponível em: <https://imp.seade.gov.br/frontend/#!/>

GALATI, Eunice A.B. "Classificação de Phlebotominae." Em: Rangel, E. F., Lainson, R., (eds). **Flebotomíneos do Brasil**. Rio de Janeiro: Fundação Oswaldo Cruz, p. 23-51, 2003.

GALATI, Eunice A.B. Morfologia e terminologia de Phlebotominae (Diptera: Psychodidae). **Classificação e identificação de táxons das Américas**. Vol I. Apostila da Disciplina Bioecologia e Identificação de Phlebotominae de Pós-Graduação em Saúde Pública. Universidade de São Paulo, São Paulo.132p, 2018

GÁLVEZ, Rosa et al. Mapping the current distribution and predicted spread of the leishmaniosis sand fly vector in the Madrid region (Spain) based on environmental variables and expected climate change. **Vector-borne and zoonotic diseases**, v. 11, n. 7, p. 799-806, 2011.

GAMERMAN, Dani; LOPES, Hedibert F. **Markov chain Monte Carlo: stochastic simulation for Bayesian inference**. Chapman and Hall/CRC, 2006.

GEISLER, Wilson S.; DIEHL, Randy L. Bayesian natural selection and the evolution of perceptual systems. **Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences**, v. 357, n. 1420, p. 419-448, 2002.

GELFAND, Alan E.; SMITH, Adrian FM. Sampling-based approaches to calculating marginal densities. **Journal of the American statistical association**, v. 85, n. 410, p. 398-409, 1990.

GEMAN, Stuart; GEMAN, Donald. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. **IEEE Transactions on pattern analysis and machine intelligence**, n. 6, p. 721-741, 1984.

GOMEZ-BARROSO, D. et al. Spatial distribution and cluster analysis of a leishmaniasis outbreak in the south-western Madrid region, Spain, September 2009 to April 2013. **Eurosurveillance**, v. 20, n. 7, p. 21037, 2015.

GOOD, Irving J. Studies in the history of probability and statistics. XXXVII AM Turing's statistical work in World War II. **Biometrika**, p. 393-396, 1979.

GRAHAM, Bryan S. Identifying social interactions through conditional variance restrictions. **Econometrica**, v. 76, n. 3, p. 643-660, 2008.

GRIMALDI, Gabriel; TESH, Robert B. Leishmaniasis of the New World: current concepts and implications for future research. **Clinical microbiology reviews**, v. 6, n. 3, p. 230-250, 1993.

GUYON, Xavier; KÜNSCH, Hans R. Asymptotic comparison of estimators in the Ising model. In: **Stochastic Models, Statistical Methods, and Algorithms in Image Analysis**. Springer, New York, NY, 1992. p. 177-198.

HALLORAN, M. Elizabeth; LONGINI, Ira; STRUCHINER, Claudio. **Design and analysis of vaccine studies**. New York: Springer, 2010.

HALLORAN, M. Elizabeth; STRUCHINER, Claudio J. Causal inference in infectious diseases. **Epidemiology**, p. 142-151, 1995.

HASTINGS, Keith. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, v. 57, n. 1, p. 97-109, 1970.

HERNÁN, Miguel. A.; ROBINS, Jamie. M. Causal inference: What if. **Boca Raton: Chapman & Hill/CRC**, 2020.

HERNÁN, Miguel A. et al. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. **American journal of epidemiology**, v. 155, n. 2, p. 176-184, 2002.

HERNÁN, Miguel A.; HERNÁNDEZ-DÍAZ, Sonia; ROBINS, James M. A structural approach to selection bias. **Epidemiology**, v. 15, n. 5, p. 615-625, 2004.

HERNÁN, Miguel A.; ROBINS, James M. Estimating causal effects from epidemiological data. **Journal of Epidemiology & Community Health**, v. 60, n. 7, p. 578-586, 2006.

HIJMANS, Robert J. et al. Package 'raster'. **R package**, 2015.

HILBORN, Ray; MANGEL, Marc. **The ecological detective: confronting models with data**. Princeton University Press, 1997.

HILL, Austin Bradford. The Environment and Disease: Association or Causation? **Proceedings of the Royal Society of Medicine**, v. 58, n. 5, p. 295, 1965.

HLAVACOVA, Jana; VOTYPKA, J.; VOLF, P. The effect of temperature on Leishmania (Kinetoplastida: Trypanosomatidae) development in sand flies. **Journal of Medical Entomology**, v. 50, n. 5, p. 955-958, 2013.

HOLLAND, Paul W. Statistics and causal inference. **Journal of the American statistical Association**, v. 81, n. 396, p. 945-960, 1986.

HUDGENS, Michael G.; HALLORAN, M. Elizabeth. Toward causal inference with interference. **Journal of the American Statistical Association**, v. 103, n. 482, p. 832-842, 2008.

HUME, David. **An Enquiry Concerning Human Understanding**, 1748.

IBGE Cidades. 2018 .Disponível em: <<http://www.ibge.gov.br/cidadesat/topwindow.htm?1>>. Acesso em: 07 dez. 2018.

JEFFREYS, Harold. **The theory of probability**. Oxford University Press: Oxford, 1961.

KADANE, Joseph B. et al. Bayesian thought in early modern detective stories: Monsieur Lecoq, C. Auguste Dupin and Sherlock Holmes. **Statistical Science**, v. 24, n. 2, p. 238-243, 2009.

KARAGIANNIS-VOULES, Dimitrios-Alexios et al. Bayesian geostatistical modeling of leishmaniasis incidence in Brazil. **PLoS neglected tropical diseases**, v. 7, n. 5, 2013.

KÉRY, Marc. **Introduction to WinBUGS for ecologists: Bayesian approach to regression, ANOVA, mixed models and related analyses**. Academic Press, 2010.

KINDERMANN, Ross. Markov random fields and their applications. **American mathematical society**, 1980.

KOCH, R. Über Bakteriologische Forschung. **Verhandlungen des X. Internationalen Medizinischen Kongresses**, p. 35-47, 1890.

KRUSCHKE, John. **Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan**. Academic Press, 2014.

LACHAUD, L. et al. Value of two PCR methods for the diagnosis of canine visceral leishmaniasis and the detection of asymptomatic carriers. **Parasitology**, v. 125, n. 3, p. 197-207, 2002.

LAINSON, Ralph.; SHAW, J. J. Evolution, classification and geographical distribution. Em: **The Leishmaniasis in Biology and Medicine**, v. 1, Peters, W., Killick-Kendrick, R. (Eds.), London: Academic Press, p. 1-120. 1987.

LAMBIN, Eric F. et al. Pathogenic landscapes: interactions between land, people, disease vectors, and their animal hosts. **International journal of health geographics**, v. 9, n. 1, p. 54, 2010.

LAPLACE, Pierre S. M. **Essai philosophique sur les probabilités**. Bachelier, 1814.

LAURITZEN, Steffen L. **Graphical models**. Clarendon Press, 1996.

LAWSON, Andrew B. **Bayesian disease mapping: hierarchical modeling in spatial epidemiology**. Chapman and Hall/CRC, 2013.

LIANG, Faming. A double Metropolis–Hastings sampler for spatial models with intractable normalizing constants. **Journal of Statistical Computation and Simulation**, v. 80, n. 9, p. 1007-1022, 2010.

LEISHMAN, William.B. On the possibility of the occurrence of trypanosomiasis in India. **British Medical Journal**. 1903;1:1252–4.

LLOYD-SMITH, James O. et al. Epidemic dynamics at the human-animal interface. **science**, v. 326, n. 5958, p. 1362-1367, 2009.

LUIZ, Ronir Raggio; STRUCHINER, Claudio José. **Inferência causal em epidemiologia: o modelo de respostas potenciais**. Editora Fiocruz, 2002.

LUKEŠ, Julius et al. Evolutionary and geographical history of the *Leishmania donovani* complex with a revision of current taxonomy. **Proceedings of the National Academy of Sciences**, v. 104, n. 22, p. 9375-9380, 2007.

MAATHUIS, Marloes et al. **Handbook of Graphical Models**. CRC Press, 2019.

MACDONALD, George. Epidemiological basis of malaria control. **Bulletin of the World Health Organization**, v. 15, n. 3-5, p. 613, 1956.

MACK, Richard N. et al. Biotic invasions: causes, epidemiology, global consequences, and control. **Ecological applications**, v. 10, n. 3, p. 689-710, 2000.

MAPBIOMAS. **Projeto MapBiomas – Coleção 3 da Série Anual de Mapas de Cobertura e Uso de Solo do Brasil**. 2018 acessado em 15 de outubro de 2018 através do link: <http://mapbiomas.org>

MARZOCHI, M.C.A.; MARZOCHI, K.B.F. Leishmanioses em áreas urbanas. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 30, p. 162-164, 1997.

MIETTINEN, Olli Sakari; KARP, I. **Epidemiological research: an introduction**. Springer Science & Business Media, 2012.

METROPOLIS, Nicholas et al. Equation of state calculations by fast computing machines. **The journal of chemical physics**, v. 21, n. 6, p. 1087-1092, 1953.

MINISTÉRIO DA SAÚDE. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. **Manual de vigilância e controle da leishmaniose visceral** / Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Vigilância Epidemiológica. – 1. ed., 5. reimpr. – Brasília : Ministério da Saúde, 2014.

MINISTÉRIO DA SAÚDE. **Sistema de Informação de Agravos de Notificação**. 2019. Disponível em: <http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sinannet/cnv/leishvmg.def> . Acesso em 12 de Março de 2019.

MINISTÉRIO DOS TRANSPORTES. **Eixo Dutoviário**. Em: Open data ArcGIS [Internet]. 2010. http://pnlt.imagem-govfed.opendata.arcgis.com/datasets/596226509f0a438797628912a780380f_0

MORENO, Javier; ALVAR, Jorge. Canine leishmaniasis: epidemiological risk and the experimental model. **Trends in parasitology**, v. 18, n. 9, p. 399-405, 2002.

MOŠKOVSKIJ, Š.D.; DUHANINA, N.N. Epidemiology of the leishmaniasis: general considerations. **Bulletin of the World Health Organization**, v. 44, n. 4, p. 529, 1971.

MUNSTERMANN, Leonard E. Phlebotomine Sand Flies and Moth Flies (Psychodidae). In: **Medical and Veterinary Entomology**. Academic Press, 2019. p. 191-211.

NEYMAN, Jerzy. Sur les applications de la thar des probabilités aux expériences Agaricales: Essay des principe. Excerpts reprinted (1990) in English. **Statistical Science**, v. 5, n. 463-472, p. 4, 1923.

NICOLLE, Charles. Sur trois cas d'infection splénique infantile a corps de Leishman observés en Tunisie. **Archives de l'Institut Pasteur**, v. 3, p. 1-26, 1908.

NICOLLE, Charles; COMTE, C. H. Origine canine du Kala-azar. **Bull Soc Pathol Exot**, v. 1, p. 299-301, 1908.

OLIVEIRA, Cláudia Di Lorenzo et al. Spatial distribution of human and canine visceral leishmaniasis in Belo Horizonte, Minas Gerais State, Brasil, 1994-1997. **Cadernos de saude publica**, v. 17, p. 1231-1239, 2001.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). **Leishmaniasis Fact sheet No375 2018**. Disponível em: <http://www.who.int/mediacentre/factsheets/fs375/en/>. Acesso em: 16 de Outubro de 2018.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). Report of a meeting of the WHO Expert Committee on the Control of Leishmaniasis, Geneva, Switzerland, 22-26 March 2010. **WHO technical report series**, n. 949, 2010.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). **The Global Burden of Disease: 2008 update**. Geneva, Switzerland: World Health Organization. 84 p.

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE (OPAS). Leishmanioses: Informe Epidemiológico nas Américas. Washington D.C OPAS; 2019 Disponível em: <http://iris.paho.org/xmlui/handle/123456789/50505>

PACE, David. Leishmaniasis. **Journal of Infection**, v. 69, s. 1, p. 10-18, 2014.

PARENT, Eric; RIVOT, Etienne. **Introduction to hierarchical Bayesian modeling for ecological data**. Chapman and Hall/CRC, 2012.

PATZ, Jonathan A. et al. Effects of environmental change on emerging parasitic diseases. **International journal for parasitology**, v. 30, n. 12-13, p. 1395-1405, 2000.

PEARL, Judea. Probabilistic reasoning in intelligent systems. 1988. **San Mateo, CA: Kaufmann**, v. 23, p. 33-34.

PEARL, Judea; GLYMOUR, Madelyn; JEWELL, Nicholas P. **Causal inference in statistics: A primer**. John Wiley & Sons, 2016.

PEARL, Judea; MACKENZIE, Dana. **The book of why: the new science of cause and effect**. Basic Books, 2018.

PEARL, Judea; ROBINS, James M. Probabilistic evaluation of sequential plans from causal models with hidden variables. **Proceedings of the Eleventh conference on Uncertainty in artificial intelligence**. Morgan Kaufmann Publishers Inc., 1995. p. 444-453.

PEARL, Judea. Causality. 2000. **Cambridge University, New York**.

PEARL, Judea. Causal diagrams for empirical research. **Biometrika**, v. 82, n. 4, p. 669-688, 1995.

PEARL, Judea. Comment: understanding Simpson's paradox. **The American Statistician**, v. 68, n. 1, p. 8-13, 2014.

PEARSON, Karl. **The grammar of science**. Londres: Adam & Charles Black, 1892.

PEBESMA, Edzer; BIVAND, Roger S. S classes and methods for spatial data: the sp package. **R news**, v. 5, n. 2, p. 9-13, 2005.

PENNA, Henrique A. Leishmaniose visceral no Brasil. **Bras Med**, v. 48, p. 949-950, 1934.

PETERS, Wallace et al. **Leishmaniasis in biology and medicine**. Academic press, 1987.

PLOWRIGHT, Raina K. et al. Causal inference in disease ecology: investigating ecological drivers of disease emergence. **Frontiers in Ecology and the Environment**, v. 6, n. 8, p. 420-429, 2008.

PRENTER, John et al. Roles of parasites in animal invasions. **Trends in ecology & evolution**, v. 19, n. 7, p. 385-390, 2004.

QGIS, Development T. Quantum GIS geographic information system. **Open Source Geospatial Foundation Project**, v. 45, 2016.

QUINNELL, Rupert J.; COURTENAY, Orin. Transmission, reservoir hosts and control of zoonotic visceral leishmaniasis. **Parasitology**, v. 136, n. 14, p. 1915-1934, 2009.

R CORE TEAM. **R: A language and environment for statistical computing**. 2013.

RANGEL Elizabete F. Flebótomos transmissores de *Leishmania* (L) infantum chagasi nas Américas e técnicas disponíveis de captura para vigilância entomológica. **Informe final de La reunión de Expertos OPS/OMS sobre Leishmaniasis Visceral em las Américas**, Rio de Janeiro, Panafiosa, 2006.

READY, P. D. Leishmaniasis emergence in Europe. **Eurosurveillance**, v. 15, n. 10, p. 19505, 2010.

REISEN, William K. Landscape epidemiology of vector-borne diseases. **Annual review of entomology**, v. 55, p. 461-483, 2010.

REITHINGER, Richard et al. Cutaneous leishmaniasis. **The Lancet infectious diseases**, v. 7, n. 9, p. 581-596, 2007.

ROBINS, James M. Data, design, and background knowledge in etiologic inference. **Epidemiology**, v. 12, n. 3, p. 313-320, 2001.

ROBINS, James. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. **Mathematical modelling**, v. 7, n. 9-12, p. 1393-1512, 1986.

RODRÍGUEZ-MORALES, Alfonso J.; RISQUEZ, Alejandro; ECHEZURIA, Luis. Impact of climate change on health and disease in Latin America. **Climate Change and Variability. Croatia: Sciyo**, p. 463-486, 2010.

ROSS, Ronald. Note on the bodies recently described by Leishman and Donovan. **British medical journal**, v. 2, n. 2237, p. 1261, 1903.

ROSS, Ronald. An application of the theory of probabilities to the study of a priori pathometry.—Part I. **Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character**, v. 92, n. 638, p. 204-230, 1916.

ROTHMAN, Kenneth J. et al. **Modern epidemiology**. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008.

ROY, Manoj et al. (Ed.). **Urban poverty and climate change: life in the slums of Asia, Africa and Latin America**. Routledge, 2016.

RUBIN, Donald B. Estimating causal effects of treatments in randomized and nonrandomized studies. **Journal of educational Psychology**, v. 66, n. 5, p. 688, 1974.

SALOMÓN, Oscar D. et al. *Lutzomyia longipalpis* urbanisation and control. **Memorias do Instituto Oswaldo Cruz**, v. 110, n. 7, p. 831-846, 2015.

SÃO PAULO (Estado). Secretaria de Estado da Saúde, Superintendência de Controle de Endemias - SUCEN e Coordenadoria de Controle de Doenças - CCD. **Manual de Vigilância e Controle da Leishmaniose Visceral Americana do Estado de São Paulo**. São Paulo: 2006.

SATTENSPIEL, Lisa; LLOYD, Alun. **The geographic spread of infectious diseases: models and applications**. Princeton University Press, 2009.

SAVAGE, Leonard J. **The foundations of statistics**. Courier Corporation, 1954.

SECRETARIA DE ESTADO DE SAÚDE DE SÃO PAULO. **Dados estatísticos da Leishmaniose Visceral Americana de 1999-2018**. Disponível em: http://www.cve.saude.sp.gov.br/htm/zoo/leishv_dados.html. Acesso em: 15 de outubro de 2018.

SEVÁ, Anaiá P. et al. Risk analysis and prediction of visceral leishmaniasis dispersion in São Paulo State, Brazil. **PLoS neglected tropical diseases**, v. 11, n. 2, p. e0005353, 2017.

SHAW, Jeffrey. The leishmaniasis-survival and expansion in a changing world: a mini-review. **Memórias do Instituto Oswaldo Cruz**, v. 102, n. 5, p. 541-547, 2007.

SHPITSER, Ilya; TCHETGEN, Eric Tchetgen; ANDREWS, Ryan. Modeling Interference Via Symmetric Treatment Decomposition. **arXiv preprint arXiv:1709.01050**, 2017.

SILVA, Alba V. et al. Ecological study and risk mapping of leishmaniasis in an endemic area of Brazil based on a geographical information systems approach. **Geospatial Health**, v. 6, n. 1, p. 33-40, 2011.

SILVA, Denise Amaro et al. Assessment of serological tests for the diagnosis of canine visceral leishmaniasis. **The Veterinary Journal**, v. 195, n. 2, p. 252-253, 2013.

SOBEL, Michael E. What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. **Journal of the American Statistical Association**, v. 101, n. 476, p. 1398-1407, 2006.

SOLANO-GALLEGO, Laia et al. Directions for the diagnosis, clinical staging, treatment and prevention of canine leishmaniosis. **Veterinary parasitology**, v. 165, n. 1-2, p. 1-18, 2009.

SPIRITES, Peter et al. Search. 1993. **Lecture Notes in Statistics**, 1993.

STEVERDING, Dietmar. The history of leishmaniasis. **Parasites & vectors**, v. 10, n. 1, p. 82, 2017.

TCHETGEN, Eric J. Tchetgen; VANDERWEELE, Tyler J. On causal inference in the presence of interference. **Statistical methods in medical research**, v. 21, n. 1, p. 55-75, 2012.

TCHETGEN, Eric J. Tchetgen; FULCHER, Isabel; SHPITSER, Ilya. Auto-g-computation of causal effects on a network. **arXiv preprint arXiv:1709.01577**, 2017.

VANDERMEER, John; PERFECTO, Ivette. The agricultural matrix and a future paradigm for conservation. **Conservation biology**, v. 21, n. 1, p. 274-277, 2007.

VICTOR, Mauro A de M. et al. Cem anos de devastação: revisitada 30 anos depois. **Ministério do Meio Ambiente, Brasília, Brazil**, 2005.

VITOUSEK, Peter M. Biological invasions and ecosystem processes: towards an integration of population biology and ecosystem studies. In: **Ecosystem management**. Springer, New York, NY, 1990. p. 183-191.

WALSH, J.F.; MOLYNEUX, D.H.; BIRLEY, M.H. Deforestation: effects on vector-borne disease. **Parasitology**, v. 106, n. S1, p. S55-S75, 1993.

WERNECK, Guilherme L. geographic spread and urbanization of visceral leishmaniasis in Brazil. **Cadernos de Saúde Pública**, v. 24, p. 2937-2940, 2008.

WERNECK, Guilherme L. et al. Multilevel modelling of the incidence of visceral leishmaniasis in Teresina, Brazil. **Epidemiology & Infection**, v. 135, n. 2, p. 195-201, 2007.

WERNECK, Guilherme L.; MAGUIRE, James H. Spatial modeling using mixed models: an ecologic study of visceral leishmaniasis in Teresina, Piauí State, Brazil. **Cadernos de Saúde Pública**, v. 18, p. 633-637, 2002.

WICKHAM, Hadley. The tidyverse. **R package ver. 1.1**, v. 1, 2017.

WILCOX, Bruce A.; GUBLER, Duane J. Disease ecology and the global emergence of zoonotic pathogens. **Environmental Health and Preventive Medicine**, v. 10, n. 5, p. 263-272, 2005.

APÊNDICE A – Descrição das notações, suposições e método de estimação.

NOTAÇÕES

Temos n indivíduos (i.e., municípios) conectados em uma rede (i.e., estado de São Paulo), ε corresponde ao conjunto de pares de vizinhos, de forma que i e j estão diretamente conectados se, e somente se, compartilham fronteira, ou seja $(i, j) \in \varepsilon$. Já $N_i = \{j : (i, j) \in \varepsilon\}$ representa o conjunto de rede de vizinhos do indivíduo i , isto é, o conjunto de municípios com o qual i faz fronteira.

Suponhamos que para cada $i = 1, \dots, n$, existem dados observados (Y_i, T_i, C_i) , onde Y_i denota o desfecho (presença do vetor, LVC, ou LVH), T_i corresponde ao tratamento ou exposição binária (i.e., status de desmatamento) recebido pelo indivíduo i , e C_i o vetor de covariáveis. Definimos $Y = (Y_1, \dots, Y_n)$, $T = (T_1, \dots, T_n)$ e $C = (C_1, \dots, C_n)$ como a alocação de desfecho, tratamento e covariáveis, respectivamente. Denotamos $O = (Y, T, C)$ como sendo o conjunto de alocação do desfecho, tratamento e covariáveis através da rede e $O_i = \{O_j, j \in N_i\}$ como os dados observados para todos os vizinhos do município i .

Seja $t = (t_1, \dots, t_n)$ denotado como a alocação do tratamento observado na rede e $t_{-j} = (t_1, \dots, t_n) \setminus t_j$ significando o subvetor $n-1$ com o j -ésimo valor deletado. Definimos $T(n)$ como o conjunto de vetores de possíveis alocações do tratamento de tamanho n . De forma que t toma um dos 2^n valores possíveis em $T(n)$. Como já é padrão em inferência causal, nós assumimos a existência de valores contrafactuais do desfecho $Y(t) = \{Y_i(t), \dots, Y_n(t)\}$, onde $Y_i(t)$ representa a resposta potencial do município i sob ação do tratamento t . Repare que sob a suposição de não interferência, e com a pressuposto de consistência (como pôde ser visto no subtópicos 2.10), o desfecho potencial para i é resumido a $Y_i(t) = Y_i(t_i)$.

O risco marginal (i.e., efeito causal geral) de presença *Lu. longipalpis*/LVC/LVH de um indivíduo sobre alocação hipotética do tratamento t é dada por $\psi_i(t) = E(Y_i(t))$ para $i = 1, \dots, n$. Ou seja, esse é o risco da presença de *Lu. longipalpis*/LVC/LVH do município i sobre a alocação hipotética do tratamento t a todo o estado de São Paulo. Perceba que também podemos escrever $\psi_i(t) = \psi_i(a_{-i}, a_i) = E(Y_i(a_{-i}, a_i))$ para distinguir o status de desmatamento do município i e o status do resto da rede (i.e., estado) excluindo i . As seguintes definições são baseadas nas definições análogas de Hudgens e Halloran (2008). Primeiro temos o efeito causal direto médio do município i ao mudar o seu status de tratamento de $t=0$ (sem desmatamento) para $t=1$ (desmatado), enquanto o status de desmatamento dos municípios restantes em t_{-i} , que é dado por:

$$DE_i(t_{-i}) = \psi_i(t_{-i}, t_i = 1) - \psi_i(t_{-i}, t_i = 0) \quad (A1)$$

O efeito indireto²⁴ médio do município i quando seu status de desmatamento é fixado em 0 (= não exposto/não tratado), quando o status de desmatamento dos outros municípios é

²⁴ também chamado de *spillover effect* (Sobel, 2006)

cambiado para o oposto de i (i.e., t_{-i}), ou seja, quando o status passa de não exposto para exposto:

$$IE_i(t_{-i}) = \psi_i(t_{-i}, t_i = 0) - \psi_i(t_{-i} = 0, t_i = 0) \quad (A2)$$

Da mesma forma que previamente descrito por Hudgens e Halloran (2008), é possível calcular os efeitos diretos e indiretos médios da estratégia de alocação, ou seja, o efeito causal (i.e., geral, direto e indireto) para uma prevalência do tratamento previamente arbitrada. Seja a estratégia de alocação definida como $g_i(t_{-i}; \gamma)$, a densidade de probabilidade de t_{-i} , indexada por γ . Ou seja:

$$DE_i(\gamma) = \sum_{t_{-i} \in T(n)} g(t_{-i}; \gamma) DE_i(t_{-i}) \quad (A3)$$

$$IE_i(\gamma) = \sum_{t_{-i} \in T(n)} g(t_{-i}; \gamma) IE_i(t_{-i}) \quad (A4)$$

Já o efeito causal geral para o município i compara o risco dos desfechos sob duas estratégias de alocação γ_1 e γ_2 ,

$$OE_i(\gamma_1, \gamma_2) = \psi_i(\gamma_1) - \psi_i(\gamma_2) = \sum_{t \in T(n)} \ddot{t} \quad \ddot{t}$$

Já as versões de rede desses efeitos são definidas como:

$$DE(\gamma) = n^{-1} \sum_i DE_i(\gamma) \quad (A6)$$

$$IE(\gamma) = n^{-1} \sum_i IE_i(\gamma) \quad (A7)$$

$$OE(\gamma_1, \gamma_2) = n^{-1} \sum_i \psi_i(\gamma_1) - \psi_i(\gamma_2) \quad (A8)$$

ASSUNÇÕES

Para estimação dos efeitos causais, nós seguimos as três assunções de Tchetgen Tchetgen et al., (2017) para os dados observados:

$$4. \quad Y_i(T) = Y_i \text{ a.e.}^{25}$$

A primeira assunção é uma versão de rede da assunção de consistência e assume que o desfecho Y observado para o indivíduo i (i.e., Y_i) é igual ao seu desfecho potencial, $Y_i(T)$, sob a alocação de tratamento observada.

$$5. \quad T_i = Y_i(t) \vee C_i \forall t \in T(n)$$

²⁵ a.e.: do inglês *for almost everywhere* e do francês *p.p (presque partout)*, se referindo a quase todos valores possíveis.

Ou seja, para todo i , o tratamento T é independente do desfecho, $Y(t)$, dadas as covariáveis, C , para todos níveis de t . Em outras palavras, assumimos que todas as informações relevantes (fatores de risco), usadas para gerar a alocação da exposição (i.e., status de desmatamento) estão contidas nas covariáveis da linha de base (ignorabilidade/permutabilidade condicional).

6. $P(O = o) > 0$ para todos os possíveis valores de o

Por último temos a assunção de positividade a nível de rede, onde o conjunto de dados (tratamento, desfecho e covariáveis) devem possuir uma probabilidade conjunta maior que zero.

Sob as assunções 1-3, pode-se definir $\psi_i(t)$ como:

$$E(Y_i(t)) = \sum_c E(Y_i \vee T=t, C=c) f(c) \quad (A9) \quad \sum_{y,c} y_i f(y \vee T=t, C=c) f(c)$$

Então o desfecho potencial para o município i sob uma estratégia de alocação parametrizada por γ pode ser expresso como:

$$\psi_i(\gamma) = \sum_{t \in T(n)} \sum_{y,c} y_i f(y \vee T=t, C=c) g(t; \gamma) f(c) \quad (A10)$$

E, por fim, pode ser calculada a média para obter efeitos causais a nível de rede (i.e., estado).

Como descrito por Tchetgen Tchetgen (2017), o desfecho potencial para o município i sob o tratamento t , como descrito na fórmula A.9 como $E(Y_i(t))$, pode ser calculado por meio de um algoritmo de amostragem de Gibbs. E assumindo que as variáveis (\mathbf{Y} , \mathbf{T} e \mathbf{C}) podem ser representadas na forma de um gráfico de cadeia G_ϵ com ϵ adjacências. Dado o nosso grafo G_ϵ , as seguintes independências condicionais seguidas por propriedade Markov global associada aos gráficos de cadeia,

$$Y_i \perp \{Y_k, T_k, C_k\} \vee (T_i, C_i, O_i) \text{ para todo } i \text{ e } k \notin N_i, k \neq i; \quad (A11)$$

$$C_i \perp C_{-i} \{O_i \vee C_{-i} \cap O_i \text{ para todo } i \text{ e } k \notin N_i, k \neq i \quad (A12)$$

A equação A11 significa que o status do desfecho (vetor/LVC/LVH) de um município é independente dos dados (exposição, covariáveis e desfecho) dos municípios com o qual ele não faz fronteira, condicional aos dados dos municípios que ele faz fronteira. Da mesma maneira, a equação A12 afirma que as covariáveis de um município são independentes das covariáveis dos municípios que ele não faz fronteira, dado as covariáveis dos seus municípios contactantes. Essas assunções permitem formalizam a ideia que os contactantes de um município criam um cobertor de Markov²⁶, deixando o município protegido da influência

²⁶ O termo cobertor de Markov, ou Markov blanket, foi criado por Pearl (1988). O cobertor Markov para um nó (i.e., variável) em um modelo gráfico contém todas as variáveis que

do restante da rede. Note, porém, que essa assunção não exclui a dependência entre os municípios conectados. Por exemplo, C_i não é independente C_j se há um k que (i, k) e (j, k) pertencem a \mathbf{E} até se $(i, j) \notin \mathbf{E}$.

Assim como assumido em Tchetgen Tchetgen (2017) e replicado por Fulcher (2019), nós assumimos que o conjunto \mathbf{Y} segue um modelo condicional, chamado de modelo automático condicional²⁷, dado \mathbf{T} e \mathbf{C} (BESAG, 1974). No nosso caso, onde o desfecho Y_i é binário, um modelo automático logístico codificando as restrições de Markov descritas em A11 e A12 é obtida especificando um modelo de verossimilhança conjunta sob a forma de

$$f(y|t, c) = \frac{h_y(y; \beta, t, c)}{Z_y(\beta, t, c)} \quad (\text{A13})$$

Onde

$$h_y(y; \beta, t, c) = \exp \left(\sum_i y_i \beta_i + \sum_{(i,j) \in \mathbf{E}} w_{ij} y_i y_j \right) \quad (\text{A14})$$

$\beta = \{\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5 : p \in 1, \dots, P\}$, onde P é o número de covariáveis (i.e., $P=5$), w_i indica o número de vizinhos do município i , já $w_i > 0$ é um indicador que o município tem que ter pelo menos um vizinho, e $Z_y(\beta, t, c) = \sum_{\bar{y}} h_y(\bar{y}; \beta, t, c)$ é uma constante normalizadora onde \bar{y} contém 2^n alocações de desfechos possíveis na rede.

Igualmente, o modelo para as covariáveis, que é indexado pelo parâmetro α .

$$f(c) = \frac{h_c(c; \alpha)}{Z_c(\alpha)} \quad (\text{A15})$$

Onde

$$h_c(c \vee \alpha) = \exp \left(\sum_p \sum_i c_{i,p} K(c_{i,p}; \tau_p) + \sum_{p \neq s} P_{ps} \sum_i c_{i,p} c_{i,s} + I(w_i > 0) \sum_p v_p \sum_{i,j \in N_i} c_{i,p} c_{j,p} w_{ij} \right) \quad (\text{A16})$$

Para uma função suave conhecida $K(c_{i,p}; \tau_p)$ com parâmetro desconhecido τ_p e $Z_c(\alpha) = \sum_{\bar{c}} h_c(\bar{c}; \alpha)$ é uma constante normalizante onde $\alpha = \{\tau_p, P_{ps}, v_p : p \in 1, \dots, P, s \in 1, \dots, P, e p \neq s\}$.

Ambos modelos do desfecho e covariáveis podem ser especificados como modelos condicionais (chamados de *auto-models*) para o desfecho de um município, dado suas covariáveis e o desfecho e covariáveis dos seus vizinhos, e também para as covariáveis de um município dado as covariáveis dos seus vizinhos. No nosso caso, a equação A13 pode ser resumido como:

protegem o nó do resto da rede. Isso significa que o cobertor Markov de um nó é o único conhecimento necessário para prever o comportamento desse nó e de seus filhos.

²⁷ tradução literal para *conditional auto-model*

$$Pr(Y_i=1|T=t, C=c, Y_{-i}=y_{-i})=Pr(y_i=1|T_i=t_i, C_i=c_i, O_i)$$

$\hookrightarrow \text{expit}$

$$\beta_4 \sum_{j \in N_i} \frac{t_j}{w_i} + \sum_p \beta_{5,p} \sum_{j \in N_i} \frac{c_{j,p}}{w_i}$$

(A17)

Já a equação para as covariáveis (A15) é definida como:

$$Pr(C_{i,p}=1|C_{i,-p}=c_{i,-p}, C_{-i}=c_{-i})=Pr(C_{i,p}=1|C_{i,-p}=c_{i,-p}, \{C_j: j \in N_i\})$$

$$\hookrightarrow \text{expit} \left(K(c_{i,p}; \tau_p) + \sum_{p \neq s} P_{ps} c_{i,s} + v_p \sum_{j \in N_i} \frac{c_{j,p}}{w_i} \right)$$

(A18)

Para uma função suave conhecida $K(c_{i,p}; \tau_p)$ com parâmetro desconhecido τ_p .

ESTIMAÇÃO

A estimação dos efeitos causais da rede se deu em dois passos, como descrito em Fulcher (2019). No primeiro passo, foram estimados os valores dos parâmetros (α e β) por meio de uma abordagem bayesiana chamada de algoritmo de Metropolis-Hastings duplo, que se trata de dois algoritmos de Metropolis-Hastings aninhados (Liang, 2010) e que podem ser vistos em maiores detalhes em Fulcher (2019).

Com os resultados da distribuição posterior do passo anterior, nós pudemos estimar a distribuição posterior do efeito causal a nível de indivíduo (i.e., município) sob dada estratégia de alocação do tratamento (i.e., exposição). Seguindo Tchetgen Tchetgen (2017) e Fulcher (2019), fizemos um amostrador de Gibbs com as densidades de A10:

$$f(y|t, c; \beta^r) \times g(t; \gamma) \times f(c; \alpha^r)$$

(A19)

O procedimento do passo 1 foi implementado por meio da função `agcParam`, já a segunda parte foi feita a partir da função `agcEffect`, ambas funções presentes no pacote `autognet` (Fulcher, 2019).

Referências:

FULCHER, Isabel Rose. **Statistical Inference for Causal Mechanisms: Mediation and Interference**. 2019. Tese (Doutorado em Bioestatística) - Departamento de Bioestatística, Universidade de Harvard.

HALLORAN, M. Elizabeth; STRUCHINER, Claudio J. Causal inference in infectious diseases. **Epidemiology**, p. 142-151, 1995.

HUDGENS, Michael G.; HALLORAN, M. Elizabeth. Toward causal inference with interference. **Journal of the American Statistical Association**, v. 103, n. 482, p. 832-842, 2008.

TCHETGEN, Eric J. Tchetgen; FULCHER, Isabel; SHPITSER, Ilya. Auto-g-computation of causal effects on a network. **arXiv preprint arXiv:1709.01577**, 2017.

APÊNDICE B – Descrição das análises de dados de uso do solo e covariáveis

Os dados do MapBiomias são disponíveis por ano (1980-2019) para todo o Brasil, dessa base foram extraídos os dados referentes ao estado de São Paulo. Os arquivos disponibilizados para download são em formato *raster*, que é uma imagem composta por uma matriz de células (pixels), onde cada pixel armazena um valor indicativo do tipo de objeto (classe) daquela área geográfica. A descrição detalhada das classes pode ser vista na página do [MapBiomias](#). Como o *raster* é de todo o estado, vamos utilizar um arquivo *shapefile* (.shp) com os limites dos municípios do estado para fazer o recorte dos municípios. Depois disso vamos criar uma tabela (.csv) com a contagem de pixels de cada classe e município.

```
# installing and loading packages:
library(raster)
library(sp)
library(tidyverse)
setwd ("/Users/Cleber/Desktop/msc/data/land_use")
ras00 <- raster("mapbiomas-saopaulo-2000.tif")
sp <- st_read("sp.shp") #shapefile de sp

#Extrair o NDVI de cada poligono
info_00 <- raster::extract(x = ras00, y = sp)

# gerar tabela (nome polígono, tipo de pixel, contagem de cada tipo de
pixel em cada polígono)

#1. Definição do tipo de data.frame (nomes colunas)
tabela <- data.frame(poligono = NA,
                     pixel = NA,
                     contagem = NA)
#2. Definição da tabela (tabela.temp=tab. temporária)
for(i in 1:length(info_00)){
  # criando a tabela temporaria para guardar os dados do polígono i
  tabela.temp <- data.frame(
    poligono = i,
    contagem = table(info_00[[i]])
  )
  # redefinindo os nomes para juntar as tabelas
  colnames(tabela.temp) <- colnames(tabela)

  # juntando as tabelas
  tabela <- rbind(tabela, tabela.temp)

  # Mostrando em que pé que anda o loop
  print(i)
}
tabela.wide <- tabela %>%
  # Mudando a tabela de formato:
  spread(key = pixel, value = contagem) %>%

  # Tirando quem é NA na coluna polígono
  filter(!is.na(poligono))%>%

  #Tirando a coluna NA
```

```
select(-`<NA>`) %>%

# Criando a coluna de identificação do IBGE
mutate(ibge7d = as.character(sp$CD_GEOCODM))

#salvando arquivo
write.csv(tabela.wide, file = "tabela_wide_2000.csv", row.names = F)
```

Após isso, os arquivos gerados apresentavam a quantidade de pixels de cada classe (apresentados nas colunas) de uso do solo para cada um dos 645 municípios (representados pelos códigos do ibge, nas linhas) para o respectivo ano. Abaixo um exemplo:

```
#Tabela com dados referentes ao ano de 2000
#para evitar repetições de código, mostraremos somente os calculos para
esse ano
tab_00 <- read.csv( "tabela_wide_2000.csv" , sep = ";" )
head(tab_00)
##      ibge7d X0   X12 X13      X15   X19   X20      X21 X23   X24 X25 X27 X29
X3
## 1 3500105   5    NA  NA 305601      NA 63665 105796   NA 7378 450   NA  NA
10594
## 2 3500204  NA    NA  NA 112689      NA    NA  45632   NA  537 169   NA  NA
18860
## 3 3500303 12 3796   NA  52882  5096 65263 332619   NA 5988  18   NA  NA
82821
## 4 3500402   2    NA  NA  39474      NA   36  81341   NA  944 455   NA  NA
49452
## 5 3500501  NA    5   NA  30788      NA    NA  27819   NA 2412 119   NA  NA
11018
## 6 3500550  NA 3633   NA 239327 13019   155 131047   NA  661 117   NA  NA
53289
##      X30 X32   X33   X4 X5      X9 total
## 1   NA  NA   444   NA NA      NA 493933
## 2   NA  NA 75065   NA NA      NA 252952
## 3   32  NA  2052 1895 NA 19411 571885
## 4  203  NA    15   NA NA    122 172044
## 5   NA  NA    55   NA NA    420  72636
## 6   NA  NA  1104 6241 NA 41936 490529
```

Os dados acima representam a quantidade de pixels para cada classe. Para facilitar, iremos trabalhar com a porcentagem que cada classe de interesse apresenta no município. Vamos criar):

```
# Calculando porcentagens de área de floresta (código 3):
porc_flo_00 <- (tab_00$X3/tab_00$total) *100
ibge <- tab_00$ibge7d
flo <-
cbind(porc_flo_99,porc_flo_00,porc_flo_01,porc_flo_02,porc_flo_03,
porc_flo_04,porc_flo_05,porc_flo_06,porc_flo_07,porc_flo_08,porc_flo_09
,porc_flo_10,porc_flo_11,porc_flo_12,porc_flo_13, porc_flo_14,
porc_flo_15, porc_flo_16, porc_flo_17, porc_flo_18)
write.csv(flo , file="porcentagem_floresta.csv" , row.names=FALSE)
```

Agora vamos fazer algumas análises com as covariáveis, para cobrir algumas suposições.

Primeiro veremos se as covariáveis estão relacionadas, por meio da matriz de correlação e seus respectivos valores-p:

```
#install.packages("Hmisc")
library(Hmisc)
## Loading required package: lattice
## Loading required package: survival
## Loading required package: ggplot2
## Attaching package: 'Hmisc'
##
data <- read.csv('covariates.csv', sep=",", dec = ".", as.is=T)
(corr <- rcorr(cbind(data$for, data$GDP, data$Gas, data$h_pop)))
##      [,1] [,2] [,3] [,4]
## [1,]  1.00 0.17 -0.03 -0.02
## [2,]  0.17 1.00  0.06  0.16
## [3,] -0.03 0.06  1.00  0.12
## [4,] -0.02 0.16  0.12  1.00
## n= 620
##
## P      [,1]      [,2]      [,3]      [,4]
## [1,]      0.0000  0.4878  0.5483
## [2,] 0.0000      0.1351  0.0000
## [3,] 0.4878 0.1351      0.0034
## [4,] 0.5483 0.0000 0.0034
##
```

De acordo com a matrix, e tomando como rule of thumb uma correlação acima de .7, que é amplamente citado na literatura, os dados não têm uma correlação forte.

Agora vamos ver se há multicolinearidade, para fazer isso de forma simples, vamos gerar um modelo logístico:

```
#criando objeto do glm
m <- glm(vec~fore_bin+Gas+GDP+h_pop, family='binomial', data=data)

#depois disso podemos ver se há multicolinearidade
install.packages('car')
library(car)
## Loading required package: carData
(vif(m))
##   fore_bin      Gas      GDP      h_pop
##  1.001500  1.054080  1.053337  1.253379
```

Na literatura é amplamente utilizado um VIF > 3 para indicar multicolinearidade. O que não foi observado nos nossos dados.

Agora vamos criar a matrix de vizinhança dos municípios de São Paulo, essa matriz será a rede que permitirá os cálculos dos efeitos geral, direto e indireto.

```
## Loading required package: carData
install.packages('maptools')
install.packages('spdep')
library(maptools)
library(spdep)

#shapefile
sp<- st_read("/Users/Cleber/Desktop/msc/data/land_usesp.shp")
```

```
#getting neighbor matrix  
sp.F0Q = poly2nb(sp, queen=TRUE)  
sp_matrix <- nb2mat(sp.F0Q, style='B', zero.policy = TRUE)  
write.csv2(sp_matrix, "sp_matrix.csv")
```

APÊNDICE C: Descrição do script dos DAGs e variáveis para ajuste

First of all, we need to install and load the ggdag package:

```
install.packages("ggdag", repos = "http://cran.us.r-project.org")
library(ggdag)

## Loading required package: ggplot2
## Attaching package: 'ggdag'
```

Then we need to create an object with the coordinates (x and y) of each node (variable):

```
coord <- list (
  x = c(Agri = 0.326, Air_p = 0.001, Alt = -0.093 , CVL = 0.741 ,
        D_pop= 0.654 ,For = 0.167 , GDP = 0.024 ,Gas = 0.582 , Humi = 0.005
        , H_pop = 0.164 , HVL = 0.986 , Prec = 0.008 , Sea_p = 0.004 , Surv
        = 0.862 , Temp = 0.409 , Urb = 0.358 , Vec = 0.636 , V_C = 0.854) ,
  y = c(Agri = 0.512, Air_p = 1.105 , Alt = 0.783 , CVL = 0.587 ,
        D_pop= -0.092 , For = 0.505 , GDP = 0.276 ,Gas = 0.176 , Humi = 0.834
        , H_pop = 0.056 , HVL = 0.565 , Prec = 0.656 , Sea_p = 0.973 , Surv
        = 0.037 , Temp = 0.829 , Urb = 0.283 , Vec = 0.579 , V_C = 0.593)
)
```

Later then, it's necessary to create an object to account the relationship between the nodes and the corresponding treatment and outcome:

```
dag <- dagify(HVL ~ Surv,
              HVL ~ Gas,
              HVL ~ H_pop,
              HVL ~ V_C,
              V_C ~ Vec,
              V_C ~ CVL,
              CVL ~ Surv,
              CVL ~ D_pop,
              CVL ~ Gas,
              CVL ~ Vec,
              Vec ~ Surv,
              Vec ~ Gas,
              Vec ~ Urb,
              Vec ~ Prec,
              Vec ~ Humi,
              Vec ~ Temp,
              Surv ~ GDP,
              D_pop ~ H_pop,
              H_pop ~ GDP,
              Gas ~ H_pop,
              Gas ~ GDP,
              Urb ~ GDP,
              Urb ~ For,
```

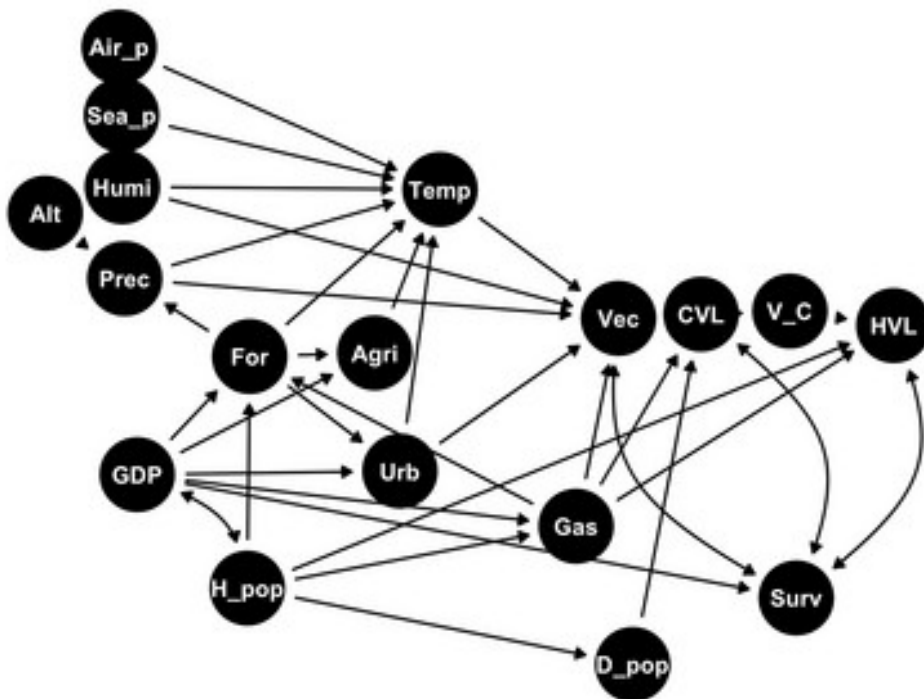
```

Agri ~ For,
Agri ~ GDP,
For ~ GDP,
For ~ Gas,
For ~ H_pop,
Temp ~ Prec,
Temp ~ Humi,
Temp ~ Sea_p,
Temp ~ Air_p,
Temp ~ Agri,
Temp ~ Urb,
Temp ~ For,
Prec ~ Alt,
Prec ~ For,
Humi ~ Alt,
Humi ~ Prec,
Humi ~ Sea_p,
coords = coord,
exposure = "For",
outcome = "Vec")

```

Now it is possible to print the dag:

```
ggdag(dag) + theme_dag_blank(base_size = 8, base_family = "")
```

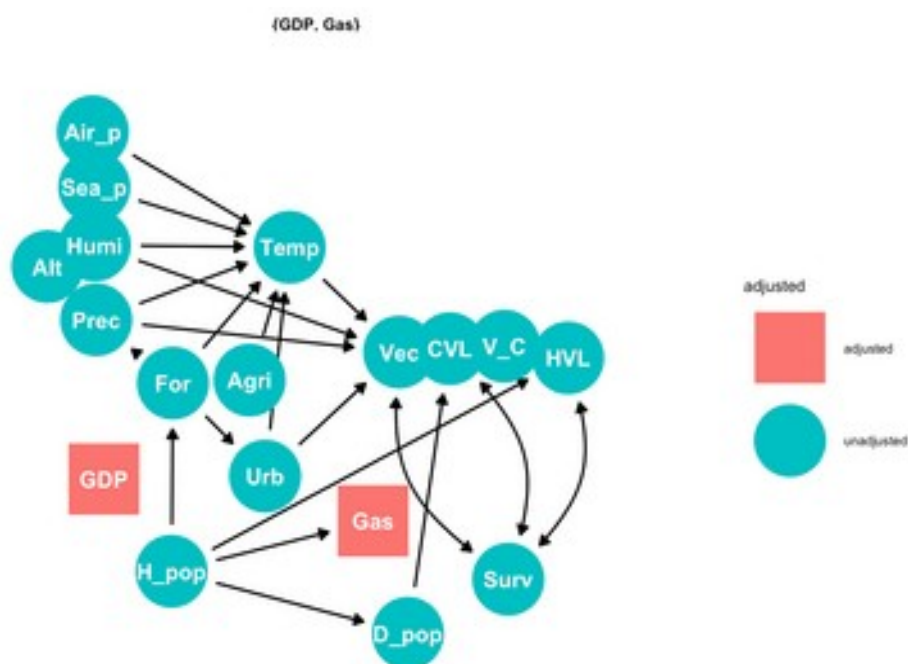


the `ggdag` package also helps to quickly get the minimally sufficient adjustment sets to adjust for when analyzing the effect of a treatment (Deforestation) on the outcomes (Vector/CVL/HVL).

As we had three outcomes, let's do it in separate:

(i) vector presence

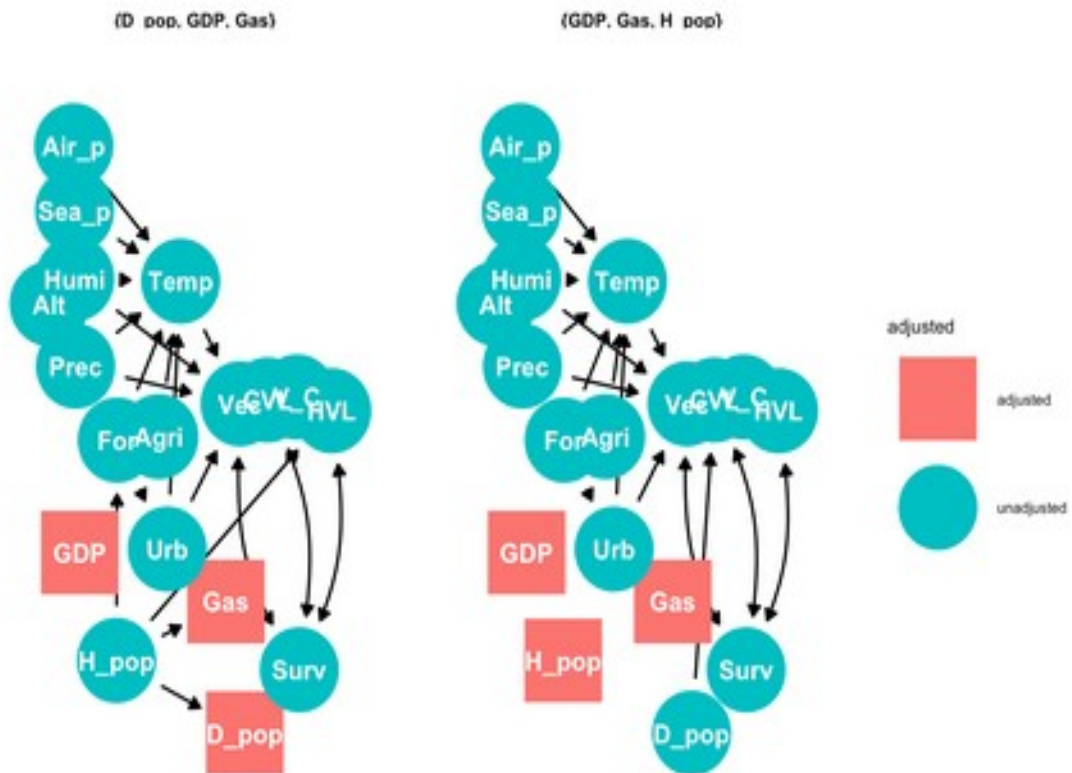
```
ggdag_adjustment_set(dag)+theme_dag_blank(base_size = 8,base_family = " ")
```



(ii) canine visceral leishmaniasis

```
dag <- dagify(HVL ~~~ Surv,
              ...
              coords = coord,
              exposure = "For",
              outcome = "CVL")
```

```
ggdag_adjustment_set(dag)+theme_dag_blank(base_size = 8,base_family = " ")
```

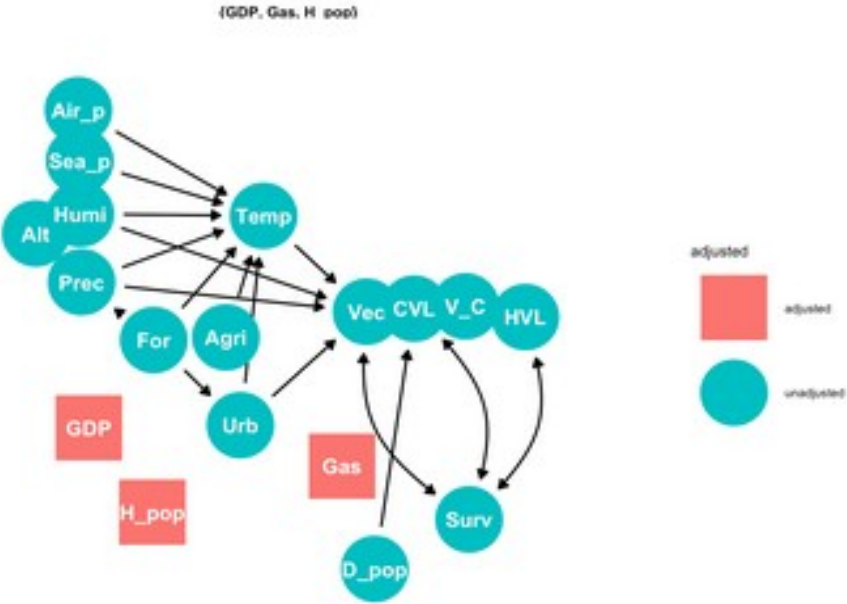



(iii) human visceral leishmaniasis

```
dag <- dagify(HVL ~ Surv,
```

```
  ...
  coords = coord,
  exposure = "For",
  outcome = "HVL")
```

```
ggdag_adjustment_set(dag) + theme_dag_blank(base_size = 8, base_family =
```



APÊNDICE D – script bayes auto-g-computation

Bayes auto-g-computation script

Cleber V B dos Santos

22/12/2019

Script of the Bayesian auto-g-computation algorithm used to evaluate the causal effects (overall, direct, and indirect) of deforestation in the vector presence, CVL, and HVL; correspondent to the dissertation: "Como o desmatamento influencia a ocorrência da leishmaniose visceral? Uma abordagem contrafactual" by Cleber Vinicius Brito dos Santos

installing packages and support files:

```
#downloading package and support files
#devtools::install_github("isabelfulcher/autognet", ref = "nuterms")

#get working directory

setwd("~/Users/Cleber/Desktop/msc/data/land_use ")

#getSheets
data <- read.csv('data.csv',header = T,sep=",")

#adjacency matrix (0/1s) specifying the network structure (i.e.,
municipalities that share borders)
sp <- read.csv('sp_matrix.csv',sep=",",header = F)
sp <- as.matrix(sp)
```

First, let's estimate the parameter for the vector presence model:

```
#naming variables
treatment <- data$fore_bin
outcome <- data$vec
cov1 <- data$Gas
cov2 <- data$GDP #gdp
cov3 <- data$H_pop #human population
adjmat <- sp

#creating a list
dat <- data.frame(treatment,outcome,cov1,cov2,cov3)
caus <- list(adjmat,dat)

#defining objects
treatment <- "treatment"
outcome <- "outcome"
```

```

B <- 25000      #number of interactions for the mcmc
R <- 15         #number of interactions for the gibbs inner loop
seed <- 1       #set.seed
s_a <- 0.005    #
s_b <- 0.005    #
p_a <- 1        #prior for alpha
p_b <- 1        #prior for beta
adjmat <- caus[[1]] # the 0/1 adjacency matrix for the state of Sao
Paulo network
datas <- caus[[2]] #dataframe containing the covariates

```

Now we're able to estimate the parameters via double metropolis hastings

```

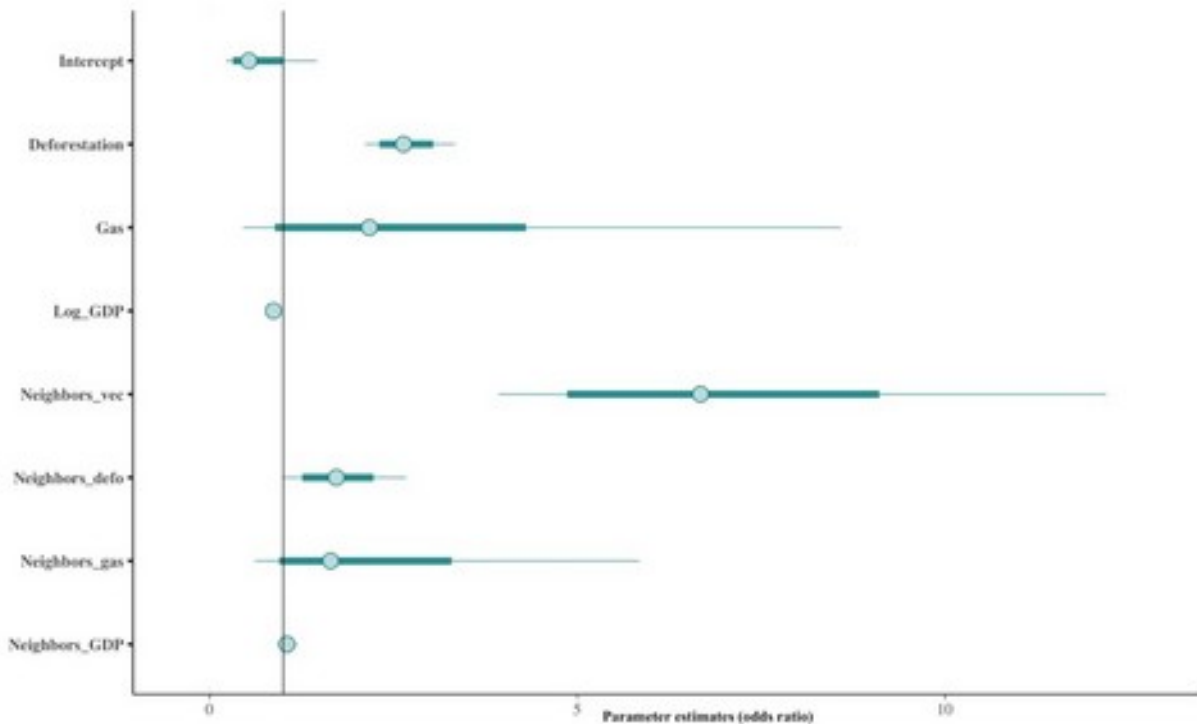
mod_vec <- agcParam(datas, treatment, outcome, adjmat, B = B, R =
R ,seed = seed, scale.alpha = s_a, scale.beta = s_b, prior.alpha =
p_a, prior.beta = p_b)

betas <- mod$beta

colnames(betas) <- c('Intercept','Deforestation','Gas','LogGDP',
'Neighbors_vec','Neighbors_defo','Neighbors_gas','Neighbors_GDP')

#posterior mean and credible intervals for parameters estimates
tiff("beta_estim_log.tiff", units="cm", width=25, height=15,
res=300)
color_scheme_set("teal")
p_vec <- mcmc_intervals(betas, point_est = "mean",prob =
.5,prob_outer = .95)
p_vec + vline_at(1, linetype = 1, size = 0.3)
dev.off()

```

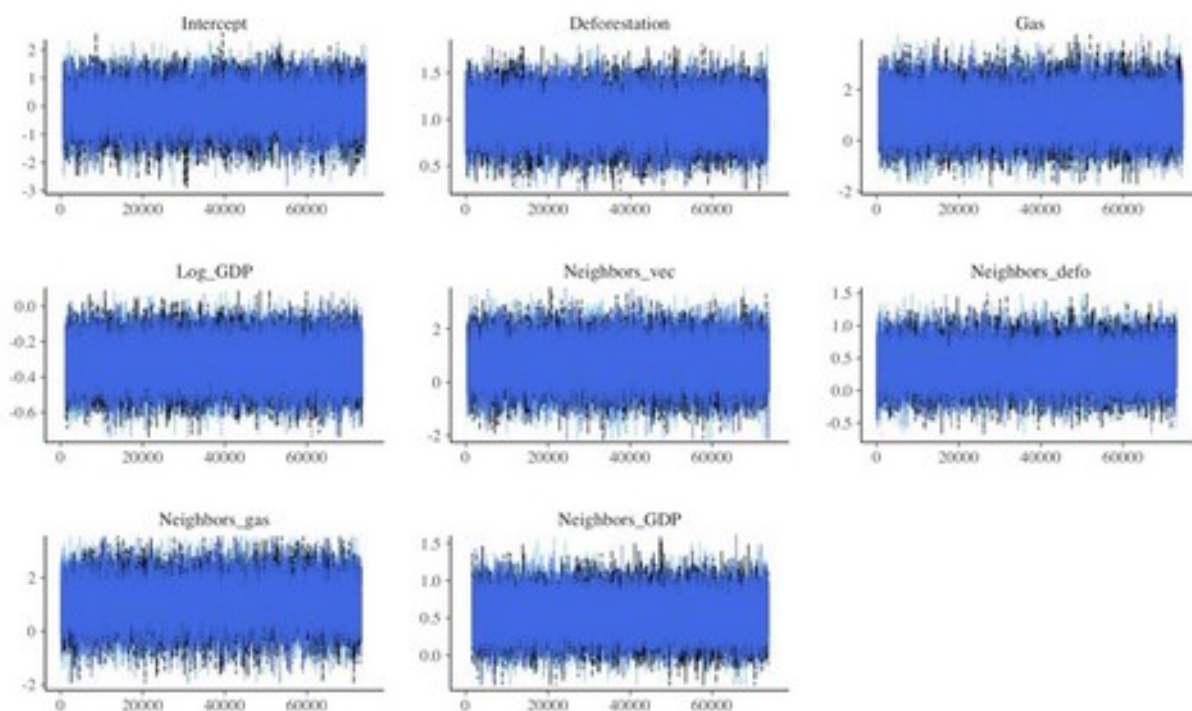


```
#parameter values and CI
vec_param <- cbind(p$data$ll,p$data$m,p$data$hh)
row.names(vec_param) <-
c('Intercept','Deforestation','Gas','Log_GDP',
  'Neighbors_vec','Neighbors_defo','Neighbors_gas','Neighbors_GDP')
colnames(vec_param) <- c('2.5%','mean','97.5%')
vec_param
```

	2.5%	mean	97.5%
Intercept	0.2150649	0.5330642	1.4579395
Deforestation	2.1159625	2.6361226	3.3386914
Gas	0.4522859	2.1717504	8.5850955
Log_GDP	0.8149061	0.8673880	0.9252537
Neighbors_vec	3.9299413	6.6748001	12.1941265
Neighbors_defo	0.9860946	1.7225879	2.6735296
Neighbors_gas	0.6147502	1.6450148	5.8440174
Neighbors_GDP	0.9176805	1.0506761	1.1897057

Agora vamos fazer análise do MCMC

```
#examining mcmc
tiff("trace.tiff", units="cm", width=25, height=15, res=300)
mcmc_trace(betas)
dev.off()
```



A inspeção da trajetória das cadeias, com os trace plots, são uma forma de fazer diagnóstico do mcmc, observando a representatividade das cadeias. Na inspeção visual as cadeias devem estar superpostas. Como pode-se ver grupo de gráficos, as cadeias parecem bem representativas

Estimates of causal effects of deforestation on *lu. longipalpis* presence in the state of sao paulo:

```
#50% p
vec_50_estim <- agcEffect(mod_vec, burnin = 5, thin = 0.5,
  treatment_allocation = 0.5)

caus_vector_50 <- apply(vec_50_estim,2,mean)
caus_ci_vector_50 <- apply(vec_50_estim,2,quantile,c(0.025,0.975))
causal_eff_vec_50 <- rbind(caus_ci_vector_50[1,], caus_vector_50,
  caus_ci_vector_50[2,])

#20% p
vec_20_estim <- agcEffect(mod_vec, burnin = 5, thin = 0.5,
  treatment_allocation = 0.2)

caus_vector_20 <- apply(vec_20_estim,2,mean)
caus_ci_vector_20 <- apply(vec_20_estim,2,quantile,c(0.025,0.975))
causal_eff_vec_20 <- rbind(caus_ci_vector_20[1,], caus_vector_20,
  caus_ci_vector_20[2,])

#0% p
vec_0_estim <- agcEffect(mod_vec, burnin = 5, thin = 0.5,
```

```

treatment_allocation = 0)

caus_vector_0 <- apply(vec_0_estim,2,mean)
caus_ci_vector_0 <- apply(vec_0_estim,2,quantile,c(0.025,0.975))
causal_eff_vec_0 <- rbind(caus_ci_vector_0[,1], caus_vector_0,
caus_ci_vector_0[,2])

#80% p
vec_80_estim <- agcEffect(mod_vec, burnin = 5, thin = 0.5,
  treatment_allocation = 0.8)

caus_vector_80 <- apply(vec_80_estim,2,mean)
caus_ci_vector_80 <- apply(vec_80_estim,2,quantile,c(0.025,0.975))
causal_eff_vec_80 <- rbind(caus_ci_vector_80[,1], caus_vector_80,
caus_ci_vector_80[,2])

#100% p
vec_100_estim <- agcEffect(mod_vec, burnin = 5, thin = 0.5,
  treatment_allocation = 1)

caus_vector_100 <- apply(vec_100_estim,2,mean)
caus_ci_vector_100 <- apply(vec_100_estim,2,quantile,c(0.025,0.975))
causal_eff_vec_100 <- rbind(caus_ci_vector_100[,1], caus_vector_100,
caus_ci_vector_100[,2])

#separating the effects for the plot
average_vec <- cbind(causal_eff_vec_0 [,1],causal_eff_vec_20 [,1],
causal_eff_vec_50 [,1], causal_eff_vec_80 [,1], causal_eff_vec_100
[,1])
colnames(average_vec) <- c('0','0.2','0.5','0.8','1')

direct_vec <- cbind(causal_eff_vec_0 [,2], causal_eff_vec_20 [,2],
causal_eff_vec_50 [,2], causal_eff_vec_80 [,2], causal_eff_vec_100
[,2])
colnames(direct_vec) <- c('0','0.2','0.5','0.8','1')

spillover_vec <- cbind(causal_eff_vec_0 [,3], causal_eff_vec_20
[,3],
causal_eff_vec_50 [,3], causal_eff_vec_80 [,3], causal_eff_vec_100
[,3])
colnames(spillover_vec) <- c('0','0.2','0.5','0.8','1')

# average_vector
-----

df_vec_ave <- data.frame(grp = c('0','0.2','0.5','0.8','1'),
fit=apply(average_vec,2,mean),yma=apply(average_vec,2,quantile,c(0.9
75)),
ymi=apply(average_vec,2,quantile,c(0.025)))

j <- ggplot(df_vec_ave, aes(grp, fit,ymax=yma,ymin=ymi))

```

```
j+ geom_errorbar(width=.2)+ geom_pointrange() + theme_classic() +
labs(y='Total effect',x='Deforestation prevalence')
df_vec_ave
```

	mean	upper	lower bound
p=0	0.5004652	0.7302823	0.2344355
p=0.2	0.5314005	0.7320968	0.3011694
p=0.5	0.5623358	0.7363306	0.3824194
p=0.8	0.5870417	0.7520968	0.4262097
p=1	0.6117475	0.7793952	0.4614113

```
# Direct_vector
```

```
-----

df_dir_ave <- data.frame(grp = c('0','0.2','0.5','0.8','1'),
fit=apply(direct_vec,2,mean),yma=apply(direct_vec,2,quantile,c(0.975
)),
ymi=apply(direct_vec,2,quantile,c(0.025)))

j_dir <- ggplot(df_dir_ave, aes(grp, fit,ymax=yma,ymin=ymi))
j_dir + geom_errorbar(width=.2)+ geom_pointrange() + theme_classic()
+ labs(y='Total effect',x='Deforestation prevalence')
df_vec_ave
```

	mean	upper	lower bound
P=0	0.0823148	0.1079225	0.05383548
P=0.2	0.0827599	0.1082034	0.05415381
P=0.5	0.0827121	0.1082258	0.05435484
P=0.8	0.0840191	0.1102016	0.05661290
P=1	0.0853261	0.1134274	0.05544355

```
# Spillover_vector
```

```
-----

df_spi_ave <- data.frame(grp = c('0','0.2','0.5','0.8','1'),
fit=apply(spillover_vec,2,mean),yma=apply(spillover_vec,2,quantile,c
(0.975)), ymi=apply(spillover_vec,2,quantile,c(0.025)))

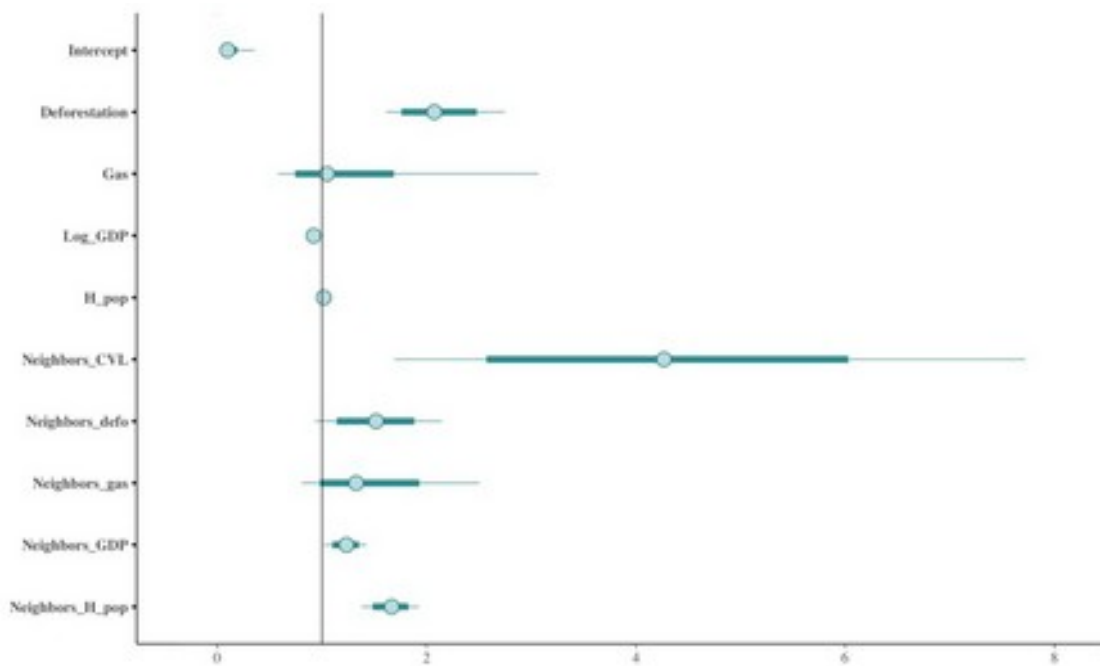
j_spi <- ggplot(df_spi_ave, aes(grp, fit,ymax=yma,ymin=ymi))
j_spi+ geom_errorbar(width=.2)+ geom_pointrange() + theme_classic()
+ labs(y='Indirect effect',x='Deforestation prevalence')
```


	mean	upper	lower bound
P=0	0.0823148	0.1079225	0.05383548
P=0.2	0.0827599	0.1082034	0.05415381
P=0.5	0.0827121	0.1082258	0.05435484
P=0.8	0.0840191	0.1102016	0.05661290
P=1	0.0853261	0.1134274	0.05544355

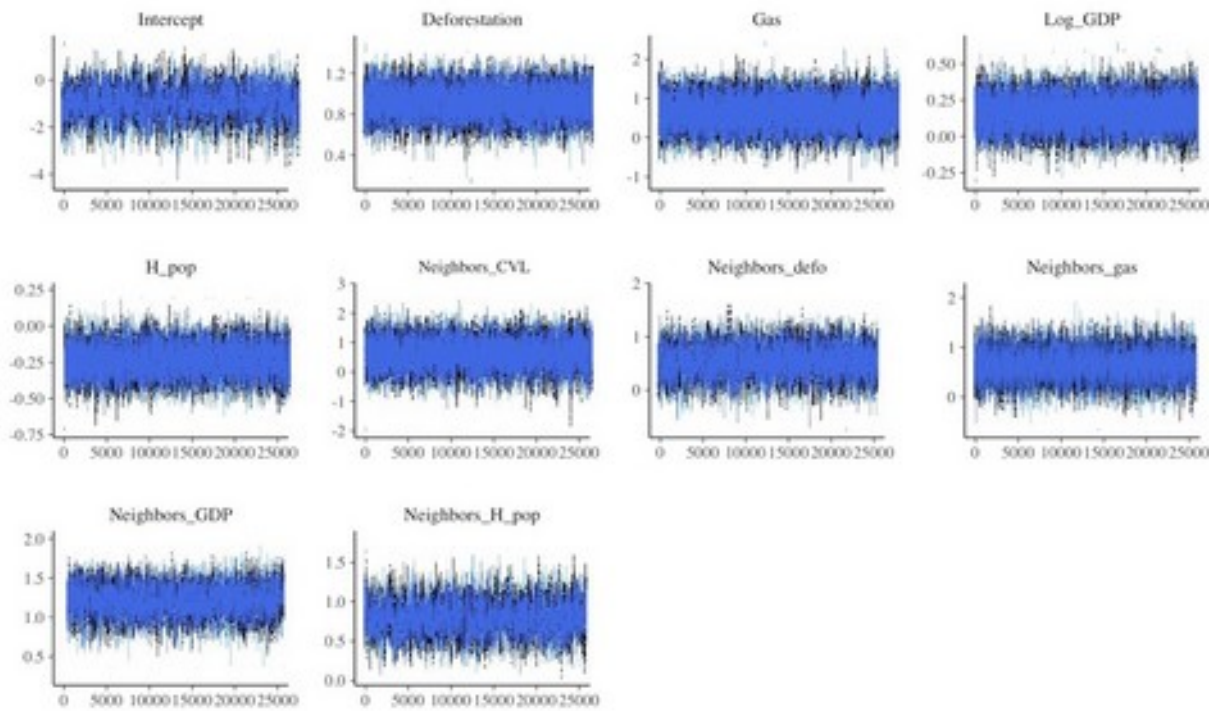
The resulting graphics were presented below, in the results section.

Now that we had illustrated all the steps, we'll show only the results of the remaining analysis, without presenting the script, since they are equal.

CVL figures:



-Trace plots (cvl)



```
# average effect_dog
```

	mean	upper	lower bound
p=0	0.1526130	0.2220161	0.05790323
p=0.2	0.1580670	0.2295161	0.06286290
p=0.5	0.1635210	0.2351613	0.06532258
p=0.8	0.1680542	0.2416129	0.06580645
p=1	0.1735082	0.2458065	0.06790323

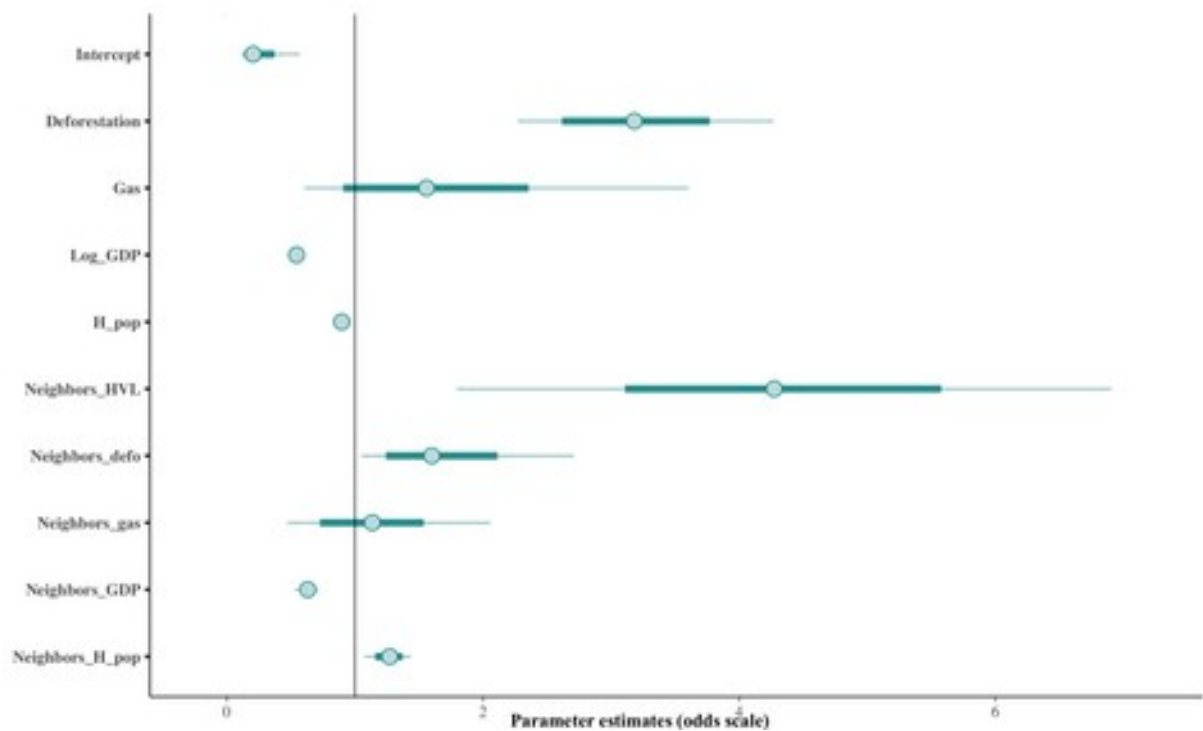
```
# direct effect_dog
```

	mean	upper	lower bound
p=0	0.02812629	0.04439516	0.003709678
p=0.2	0.02712026	0.04112903	0.005646133
p=0.5	0.02611423	0.03903226	0.005645161
p=0.8	0.02569658	0.03758006	0.005641275
p=1	0.02469055	0.03758065	0.004032258

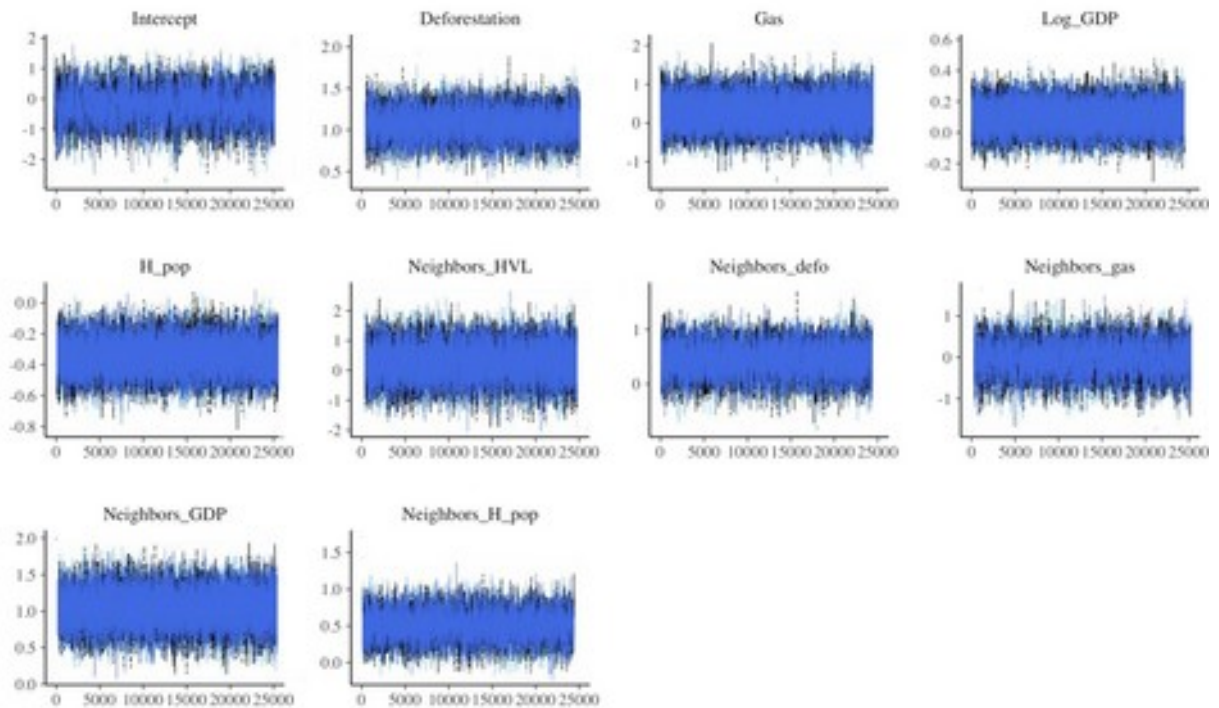
direct effect_dog

	mean	upper	lower bound
p=0	0.003459035	0.013709680	-0.007258063
p=0.2	-0.001376320	0.003225811	-0.005645155
p=0.5	-0.006211675	0.006451614	-0.017741931
p=0.8	-0.009087514	0.010000004	-0.026774190
p=1	-0.013922868	0.015000005	-0.040362902

HVL figures:



Trace plots (hvl)



```
# average effect_human
```

	mean	upper	lower bound
p=0	0.2060936	0.2912500	0.07709677
p=0.2	0.2500016	0.3614919	0.12370968
p=0.5	0.2939096	0.4318347	0.16250000
p=0.8	0.3378176	0.4895565	0.18947581
p=1	0.4695416	0.6651008	0.25526210

```
# direct effect_human
```

	mean	upper	lower bound
p=0	0.1031735	0.1396976	0.06167339
p=0.2	0.1053091	0.1419577	0.06338041
p=0.5	0.1053101	0.1419355	0.06338710
p=0.8	0.1053116	0.1419353	0.06331098
p=1	0.1074466	0.1459073	0.06258065

```
# indirect effect_human
```

	mean	upper	lower bound
p=0	-0.004206093	0.001858521	-0.01056083
p=0.2	-0.004206093	0.008951615	-0.01669355
p=0.5	0.006593004	0.023064519	-0.01241935
p=0.8	0.011992552	0.041451616	-0.02177419
p=1	0.028191197	0.095564520	-0.05050403