



Universidade do Estado do Rio de Janeiro

Centro de Tecnologia e Ciências

Instituto de Matemática e Estatística

Marcelo Goulart Correia


**Desenvolvimento de interfaces gráficas para estatística Bayesiana
aplicada à comparação mista de tratamentos**

Rio de Janeiro

2013

Marcelo Goulart Correia

**Desenvolvimento de interfaces gráficas para estatística Bayesiana
aplicada à comparação mista de tratamentos**



Dissertação apresentada, como requisito parcial para obtenção do Título de Mestre, ao Programa de Pós-Graduação em Ciências Computacionais, da Universidade do Estado do Rio de Janeiro.

Orientadores: Prof.^a Dra. Regina Serrão Lanzillotti
Prof. Dr. Bernardo Rangel Tura

Rio de Janeiro
2013

CATALOGAÇÃO NA FONTE
UERJ / REDE SIRIUS / BIBLIOTECA CTC/A

C824 Correia, Marcelo Goulart.
Desenvolvimento de interfaces gráficas para estatística Bayesiana aplicada à comparação mista de tratamentos/ Marcelo Goulart Correia. - 2013.
60 f.

Orientadores: Regina Serrão Lanzillotti, Bernardo Rangel Tura.
Dissertação (Mestrado) - Universidade do Estado do Rio de Janeiro, Instituto de Matemática e Estatística.

1. Interfaces gráficas de usuário (Sistema de computador) - Teses. 2. Teoria bayesiana de decisão estatística - Teses. 3. Medicamentos - Equivalência terapêutica – Teses. I. Lanzillotti, Regina Serrão. II. Tura, Bernardo Rangel. III. Universidade do Estado do Rio de Janeiro. Instituto de Matemática e Estatística. IV. Título.

CDU 004.514

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

Assinatura

Data

Marcelo Goulart Correia

**Desenvolvimento de interfaces gráficas para estatística Bayesiana
aplicada à comparação mista de tratamentos**

Dissertação apresentada, como requisito parcial para obtenção do Título de Mestre, ao Programa de Pós-Graduação em Ciências Computacionais, da Universidade do Estado do Rio de Janeiro.

Aprovado em 12 de setembro de 2013

Orientadora:

Prof.^a Dra. Regina Serrão Lanzillotti
Instituto de Matemática e Estatística - UERJ

Banca Examinadora:

Prof. Dr. Bernardo Rangel Tura (Orientador)
Instituto Nacional de Cardiologia

Prof. Dr. Paulo Eustáquio Duarte Pinto
Instituto de Matemática e Estatística - UERJ

Prof.^a Dra. Mariane Branco Alves
Universidade Federal do Rio de Janeiro

Rio de Janeiro

2013

RESUMO

CORREIA, Marcelo Goulart. **Desenvolvimento de interfaces gráficas para estatística Bayesiana aplicada à comparação mista de tratamentos**. 2013. 66 f. Dissertação (Mestrado em Ciências Computacionais) – Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2013.

A partir dos avanços obtidos pela indústria farmacêutica surgiram diversos medicamentos para o combate de enfermidades. Esses medicamentos possuem efeito tópico similar porém com suaves modificações em sua estrutura bioquímica, com isso a concorrência entre as indústrias farmacêuticas se torna cada vez mais acirrada. Como forma de comparar a efetividade desses medicamentos, surgem diversas metodologias, com o objetivo de encontrar qual seria o melhor medicamento para uma dada situação. Uma das metodologias estudadas é a comparação mista de tratamentos, cujo objetivo é encontrar a efetividade de determinadas drogas em estudos e/ou ensaios clínicos que abordem, mesmo que de maneira indireta, os medicamentos estudados. A utilização dessa metodologia é demasiadamente complexa pois requer conhecimento de linguagens de programação em ambientes estatísticos além do domínio sobre as metodologias aplicadas a essa técnica. O objetivo principal desse estudo é a criação de uma interface gráfica que facilite a utilização do MTC para usuários que não possuam conhecimento em linguagens de programação, que seja de código aberto e multiplataforma. A expectativa é que, com essa interface, a utilização de técnicas mais abrangentes e avançadas seja facilitada, além disso, venha tornar o ensinamento sobre o tema mais facilitado para pessoas que ainda não conhecem o método.

Palavras Chaves: MTC. Estatística Bayesiana. Medicamentos. Interface gráfica.

ABSTRACT

CORREIA, Marcelo Goulart. **Graphical User Interface development for Bayesian Statistics applied to Mixed Treatment Comparison**. 2013. 66 f. Dissertação (Mestrado em Ciências Computacionais) – Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2013

Based on the progress made by the pharmaceutical industry, several medications have emerged to combat diseases. These drugs have similar topic effects but with subtle changes in their biochemical structure, thus competition between the pharmaceutical industry becomes increasingly fierce. In order to compare the effectiveness of these drugs, appear different methodologies with the objective of find what would be the best medicine for a given situation. One of the methods studied is the Mixed Treatment Comparison (MTC) whose objective is to find the effectiveness of certain drugs in studies and / or clinical trials that address, even if indirectly, the drugs studied. The use of this methodology is too complex because it requires knowledge of programming languages in statistical environments, beyond the mastery of the methodologies applied to this technique. The main objective of this study is to create a graphical user interface (GUI) that facilitates the use of MTC for users who have no knowledge in programming languages, which is open source and cross-platform. The expectation about this interface is that the use of more comprehensive and advanced techniques is facilitated, moreover, make the teaching about the topic easier for people who do not know the method.

Keywords: MTC. Bayesian Statistics. Drugs.GUI.

LISTA DE ILUSTRAÇÕES

Figura 1	–	Representação gráfica do modelo hierárquico – (a) condensado e (b) expandido com os relacionamentos condicionais independentes	16
Quadro 1	–	Processo descritivo do funcionamento da fórmula da Binomial	20
Quadro 2	–	Processo descritivo do funcionamento da fórmula do Logit.....	21
Figura 2	–	Esquema de funcionamento da ferramenta proposta	25
Figura 3	–	Interface gráfica construída	28
Figura 4	–	Arquivo Samples_d	30
Figura 5	–	Arquivo Gelman_T	31
Figura 6	–	Arquivo Interações	32
Quadro 3	–	Dados do estudo	33
Tabela 1	–	Variáveis do banco de dados (Arquivo CSV)	34
Figura 7	–	Preenchimento das informações no aplicativo	34
Quadro 4	–	Indicadores de qualidade para os tratamentos	35
Figura 8	–	Tabela de razão de chance com gráfico de floresta	35
Figura 9	–	Gráfico com as comparações existentes no trabalho	36
Quadro 5	–	Resultados acerca do logaritmo da razão de chance dos tratamentos (d)	37
Quadro 6	–	Eficácia absoluta do tratamento (T)	37
Figura 10	–	Gráficos de monitoramento da variável d	38
Figura 11	–	Gráficos de monitoramento da variável T	39
Figura 12	–	Gráfico de Gelman-Rubin para variável d	40
Figura 13	–	Gráfico de Gelman-Rubin para variável T	41

LISTA DE ABREVIATURAS E SIGLAS

BUGS	Bayesian inference Using Gibbs Sampling
CMT	Comparação mista de tratamentos
ECR	Ensaio clínico randomizado
GNU	GNU's Not Unix
IG	Interface gráfica
JAGS	Just Another Gibbs Sampler
MCMC	Monte Carlo via cadeias de Markov
REBRATS	Rede Brasileira de Avaliação de Tecnologias em Saúde
SST	Smith-Spiegelhalter-Thomas

SUMÁRIO

	INTRODUÇÃO	8
1	METODOLOGIA	11
1.1	Ensaio clínico randomizado	11
1.2	Metanálise	11
1.3	Comparação indireta	12
1.4	Inferência Bayesiana	13
1.5	Distribuição a priori não informativa	15
1.6	Modelo hierárquico Bayesiano	15
1.7	Monte Carlo via cadeias de Markov	17
1.8	Amostragem de Gibbs	18
1.9	Comparação Mista de Tratamentos	19
1.10	Análise de consistência	24
1.11	Estruturação do software	24
1.12	Resultados do aplicativo	29
2	APLICAÇÃO	33
	CONCLUSÃO	42
	REFERÊNCIAS	44
	ANEXO	47

INTRODUÇÃO

CONSIDERAÇÕES INICIAIS

O avanço das pesquisas na área farmacêutica e a quebra de patentes de alguns medicamentos influenciaram o mercado de drogas em escala mundial. Diversas indústrias investiram na criação de medicamentos com o objetivo de atender as demandas crescentes de diversos países, porém esse processo trouxe também como consequência o acirramento da concorrência entre as indústrias farmacêuticas. Esse acontecimento desencadeou a criação de diversos medicamentos que possuem efeito tópico similar porém com suaves alterações em suas estruturas bioquímicas. O efeito desse processo propiciou desenvolver diversas pesquisas na tentativa de provar a eficiência ou eficácia dessas drogas.

A maior parte dessas pesquisas busca comparar outros estudos desenvolvidos sobre drogas similares através do processo chamado metanálise, que compara drogas analisadas em relação às drogas de mesmo efeito tópico, ou seja, avalia se uma determinada droga tem desempenho pior, igual ou melhor que as drogas que estão no mercado.

Um dos maiores desafios nas análises das pesquisas desenvolvidas para comparação de drogas é encontrar estudos que compartilhem os mesmos objetivos, além de possuir parâmetros populacionais semelhantes. Essa dificuldade acaba limitando a execução das análises, bem como a comparação das drogas analisadas. Para contornar essa dificuldade, são utilizadas metodologias para comparação de tratamentos, mesmo que de maneira indireta, a partir de dados coletados nos estudos clínicos com populações que possuam características clínicas similares.

Revisão Bibliográfica

Já existem algoritmos desenvolvidos para a análise de tratamentos de saúde¹⁻³, porém a forma de utilização dos mesmos requer um conhecimento prévio em linguagens de programação em ambientes estatísticos. Isso acaba afastando possíveis usuários dessas metodologias, mesmo aqueles que possam ter conhecimento prévio dos métodos.

As técnicas computacionais para facilitar a execução de algoritmos tornam-se cada vez mais simplificadas através de interfaces gráficas e codificações estruturadas⁴. Nesse contexto, a criação de um ambiente que permita a utilização de técnicas multivariadas de comparação de estudos é de extrema relevância para um melhor julgamento sobre as decisões a serem tomadas para aquisição de novas tecnologias de saúde.

Algumas ferramentas propõem a facilitação no uso dessas metodologias⁵⁻⁶, porém limitam-se à criação do código de programação para a execução do modelo em ferramentas estatísticas. Além disso, algumas ferramentas possuem seu uso limitado a determinados sistemas operacionais.

Justificativa

Com o surgimento da Rede Brasileira de Avaliação de Tecnologias em Saúde (REBRATS)⁷, a demanda por pareceres sobre medicamentos e dispositivos de saúde em âmbito nacional passou a desempenhar um papel importante na tomada de decisão sobre a compra e a avaliação do desempenho dos mesmos.

A legislação brasileira para a elaboração de um parecer sobre avaliações de tecnologias em saúde propõe o seguinte cronograma: ⁷

- a) em 15 dias: envio de resposta sobre o tipo de estudo necessário, mencionando a possibilidade de realizá-lo;
- b) em 2 meses: elaboração de um parecer técnico-científico, contendo síntese da disponibilidade de evidências de qualidade;
- c) entre 6 e 12 meses: revisões sistemáticas ou estudos de avaliação econômica são elaborados por instituições de ensino e pesquisa.

Dentro dos prazos expostos são realizadas pesquisas bibliográficas sobre o tema abordado e a realização de possíveis estudos para a verificação da qualidade dos fármacos e dispositivos clínicos.

Nesse contexto, a elaboração de uma ferramenta que auxilie na condução das análises estatísticas com menor tempo para a execução pode contribuir para a elaboração de pareceres, tornando a tomada de decisão sobre a análise dos resultados um processo mais exequível.

Atualmente para o desenvolvimento de análises de comparação indireta nos institutos de ensino e pesquisa é necessário programar toda a rotina de análises num software de estatística Bayesiana, processo esse que consome maior tempo para a elaboração dos pareceres. Além disso, esse processo requer um profissional qualificado em linguagem de programação e estatística Bayesiana.

Objetivo

O principal objetivo do estudo é fornecer acesso às técnicas multivariadas de maneira simplificada, livre, gratuita e irrestrita através da construção de interfaces gráficas que permitam o uso desses algoritmos sem a necessidade de conhecimento em linguagens de programação.

Disseminar a modelagem abordada, fazendo com que pessoas que possuam pouca familiaridade com linguagens de programação, possam ter acesso às técnicas avançadas de comparação de tratamentos de saúde.

O uso de um banco de dados que agregue artigos selecionados para gerar testes de convergência, facilitando à obtenção de informações básicas para a escolha do modelo a ser escolhido em uma dada análise.

A abertura do código-fonte, bem como a distribuição da aplicação de forma gratuita, e a possibilidade de portabilidade da aplicação para diversos dispositivos é uma característica que contribui para a disseminação da metodologia proposta.

1 METODOLOGIA

1.1 Ensaio clínico randomizado

Durante a construção de um ensaio clínico, diversas premissas precisam ser respeitadas para que sua execução seja bem sucedida. Uma das etapas mais importantes é a randomização do estudo. Um ensaio clínico randomizado é um estudo em que os indivíduos são alocados ao acaso para receber uma determinada intervenção. Uma das principais razões para a randomização das intervenções para os pacientes é assegurar que a escolha do indivíduo não seja previsível⁸⁻⁹.

Em qualquer ensaio clínico que intervêm no curso natural dos acontecimentos, uma decisão precisa ser tomada sobre qual intervenção deve ser aplicada a qual indivíduo. Em geral, seja qual for o ensaio desenvolvido, deve-se planejar a estrutura de análise para responder às questões abordadas, fazendo com que a escolha da intervenção seja a mais aleatória possível⁸.

A validade dos testes de randomização é motivada pela intercambialidade de atribuições de tratamento em um ensaio clínico. Sob um modelo populacional, duas amostras da mesma população são intercambiáveis¹⁰.

1.2 Metanálise

Glass (1976)¹¹ desenvolveu a formalização do termo metanálise como sendo “a análise estatística de um grande conjunto dos resultados de estudos individuais com o objetivo de agrupá-los num único resultado”.

As metanálises são frequentemente baseadas em análises estatísticas obtidas em trabalhos publicados. Por esse motivo, existe um grande número de potenciais problemas potenciais que podem afetar a validade destas¹².

Uma limitação importante é que uma metanálise só pode incluir estudos com dados recuperáveis. Se apenas estudos publicados podem ser utilizados, então existe também a preocupação com o viés de publicação. Esse viés ocorre quando existe algum tipo de suspeita de ocultação ou seleção de resultados positivos que favoreçam o objeto estudado¹².

O objetivo de uma metanálise é estimar e realizar inferências sobre a diferença entre os efeitos estudados de dois tratamentos. Isso envolve a escolha de uma medida apropriada da diferença entre os tratamentos. Por exemplo, o logaritmo da razão de chance para desfecho dicotômico ou a diferença entre as médias para dados que seguem distribuição normal e assim calcular as estimativas de cada estudo analisado e a estimativa geral da diferença entre os tratamentos estudados¹².

Uma controvérsia que reside na metanálise é quanto à escolha entre o modelo de efeito fixo e o modelo de efeito aleatório para calcular a estimativa geral da diferença entre os tratamentos. No modelo de efeito fixo, a diferença entre os tratamentos é considerada a mesma para todos os ensaios. O erro padrão estimado de cada ensaio é baseado na variação amostral dentro do ensaio. No modelo de efeito aleatório, a diferença entre os tratamentos em cada ensaio é assumida como uma variável aleatória com distribuição normal. O erro padrão de cada ensaio é aumentado devido à introdução dessa variação entre os ensaios¹².

1.3 Comparação indireta

O uso das técnicas de comparação indireta de um cenário que possua diversos tratamentos surge com o objetivo de auxiliar na tomada de decisão sobre a incorporação de um tratamento para um objetivo determinado.

Grande parte dos ensaios clínicos randomizados não inclui todas as técnicas possíveis no cenário de comparação, fazendo com que um tratamento seja comparado contra o placebo ou contra uma terapia padrão.

Assim, para certas indicações, mesmo se houvesse um incentivo para incorporar intervenções competindo em um ensaio clínico randomizado, as intervenções de interesse poderiam variar por país ou mudar ao longo do tempo devido a novas evidências e novas abordagens de tratamento¹³.

Na ausência de ensaios que justifiquem uma comparação direta dos tratamentos em intervenções concorrentes, uma comparação indireta pode fornecer um indicador útil da diferença entre os efeitos e ajudar na seleção de uma melhor escolha de tratamento¹³.

1.4 Inferência Bayesiana

Nos últimos anos, o uso de métodos Bayesianos em estatística aplicada tem elevado crescimento¹⁴. A estatística Bayesiana tem a vantagem, em comparação à estatística tradicional, de ser facilmente estabelecida e trabalhada. Os problemas relacionados aos testes de hipótese ou estimativas de intervalos de confiança para parâmetros desconhecidos podem ser facilmente superados com estatística Bayesiana¹⁵. A abordagem Bayesiana considera as incertezas associadas com todas as variáveis desconhecidas mesmo que elas tenham sido observadas ou não¹⁶.

A estatística Bayesiana possui alguns conceitos que necessitam ser melhor explicitados para permitir um correto entendimento das análises realizadas.

A inferência é desenhada através da construção de distribuições de probabilidade conjunta de todas as variáveis não observadas, baseado em todo o conhecimento sobre elas. Esse conhecimento incorpora informações prévias sobre o fenômeno estudado quando disponível¹⁶.

A probabilidade é uma medida de crença em uma proposição ou evento, isto significa que pessoas diferentes podem ter valores diferentes para probabilidade de uma mesma proposição ou evento.

Da mesma forma, todos os valores desconhecidos (por exemplo, prevalência, custo, efetividade, etc), mesmo que sejam observados posteriormente, são considerados como incertos e devem expressos na forma de uma distribuição de probabilidade.

O objetivo geral da inferência Bayesiana é o conhecimento da posteriori. A ideia fundamental por trás de quase todos os métodos estatísticos é que, conforme o tamanho da amostra aumenta, a distribuição de uma amostra populacional se aproxima da distribuição populacional¹⁴. É chamada posteriori uma distribuição de probabilidade relacionada a uma quantidade que foi modificada a partir de uma nova informação.

A inferência Bayesiana possui dois ingredientes: a distribuição observacional (ou amostragem) $f(y|\theta)$ e a distribuição $p(\theta)$, denominada distribuição a priori. Essa última distribuição pode também ser especificada com a ajuda de constantes assim como na distribuição de y . Essas constantes são chamadas de hiperparâmetros, visto que elas são os parâmetros da distribuição de parâmetros¹⁶.

A distribuição a priori sintetiza as crenças sobre um determinado parâmetro

antes da observação dos dados. A densidade a priori fornece os pesos relativos da crença para todos os valores possíveis do parâmetro θ antes de observar os dados²⁷. A priori é resultado das informações que cada pessoa tem sobre o assunto.

Toda a informação sobre o parâmetro θ que está presente no vetor x de dados coletados está contida na função de verossimilhança $f(y|\theta)$ ¹⁴.

É chamada verossimilhança, a distribuição de probabilidade que exprime a credibilidade de um valor a partir de uma informação, ou seja, não está sendo calculada a possibilidade da medida de interesse ter um determinado valor baseado num estudo e sim quanto se acredita que a medida de interesse tenha um determinado valor baseado num estudo.

Uma vez o problema evocado dessa forma, é natural que a inferência deva ser baseada na distribuição de probabilidade de θ depois de observar os valores de y . Essa distribuição é chamada de distribuição a posteriori, podendo ser obtida através do teorema de Bayes¹²:

$$p(\theta|y) = f(y|\theta) * p(\theta) / f(y) \quad (1)$$

Com

$$f(y) = \int f(y|\theta) p(\theta) d\theta \quad (2)$$

Os conceitos de distribuição a priori e a posteriori são sempre relativos à observação considerada no dado momento. É possível que, após observar y e obter a distribuição a posteriori, uma nova observação x também relacionada à θ através de uma eventual função de verossimilhança torne-se disponível¹⁶.

Como forma de ilustrar o uso da técnica, suponha uma moeda honesta onde se deseja verificar qual a probabilidade de num lançamento o resultado ser cara. Depois de alguns lançamentos, é obtida uma probabilidade a priori de 0,60, ou seja, $P(K) = 0,60$ e $P(C) = 0,40$.

Suponha uma amostra com $Y = (Y_1, Y_2, \dots, Y_n)$ de tamanho n e k sendo os desfechos Y_1, Y_2, \dots, Y_n que resultam em cara.

Para determinar a probabilidade posteriori $P(K|Y)$ e $P(C|Y)$ é utilizado o teorema de Bayes. O cálculo da probabilidade reversa é dado por:

$$\begin{aligned} P(Y|K) &= (0,5)^k * (0,5)^{n-k} \\ P(Y|C) &= (0,5)^{k-n} * (0,5)^k \end{aligned} \quad (3)$$

Logo, para o cálculo da probabilidade posteriori é feito:

$$\begin{aligned} P(K|Y) &= \frac{P(Y|K) * P(A)}{[P(Y|K) * P(A) + P(Y|C) * P(C)]} \\ P(K|Y) &= \frac{(0,5)^k * (0,5)^{n-k} * 0.60}{[(0,5)^k * (0,5)^{n-k} * 0.60 + (0,5)^{k-n} * (0,5)^k * 0.40]} \\ P(C|Y) &= 1 - P(K|Y) \end{aligned} \quad (4)$$

Ou seja, se for realizado nesse cenário uma simulação com 20 lançamentos que resultam em 15 caras, a probabilidade posteriori será:

$$\begin{aligned} P(K|Y) &= \frac{(0,5)^{15} * (0,5)^5 * 0.60}{[(0,5)^{15} * (0,5)^5 * 0.60 + (0,5)^5 * (0,5)^{15} * 0.40]} \\ P(K|Y) &= 0,5997 \\ P(C|Y) &= 1 - 0,5997 = 0,4003 \end{aligned} \quad (5)$$

1.5 Distribuição a priori não informativa

A ideia por trás dessa distribuição a priori vem do desejo de realizar inferência estatística baseada no mínimo de informações a priori subjetivas¹⁷.

Considere uma observação y com função de probabilidade $p(y|\theta)$. A função de probabilidade (densidade) da distribuição a priori não informativa de Jeffreys (1961)¹⁸ tem a densidade dada por:

$$p(\theta) \propto \sqrt{I(\theta)}. \theta \in \Theta \quad (6)$$

Sendo $I(\theta)$ a matriz de informação de Fisher esperada sobre a variável θ , que é escrita da seguinte forma:

$$I(\theta) = E \left[\frac{-\partial^2 \log f(x|\theta)}{\partial \theta \partial \theta'} \middle| \theta \right]$$

A distribuição a priori não informativa é frequentemente obtida através da distribuição a priori conjugada, deixando os parâmetros de escala reduzidos à zero e mantendo os outros parâmetros constantes¹⁷.

1.6 Modelo hierárquico Bayesiano

Modelagem hierárquica é uma abordagem amplamente utilizada para a construção de modelos complexos, através da especificação de uma série de distribuições condicionais simples.

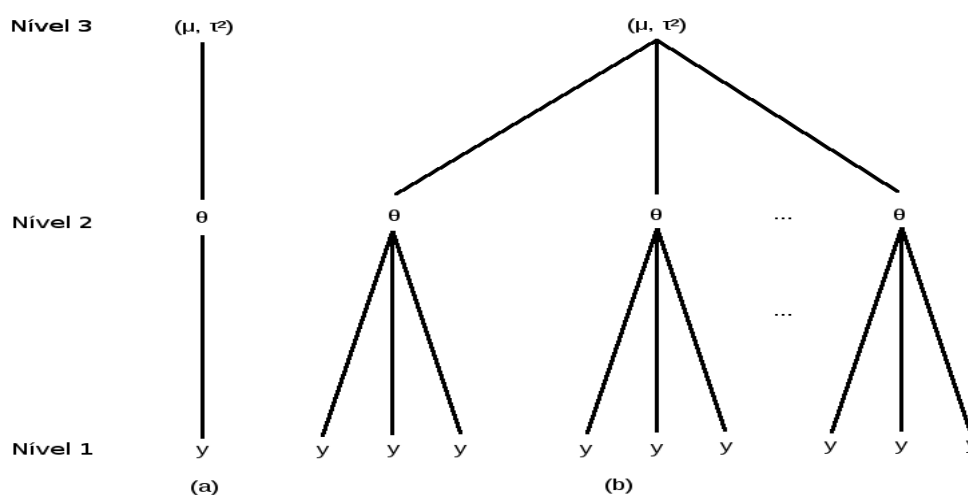
Uma estratégia para especificação da distribuição a priori ou para melhor descrição de uma situação experimental é dividir a distribuição em estágios ou em hierarquias. Dessa forma, a especificação da distribuição a priori pode ser feita em duas fases¹⁷: estrutural, para a divisão em estágios e subjetiva, para a especificação quantitativa em cada estágio.

Embora seja difícil formular uma definição que inclui todas as várias aplicações do conceito de modelagem hierárquica, as ideias chave são capturadas nas seguintes características comuns¹⁹ parametrização multinível; dados hierarquicamente estruturados; correspondência de parâmetros para estruturas populacionais; independência condicional.

Quando os parâmetros e observações estão relacionados em um modelo hierárquico, as relações de dependência entre todos os parâmetros e as observações seguem um fluxo de única direção¹⁴. Para estes modelos hierárquicos, pode-se facilmente encontrar as distribuições condicionais para cada parâmetro. A representação gráfica do modelo hierárquico sob a ótica Bayesiana pode ser visualizada na Figura 1.

Naturalmente o método é uma modelagem que se presta a inferência Bayesiana, utilizando especialmente ferramentas modernas para a computação Bayesiana¹⁹.

Figura 1 – Representação gráfica do modelo hierárquico – (a) condensado e (b) expandido com os relacionamentos condicionais independentes.



Em geral, quanto mais elevado for um estado, mas difícil é a especificação de sua distribuição. Raramente, os modelos possuem mais que três estados e é muito comum que distribuição a priori no estado maior seja não informativa¹⁶.

A estrutura linear do modelo combina com a modelagem hierárquica no contexto de modelos lineares normais e essa estrutura pode ser explorada em mais detalhes. Especificamente, a estrutura hierárquica é utilizada observando à suposição de linearidade e normalidade. Essa configuração preserva a linearidade do modelo como um todo e a conjugação em qualquer estágio do modelo¹⁷.

De modo a especificar o modelo e suas distribuições a priori, pode-se reescrevê-lo como:

$$\begin{aligned} Y|\beta_1, \varphi &\sim N(X_1\beta_1, \varphi^{-1}I_n), \\ \beta_1|\beta_2 &\sim N(X_2\beta_2, C), \\ \beta_2 &\sim N(b, B), \\ \varphi &\sim G(n_0/2, n_0\sigma_0^2/2) \end{aligned} \tag{7}$$

onde a matriz com variáveis explicativas é renomeada para X_1 devido à presença de outra matriz incluindo o segundo estágio de variáveis explicativas, X_2 . Essa matriz inclui os valores que explicam as variações em β_1 , e o coeficiente dessa variáveis é dada por β_2 .

1.7 Monte Carlo via cadeias de Markov

Uma cadeia de Markov é um tipo especial de processo estocástico, que lida com a caracterização de sequências de variáveis aleatórias, em que, os dados estados presente, passados e futuro são independentes. Um processo estocástico pode ser definido como uma coleção de variáveis aleatórias $\{\theta^{(t)} : t \in T\}$ para alguns conjuntos T . O conjunto $\{\theta^{(t)} : t \in T\}$ é dito ser um processo estocástico com espaço de estado S e conjunto índice (ou parâmetro) T ¹⁶.

Essa propriedade pode ser formalmente estabelecida através de:

$$\begin{aligned} P_r(\theta^{(n+1)} \in A | \theta^{(n)} = y, \theta^{(n-1)} \in A_{n-1}, \dots, \theta^{(0)} \in A_0) \\ P_r(\theta^{(n+1)} \in A | \theta^{(n)} = y) \end{aligned} \tag{8}$$

para todos os conjuntos $A_0, \dots, A_{n-1}, A \subset S$ e $y \in S$.

Métodos a posteriori $p(\theta|y)$ são dados através dos métodos de Monte Carlo

via cadeias de Markov¹⁷ gerais para geração de variáveis aleatórias para a função de probabilidade (ou densidade).

Para o desenvolvimento do trabalho, o método de Monte Carlo via cadeias de Markov utilizado foi a técnica de amostradores de Gibbs.

1.8 Amostragem de Gibbs

O amostrador de Gibbs foi desenvolvido como um método Bayesiano para restauração de imagens e foi frequentemente utilizado para diversas tarefas de análise de imagens digitais¹⁵.

A técnica tem como objetivo decompor parâmetros espaciais através de amostragem de distribuições condicionais da distribuição de probabilidade a posteriori para cada parâmetro x_k do vetor $\mathbf{x} = (x_k)$ de parâmetros desconhecidos, assim diminuindo o problema da alta dimensionalidade¹⁵.

No contexto dos métodos de Monte Carlo via cadeia de Markov, a amostragem de Gibbs é um esquema onde o núcleo de transformação é formado por distribuições condicionais completas. Assume-se que a distribuição de interesse é $\pi(\theta)$ onde $\theta = (\theta_1, \dots, \theta_d)'$. Cada um dos componentes θ_i pode ser escalar, um vetor ou uma matriz. Considera-se também que as distribuições condicionais completas $\pi_i(\theta_i) = \pi_i(\theta_i|\theta_{-i})$, com $i = 1, \dots, d$ estão disponíveis¹⁶.

O problema a ser resolvido é extrair π por conta dos esquemas de geração direta serem caros, complicados ou indisponíveis, porém a geração a partir de π_i torna-se viável¹⁶. Isso ocorre devido a complexidade da função $\pi(\theta)$ ou pela alta dimensionalidade de θ ²¹.

A técnica fornece um esquema de criação baseado em sucessivas gerações a partir das distribuições condicionais completas²⁰. Esse processo é desenvolvido da seguinte forma: Inicia-se o contador da iteração da cadeia com $j = 1$ e ajustam-se os valores iniciais $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})'$; obtém-se um novo valor $\theta^{(j)} = (\theta_1^{(j)}, \dots, \theta_d^{(j)})'$ a partir da iteração anterior $\theta^{(j-1)}$ através de gerações sucessivas de valores:

$$\begin{aligned}
\theta_1^{(j)} &\sim \pi(\theta_1 | \theta_2^{(j-1)}, \dots, \theta_d^{(j-1)}), \\
\theta_2^{(j)} &\sim \pi(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_d^{(j-1)}), \\
&\vdots \\
&\vdots \\
&\vdots \\
\theta_d^{(j)} &\sim \pi(\theta_d | \theta_1^{(j)}, \dots, \theta_{d-1}^{(j)});
\end{aligned} \tag{9}$$

Muda-se o contador j para $j+1$ e retorna-se ao passo 2 até que a convergência seja atingida.

Quando a convergência é alcançada, o valor resultante $\theta^{(j)}$ é uma extração de π . Conforme o número de iterações aumenta, a cadeia aproxima-se de sua condição de equilíbrio. Assim a convergência é então assumida por aproximação.

1.9 Comparação Mista de Tratamentos

Sutton et al.²² indica que a Comparação Mista de Tratamento (CMT) possui um grande potencial para estimativa de efetividade comparativa utilizando uma base de evidências de ensaios. Entende-se como evidências de ensaios, a incidência de casos num determinado cenário analisado, que individualmente não compara todas as opções de tratamento. Lu e Ades³ dizem que a metodologia sintetiza evidências comparativas sobre múltiplos tratamentos de Ensaios Clínicos Randomizados (ECR) disponíveis em uma determinada área terapêutica e tem sido cada vez mais utilizado na tomada de decisão médica e na avaliação das tecnologias da saúde.

A construção metodológica da CMT é desenvolvida através das técnicas de estatística Bayesiana, com base principalmente nos estudos desenvolvidos por Bayes²³. O método mais utilizado para a construção de um modelo para avaliação de tecnologias em saúde é o hierárquico Bayesiano³. Existem dois modelos para o desenvolvimento de análises. Um dos modelos foi proposto por Smith et al.²⁴ para metanálises comparativas de dois tratamentos e uma extensão, proposta por Lu e Ades³, para K tratamentos.

Para a utilização do modelo Smith-Spiegelhalter-Thomas (SST)²⁴ para comparação de dois tratamentos, assume-se $(r_i^T, n_i^T, r_i^C, n_i^C)$, onde $i = 1, \dots, n$, sejam dados obtidos a partir de n ensaios clínicos randomizados, onde r_i^T e r_i^C são o número de eventos no grupo tratamento e no grupo controle respectivamente, em i ensaios com distribuição de probabilidade em:

- $r_i^T \sim \text{Binomial}(p_i^T, n_i^T)$
- $r_i^C \sim \text{Binomial}(p_i^C, n_i^C)$
- n_i^T e n_i^C o número de casos nos respectivos grupos.

Por se tratar de uma distribuição de probabilidade binomial, as variáveis acima descritas pelos autores²⁴ podem ser apresentadas conforme as fórmulas abaixo:

Quadro 1 - Processo descritivo do funcionamento da fórmula da Binominal

i	p_i^T	n_i^T	r_i^T	p_i^C	n_i^C	r_i^C
1	p_1^T	n_1^T	$C_{n_1^T}^x * (p_1^T)^x * \left((1 - p_1^T) \right)^{(n_1^T - x)}$	p_1^C	n_1^C	$C_{n_1^C}^x * (p_1^C)^x * \left((1 - p_1^C) \right)^{(n_1^C - x)}$
2	p_2^T	n_2^T	$C_{n_2^T}^x * (p_2^T)^x * \left((1 - p_2^T) \right)^{(n_2^T - x)}$	p_2^C	n_2^C	$C_{n_2^C}^x * (p_2^C)^x * \left((1 - p_2^C) \right)^{(n_2^C - x)}$
3	p_3^T	n_3^T	$C_{n_3^T}^x * (p_3^T)^x * \left((1 - p_3^T) \right)^{(n_3^T - x)}$	p_3^C	n_3^C	$C_{n_3^C}^x * (p_3^C)^x * \left((1 - p_3^C) \right)^{(n_3^C - x)}$
...
n	p_n^T	n_n^T	$C_{n_n^T}^x * (p_n^T)^x * \left((1 - p_n^T) \right)^{(n_n^T - x)}$	p_n^C	n_n^C	$C_{n_n^C}^x * (p_n^C)^x * \left((1 - p_n^C) \right)^{(n_n^C - x)}$

Fonte: O AUTOR, 2013.

O modelo sob a ótica da função logit faz uso da logaritmização da razão de chance. Ou seja, $\text{logit}(p) = \log [p/1-p]$. Os efeitos do tratamento no nível do estudo em relação ao grupo controle (δ_i) correspondem a:

$$\begin{aligned} \delta_i &= \text{logit}(p_i^T) - \text{logit}(p_i^C) \\ \delta_i &= \log\left(\frac{p_i^T}{1-p_i^T}\right) - \log\left(\frac{p_i^C}{1-p_i^C}\right) \\ \delta_i &= \log\left[\left(\frac{p_i^T}{1-p_i^T}\right) / \left(\frac{p_i^C}{1-p_i^C}\right)\right] \\ \delta_i &= \log\left[\left(\frac{p_i^T}{1-p_i^T}\right) * \left(\frac{1-p_i^C}{p_i^C}\right)\right] \\ \delta_i &= \log\left[\left(\frac{p_i^T}{p_i^C}\right) * \left(\frac{1-p_i^C}{1-p_i^T}\right)\right] \end{aligned}$$

(10)

com $i = 1, 2, \dots, n$.

Assume-se que as variáveis δ_i seguem a distribuição de probabilidade Normal (d, s^2), sendo que d é o efeito do tratamento da população.

A taxa média (μ_i) do i -ésimo ensaio, em escala logit, tem a expressão:

$$\mu_i = \frac{\text{logit}(p_i^t) + \text{logit}(p_i^c)}{2} \quad (11)$$

Logo,

$$2\mu_i = \text{logit}(p_i^t) + \text{logit}(p_i^c) \quad (12)$$

Substituindo $\text{logit}(p_i^t)$ por $\delta_i + \text{logit}(p_i^c)$, a equação toma a expressão:

$$\begin{aligned} 2\mu_i &= \delta_i + \text{logit}(p_i^c) + \text{logit}(p_i^c) \\ 2\mu_i &= \delta_i + 2\text{logit}(p_i^c) \\ \mu_i &= \delta_i/2 + \text{logit}(p_i^c) \\ \text{logit}(p_i^c) &= \mu_i - \delta_i/2 \end{aligned} \quad (13)$$

Analogamente,

$$\text{logit}(p_i^t) = \mu_i + \delta_i/2 \quad (14)$$

com $i = 1, 2, \dots, n$

Quadro 2 – Processo descritivo do funcionamento da fórmula do Logit

i	$\text{logit}(p_i^T)$	$\text{logit}(p_i^C)$
1	$\mu_1 + \delta_1/2$	$\mu_1 - \delta_1/2$
2	$\mu_2 + \delta_2/2$	$\mu_2 - \delta_2/2$
3	$\mu_3 + \delta_3/2$	$\mu_3 - \delta_3/2$
...
n	$\mu_n + \delta_n/2$	$\mu_n - \delta_n/2$

Fonte: O AUTOR, 2013.

Os parâmetros μ_i , d e σ^2 seguem distribuição a priori. Nesse modelo proposto,

δ_i e d são os parâmetros de interesse primário e a variável μ_i é tratada como um parâmetro perturbador.

O modelo desenvolvido pelos autores foi estruturado para a comparação entre dois tratamentos, porém em algumas terapias dispõe-se de mais de duas alternativas para tratamento clínico. Visando ampliação das comparações, Lu e Ades² desenvolveram a extensão do modelo SST²⁴. É necessário escolher um tratamento (k) como referência para a análise dos demais.

Seja r_{ik} o número de ventos estudados do tratamento de referência k para o ensaio i , $r_{ik} \sim \text{Binomial}(p_{ik}, n_{ik})$. Então,

$$\begin{aligned} \text{logit}(p_{i1}) &= \mu_i - \delta_{i2}/K - \delta_{i3}/K - \dots - \delta_{iK}/K \\ \text{logit}(p_{i2}) &= \mu_i + (K-1)\delta_{i2}/K - \delta_{i3}/K - \dots - \delta_{iK}/K \\ &\vdots \\ \text{logit}(p_{iK}) &= \mu_i - \delta_{i2}/K - \delta_{i3}/K - \dots + (K-1)\delta_{iK}/K \end{aligned} \quad (15)$$

Com $(\delta_{i2}, \dots, \delta_{iK})^T \sim \text{Normal}(\delta, \Sigma)$. Onde p_{ik} é a probabilidade de um determinado evento no ensaio i sob o tratamento k e n_{ik} é o número de elementos no ensaio i sob tratamento k .

A variável d é o hiperparâmetro de dimensão $(K-1)$ da média dos efeitos do tratamento no nível do estudo em relação ao tratamento referência e Σ é a matriz de covariância com $(K-1)$ linhas e $(K-1)$ colunas. Explicitando o sistema de equações para $\text{logit}(p_{iK})$ sob a forma matricial:

$$\begin{bmatrix} \text{logit}(P_{i1}) \\ \text{logit}(P_{i2}) \\ \text{logit}(P_{i3}) \\ \dots \\ \text{logit}(P_{iK}) \end{bmatrix} = \begin{bmatrix} \mu_i \\ \mu_i \\ \mu_i \\ \dots \\ \mu_i \end{bmatrix} - \begin{bmatrix} 1 & 1 & \dots & 1 \\ -(K-1) & 1 & \dots & 1 \\ 1 & -(K-1) & \dots & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & -(K-1) \end{bmatrix} * \begin{bmatrix} \delta_{i2}/K \\ \delta_{i3}/K \\ \delta_{i4}/K \\ \dots \\ \delta_{iK}/K \end{bmatrix} \quad (16)$$

Ou, $\mathbf{Y}_i = \mathbf{M} - \mathbf{X} * \boldsymbol{\theta}_i$.

$$\text{Onde } \mathbf{Y}_i = \begin{bmatrix} \text{logit}(P_{i1}) \\ \text{logit}(P_{i2}) \\ \text{logit}(P_{i3}) \\ \dots \\ \text{logit}(P_{iK}) \end{bmatrix}, \mathbf{M} = \begin{bmatrix} \mu_i \\ \mu_i \\ \mu_i \\ \dots \\ \mu_i \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ -(K-1) & 1 & \dots & 1 \\ 1 & -(K-1) & \dots & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & -(K-1) \end{bmatrix}, \boldsymbol{\theta}_i = \begin{bmatrix} \delta_{i2}/K \\ \delta_{i3}/K \\ \delta_{i4}/K \\ \dots \\ \delta_{iK}/K \end{bmatrix} \quad (17)$$

A estimativa da taxa média do k-ésimo evento no i-ésimo ensaio é dada por:

$$\mu_i = \sum_k \text{logit}(p_{ik})/K$$

Os efeitos de cada tratamento em relação ao tratamento referência (tratamento k), em escala logit, são obtidos através de:

$$\delta_{ik} = \text{logit}(p_{ik}) - \text{logit}(p_{il}), k=2, \dots, K$$

$$\delta_{ik} = \log\left[\left(\frac{p_{ik}}{1-p_{ik}}\right) / \left(\frac{p_{il}}{1-p_{il}}\right)\right]$$

$$\delta_{ik} = \log\left[\left(\frac{p_{ik}}{1-p_{ik}}\right) * \left(\frac{1-p_{il}}{p_{il}}\right)\right]$$

$$\delta_{ik} = \log\left[\left(\frac{p_{ik}}{p_{il}}\right) * \left(\frac{1-p_{il}}{1-p_{ik}}\right)\right]$$

(18)

Para o desenvolvimento da interface gráfica (IG) foi utilizada uma versão modificada do modelo apresentado acima com base no trabalho desenvolvido por Higgins e Whitehead²⁵. O modelo desenvolvido pelos autores é:

$$\text{logit}(p_{jk}) = \begin{cases} \mu_{jb} & \text{se } k=b; \\ \mu_{jb} + \delta_{jkb} & \text{se } k \neq b; \end{cases} \quad (19)$$

onde p_{jk} é a probabilidade de um determinado evento no ensaio j sob o tratamento k, μ_{jb} é o logaritmo da chance de um determinado evento no tratamento referência b no ensaio j e δ_{jkb} é o logaritmo da razão de chance num ensaio específico do tratamento k em relação ao tratamento referência³.

Os efeitos do estudo são tratados como sendo parâmetros de perturbação com distribuição a priori, $\mu_{jb} \sim N(0, 10000)$. Assumindo-se o tratamento 1 como referência, os efeitos dos outros tratamentos em relação ao tratamento referência como sendo o parâmetro base com distribuição a priori vaga³:

$$d_{1j}, d_{1(j+1)}, \dots, d_{1(j+k)} \sim \text{Normal}(0, 10000)$$

As comparações restantes podem ser expressas conforme abaixo:

$$d_i = d_{1j} - d_{1i} \text{ com } j = i+1, i+2, \dots, i+k$$

Os logaritmos da razão de chance num ensaio específico do tratamento k em relação ao tratamento referência são desenvolvidos a partir de uma das distribuições de efeito aleatório³.

$$\delta_{jXY} \sim N(d_{XY}, \sigma_{XY}^2)$$

Se $\sigma^2 = 0$, então é obtido o modelo de efeito fixo. Para modelos de efeito aleatório, é assumida a hipótese de variância homogênea, $\sigma_{XY}^2 = \sigma^2$. Uma

distribuição a priori vaga é fornecida para a variância, sendo $\sigma \sim \text{Uniforme}(0,2)$. Para ensaios com mais de duas comparações a variância homogênea, $\sigma^2/2$.

1.10 Análise de consistência

Uma das formas de se diagnosticar a consistência do modelo é através da análise proposta por Gelman-Rubin²⁶ que propõem dois passos para a construção do modelo de diagnóstico de convergência. Em primeiro lugar, uma estimativa da distribuição referência é criada, centrada em torno de sua moda e as distribuições são sobrepostas de forma a verificar a disposição delas em relação à distribuição referência. A distribuição aproximada é utilizada para iniciar as várias sequências independentes da simulação iterativa. O segundo passo é a análise das múltiplas sequências de modo a formar uma distribuição estimada da variável aleatória referência. Utilizando os resultados das simulações realizadas, a estimativa seguindo a distribuição de t de Student para cada parâmetro e que vai sendo ajustada ao longo das simulações, obtendo assim, a base para uma estimativa de quanto aproximada está de sua convergência, ou seja, se mais simulações forem realizadas melhor a estimativa da distribuição ajustada.

Os autores concluem que quando a redução da medida de Gelman-Rubin é próxima de 1, então cada um dos m conjuntos com n valores simulados estão próximos da distribuição referência. Logo, o modelo apresenta convergência.

1.11 Estruturação do software

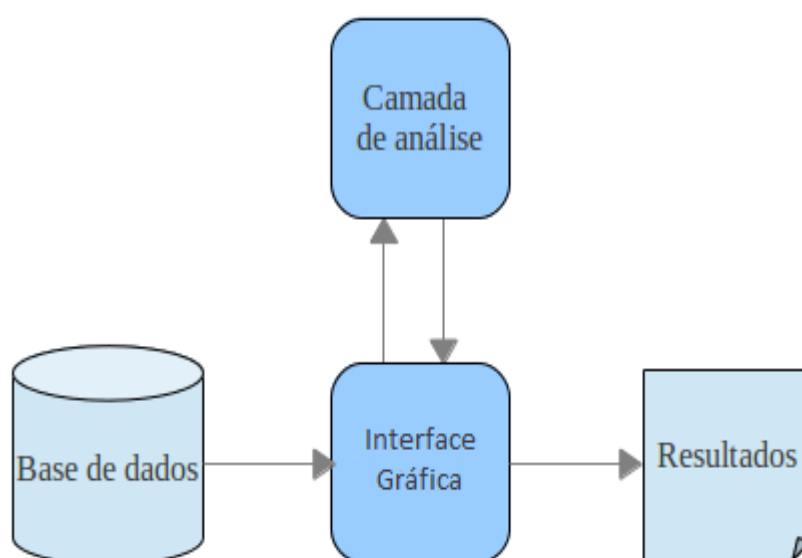
Dada a complexidade da técnica para comparação de K tratamentos, aplicou-se diversos processos de tratamentos no ambiente computacional, permitindo que os usuários tenham um acesso simplificado às técnicas. A efetivação do trabalho seguiu as seguintes etapas:

- a) planejamento do aplicativo (software), para delimitar o escopo, definindo metas e procedimentos. Foi definido no escopo que o software deveria apresentar os resultados obtidos na simulação, bem como gráficos de performance das simulações e as iterações entre os estudos analisados;

- b) algoritmização do modelo proposto. Este é um passo importante para estudar todo o processo teórico que sirva como base para as técnicas de comparação de tratamentos, trabalho crucial para a codificação dos modelos para uma estrutura de algoritmos que permitam o funcionamento em um ambiente computacional;
- c) construção da interface gráfica, que deve ser livre, aberta e intuitiva para uso em diversas plataformas, desde aos menos experientes em programação até os mais familiarizados, mas que conheçam as técnicas estatísticas abordadas;
- d) validação do software, como forma de corrigir erros e eventuais problemas que a ferramenta possa apresentar durante a construção e testes. Essa etapa foi realizada através da comparação dos resultados obtidos no modelo original com o modelo que foi estruturado nesse trabalho.

A interface gráfica funcionará como elo de comunicação dos processos de análises e procedimentos inerentes à execução, ou seja, esta interface receberá o banco de dados, enviará o mesmo para uma camada de análises estatísticas que devolverá o resultado construindo relatórios que facilitam a interpretação para o usuário final. Esse processo está esquematizado na figura abaixo.

Figura 2 – Esquema de funcionamento da ferramenta proposta



A interface gráfica desempenha um papel centralizador no processo de geração das análises estatísticas. Sendo assim, a ferramenta busca as bases de dados, distribui a informações para o motor analítico e as apresenta.

O desenvolvimento do trabalho exigiu dois aplicativos (R⁴ e JAGS²⁷) e de seis extensões do aplicativo R (*fgui*²⁸, *lattice*²⁹, *coda*³⁰, *igraph*³¹, *rmeta*¹⁴, *rjags*²⁷ e *gplots*³², que são de código aberto e sob licença GNU's Not Unix (GNU).

O software R fornece uma variada modelagem da Estatística e de interface gráfica, opções de representação gráfica e é altamente abrangente. R pode ser considerado como uma implementação diferenciada do S, porém muitos códigos escrito em S são executáveis sob R. A linguagem S é muitas vezes o veículo de escolha para pesquisas em metodologia estatística. R é um projeto GNU, similar à linguagem S, que foi desenvolvida nos Laboratórios Bell (ex-AT&T, agora Lucent Technologies), fornecendo uma rota de código aberto para a prática de modelagem.

Just Another Gibbs Sampler (JAGS) é um programa para análise de modelos hierárquicos Bayesianos^{19,14} utilizando o modelo de simulação de Monte Carlo via cadeias de Markov. Foi escrito objetivando ter um motor multiplataforma para a linguagem Bayesian inference Using Gibbs Sampling (BUGS), ser extensível, para permitir aos usuários escreverem suas próprias funções, distribuições e amostragens e ser uma plataforma para experimentação de ideias em modelagem Bayesiana.

O pacote de R *fgui* é projetado para desenvolvedores de pacotes R, com o objetivo de auxiliar, e por vezes, de forma totalmente automatizada, na criação de interfaces gráficas para um comando de programação em R. A interface é construída sobre a interface gráfica Tcl/Tk³³ incluída no R. O pacote facilita ainda mais o desenvolvedor pois carrega arquivos de ajuda das funções de linha de comando facilitando a tarefa do usuário. Ao passar uma função como argumento para as rotinas do pacote *fgui*, cria-se uma interface gráfica para esta função e outras opções estão disponíveis para ajustar esta interface com maior flexibilidades.

CODA é um conjunto de funções do S que fornecem uma infraestrutura baseada em objeto para analisar a saída da simulação do modelo de Monte Carlo via cadeias de Markov (MCMC).

O pacote de extensão *lattice* é uma implementação de gráficos de treliça para R. Nesse contexto, o pacote foi utilizado para a geração dos gráficos de monitoramento dos dados gerados pelo modelo. É um sistema poderoso, com alto

nível de visualização de dados com ênfase em modelos multivariados, sendo projetado para atender às necessidades gráficas mais peculiares de ajuste mínimo, mas também pode ser facilmente estendido quando ocorrem requisitos mais complexos.

O *igraph* permite criar e manipular grafos direcionados. O pacote inclui implementações gráficas para fluxos de redes e árvores geradoras mínimas. Também implementa algoritmos para alguns métodos recentes de análise de rede, como a busca estruturada da comunidade de objetos sob a ótica da Informática.

O pacote *rmeta* trabalha com funções para metanálise simples de efeitos fixos e aleatórios na comparação de duas amostragens e metanálises cumulativas. O pacote desenha gráficos padronizados de síntese, seguindo regras pré-definidas, lotes de funil e testes de associação e heterogeneidade.

O *gplots* é um pacote que disponibiliza diversas ferramentas para formatação e criação de relatórios a partir dos resultados obtidos no R.

Nesse trabalho, as versões utilizadas foram: *R* 2.15.1, *JAGS* 3.1.0, *rjags* 3-5, *coda* 0.15-2, *lattice* 0.20-10, *fgui* 1.0-2, *igraph* 0.6-3, *rmeta* 2.16 e *gplots* 2.11.0.

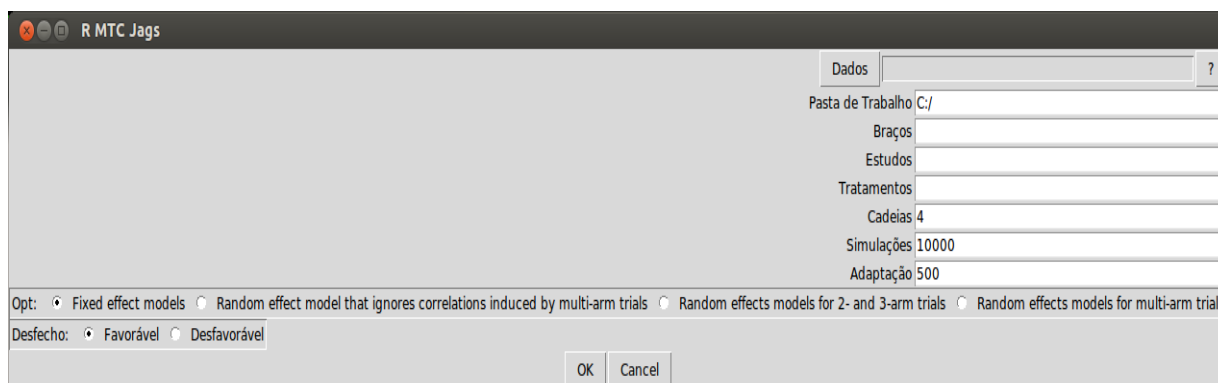
Os modelos para análise foram obtidos no sítio da Escola de Medicina Social e Comunitária da Universidade de Bristol³⁴, atribuídos a Higgins e Whitehead^{25,12}, os modelos foram construídos em *WinBUGS*, logo foi necessário convertê-los para *JAGS*.

A conversão foi facilitada pois muitos comandos existiam nas duas ferramentas e possuíam a mesma operabilidade e os que não a possuíam foram portados para o r. na criação da interface gráfica foi utilizado o comando *gui* do pacote *fgui* e a programação que foi desenvolvida (anexos) no r e no *jags* executa: criação da base de dados; seleção do modelo; cálculo das estatísticas descritivas e das probabilidades de melhor tratamento; geração de gráficos.

A execução das análises a partir da interface gráfica (Figura 3) torna o trabalho simplificado pois, a partir das informações básicas, toda a programação é executada de maneira que seja ajustada na própria interface. É necessário incluir as informações sobre o local do arquivo do banco de dados, o número total de braços, ou seja, número total de comparações em todos os estudos analisados, o número total de estudos, quantos tratamentos estão sendo analisados, quantas cadeias serão construídas na simulação, quantas simulações serão feitas, quantas simulações serão realizadas para o período de aquecimento, qual modelo será

aplicado e qual o desfecho clínico observado.

Figura 3 – Interface gráfica construída



Fonte: O AUTOR, 2013.

A interface possui os seguintes parâmetros de entrada:

- a) Dados → Abre uma janela para a busca do arquivo de dados que será utilizado.
- b) Pasta de Trabalho → Indica o local do disco rígido onde serão salvos os resultados da análise.
- c) Braços → Soma de todos os braços de comparação incluídos em todos os estudos analisados. Por exemplo, um estudo que compara três tratamentos, possui três braços de comparação.
- d) Estudos → Quantidade de estudos analisados.
- e) Tratamentos → Quantos tratamentos estão sendo estudados no total.
- f) Cadeias → Número de cadeias que serão geradas na simulação.
- g) Simulações → Número de simulações que serão realizadas em cada cadeia.
- h) Adaptação → Número de amostras para adaptação (ou aquecimento) das cadeias

Além disso, pode-se trabalhar com quatro modelos:

1. Modelo de efeito fixo para estudos com dois braços de comparação.
2. Modelo de efeito aleatório para estudos com dois braços de comparação.
3. Modelo de efeito aleatório para estudos com até três braços de comparação.
4. Modelo de efeito aleatório para estudos com múltiplos braços de

comparação.

É possível trabalhar com comparações que possuem dois desfechos:

1. Favorável → Adesão à tratamento, redução de intercorrências, etc.
2. Desfavorável → Mortalidade, aumento de intercorrências, etc.

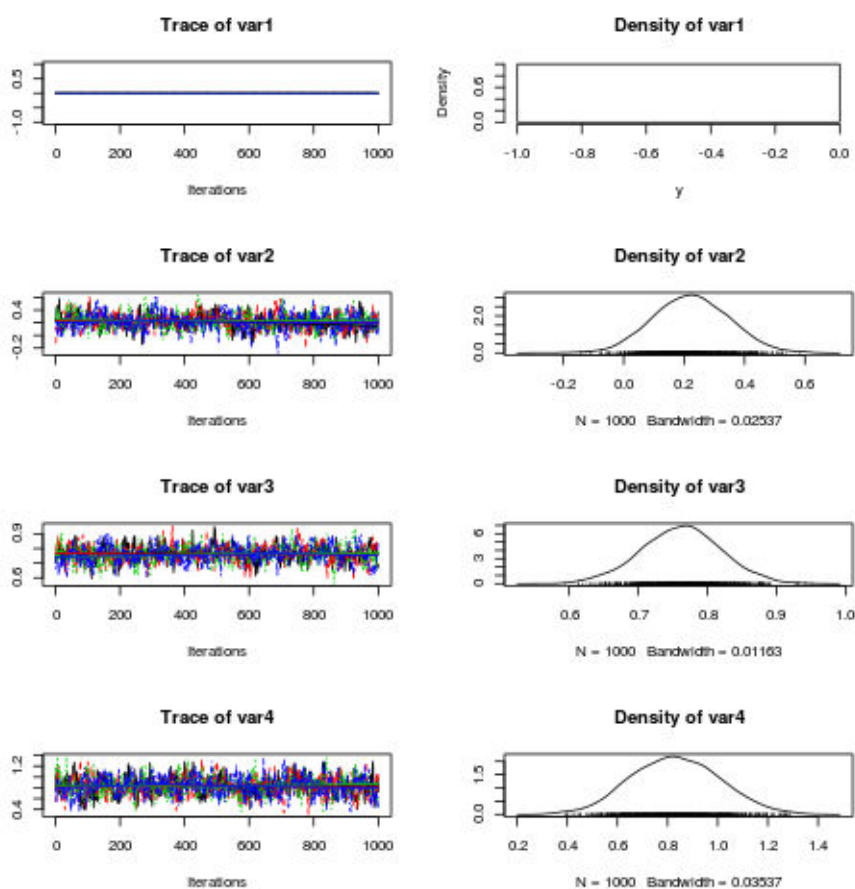
1.12 Resultados do aplicativo

Na execução do modelo na interface gráfica serão criados sete arquivos para análise:

- 1 - Estatísticas.txt;
- 2 - Samples_d.png;
- 3 - Samples_T.png;
- 4 - Gelman_d.png;
- 5 - Gelman_T.png;
- 6 - Iterações.png;
- 7 - Relatório.pdf.

Os arquivos Estatísticas.txt, Samples_d.png e Samples_T.png apresentam as estatísticas descritivas acerca das variáveis d (Logaritmo da razão de chance dos tratamentos), T (Eficácia absoluta dos tratamentos) é uma variável calculada a partir da soma das distribuições a priori vagas de cada do logaritmo da razão de chance dos tratamentos com um parâmetro obtido através da metanálise dos estudos do tratamento referência, a comparação das razões de chances entre os tratamentos e a probabilidade de um determinado tratamento ser o melhor/pior dentre os demais tratamentos ao longo da simulação, tendo como saída gráficos de desempenho.

Figura 4 – Arquivo Samples_d

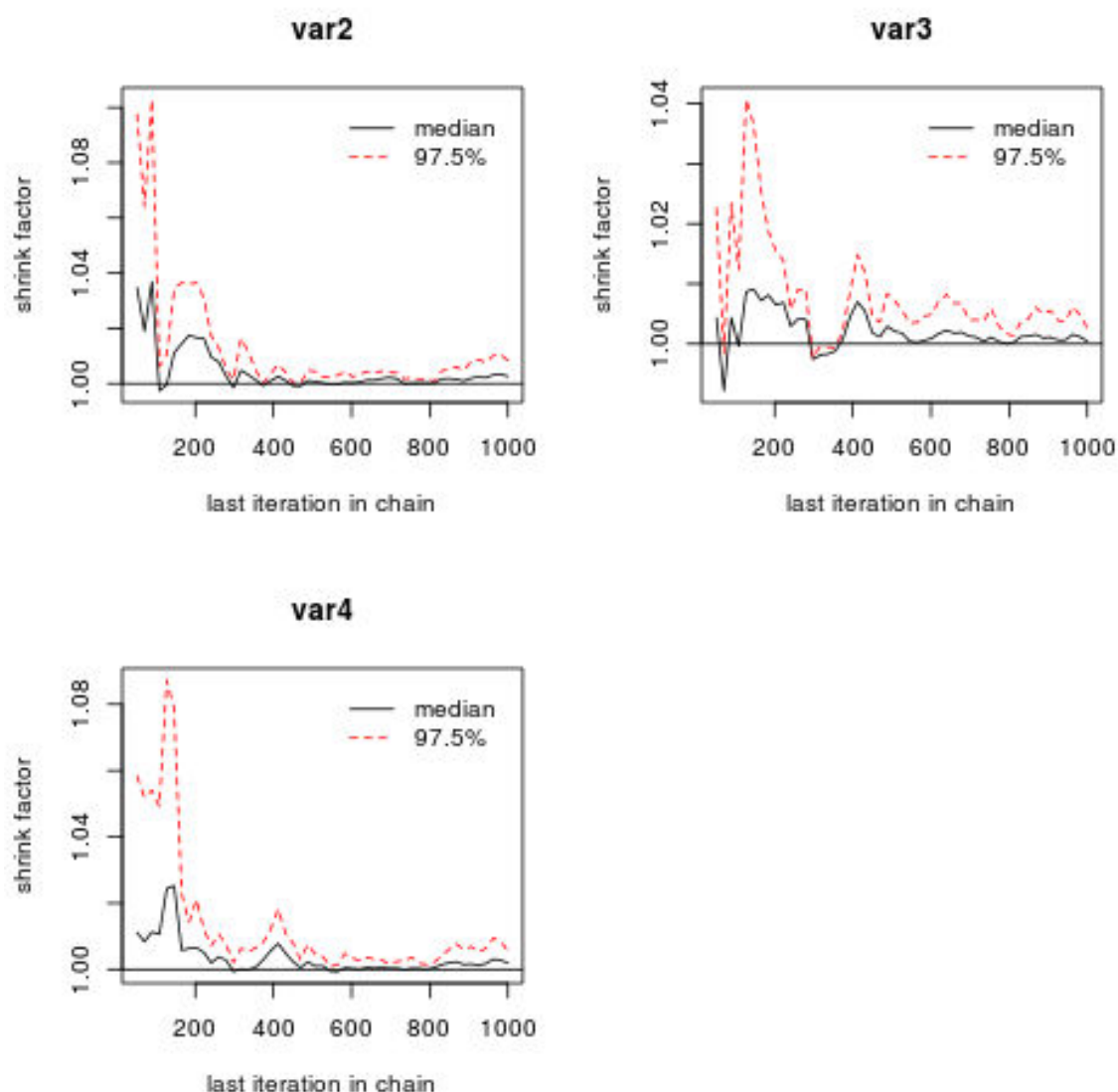


Fonte: O AUTOR, 2013. Gerado pelo software.

Na Figura 5, a coluna da esquerda apresenta gráficos com o comportamento de cada um dos tratamentos para a variável d. (Var1 igual a tratamento 1, Var2 igual a tratamento 2, Var3 igual a tratamento 3 e Var4 igual a tratamento 4).

Na Figura 6, são apresentados gráficos de Gelman-Rubin para a verificação de convergência do modelo para variáveis d e T, que estão disponíveis nos arquivos arquivos Gelman_d e Gelman_T. Cada arquivo mostra o comportamento das variáveis d e T ao longo da simulação executada, que verificará se ocorreu convergência entre os espectros obtidos pelo modelo e os dados observados. Os valores ajustados correspondem ao limite superior do intervalo de credibilidade para mediana adotada em risco probabilístico de 5%.

Figura 5 – Arquivo Gelman_T

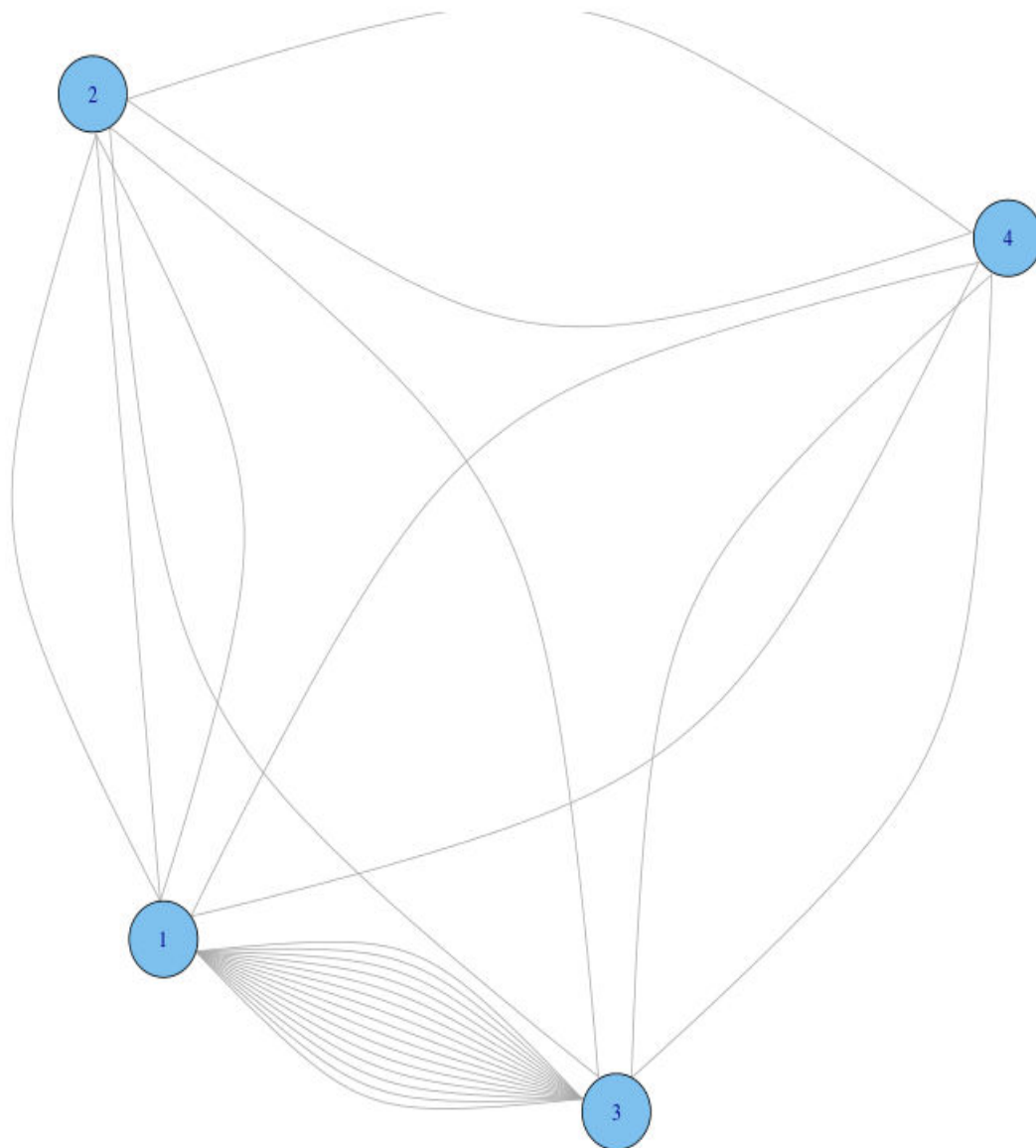


Fonte: O AUTOR, 2013. Gerado pelo software.

O tratamento 1 foi considerado o tratamento de referência para a enfermidade analisada, não sendo necessário obter o gráfico.

O próximo grafo gerado, Figura 6 (Iterações.png), mostra a interatividade entre os tratamentos através da quantidade de comparações. Cada linha que interliga os tratamentos representa a interatividade entre eles. Por exemplo, a comparação entre os tratamentos 1 e 3 foi a mais estudada, enquanto a comparação entre 1 e 2 teve comportamento oposto.

Figura 6 – Arquivo Interações



Fonte: O AUTOR, 2013. Gerado pelo software.

O arquivo “Relatório.pdf”, condensa todas as informações mostradas acima num único relatório, facilitando a divulgação dos resultados.

2 APLICAÇÃO

A funcionalidade da ferramenta foi mostrada a partir do exemplo que visava indicar qual o medicamento que reduz a incidência de diabetes na população³⁵. Pelo fato do desfecho estudado ser desfavorável (incidência de diabetes na população), os resultados obtidos para a variáveis T representam a ineficácia do tratamento.

Os medicamentos a serem comparados são:

- a) Diurético;
- b) Beta Bloqueador;
- c) Bloqueador do canal de cálcio (BCC);
- d) Placebo;
- e) Inibidor de enzimas conversoras de angiotensina (IECA);
- f) Bloqueador dos receptores da angiotensina (BRA).

O trabalho possui 22 estudos com 48 braços de comparações, correspondendo ao total de comparações em todos os estudos selecionados, conforme quadro abaixo.

Quadro 3 – Dados do estudo

Estudo	Diurético (Eventos/n)	Beta bloqueador (Eventos/n)	Placebo (Eventos/n)	BCC (Eventos/N)	IECA (Eventos/n)	BRA (Eventos/n)
HAPPHY	75/3272	86/3297				
MRC-E	43/1081	37/1102	34/2213			
EWPHE	29/416		20/424			
SHEP	140/1631		118/1578			
INSIGHT	176/2511			136/2508		
ALLHAT	302/6766			154/3954	119/4096	
ANBP-2	200/2826				138/2800	
ALPINE	8/196					1/196
INVEST		665/8078		569/8098		
NORDIL		251/5059		216/5095		
ASCOT		799/7040		567/7072		
AASK		70/405		32/202	45/410	
STOP-2		97/1960		95/1965	93/1970	
CAPP		380/5230			337/5183	
LIFE		320/3979				242/4020
FEVER			154/4870	177/4841		
HOPE			155/2883		102/2837	
PEACE			399/3472		335/3432	
DREAM			489/2646		449/2623	
SCOPE			115/2175			93/2167
CHARM			202/2721			163/2715
VALUE				845/5074		690/5087

Fonte: WELLS et al., 2006.

Os dados precisam ser tabulados em formato executável exigido pelo programa em arquivo na extensão CSV (valores separados por vírgula), para as variáveis, número identificador do estudo (s), tratamento estudado(t), número de casos (r), número total da população(n), tratamento de referência (b) e quantificador de braços (m), conforme tabela abaixo.

Tabela 1 – Variáveis do banco de dados

s	t	r	n	b	m
1	1	75	3272	1	1
1	2	86	3297	1	2
2	1	43	1081	1	1
2	2	37	1102	1	2
2	3	34	2213	1	3
3	1	29	416	1	1
3	3	20	424	1	2

Fonte: O AUTOR, 2013.

No arquivo CSV o estudo 2 possui três tratamentos com 43, 37 e 34 casos (incidência), com a população total do braço 1081, 1102 e 2213, respectivamente, sendo o tratamento 1 o de referência tendo 3 braços (1, 2 e 3). Na Figura 7, está a imagem da entrada das informações da aplicação.

Após a construção deste arquivo, foi possível realizar as análises no aplicativo.

Figura 7 – Preenchimento das informações no aplicativo



Fonte: O AUTOR, 2013.

Optou-se por executar as análises no modelo de efeito rândomico para estudos que possuam até três braços, pois alguns estudos possuem essa característica.

O primeiro resultado diz respeito ao medicamento considerado melhor dentro das simulações realizadas. A simulação apontou o tratamento 6 como o mais eficaz com índice de 76%. Os demais tratamentos apresentaram resultados de eficácia igual à 24%.

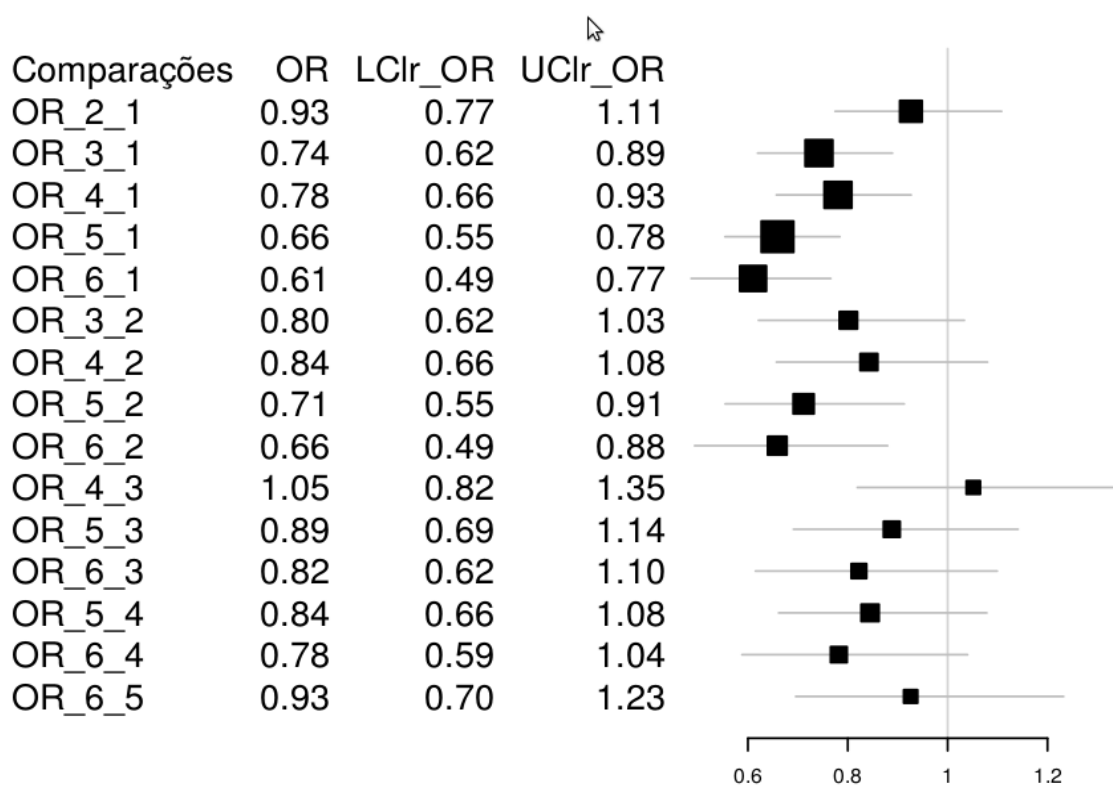
Quadro 4 – Indicadores de qualidade para os tratamentos

Tratamento 1	0,00%
Tratamento 2	0,00%
Tratamento 3	0,61%
Tratamento 4	0,11%
Tratamento 5	23,28%
Tratamento 6	76,00%

Fonte: O AUTOR, 2013.

As razões de chance das comparações entre os tratamentos foram discriminadas na figura abaixo, juntamente com o gráfico de floresta:

Figura 8 – Tabela de razão de chance com gráfico de floresta

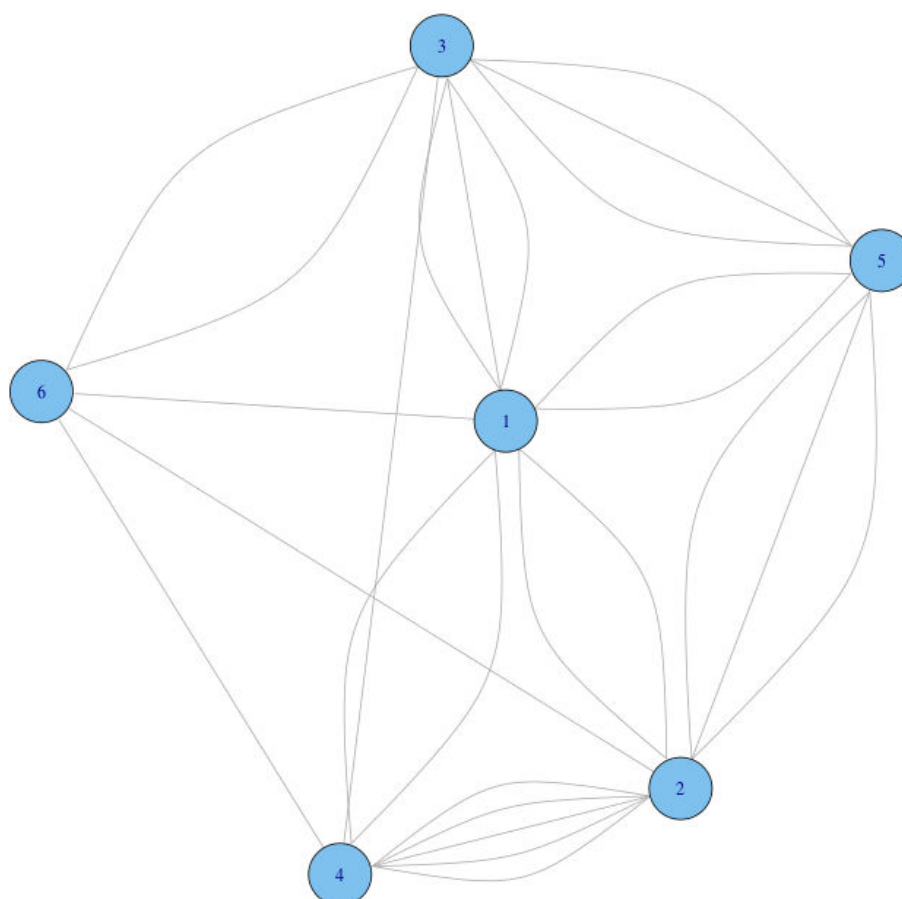


Fonte: O AUTOR, 2013.

Observou-se que todos os tratamentos, quando comparados com Diurético tiveram desempenho melhor, com exceção da comparação de Diurético e Beta Bloqueador (0,93) que obteve índice similar à comparação dos tratamentos IECA e BRA. As comparações entre o tratamento Diurético contra IECA e Diurético contra BRA, foram as que tiveram maior discrepância.

É importante também verificar como o processo foi desenhado, ou seja, como as comparações foram delineadas sob a ótica gráfica. Assim sendo, o grafo abaixo ilustra as comparações existentes.

Figura 9 – Grafo com as comparações existentes no trabalho



O grafo mostra que a maior parte das comparações feitas foram entre o tratamento Beta Bloqueador e Placebo (5 comparações), de maneira oposta só existe uma comparação direta entre Diurético contra BRA, Beta Bloqueador contra BRA e Placebo contra BRA.

Também são disponibilizadas as estatísticas descritivas e os gráficos de monitoramento da simulação das variáveis d (Logaritmo da razão de chance dos tratamentos) e T (Ineficácia absoluta do tratamento) no Quadro 5 e 6 e nas Figuras 10 e 11, respectivamente.

Quadro 5 – Resultados acerca do logaritmo da razão de chance dos tratamentos (d)

Tratamentos	Média	Desvio Padrão	Erro padrão ingênuo	Erro padrão da série histórica
1	0,0000	0,0000	0,0000	0,0000
2	-0,0764	0,0913	0,0005	0,0018
3	-0,2980	0,0923	0,0005	0,0020
4	-0,2479	0,0878	0,0004	0,0017
5	-0,4167	0,0885	0,0004	0,0018
6	-0,4937	0,1159	0,0006	0,0022

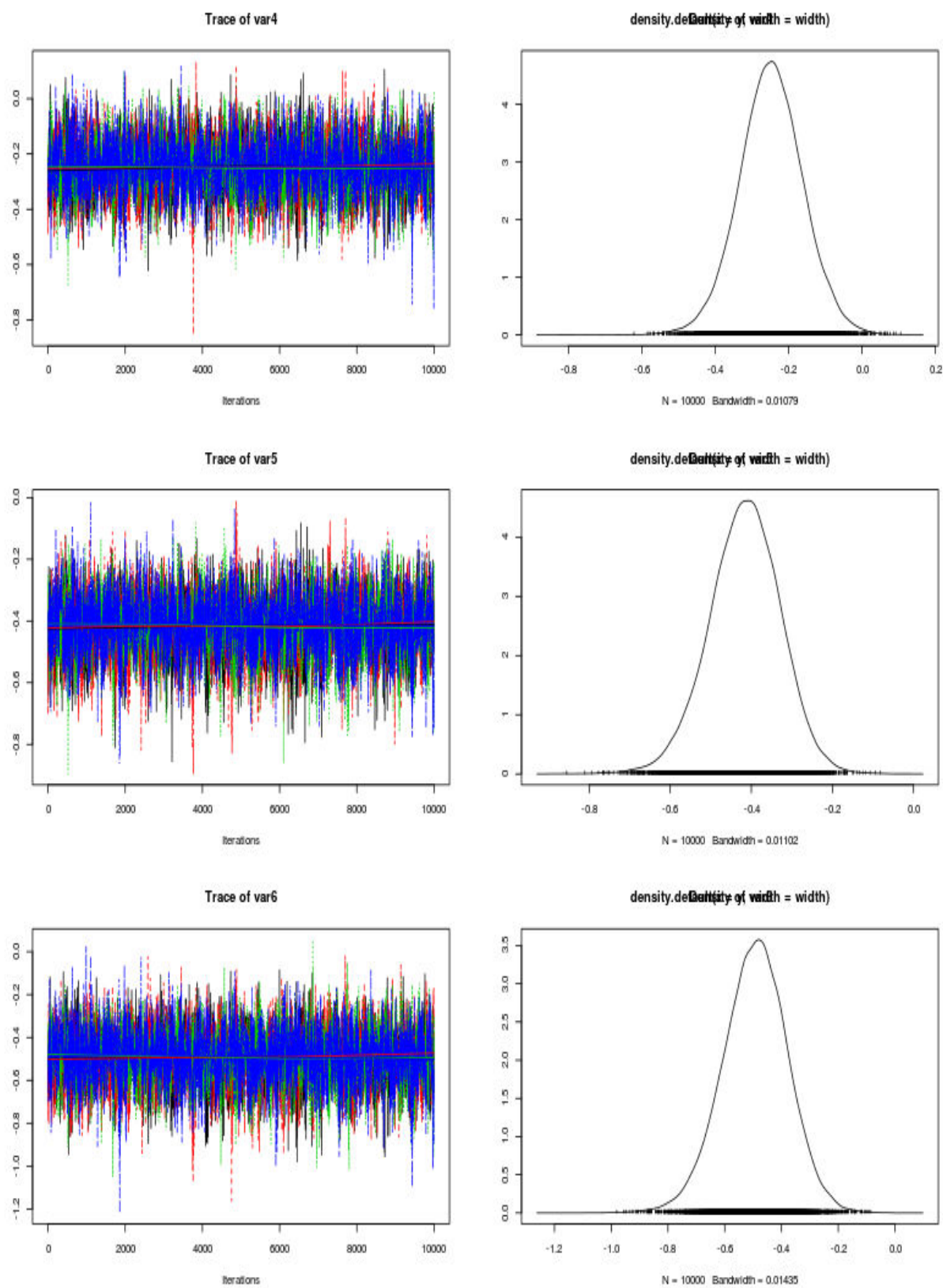
Fonte: O AUTOR, 2013.

Quadro 6 – Eficácia absoluta do tratamento (T)

Tratamentos	Média	Desvio Padrão	Erro padrão ingênuo	Erro padrão da série histórica
1	0,0546	0,0232	0,00001	0,0001
2	0,0510	0,0223	0,00001	0,0001
3	0,0414	0,0184	0,000009	0,0001
4	0,0434	0,0192	0,000009	0,0001
5	0,0369	0,0165	0,000008	0,0001
6	0,0344	0,0156	0,000008	0,0001

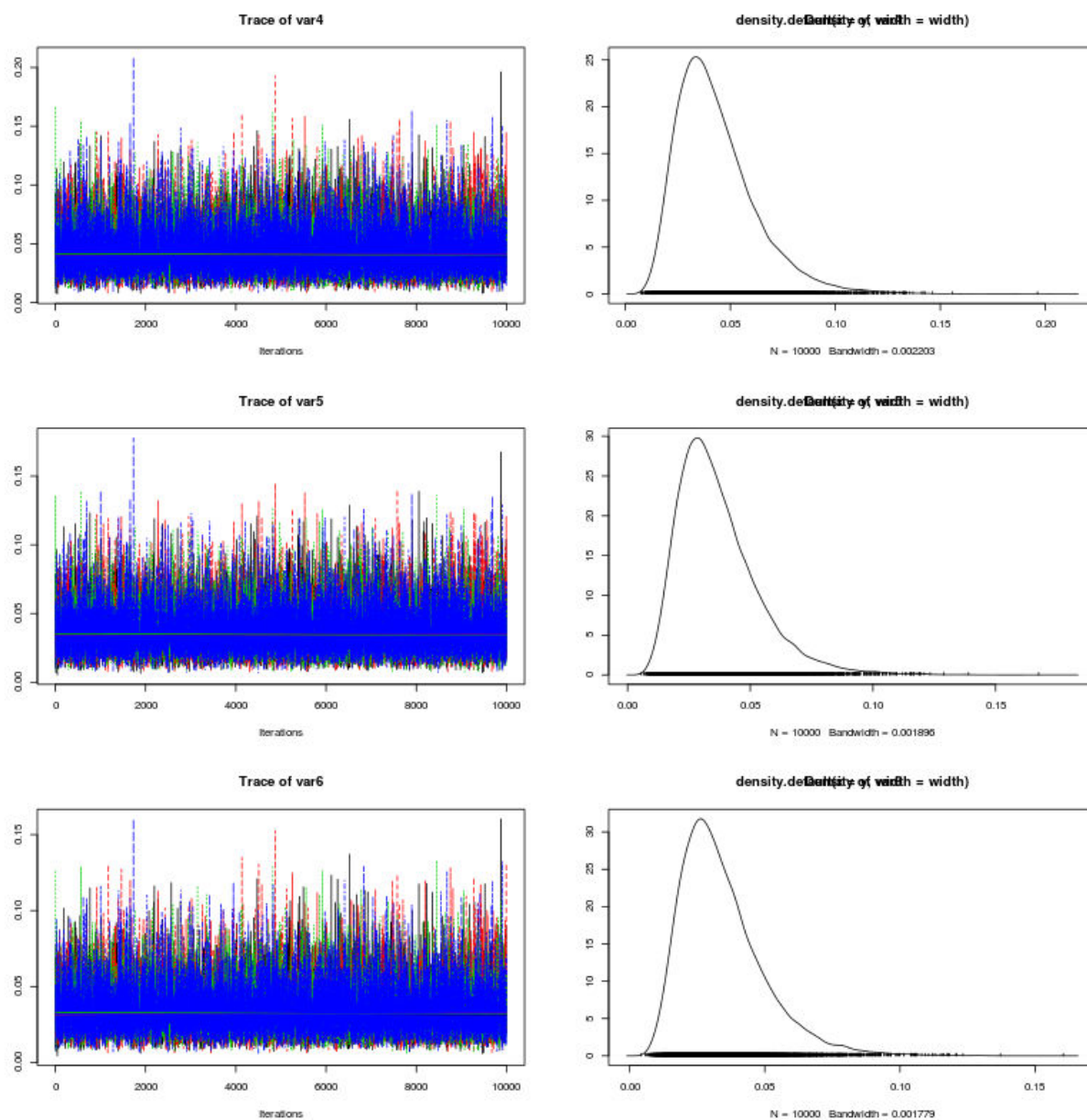
Fonte: O AUTOR, 2013.

Figura 10 – Gráficos de monitoramento da variável d



Fonte: O AUTOR, 2013.

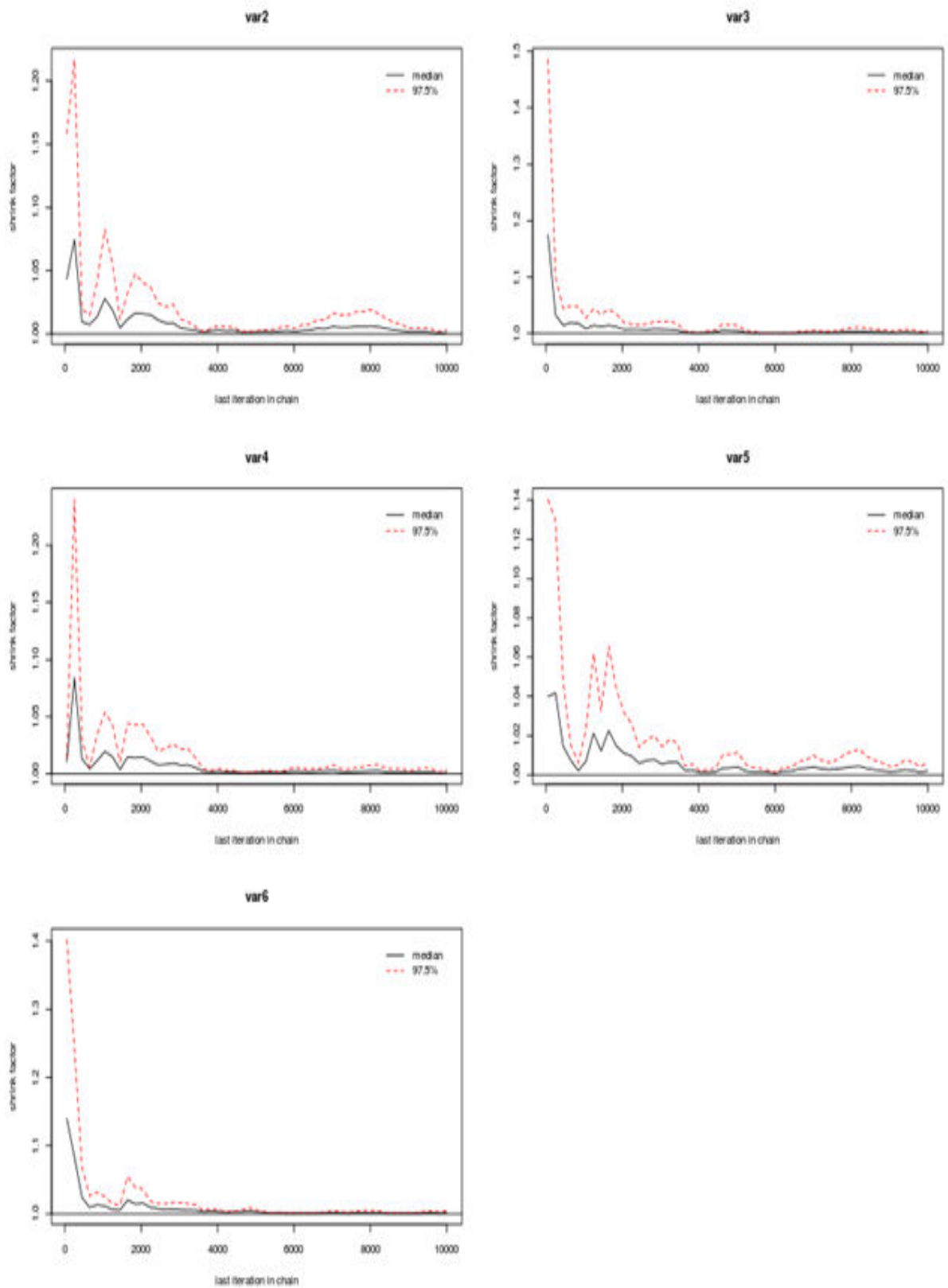
Figura 11 – Gráficos de monitoramento da variável T



Fonte: O AUTOR, 2013.

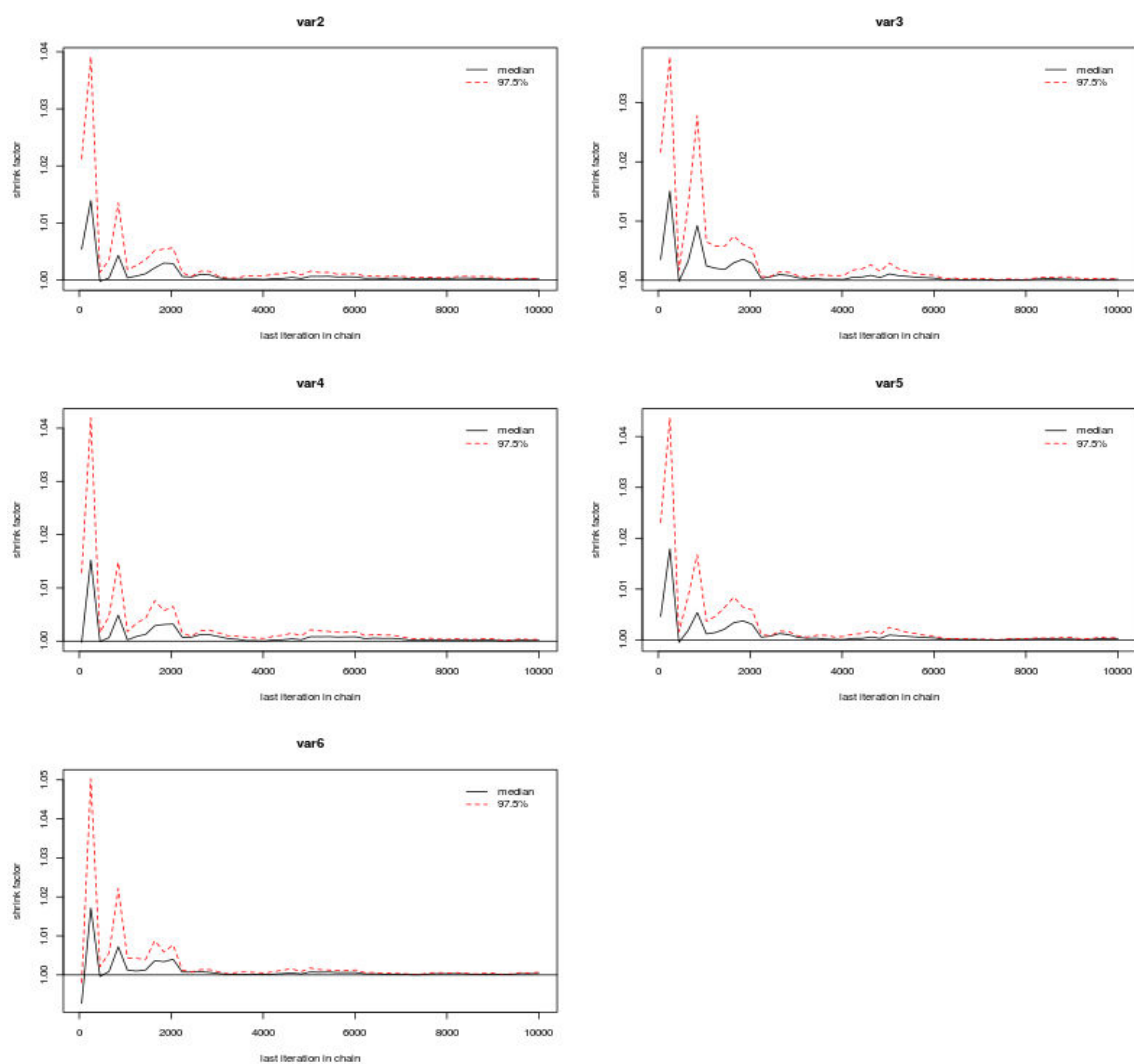
Como forma de verificar a convergência dos modelos, foram analisados os gráficos de Gelman-Rubin para a determinação da presença de convergência nos modelos. As Figuras 12 e 13 apresentam os espectros para a mediana com respectivo limite superior do intervalo de credibilidade do fator de contração da convergência.

Figura 12 – Gráfico de Gelman-Rubin para variável d



Fonte: O AUTOR, 2013.

Figura 13 – Gráfico de Gelman-Rubin para variável T



Fonte: O AUTOR, 2013.

A título de exemplo, os gráficos mostrados acima indicam que houve convergência do modelo de simulação. Essa condição é possível de ser verificada através do comportamento da mediana, se ao final da simulação o valor obtido for próximo a 1 e menor que 1,1 então indica convergência do modelo²⁶.

CONCLUSÃO

O aplicativo mostrou-se eficaz na execução das análises, bem como nas modelagens desenvolvidas. As principais características são:

- a) software livre, aberto e gratuito;
- b) programável para Windows, Mac e Linux;
- c) possui interface gráfica simplificada;
- d) utiliza o R e o Jags como motor analítico;

Existe apenas um aplicativo para criação de modelos de Comparação Mista de Tratamentos chamado GeMTC³⁶. O programa executa algumas das análises que foram propostas por esse trabalho, porém não aborda o monitoramento de variáveis, interações entre os estudos e simulação para busca do melhor tratamento.

O WinBUGS auxilia na execução das análises porém não é um programa exclusivo para a criação de análises em Comparação Mista de Tratamentos.

O principal impacto causado pela implementação desse aplicativo foi no aspecto econômico, pois propiciou comparar diversas terapias e posteriormente verificar as oportunidades de compras. Os resultados obtidos pelos usuários da ferramenta serão utilizados em futuras tomadas de decisão.

Nesse trabalho não foi priorizada a construção do aplicativo para dispositivos móveis (tablets e celulares), porém, através de um trabalho de portabilidade dos códigos-fontes criados para as linguagens de programação desses dispositivos móveis, será possível a replicação do aplicativo para tais dispositivos.

Não foram aplicados testes de desempenho nem aferição de consumo de recursos nos computadores onde foram realizados os testes.

Por se tratar de uma aplicação de código livre e extensível para outras plataformas é possível modificar diversos aspectos do aplicativo, como por exemplo: A reestruturação da interface gráfica; implementação de novos métodos; customização do relatório de saída.

Existe um amplo processo para melhorias do software construído através da abertura do código-fonte. Esse processo pode ser compartilhado com outros usuários tornando o trabalho mais ágil e validado para diversas realidades de utilização.

O processo de criação de modelos também se configura como tarefa importante para futuros trabalhos, visto que nesse processo foram utilizados

modelos já criados por outros autores. A construção de um novo modelo dentro da realidade abordada pode contribuir ainda mais com o processo de desenvolvimento da Comparação Mista de Tratamentos.

REFERÊNCIAS

1. BUCHER HC et al. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of Clinical Epidemiology*. Philadelphia, 1997, 50 (6), p.683-91.
2. LUMLEY T. Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine*. Malden, 2002, 21 (6), p. 2313-24.
3. LU, G.; ADES, A.E. Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine*. Malden, 2004, 23 (30), p. 3105-3124.
4. R FOUNDATION FOR STATISTICAL COMPUTING. R Core Team. *R: A language and environment for statistical computing*. Viena. Disponível em <<http://www.R-project.org/>> Acesso em 2012.
5. VAN VALKENHOEF G. et.al. Automating network meta-analysis. *Research Synthesis Methods*. Malden, 2012, 3 (4), p.285-299.
6. MRC BIOSTATISTICS UNIT. The BUGS Project. Apresenta informações gerais sobre o projeto BUGS. Disponível em <<http://www.mrc-bsu.cam.ac.uk/bugs/>> Acesso em outubro de 2012.
7. PORTAL DA SAÚDE/ Avaliação de tecnologias em saúde no Brasil. Desenvolvido pelo Ministério da Saúde. Apresenta as regras e diretrizes sobre estudos de avaliação de tecnologias em saúde no Brasil. Disponível em <http://portal.saude.gov.br/portal/saude/profissional/area.cfm?id_area=1026> Acesso em outubro de 2012.
8. MACHIN, D.; FAYERS P. *Randomized clinical trials: design, practice and reporting*. Chichester: Wiley-Blackwell, 2010.
9. JADAS, A.R. *Randomized controlled trials: questions, answers, and musings*. 2th ed. Malden: Blackwell, 2007.
10. PIANTADOSI S. *Clinical trials: a methodologic perspective*. 2th ed. Hoboken: Wiley, 2005.
11. GLASS, G.V. Primary, secondary and meta-analysis of research. *Educational Researcher*. Washington, 1976, 5 (10), p. 3-8.
12. WHITEHEAD A. *Meta-analysis of controlled clinical trials*. Chichester: John Wiley & Sons, 2003.
13. JENSEN et al. Interpreting Indirect Treatment Comparisons and Network Meta-Analysis for Health-Care Decision Making: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 1. *Value in Health*. Lawrenceville, 2011, vol.14, p. 417-428.

14. BOLSTAD, W.M. *Understanding Computational Bayesian statistics*. Hoboken: Wiley, 2010.
15. KOCH, K.C. *Introduction to Bayesian Statistics*. 2th ed. Berlim: Springer, 2007.
16. GAMERMAN, D.; LOPES, H.F. *Markov Chain Monte Carlo*. 2th ed. Boca Raton: Chapman & Hall, 2006.
17. MIGON H.S.; GAMERMAN D. *Statistical Inference: An Integrated Approach*. Londres: Arnold, 1999.
18. JEFFREYS, H. *Theory of Probability*. 3th ed. Oxford: Oxford University Press, 1961.
19. PRESS S.J. *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications*. 2th ed. Hoboken: Wiley-Interscience, 2002.
20. BROOKS, S. et al. *Chain Monte Carlo*. Boca Raton: Chapman and Hall, 2011.
21. JING L. Hastings-within-Gibbs Algorithm: *Introduction and Application on Hierarchical Model*. Santo Antonio. Universidade do Texas, 2010. Disponível em <<http://georgism.rforge.r-project.org/site-projects/computation.html>> Acesso em setembro de 2012.
22. SUTTON, A. et al. Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics*. Leicester: University of Leicester, 2008, 26 (9), p.753-767.
23. BAYES, T.; PRICE, R. An Essay towards solving a Problem in the Doctrine of Chance. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F.R.S. *Philosophical Transactions of the Royal Society of London*. Londres, 1763, vol. 53, p.370-418.
24. SMITH, T.C. et al. Bayesian meta-analysis of randomised trials using graphical models and BUGS. *Bayesian Biostatistics*, p. 411-428, Berry DA, Stangl DK (eds.). New York: Marcel Dekker, 1996.
25. HIGGINS J.P.T.; WHITEHEAD J. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine*. Malden, 1996, 15 (24), p. 2733-2749.
26. BROOKS, S.P.; GELMAN, A. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*. Alexandria, 1998, 7 (4), p.434-455.
27. PLUMMER, M. JAGS: *A program for analysis of Bayesian graphical models using Gibbs sampling*. Viena. Distributed Statistical Computing, 2003. Programa. Disponível em <<http://mcmc-jags.sourceforge.net>> Acesso em setembro de 2012.

28. THOMAS J.H.; NAN M.L. *fgui: A Method for Automatically Creating Graphical User Interfaces for Command-Line R Packages*. Journal of Statistical Software. Alexandria. Vol. 30, 2 (14), p.1-14. Disponível em <<http://www.jstatsoft.org/v30/i02/>>. Acesso em setembro de 2012.
29. SARKAR, D. *Lattice: Multivariate Data Visualization with R*. New York: Springer, 2008.
30. PLUMMER et.al. CODA: *Convergence Diagnosis and Output Analysis for MCMC*, 6 (7). Viena: R News, 2006.
31. CSARDI G.; NEPUSZ T. *The igraph software package for complex network research*, *InterJournal, Complex Systems* 1695, 2006. Disponível em <<http://igraph.sf.net>>. Acesso em setembro de 2012.
32. WARNES et al. *gplots: Various R programming tools for plotting data*. Disponível em <<http://CRAN.R-project.org/package=gplots>>. Acesso em novembro de 2012.
33. TCL CORE TEAM TCL. Developer Xchange. Apresenta informações gerais sobre o software Tcl/Tk. Disponível em <<http://www.tcl.tk/>>. Acesso em setembro de 2012.
34. UNIVERSIDADE DE BRISTOL. Apresenta conteúdo sobre diversos cursos e trabalhos desenvolvidos na instituição. <Disponível em <http://www.bris.ac.uk/>>. Acesso em setembro de 2012.
35. WELLS et al. *Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis*. Ottawa. Canadian Agency for Drugs and Technologies in Health, 2009. Disponível em <<http://cadth.ca/>>. Acesso em setembro de 2012.
36. VALKENHOEF, G.V. et al. Automating Network Meta-analysis. *Research Synthesis Methods*. Malden, 2012, 3 (4), p. 285-299.

ANEXO - Código para criação da interface gráfica

```
require(tcltk) # Chama o pacote tcltk
```

```
funcao <- function(){ #
options("repos"="http://cran.us.r-project.org") #
install.packages('fgui') # → Cria a função para instalação do
fgui
require(fgui) #
}
```

```
ifelse(any(grepl('fgui',installed.packages()[,1])),require(fgui),funcao()) # Chama e
instala (Se necessário) o pacote fgui
```

```
Teste <- function(Dados, Pasta_de_Trabalho = "C:/", Braços, Estudos, Tratamentos,
Cadeias = "4", Simulações = "10000", Adaptação = "500", Opt, Desfecho){ # Inicia a
programação da interface gráfica
```

```
Modelo <- NULL # Cria a variável "Modelo"
```

```
V <- NULL # Cria a variável "V"
```

```
W <- NULL # Cria a variável "W"
```

```
summary_OR <- NULL # Cria a variável "summary_OR"
```

```
best <- NULL # Cria a variável "best"
```

```
rk <- NULL # Cria a variável "rk"
```

```
setwd(Pasta_de_Trabalho) # Parâmetro que indica o local onde serão salvos os
relatórios
```

```
funcaoggm <- function(){ #
source("http://bioconductor.org/biocLite.R") #
biocLite('ggm', dep=T, clean=T) # → Cria a função de instalação do
pacote
require(ggm) # ggm
}
```

```
ifelse(any(grepl('ggm',installed.packages())[,1])),require(ggm),funcaoggm()) # Chama
e instala (Se necessário) o pacote ggm
```

```
funcaorjags <- function(){ #
options("repos"="http://cran.us.r-project.org") #
install.packages('rjags') # → Cria a função de instalação do
pacote
require(rjags) # rjags
} #
```

```
ifelse(any(grepl('rjags',installed.packages())[,1])),require(rjags),funcaorjags()) #
Chama e instala (Se necessário) o pacote rjags
```

```
funcaoigraph <- function(){ #
options("repos"="http://cran.us.r-project.org") #
install.packages('igraph') # → Cria a função de instalação do
pacote
require(igraph) # igraph
} #
```

```
ifelse(any(grepl('igraph',installed.packages())[,1])),require(igraph),funcaoigraph()) #
Chama e instala (Se necessário) o pacote igraph
```

```
funcaormeta <- function(){ #
options("repos"="http://cran.us.r-project.org") #
install.packages('rmeta') # → Cria a função de instalação do
pacote
require(rmeta) # rmeta
} #
```

```
ifelse(any(grepl('rmeta',installed.packages())[,1])),require(rmeta),funcaormeta()) #
Chama e instala (Se necessário) o pacote rmeta
```

```
funcaogplots <- function(){ #
```

```

options("repos"="http://cran.us.r-project.org") #
install.packages('gplots')                    # → Cria a função de instalação do
pacote
require(gplots)                               # gplots
}                                               #

ifelse(any(grepl('gplots',installed.packages()[,1])),require(gplots),funcaogplots()) #
Chama e instala (Se necessário) o pacote gplots

ensaios <- read.csv(Dados, header=T, sep = " ") # Cria a base de dados a partir do
arquivo de dados CSV

# Formatação da base de dados para ensaios com múltiplos braços
# Início

ifelse(Opt=="Random effects models for multi-arm trials",
{newdata <- subset(ensaios, t..1. == 1)
newdata$Prop <- with(newdata,ifelse(r..1. > 0, r..1./n..1., 'NA'))
newdata$Prop <- as.numeric(newdata$Prop)
newdata <- newdata[complete.cases(newdata),]
newdata$LN_Prop <- with(newdata,ifelse(r..1. > 0, log(r..1./n..1.), 'NA'))
newdata$LN_Prop_SE <- with(newdata,ifelse(r..1. > 0, log(sqrt(Prop*(1-Prop)/n..1.)),
'NA'))
meta_mA <- with(newdata,meta.summaries(LN_Prop,
LN_Prop_SE,method=c("random")), logeffect=T)
mean_mA <- meta_mA$summary
prec_mA <- 1/with(newdata,var(LN_Prop,na.rm=T))
output_mA <- capture.output(cat('mA ~ dnorm(',mean_mA,',',prec_mA,')'))},
{newdata <- subset(ensaios, t == 1)
newdata$Prop <- with(newdata,ifelse(r > 0, r/n, 'NA'))
newdata$Prop <- as.numeric(newdata$Prop)
newdata <- newdata[complete.cases(newdata),]
newdata$LN_Prop <- with(newdata,ifelse(r > 0, log(r/n), 'NA'))
newdata$LN_Prop_SE <- with(newdata,ifelse(r > 0, log(sqrt(Prop*(1-Prop)/n)), 'NA'))

```

```

meta_mA <- with(newdata,meta.summaries(LN_Prop,
LN_Prop_SE,method=c("random")), logeffect=T)
mean_mA <- meta_mA$summary
prec_mA <- 1/with(newdata,var(LN_Prop,na.rm=T))
output_mA <- capture.output(cat('mA ~ dnorm(',mean_mA,',',prec_mA,')'))
# Fim

# Passagem de parâmetro para a programação que executa o modelo de efeito fixo
# Início

if(Opt=="Fixed effect models") {Modelo <- file("/usr/bin/FE.bug")
writeLines(c("model{","for(i in 1:N){", "logit(p[s[i],t[i]])<-mu[s[i]]+ d[t[i]] - d[b[i]]",
"r[i]~dbin(p[s[i],t[i]],n[i])}", "for(j in 1:NS){", "mu[j]~dnorm(0,.0001)}", "for (k in 2:NT){",
"d[1] <- 0", "d[k] ~ dnorm(0,.0001)}", output_mA, "for (l in 1:NT) { logit(T[l]) <- mA +
d[l] }"}),con=Modelo,sep="\n")
# Fim

# Passagem de parâmetro para a programação que executa o modelo de efeito
aleatório para dois braços de comparação
# Início
if(Opt=="Random effect model that ignores correlations induced by multi-arm trials")
{Modelo <- file("/usr/bin/REM.bug")
writeLines(c("model{","for(i in 1:N) { logit(p[i])<-mu[s[i]]+delta[i] * (1-
equals(t[i],b[i]))", "r[i]~dbin(p[i],n[i])", "delta[i] ~ dnorm(md[i],tau)", "md[i] <- d[t[i]] -
d[b[i]]", "for(j in 1:NS){ mu[j]~dnorm(0,.0001)}", "d[1]<-0", "for (k in 2:NT) {d[k] ~
dnorm(0,.0001)}", "sd~dunif(0,2)", "tau<-1/pow(sd,2)", output_mA, "for (l in 1:NT) {
logit(T[l]) <- mA + d[l] }"}),con=Modelo,sep="\n")
# Fim

# Passagem de parâmetro para a programação que executa o modelo de efeito
aleatório com até três braços
# Início
if(Opt=="Random effects models for 2- and 3-arm trials") {Modelo <-
file("/usr/bin/REM23T.bug")

```

```
writeLines(c("model{","sw[1] <- 0","for(i in 1:N){","logit(p[i])<-mu[s[i]] + delta[i] * (1-
equals(t[i],b[i]))","r[i]~dbin(p[i],n[i])","delta[i] ~ dnorm(md[i],taud[i])","taud[i] <- tau * (1 +
equals(m[i],3)/3)","md[i] <- d[t[i]] - d[b[i]] + equals(m[i],3) * sw[i]","for (i in 2:N) {sw[i]
<- (delta[i-1] - d[t[i-1]] + d[b[i-1]])/2)","for(j in 1:NS){ mu[j]~dnorm(0,.0001)}","d[1]<-
0","for (k in 2:NT) {d[k] ~ dnorm(0,.0001)}","sd~dunif(0,2)","tau<-
1/pow(sd,2)","output_mA","for (l in 1:NT) { logit(T[l]) <- mA + d[l]
}}"),con=Modelo,sep="\n")
```

```
# Fim
```

```
# Passagem de parâmetro para a programação que executa o modelo de efeito
aleatório com múltiplos braços
```

```
# Início
```

```
if(Opt=="Random effects models for multi-arm trials") {Modelo <-
file("/usr/bin/REMMAT.bug")
```

```
writeLines(c("model{","for(i in 1:NS){","w[i,1] <-0","delta[i,t[i,1]]<-0","mu[i] ~
dnorm(0,.0001)","for (k in 1:na[i]){","r[i,k] ~ dbin(p[i,t[i,k]],n[i,k])","logit(p[i,t[i,k]])<-mu[i]
+ delta[i,t[i,k]]","for (k in 2:na[i]) {"delta[i,t[i,k]] ~
dnorm(md[i,t[i,k]],taud[i,t[i,k]]","md[i,t[i,k]] <- d[t[i,k]] - d[t[i,1]] + sw[i,k]","taud[i,t[i,k]]
<- tau *2*(k-1)/k","w[i,k] <- (delta[i,t[i,k]] - d[t[i,k]] + d[t[i,1]])","sw[i,k] <-sum(w[i,1:k-
1])/(k-1)}}","d[1]<-0","for (k in 2:NT){d[k] ~ dnorm(0,.0001)}","sd~dunif(0,2)","tau<-
1/pow(sd,2)","output_mA","for (l in 1:NT) { logit(T[l]) <- mA + d[l]
}}"),con=Modelo,sep="\n")
```

```
# Fim
```

```
# Formatação da base de dados e passagens de parâmetro para a execução do
modelo selecionado
```

```
# Início
```

```
if(Opt == 'Random effects models for multi-arm trials') n <-
as.matrix(ensaios[,substr(colnames(ensaios), 1, 2) == 'n.'])
if(Opt == 'Random effects models for multi-arm trials') t <-
as.matrix(ensaios[,substr(colnames(ensaios), 1, 2) == 't.'])
if(Opt == 'Random effects models for multi-arm trials') r <-
as.matrix(ensaios[,substr(colnames(ensaios), 1, 2) == 'r.'])
```

```

ifelse(Opt=='Fixed effect models',jags <- jags.model(Modelo,data = list('s'=ensaios$s,
't'=ensaios$t, 'r'=ensaios$r, 'n'=ensaios$n,
'b'=ensaios$b,'N'=Braços,'NS'=Estudos,'NT'=Tratamentos),n.chains =
Cadeias,n.adapt = Adaptação),
ifelse(Opt=='Random effect model that ignores correlations induced by multi-arm
trials',jags <- jags.model(Modelo,data = list('s'=ensaios$s, 't'=ensaios$t, 'r'=ensaios$r,
'n'=ensaios$n, 'b'=ensaios$b,'N'=Braços,'NS'=Estudos,'NT'=Tratamentos),n.chains =
Cadeias,n.adapt = Adaptação),
ifelse(Opt=='Random effects models for 2- and 3-arm trials',jags <-
jags.model(Modelo,data = list('s'=ensaios$s, 't'=ensaios$t, 'r'=ensaios$r,
'n'=ensaios$n, 'b'=ensaios$b,
'm'=ensaios$m,'N'=Braços,'NS'=Estudos,'NT'=Tratamentos),n.chains =
Cadeias,n.adapt = Adaptação),jags <- jags.model(Modelo,data = list('t'=t, 'r'=r, 'n'=n,
'na'=ensaios$na, 'NS'=Estudos,'NT'=Tratamentos),n.chains = Cadeias,n.adapt =
Adaptação))))

```

```
# Fim
```

```

samples <- jags.samples(jags,c('d','T'),Simulações) #
mcmc_samples_d <- as.mcmc.list(samples$d) #
mcmc_samples_T <- as.mcmc.list(samples$T) #
sumario_d <- summary(mcmc_samples_d)$statistics # → Extração das amostras
das
sumario_T <- summary(mcmc_samples_T)$statistics # simulações realizadas nas
output_d <- capture.output(print(sumario_d)) # variáveis d e T
output_T <- capture.output(print(sumario_T)) #
samples_T <- jags.samples(jags,c('T'),Simulações) #
lod <- summary(mcmc_samples_d)$statistics[,1] #
sd_lod <- summary(mcmc_samples_d)$statistics[,2] #
# Cálculo de qual tratamento foi o melhor nas simulações para o desfecho favorável
# Início

if(Desfecho == 'Favorável')
{

```

```

base          <-          matrix(sapply(samples_T,['1',
1:(Cadeias*Simulações)),byrow=T,ncol=Tratamentos)
max_base <- apply(base,1,max)
p_best <- ifelse(base == max_base, 1, 0)
sumario_p_best <- summary(p_best)
output_p_best <- capture.output(print(sumario_p_best))
}
# Fim

```

```

# Cálculo de qual tratamento foi o melhor nas simulações para o desfecho
desfavorável

```

```

# Início

```

```

if(Desfecho == 'Desfavorável')
{
base          <-          matrix(sapply(samples_T,['1',
1:(Cadeias*Simulações)),byrow=T,ncol=Tratamentos)
min_base <- apply(base,1,min)
p_best <- ifelse(base == min_base, 1, 0)
sumario_p_best <- summary(p_best)
output_p_best <- capture.output(print(sumario_p_best))
}
# Fim

```

```

# Cálculo das razões de chances e dos intervalos de credibilidade

```

```

# Início

```

```

matriz_ior <- outer(lod,lod,'-')
LOR <- matriz_ior[lower.tri(matriz_ior)]
matriz_sd_ior <- outer(sd_lod,sd_lod,function(a,b) {sqrt(a^2+b^2)})
SD_LOR <- matriz_sd_ior[lower.tri(matriz_sd_ior)]
OR <- exp(LOR)

```

```

nomes <- paste('OR',col(matriz_ior),row(matriz_ior), sep = "_")

```

```

nomes <-
matrix(nomes,ncol=Tratamentos,byrow=T,dimnames=list(letters[1:Tratamentos],letter
s[1:Tratamentos]))
nomes <- nomes[lower.tri(nomes)]

LCIr_OR <- exp(LOR - 1.96 * (SD_LOR))
UCIr_OR <- exp(LOR + 1.96 * (SD_LOR))
summary_OR <- t(rbind(OR, LCIr_OR, UCIr_OR))

m <- c(OR)
l <- c(LCIr_OR)
u <- c(UCIr_OR)

tabletext<-cbind(
c("Comparações",nomes),
c("OR",format(m,digits=2)),
c("LCIr_OR",format(l,digits=2)),
c("UCIr_OR",format(u,digits=2))
)

output_OR <- capture.output(print(summary_OR))
# Fim

# Geração dos relatórios e gráficos dos resultados obtidos
# Início

Estat_Hora <- gsub(':', '_',Sys.time())
cat(output_d, output_T, output_p_best, output_OR, file =
paste('Estatisticas',Estat_Hora,'txt',sep='.'), type = 'txt', fill=TRUE, sep='\n',
append=TRUE)
samples_d_Hora <- gsub(':', '_',Sys.time())
png(paste('samples_d', samples_d_Hora,'png',sep='.'), width = 800, height = 800)
plot(mcmc_samples_d)
dev.off()

```



```

samples_T_Hora <- gsub(':', '_', Sys.time())
png(paste('samples_T', samples_T_Hora, 'png', sep='.'), width = 800, height = 800)
plot(mcmc_samples_T)
dev.off()

gelman_d_Hora <- gsub(':', '_', Sys.time())
png(paste('gelman_d', gelman_d_Hora, 'png', sep='.'), width = 800, height = 800)
gelman.plot(mcmc_samples_d[,2:Tratamentos])
dev.off()

gelman_T_Hora <- gsub(':', '_', Sys.time())
png(paste('gelman_T', gelman_T_Hora, 'png', sep='.'), width = 800, height = 800)
gelman.plot(mcmc_samples_T[,2:Tratamentos])
dev.off()

if(Opt == "Random effects models for multi-arm trials")
{
final <- 0
for (x in 2:ncol(t)-1)
{
temp <- as.data.frame(cbind(t[,x], t[,x+1]))
final <- rbind(final, temp)
}
plota <- cbind(final$V1, final$V2)
plota <- as.data.frame(plota)
plota <- plota[plota$V1 != plota$V2,]
plota <- plota[complete.cases(plota$V1),]
g <- graph.data.frame(plota, directed=FALSE)
summary(g)
g$layout <- layout.fruchterman.reingold(g)
Intera_Hora <- gsub(':', '_', Sys.time())
png(paste('Interações', Intera_Hora, 'png', sep='.'), width = 800, height = 800)
plot(g)
dev.off()
}

if(Opt == "Fixed effect models" | Opt == "Random effect model that ignores

```

correlations induced by multi-arm trials" | Opt == "Random effects models for 2- and 3-arm trials")

```

plota <- cbind(ensaios$b, ensaios$t)
plota <- as.data.frame(plota)
plota <- plota[plota$V1 != plota$V2,]
g <- graph.data.frame(plota, directed=FALSE)
summary(g)
g$layout <- layout.fruchterman.reingold(g)
Intera_Hora <- gsub(':', '_', Sys.time())
png(paste('Interações', Intera_Hora, 'png', sep='.'), width = 800, height = 800)
plot(g)
dev.off()

```

```

colnames(p_best) <- c(1:Tratamentos)
medias_melhor <- sort(colMeans(p_best), decreasing=T)
melhor_p <- paste(round(medias_melhor[1]*100, digits=2), '%')
nome_melhor <- names(medias_melhor)[1]
texto_melhor <- capture.output(cat('O tratamento', nome_melhor, 'é o melhor
tratamento em', melhor_p, 'das vezes.\n'))

```

```

Relatorio_Hora <- gsub(':', '_', Sys.time())
pdf(file=paste('Relatorio', Relatorio_Hora, 'pdf', sep='.'),
family='Helvetica', paper='A4', pointsize=12)
par(mfrow=c(2,1))
textplot(texto_melhor, cex=1.2)
textplot(output_p_best, cex=0.9)
forestplot(tabletext, c(NA, OR), c(NA, LCIr_OR), c(NA, UCIr_OR), zero=1, clip=c(0,5))
par(mfrow=c(1,1), din=10)
plot(g)
par(mfrow=c(3,1))
textplot(output_d, cex=1.2)
plot(mcmc_samples_d)
gelman.plot(mcmc_samples_d[,2:Tratamentos])
par(mfrow=c(3,1))

```

```

textplot(output_T, cex=1.2)
plot(mcmc_samples_T)
gelman.plot(mcmc_samples_T[,2:Tratamentos])
dev.off()

rm(list=ls(all=TRUE))

}
# Fim

# Programação que executa a abertura da interface gráfica
# Início

res <- gui(Teste, argFilename=list(Dados=NULL), helps=list(Dados="Tabela de dados
(Formato Rdata) com os estudos"), output=NULL, argOption=list(Opt=c("Fixed effect
models", "Random effect model that ignores correlations induced by multi-arm trials",
"Random effects models for 2- and 3-arm trials", "Random effects models for multi-
arm trials"), Desfecho=c("Favorável", "Desfavorável")), title="R MTC Jags")

# Fim

```

Código do modelo de efeito fixo

```

model{
#
for(i in 1:N){
#
logit(p[s[i],t[i]])<-mu[s[i]]+ d[t[i]] - d[b[i]] # → Modelo de efeito fixo
r[i]~dbin(p[s[i],t[i]],n[i])
#
}
#

for(j in 1:NS){
#
mu[j]~dnorm(0,.0001)
# → Geração da distribuição a priori vaga μ
#
}

```

```

d[1] <- 0 #
for (k in 2:NT){ #
  d[k] ~ dnorm(0,.0001) # → Geração da distribuição a priori vaga d
} #

for (l in 1:NT) { logit(T[l]) <- mA + d[l] } # Cálculo da variável T

```

Código do modelo de efeito aleatório para estudos com até dois braços

```

model{

for(i in 1:N) { logit(p[i])<-mu[s[i]]+delta[i] * (1-equals(t[i],b[i])) #
                r[i]~dbin(p[i],n[i]) # → Modelo de efeito
                delta[i] ~ dnorm(md[i],tau) # aleatório para dois
                md[i] <- d[t[i]] - d[b[i]] } # braços

for(j in 1:NS){ mu[j]~dnorm(0,.0001) } # Cálculo da distribuição a priori μ

d[1]<-0 # → Cálculo da distribuição a priori vaga d
for (k in 2:NT) {d[k] ~ dnorm(0,.0001) } #

sd~dunif(0,2) #→ Cálculo da distribuição a priori vaga sd
tau<-1/pow(sd,2) #→ Cálculo de tau

for (k in 1:NT) { logit(T[k])<- mA +d[k] } # Cálculo da variável T

}

```

Código do modelo aleatório para estudos com até três braços

```

model{

sw[1] <- 0 #
for(i in 1:N) { #
  logit(p[i])<-mu[s[i]]+ delta[i] * (1-equals(t[i],b[i])) # → Modelo de efeito aleatório

```

```

para
  r[i]~dbin(p[i],n[i]) # até três braços
  delta[i] ~ dnorm(md[i],taud[i]) #
  taud[i] <- tau * (1 + equals(m[i],3) /3) #
  md[i] <- d[t[i]] - d[b[i]] + equals(m[i],3) * sw[i] #
} #

for (i in 2:N) {sw[i] <- (delta[i-1] - d[t[i-1]] + d[b[i-1]])/2} # Cálculo de sw
for(j in 1:NS){ mu[j]~dnorm(0,.0001) } # Cálculo da distribuição a priori vaga  $\mu$ 

d[1]<-0 # → Cálculo da distribuição a priori vaga d
for (k in 2:NT) {d[k] ~ dnorm(0,.0001) } #

sd~dunif(0,2) # Cálculo da distribuição a priori vaga sd
tau<-1/pow(sd,2) # Cálculo da variável tau

for (k in 1:NT) { logit(T[k])<- mA +d[k] } # Cálculo da variável T

}

Código do modelo aleatório para estudos com qualquer quantidade de braços de comparação

model{
for(i in 1:NS){ #
  w[i, 1] <-0 #
  delta[i,t[i, 1]]<-0 #
  mu[i] ~ dnorm(0,.0001) #
  for (k in 1:na[i]) { #
    r[i,k] ~ dbin(p[i,t[i,k]],n[i,k]) #
    logit(p[i,t[i,k]])<-mu[i] + delta[i,t[i,k]] } #
    for (k in 2:na[i]) { # → Modelo de efeito
aleatório
  delta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],taud[i,t[i,k]]) # para múltiplos braços

```

```

md[i,t[i,k]] <- d[t[i,k]] - d[t[i,1]] + sw[i,k]      #
taud[i,t[i,k]] <- tau *2*(k-1)/k                    #
w[i,k] <- (delta[i,t[i,k]] - d[t[i,k]] + d[t[i,1]])  #
sw[i,k] <-sum(w[i,1:k-1])/(k-1) }                    #
}

d[1]<-0 # → Cálculo da distribuição a priori vaga d
for (k in 2:NT){d[k] ~ dnorm(0,.0001) } #

sd~dunif(0,2) # Cálculo da distribuição a priori vaga sd
tau<-1/pow(sd,2) # Cálculo da variável tau

for (k in 1:NT) { logit(T[k])<- mA +d[k] } # Cálculo da variável T
}

```