

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE MEDICINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA**



**DISSERTAÇÃO DE MESTRADO**

**Construção de ferramenta computacional para estimação de  
custos na presença de censura utilizando o método da  
Ponderação pela Probabilidade Inversa**

Paula Marques Sientchkovski

Orientadora: Prof. Patrícia Klarmann Ziegelmann

Co-Orientadora: Prof. Vanessa Bielefeldt Leotti Torman

Porto Alegre, 06 de abril de 2016.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE MEDICINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA



**DISSERTAÇÃO DE MESTRADO**

**Construção de ferramenta computacional para estimação de  
custos na presença de censura utilizando o método da  
Ponderação pela Probabilidade Inversa**

Paula Marques Sientchkovski

**Orientador: Prof<sup>ª</sup>. Dra. Patrícia Klarmann Ziegelmann**

**Co-Orientadora: Prof<sup>ª</sup>. Dra. Vanessa Bielefeldt Leotti Torman**

A apresentação desta dissertação é exigência do Programa de Pós-graduação em Epidemiologia, Universidade Federal do Rio Grande do Sul, para obtenção do título de Mestre.

Porto Alegre, Brasil.  
2016

## **BANCA EXAMINADORA**

**Prof<sup>a</sup>. Dra. Luciana Neves Nunes**, Departamento de Estatística, Programa de Pós-Graduação em Epidemiologia, Universidade Federal do Rio Grande do Sul.

**Prof<sup>a</sup>. Dra. Carisi Anne Polanczyck, Faculdade de Medicina**, Programa de Pós-Graduação em Epidemiologia, Universidade Federal do Rio Grande do Sul.

**Dra. Jeruza Lavanholi Neyeloff**, Epidemiologista, Hospital de Clínicas de Porto Alegre.

## **AGRADECIMENTOS**

Ao Programa de Pós-Graduação em Epidemiologia da Faculdade de Medicina da Universidade Federal do Rio Grande do Sul, pela oportunidade e aprendizado proporcionado.

À minha família, pelo apoio e compreensão.

Às professoras Patrícia Ziegelmann e Vanessa Torman, pelos ensinamentos, disponibilidade e dedicação.

A Isaías Prestes, pela inestimável colaboração.

À Natália, pela amizade e parceria desde a graduação.

## SUMÁRIO

Abreviaturas e siglas.....	6
Resumo.....	7
1 APRESENTAÇÃO.....	9
2 INTRODUÇÃO.....	10
3 REVISÃO DE LITERATURA.....	12
3.1 Análise de Custo-Efetividade.....	12
3.2 Dados longitudinais de custo e a ocorrência de censura.....	14
3.2.1 Métodos para a estimação de custos na presença de censura.....	15
3.2.2 Comparação entre os métodos de LIN A, LIN B, LIN T, BT, BTp e ZT.....	17
3.3 Estimador Ponderado pela Probabilidade Inversa (IPW).....	19
3.4 Razão de Custo-Efetividade Incremental (ICER).....	21
3.5 Estimação da ICER utilizando dados primários.....	22
3.6 Método bootstrap para estimação da ICER.....	22
3.7 Aplicativo computacional.....	24
4 EXEMPLO DE APLICAÇÃO DO ESTIMADOR IPW.....	25
4.1 Descrição do exemplo.....	25
4.2 Calculando a estimativa IPW para o tratamento alternativo.....	27
4.3 Estimador IPW visto como um método de imputação.....	30
4.4 Entendendo o estimador IPW na prática.....	31
5 OBJETIVOS.....	35
5 REFERÊNCIAS BIBLIOGRÁFICAS.....	36
6 ARTIGO ORIGINAL.....	39
7 CONCLUSÕES E CONSIDERAÇÕES FINAIS.....	53
8 APÊNDICES.....	54

## Abreviaturas e siglas

BCa – *Bias corrected and accelerated* (Viés corrigido e acelerado)

BT – Estimador simples ponderado de caso completo

BTme – Estimador simples melhorado

BTp – Estimador particionado

BTpme - Estimador particionado melhorado

CEA - *Cost-Effectiveness Analysis* (Análise de Custo-Efetividade)

CMA - *Cost-Minimization Analysis* (Análise de Custo-Minimização)

ICER - *Incremental Cost-Effectiveness Ratio* (Razão de Custo-Efetividade Incremental)

HTA - *Health Technology Assessment* (Avaliação de Tecnologias em Saúde)

INB - *Incremental Net Benefit* (Benefício Líquido Incremental)

IPW – *Inverse Probability Weighting* (Ponderação pela Probabilidade Inversa)

RCT – *Randomized Clinical Trial* (Ensaio Clínico Randomizado)

## Resumo

**Introdução:** Dados de custo necessários na Análise de Custo-Efetividade (CEA) são, muitas vezes, obtidos de estudos longitudinais primários. Neste contexto, é comum a presença de censura caracterizada por não se ter os dados de custo a partir de certo momento, devido ao fato de que indivíduos saem do estudo sem esse estar finalizado. A ideia da Ponderação pela Probabilidade Inversa (IPW – do inglês, *Inverse Probability Weighting*) vem sendo bastante estudada na literatura relacionada a esse problema, mas é desconhecida a disponibilidade de ferramentas computacionais para esse contexto.

**Objetivo:** Construir ferramentas computacionais em *software* Excel e R, para estimação de custos pelo método IPW conforme proposto por Bang e Tsiatis (2000), com o objetivo de lidar com o problema da censura em dados de custos. **Métodos:** Através da criação de planilhas eletrônicas em *software* Excel e programação em *software* R, e utilizando-se bancos de dados hipotéticos com situações diversas, busca-se propiciar ao pesquisador maior entendimento do uso desse estimador bem como a interpretação dos seus resultados. **Resultados:** As ferramentas desenvolvidas, ao proporcionarem a aplicação do método IPW de modo intuitivo, se mostraram como facilitadoras para a estimação de custos na presença de censura, possibilitando calcular a ICER a partir de dados de custo. **Conclusão:** As ferramentas desenvolvidas permitem ao pesquisador, além de uma compreensão prática do método, a sua aplicabilidade em maior escala, podendo ser considerada como alternativa satisfatória às dificuldades postas pelo problema da censura na CEA.

**Palavras-chave:** Ferramentas computacionais; Dados de Custos; Censura; Ponderação pela Probabilidade Inversa, Razão de Custo Efetividade Incremental;

## Abstract

**Introduction:** Cost data needed in Cost-Effectiveness Analysis (CEA) are often obtained from longitudinal primary studies. In this context, it is common the presence of censoring characterized by not having cost data after a certain point, due to the fact that individuals leave the study without this being finalized. The idea of Inverse Probability Weighting (IPW) has been extensively studied in the literature related to this problem, but is unknown the availability of computational tools for this context. **Objective:** To develop computational tools in *software* Excel and *software* R, to estimate costs by IPW method, as proposed by Bang and Tsiatis (2000), in order to deal with the problem of censorship in cost data. **Methods:** By creating spreadsheets in Excel *software* and programming in R *software*, and using hypothetical database with different situations, we seek to provide to the researcher most understanding of the use of IPW estimator and the interpretation of its results. **Results:** The developed tools, affording the application of IPW method in an intuitive way, showed themselves as facilitators for the cost estimation in the presence of censorship, allowing to calculate the ICER from more accurate cost data. **Conclusion:** The developed tools allow the researcher, besides a practical understanding of the method, its applicability on a larger scale, and may be considered a satisfactory alternative to the difficulties posed by the problem of censorship in CEA.

**Key words:** Computacional Tools; Costs Data; Censored; Censorship; Inverse Probability Weighting; Incremental Cost-Effectiveness Ratio.



## 1 APRESENTAÇÃO

Este trabalho consiste na Dissertação de Mestrado intitulada “**Construção de ferramenta computacional para estimação de custos na presença de censura utilizando o método da Ponderação pela Probabilidade Inversa**”, apresentada ao Programa de Pós-Graduação em Epidemiologia da Universidade Federal do Rio Grande do Sul, em 06 de abril de 2016.

O trabalho é apresentado em três partes, na ordem que segue:

1. Introdução, Revisão da Literatura, Exemplo e Objetivos.
2. Artigo
3. Conclusões e Considerações Finais.

Documentos de apoio estão apresentados no apêndice.

## 2 INTRODUÇÃO

A Avaliação de Tecnologias em Saúde (HTA - do inglês, *Health Technology Assessment*) tem sido assunto de interesse crescente, seja pelo aumento de oferta em tecnologias nessa área, seja pelos gastos cada vez maiores com a saúde em oposição às condições de restrições de recursos geralmente enfrentadas por indivíduos, empresas e governos. Nos últimos anos, o crescimento acelerado dos gastos em saúde se tornou um grande problema em países desenvolvidos e especialmente nos países em desenvolvimento (Vanni et al. 2009). Em razão do elevado custo de algumas tecnologias de saúde em relação aos seus benefícios percebidos, apresentam-se decisões difíceis a serem tomadas visando à aplicação de recursos nesse contexto. Desse modo, quando diferentes tecnologias em saúde estão sendo comparadas, faz-se necessária a consideração não só de suas efetividades, como de seus custos associados. Este processo completo de avaliação é parte essencial para a tomada racional de decisão sobre o uso eficiente e justificado dos recursos em saúde.

A Análise de Custo-Efetividade (CEA – do inglês, *Cost-Effectiveness Analysis*) é uma das metodologias utilizadas em avaliações econômicas para comparar duas ou mais tecnologias de cuidado com a saúde. A comparação se baseia nos custos (medidos em unidades monetárias) e nas efetividades (medidas em unidades naturais, tais como cura da doença, mortes evitadas, anos de vida ganhos, entre outras) das tecnologias.

As estimativas de custo de um tratamento são, muitas vezes, obtidas através de estudos longitudinais prospectivos. Esses estudos podem ser do tipo observacional – onde o pesquisador apenas acompanha os grupos de pacientes sem impor um tratamento a cada um desses grupos – ou do tipo Ensaio Clínico, onde o pesquisador aloca aleatoriamente os pacientes em cada grupo de tratamento. Em relação a esse último, destaca-se a metodologia *piggyback*, onde uma avaliação econômica é realizada ao se combinar estudos de custo com os Ensaios Clínicos Randomizados (RCT – do inglês, *Randomized Clinical Trial*). Alternativamente, existem ainda os *cost-effectiveness trials*, que são ensaios desenhados expressivamente para propósitos de custo-efetividade, ao invés da eficácia, com poucas restrições adicionais impostas e maior validade externa que os estudos *piggyback* (Gold et al 1996).

Ocorre que, como os custos em estudos longitudinais prospectivos podem ser

coletados uma única ou várias vezes ao longo do tempo, é possível que alguns pacientes não sejam acompanhados até o ponto final de interesse do estudo, podendo ocorrer a censura, nesse caso nomeada de censura à direita, implicando na subestimação do custo. A censura à direita é em geral informativa. Essa ocorre quando o custo total no tempo do evento está positivamente correlacionado com o custo total no tempo de censura (Lin *et al* 1997). Essa correlação ocorre porque os pacientes que acumulam custos a taxas mais elevadas tendem a gerar custos totais maiores em ambos os tempos, de sobrevivência e de censura. A ocorrência da censura informativa impede que a estimação do custo médio seja feita por meio dos métodos tradicionais de análise de sobrevivência (Lin *et al* 1997; Bang e Tsiatis 2000). Para lidar com esta dificuldade, Bang e Tsiatis (Bang e Tsiatis 2000) propuseram um estimador particionado baseado na ideia da Ponderação pela Probabilidade Inversa (IPW –do inglês, *Inverse Probability Weighting*). Embora muitos trabalhos façam a utilização do estimador IPW de Bang e Tsiatis, até onde se tem conhecimento, não há ferramenta computacional disponível para sua aplicação.

Uma maneira para comparar as estimativas de custo e efetividade de duas tecnologias concorrentes, muito utilizada na literatura, é a Razão de Custo-Efetividade Incremental (ICER – do inglês, *Incremental Cost-Effectiveness Ratio*), definida pela razão entre a diferença de custos e a diferença de efetividades. Porém, para estimar a ICER, é preciso conhecer os dados de custo de todos os pacientes, o que nem sempre é possível em razão da censura. Em atenção a esse fato, o estimador IPW oferece a possibilidade de calcular a ICER com dados de custo, mostrando-se como possível alternativa às dificuldades postas pelo problema da censura na Análise de Custo-Efetividade.

Assim, este trabalho tem como principal objetivo a construção de uma ferramenta computacional para estimar custos médios utilizando o estimador IPW de Bang e Tsiatis (2000), através da criação de planilhas eletrônicas em *software* Excel 2016 e programação em *software* R, versão 3.2.2. Ainda, utilizando-se de bancos de dados hipotéticos, com situações diversas, mas análogas a cenários encontrados em bancos de dados reais, busca-se propiciar ao leitor um maior entendimento do uso desse estimador bem como a interpretação dos resultados.

### 3 REVISÃO DE LITERATURA

Nas Seções a seguir são apresentadas considerações sobre CEA e métodos para a estimação de custos na presença de censura, além de comparações entre esses, com ênfase ao método IPW proposto por Bang e Tsiatis (2000). Algumas considerações são também feitas a ICER com utilização de dados primários e pelo método *bootstrap*. Por fim, faz-se referência a aplicativos computacionais que utilizam o método IPW em contexto diverso do da estimação de custo.

#### 3.1 Análise de Custo-Efetividade

O principal objetivo dos estudos de CEA é comparar o valor relativo de diferentes intervenções direcionadas em promover a saúde e /ou prolongar a vida (Polanczyk 2002). Quando duas ou mais tecnologias representam diferentes alternativas de tratamento a um mesmo caso, elas devem permitir a avaliação de um parâmetro equivalente de resultados clínicos, mensurável em unidades comuns a essas intervenções. Nesse contexto, a Análise de Custo-efetividade é uma metodologia estruturada para estabelecer uma comparação baseada nos custos e nas efetividades de duas ou mais tecnologias, e permite identificar qual delas consegue obter o melhor resultado por unidade monetária aplicada (Willan e Briggs, 2006). Esse tipo de análise tem sido enfatizado no contexto de pesquisa de saúde comparativa, bem como se mostrado tópico de pesquisa crescente nos últimos anos<sup>1</sup>.

A CEA se baseia nos custos medidos em unidades monetárias e nas efetividades medidas em unidades naturais experienciadas pelo paciente (cura da doença, mortes evitadas, anos de vida ganhos). Os custos são então comparados com as efetividades. De maneira ilustrativa, pense em um tratamento considerado como a principal opção em um caso de cuidado com a saúde, chamado de tratamento padrão. Pense que um novo tratamento será comparado com o tratamento padrão através de uma CEA. Sabe-se que o tratamento novo pode ser mais caro, menos caro, ou equivalente em custo. De modo análogo, pode ser mais efetivo, menos efetivo, ou equivalente em efetividade, o que

---

<sup>1</sup> Conforme o site PUBMED, houve um aumento duas vezes maior na publicação de artigos sobre o tópico "custo-efetividade" na relação entre os quinquênios 2004-2009 / 2009-2014 e 1999-2004 / 2004-2009.

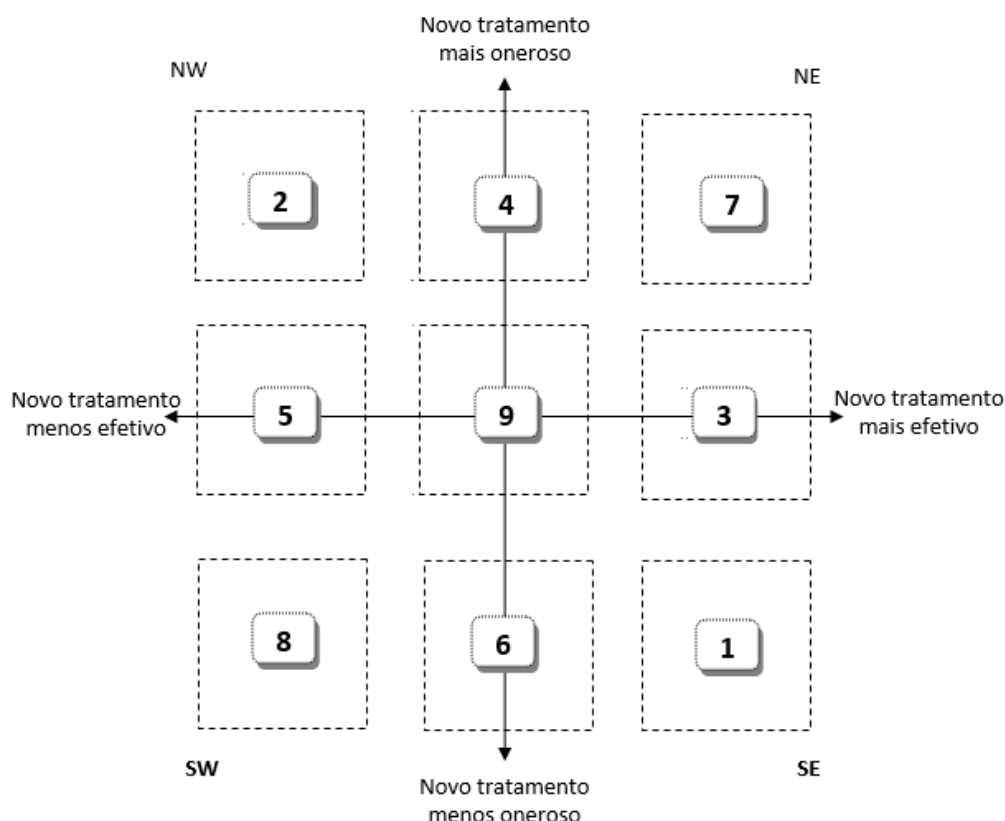
conduz a nove situações possíveis que podem surgir quando os dados de custo e efetividade são comparados (Briggs e O'Brien 2001).

Na matriz apresentada por esses autores, conforme Figura 1, e que é colocada sobreposta por eles ao chamado Plano de Custo-Efetividade, cada um dos seus nove quadros representa uma área que delimita um intervalo de 95% de confiança estimado para a verdadeira diferença entre os custos e as efetividades dos dois tratamentos. Os quadros que cruzam os eixos do Plano de Custo-Efetividade representam uma diferença não significativa ( $\alpha = 0,05$ ) na diferença dos custos e ou das efetividades. Os quadros são então classificados em:

- Casos de dominância forte (Quadro 1 e Quadro 2): quando um tratamento se apresenta significativamente mais efetivo ( $p\text{-valor} < 0,05$ ) e significativamente menos custoso ( $p\text{-valor} < 0,05$ ) do que o outro. Nesse caso a escolha do melhor tratamento é inequívoca;
- Casos de dominância fraca do efeito (Quadro 4 e Quadro 6): quando a diferença de efeito não é estatisticamente significativa ( $p\text{-valor} > 0,05$ ), mas a diferença de custo é ( $p\text{-valor} < 0,05$ ). Nessas situações é indicada uma Análise de Custo-Minimização (CMA -do inglês, *Cost-Minimization Analysis*) convencional, se assume que a diferença nos efeitos é zero e o tratamento menos oneroso é considerado como sendo o tratamento de escolha. Esses estudos raramente são utilizados, uma vez que exigem que diferentes intervenções produzam iguais consequências, de modo que, apenas os custos são comparados (Vanni et al. 2009);
- Casos de dominância fraca do custo (Quadro 3 e Quadro 5): quando a diferença de custo não é significativa ( $p\text{-valor} > 0,05$ ), mas a diferença de efeito é ( $p\text{-valor} < 0,05$ ). Nessas situações a escolha do melhor tratamento é baseada na comparação de efetividade;
- Casos sem dominância (Quadro 9): quando não são observadas diferenças estatisticamente significativas nos custos nem nos efeitos;
- Casos de CEA (Quadro 7 e Quadro 8): quando um tratamento demonstra ser significativamente mais efetivo ( $p\text{-valor} < 0,05$ ) e significativamente mais custoso ( $p\text{-valor} < 0,05$ ). Nesse cenário, a decisão de qual

tratamento escolher depende de quanto se está disposto a pagar (em inglês *willingness to pay*) para se ter maior efetividade.

Nesse contexto, de comparação de resultados e custos de diferentes tratamentos, a Razão de Custo-efetividade Incremental (ICER – do inglês, *Incremental Cost-Effectiveness Ratio*) e o Benefício Líquido Incremental (INB – do inglês, *Incremental Net Benefit*) apresentam-se como as duas estatísticas mais utilizadas, com maior ocorrência da primeira (Briggs *et al*, 1997; Bang e Zhao, 2012; Chen e Zhao, 2013). A estimativa da ICER, por sua vez, é diretamente influenciada pelos custos de um tratamento e esses, quando coletados ao longo do tempo, podem estar sujeitos a presença de censura. A influência da censura na análise dos dados longitudinais de custos e alguns métodos para lidar com esse problema é abordada na próxima Seção.



**Figura 1-** Plano de Custo-Efetividade ilustrando nove possíveis situações que podem surgir quanto à significância das diferenças de custo e efetividade (adaptada de Briggs e O'Brien 2001).

### 3.2 Dados longitudinais de custo e a ocorrência de censura

Estudos longitudinais desempenham um papel de destaque na saúde, sendo indispensáveis para a análise da mudança de um resultado ao longo do tempo. Ao medir

os custos dos participantes do estudo repetidamente ao longo do tempo, é possível realizar a análise direta de mudanças temporais nos indivíduos e os fatores que influenciam essa mudança (Crowder 2009). No entanto, a análise de dados de custo pode ser desafiadora em razão das complexidades introduzidas pela censura (Başer *et al* 2006), a qual é comum nesse tipo de estudo (Jaffa *et al* 2014). Essa censura é geralmente denominada de censura à direita e se divide em três tipos: *a*) Censura do tipo I, que ocorre quando os indivíduos são acompanhados até um período de tempo pré-estabelecido; *b*) Censura do tipo II, quando os indivíduos são acompanhados até se obter um número pré-determinado de eventos e *c*) Censura do tipo III, que é a censura aleatória à direita e acontece quando um indivíduo é perdido no decorrer do estudo sem que tenha ocorrido o evento, ou se indivíduo morre por uma razão não relacionada ao estudo. Os tipos I e III são o foco deste trabalho.

A censura pode ainda ser classificada como informativa e não informativa. A censura não informativa ocorre quando não há razão para suspeitar que o motivo da perda de informação ocorreu em virtude do evento, ou seja, ela ocorre ao acaso. A censura informativa, por outro lado, ocorre quando o indivíduo se afasta do estudo por motivo associado ao evento, como no caso em que o abandono do tratamento está relacionado com a piora do indivíduo (Carvalho *et al* 2011).

### **3.2.1 Métodos para a estimação de custos na presença de censura**

Foi de Lin *et al* a primeira tentativa realizada com o objetivo de incorporar adequadamente o efeito da censura informativa na estimativa de custos médios. Os autores propuseram três estimadores, denominados Lin A, Lin B e Lin T (Lin *et al* 1997). Os dois primeiros utilizam a história de custo dos pacientes, característica que os torna mais eficientes que o terceiro, o qual faz uso apenas do custo total do período de cada paciente. Por sua vez, a diferença entre Lin A e Lin B está relacionada com a consistência dos estimadores em relação à posição de ocorrência da censura. Lin A é consistente quando a censura ocorre no fim do intervalo de tempo, e Lin B é consistente quando ela ocorre no interior e no início do intervalo de tempo.

Partindo do pressuposto de que a consistência dos estimadores Lin A/B só é garantida sob a suposição de que a censura ocorre discretamente em cada intervalo, Bang e Tsiatis propuseram outros quatro estimadores (Bang e Tsiatis 2000). Todos eles utilizam o conceito de ponderação pela probabilidade inversa, usado pela primeira vez

por Horvitz e Thompson no contexto de amostragem com probabilidades desiguais (Horvitz e Thompson 1952). Esses quatro estimadores são consistentes e foram nomeados como: estimador simples ponderado de caso completo (BT), estimador particionado (BTp), estimador simples melhorado (BTme) e estimador particionado melhorado (BTpme). Os estimadores BTme e BTpme são estimadores semi-paramétricos e foram criados para melhorar a eficiência dos estimadores BT e BTp, respectivamente. Porém, são pouquíssimo utilizados devido a sua maior complexidade e difícil aplicação, decorrente da falta de código computacional disponível, tornando o seu uso quase inacessível (Pullenayegum e Willan 2011). Além disso, quando a preocupação do estudo estava voltada para o critério da eficiência das estimativas, o estimador particionado foi o mais utilizado, devido a sua menor complexidade em relação aos estimadores BTme e BTpme (Pullenayegum e Willan 2011).

Posteriormente, Zhao e Tian propuseram o estimador ZT, considerado por eles como mais eficiente do que BT (Zhao e Tian 2001), já que esse usa as informações da história de custo de observações tanto completas como incompletas, e mais conveniente do que BTme, pois não necessita que o histórico de custos seja particionado em subintervalos. Porém, quando a história de custo está disponível, esses autores sugerem a utilização do estimador BTp, proposto por Bang e Tsiatis.

Outra possibilidade de utilização da ideia da ponderação pelo inverso da probabilidade, a qual não é objeto desse estudo, é a sua associação a modelos de regressão para estimação dos custos na ocorrência de censura. O raciocínio utilizado parte do fato de que, na análise de regressão, um modelo preditivo é ajustado aos dados, e então se usa esse modelo para prever valores da variável dependente (VD) a partir de uma ou mais variáveis independentes (VIs). Quando o problema se restringe a apenas uma variável independente, temos o caso de regressão simples. Quando as variáveis independentes são duas ou mais, estamos diante de um caso de regressão múltipla (Field 2009). Desse modo, o modelo de regressão é utilizado na estimação dos custos a fim de que as variáveis independentes, cujos valores são conhecidos, sejam utilizadas para prever os custos (variável dependente) de um determinado tratamento. Contudo, esses dados de custo também estão sujeitos ao problema posto pela ocorrência de censura, e é nesse contexto que alguns autores têm aliado as técnicas de regressão ao método IPW.

Lin, utilizando a ideia da ponderação das observações completas pela probabilidade inversa, modificou de várias formas as equações de mínimos quadrados, a



fim de prever adequadamente os dados faltantes (Lin 2000). Mais tarde, levando em consideração a assimetria dos custos, Lin combinou o método da ponderação pela probabilidade inversa com a abordagem de equações de estimação generalizadas para dados longitudinais na presença de censura, a fim de dar conta da assimetria dos dados (Lin 2003). Willan *et al* propuseram usar o mesmo método para fornecer modelos de regressão para o ajuste das covariáveis e análise de subgrupo na análise de custo-efetividade na presença de censura (Willan *et al* 2005). Baser *et al* aplicaram o IPW no método dos mínimos quadrados para avaliar os efeitos das covariáveis em dados de custos médicos com presença de censura (Başer *et al* 2006). Tian e Huang propuseram um modelo flexível de duas partes (um modelo paramétrico, de regressão logística para a probabilidade de ter custos positivos na primeira parte e um modelo de regressão linear para os custos positivos na segunda parte) e usaram o método de ponderação pela probabilidade inversa para analisar os custos médicos totais submetidos à censura informativa à direita (Tian e Huang 2007). Por sua vez, Griffiths *et al* utilizaram o modelo de regressão de mínimos quadrados associado à ponderação IPW (Griffiths *et al* 2012).

### **3.2.2 Comparação entre os métodos de LIN A, LIN B, LIN T, BT, BTP e ZT.**

Diferentes estudos foram dedicados à comparação dos estimadores propostos por Lin *et al* e Bang e Tsiatis. Esses últimos autores simularam condições de censura leve (25-35%) e pesada (40-45%) para comparar seus estimadores entre si e com o estimador Lin B, e consideraram ele como o melhor estimador entre os propostos por Lin *et al* (Bang e Tsiatis 2000). Com censura leve, todos os estimadores propostos por Bang e Tsiatis, bem como o estimador Lin B, tiveram desempenhos semelhantes. O erro padrão empírico foi utilizado como um indicador para a eficiência e sugeriu que o BTP, BTme e BTPme foram ligeiramente mais eficientes do que o estimador BT e Lin B. Por sua vez, em todas as simulações, as eficiências de BT e Lin B foram muito semelhantes, enquanto que o estimador BTP obteve os melhores resultados nesse aspecto. Com censura pesada, os vieses de todos os estimadores tiveram aumento moderado, mas o estimador BTP continuou apresentando os melhores resultados.

Raikou e McGuire utilizam um conjunto de dados reais de um ECR de dois braços, ambos sob condições extremas de censura (82%), com o objetivo de comparar dois dos três estimadores propostos por Lin *et al*, no caso Lin B e Lin T, com os quatro

estimadores propostos por Bang e Tsiats (BT, BTP, BTme e BTPme) (Raikou e McGuire 2004). Os estimadores BTme e BTPme foram afetados pelo alto nível de censura e não apresentaram bons resultados, sendo os melhores resultados apresentados pelos estimadores Lin B e BTP, apresentando estimativas com magnitudes coerentes e erros padrões razoáveis. No mesmo estudo, os autores avaliaram novamente os estimadores acima sob as seguintes condições: *a)* utilizaram os mesmos dados, mas excluíram valores de custos extremos, *b)* usaram os mesmos dados, mas variaram o tempo de duração das análises e *c)* usaram um conjunto de dados artificial, construído para gerar custos e tempos de sobrevivência aleatórios, variando os níveis de censura. Os resultados para os tipos de dados, reais e simulados, foram semelhantes entre si, e os estimadores Lin B e BTP continuaram apresentando os melhores resultados.

Numa tentativa de relacionar teoricamente os estimadores propostos por Lin *et al* e Bang e Tsiatis, O'Hagan e Stevens examinaram as propriedades matemáticas desses estimadores sob dois diferentes cenários (O'Hagan e Stevens 2004). No primeiro, denominado pelos autores como cenário de dados de custo mínimo, Lin T e BT foram comparados quando apenas os custos totais estavam disponíveis para cada paciente até o momento de sua morte, censura ou tempo final do estudo. Nesses casos, os autores recomendam a utilização de BT, já que os estimadores demonstraram ser equivalentes. No segundo, denominado como cenário de dados de custo intervalar, e onde os custos acumulados dos pacientes são observados para cada período de tempo fixo, os estimadores LinB e BTP foram comparados, sendo o estimador BTP recomendado pela sua consistência.

Young, por sua vez, simulou uma série de mecanismos de censura em uma coorte completa de pacientes submetidos a transplante de fígado, e comparou o desempenho de nove métodos para estimar os custos totais médios e erros padrão na presença de censura, entre eles os métodos de Lin e Bang e Tsiatis, ambos com e sem história do custo, sendo o estimador BTP considerado como o método mais preciso para prever os custos totais médios e erros padrão (Young 2005).

Em outro estudo, Zhao *et al* estabeleceram, sob certas condições de partições, a equivalência matemática dos estimadores ZT, Lin A e Lin B e BTP, bem como entre Lin T e BT (Zhao *et al* 2007). Utilizando um número finito de intervalos de tempo, os autores destacam que quando os tempos de censura são escolhidos como limites de partição, os estimadores de Lin T e BT se tornam idênticos, bem como os estimadores

Lin A, Lin B, BTP, e ZT.

Em resumo, os principais estimadores para a estimação de custos na presença de censura foram Lin A, Lin B e BTP (O'Hagan e Stevens 2004). Na literatura, os estimadores Lin A/B são também conhecidos como estimador Kaplan-Meier de Média Amostral (KMSA- *Kaplan-Meier survival analysis*), enquanto o estimador BTP é também conhecido como estimador Ponderado pela Probabilidade Inversa (IPW – *Inverse Probability Weithing*), sendo essa a abordagem mais geral e adotada por diversos pesquisadores (Griffiths *et al* 2012), o que explica a nomenclatura adotada na continuidade deste trabalho.

### 3.3 Estimador Ponderado pela Probabilidade Inversa (IPW)

Nesta seção serão apresentados dois dos quatro estimadores baseados na ideia da ponderação pela probabilidade inversa, conforme proposto por Bang e Tsiatis (Bang e Tsiatis 2000): BT e BTP. O estimador BT precede o estimador BTP (IPW), sendo o último uma versão particionada a partir do primeiro, razão pela qual são os dois apresentados a seguir. Seguem algumas notações necessárias.

Considere que o tempo máximo de seguimento de um paciente é  $\tau$ . Utilize  $X_i$  para indicar o tempo observado de seguimento do paciente  $i$ , sendo  $X_i = \min(T_i, C_i)$ , onde  $T_i$  indica que o tempo observado é o tempo de ocorrência do evento de interesse e  $C_i$  indica que o tempo observado é o tempo de censura. Assume-se que  $T$  e  $C$  nunca ocorrem simultaneamente. Ou seja, para os pacientes cujo evento de interesse foi observado, o tempo de seguimento é denotado  $T_i$ . E para os pacientes onde não se observou o evento, o tempo é  $C_i$ . Denote  $M(u)$  como a história de custo acumulado até o tempo  $u$  e  $M_i(X_i)$  o custo final para cada indivíduo. Ainda, utilize  $K(u) = \Pr(C > u)$ , onde  $0 < u < \tau$ , para representar a probabilidade de um indivíduo não ser censurado até o tempo  $u$ . Então, para todo paciente  $i$  ( $i=1, \dots, n$ ), os dados observados serão  $\{X_i, \Delta_i, M_i\}$ , onde  $\Delta_i$  é a função indicadora, ou seja,  $\Delta_i = 0$  para pacientes censurados e  $\Delta_i = 1$  para pacientes não censurados.

**Estimador simples ponderado de caso completo (BT):** este estimador utiliza apenas o custo total de um indivíduo, que é dado por:

$$\hat{\mu}_{BT} = \frac{1}{n} \sum_{i=1}^n \frac{\Delta_i M_i}{\hat{K}(T_i)}$$

onde  $\hat{K}(\cdot)$  é o estimador de Kaplan-Meier (Kaplan & Meier, 1958) para  $K(\cdot)$ , com os papéis de tempo de censura  $C_i$  e tempo de sobrevivência  $T_i$  invertidos. Com os papéis invertidos, o estimador de KM fornece uma estimativa para a sobrevivência à censura, ou seja, a probabilidade de não ser censurado até um tempo  $u$ . Lembre-se que pacientes censurados tem custo observado até o tempo de censura. Note, porém, que o estimador BT desconsidera completamente qualquer custo observado para pacientes censurados. Para incorporar os custos observados até tempo de censura e, desta forma, melhorar a estimativa do custo total médio, foi criada uma versão particionada do estimador BT chamada, no artigo seminal de Estimador Particionado (Bang e Tsiatis 2000) ou, em literatura mais recente, de IPW (Gray *et al* 2010a).

**Estimador particionado (BTp) ou Estimador Ponderado pela Probabilidade Inversa (IPW):** o estimador particionado utiliza a história de custo dos pacientes e deriva o custo estimado incorrido em cada um dos  $K$  subintervalos  $(t_j, t_{j+1}] (j = 0, \dots, K-1)$  de  $(0, \tau]$ , a partir do estimador simples ponderado. Note que esses intervalos são definidos para serem fechados à direita. O estimador particionado é dado por

$$\hat{\mu}_{BTp} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^K \frac{\Delta_i^j \{M_i(t_j) - M_i(t_{j-1})\}}{\hat{K}(T_i^j)} = \frac{\Delta_i^i M_{ij}}{\hat{K}(T_i^j)}.$$

Então, para o  $i$ -ésimo indivíduo,  $\Delta_i^j = \{\min(T_i, t_j) \leq C_i\}$ ,  $M_i(t_j)$  é o custo cumulativo até o momento  $t_j$ , e  $\hat{K}_j(T_i^j)$  é o estimador de Kaplan-Meier para a censura com base no conjunto de dados  $\{X_i^j, \Delta_i^j, i = 1, \dots, n\}$  onde  $X_i^j = \min(T_i^{t_j}, C_i)$ . Para facilitar o entendimento da notação, considere  $M_i(t_j) - M_i(t_{j-1}) = M_{ij}$ .

A eficiência e a normalidade assintótica desses estimadores são demonstradas detalhadamente por Bang Tsiatis, e a sua vantagem está no fato dele considerar todos os dados de custo observados, mesmo que estes sejam oriundos de indivíduos censurados (Bang e Tsiatis 2000).

Desse modo, como mencionado anteriormente, se a estimativa da ICER é diretamente influenciada pelos custos de um tratamento, e esses são passíveis de censura, o método IPW se coloca como possível alternativa para a obtenção de uma ICER mais precisa. A definição da ICER, a sua estimação utilizando dados primários, bem como pelo método *bootstrap*, são abordadas nas próximas Seções.

### 3.4 Razão de Custo-Efetividade Incremental (ICER)

Considere o cenário onde duas tecnologias (A e B) para o tratamento de uma determinada doença estão sendo comparadas, ou seja, uma intervenção alternativa *versus* uma intervenção padrão ou uma intervenção alternativa *versus* nenhuma intervenção. A ICER é, então, definida como a razão entre a diferença dos custos de A e B e a diferença de suas efetividades. Desta forma, tem-se:

$$ICER = \frac{C_A - C_B}{E_A - E_B}, \quad (1)$$

onde  $C_A$  e  $C_B$  representam, respectivamente, o custo médio do tratamento A e do tratamento B, e  $E_A$  e  $E_B$  representam as efetividades médias dos tratamentos A e B.

A ICER é, portanto, o custo adicional de realização de uma medida de efetividade do uso de um tratamento em relação a outro (Willan e Briggs 2006), quantificando os balanços entre os resultados alcançados pelos pacientes e os recursos gastos, onde um custo maior geralmente implica um melhor resultado. Assim, se a medida de efetividade é a probabilidade de sobrevivência, a ICER é o custo de salvar uma vida (ou prevenir uma morte); se a medida da efetividade é a média de sobrevivência ou a média de sobrevivência ajustada pela qualidade, então a ICER é o custo de alcançar um ano a mais de vida ou de qualidade ajustada de vida (Willan e Briggs 2006). Em resumo, a ICER pode ser interpretada como o custo necessário para uma unidade a mais de efetividade.

### 3.5 Estimação da ICER utilizando dados primários

O verdadeiro valor populacional da ICER é desconhecido e tem sido estimado de duas maneiras distintas, dependendo dos dados de custo e de efetividade disponíveis. O padrão-ouro para estimativas de efetividade são os ECRs. Em geral, esses são delineados para avaliar a eficácia, mas estudos econômicos podem estar aninhados a esses ensaios clínicos, sendo então chamados de estudos *piggyback* (Gold *et al* 1996). Nessas situações, dados primários de custo e efetividade estão disponíveis e podem ser utilizados na estimação da ICER. A outra maneira de estimação da ICER é baseada em dados oriundos de modelagem (Drummond 2005). Neste trabalho apenas o primeiro modo de estimação é abordado.

A ideia básica da estimação da ICER é utilizar os dados de custo e efetividade oriundos de duas amostras independentes (amostra de pacientes tratados com A e amostra de pacientes tratados com B). Nesse caso, as médias dos custos e das efetividades de cada tratamento são estimadas pelas médias amostrais e o estimador pontual da ICER é dado por

$$\hat{ICER} = \frac{\bar{C}_A - \bar{C}_B}{\bar{E}_A - \bar{E}_B} = \frac{\Delta \bar{C}}{\Delta \bar{E}} \quad (2)$$

onde  $\bar{C}_A$  e  $\bar{E}_A$  são, respectivamente, a média amostral dos custos e a média amostral das efetividades no grupo de tratamento alternativo, e  $\bar{C}_B$  e  $\bar{E}_B$  são, respectivamente, a média amostral dos custos e das efetividades do tratamento padrão.

Sabe-se que a distribuição amostral da ICER é desconhecida (Briggs *et al* 1997). Dessa forma, técnicas paramétricas para estimar intervalos de confiança devem ser evitadas. Uma abordagem alternativa que tem sido muito utilizada é o método *bootstrap* não paramétrico (Briggs *et al* 1997).

### 3.6 Método *bootstrap* para estimação da ICER

O método *bootstrap* é uma técnica computacional intensiva, que possibilita que a ICER seja construída empiricamente e realiza a avaliação da incerteza CEArca das diferenças nas medidas estimadas de custo e efetividade. Sua aplicação, para o caso de uma única amostra, consiste em selecionar aleatoriamente  $n$  valores com reposição da amostra de dados originais. O fato da amostragem ser com reposição, permite que uma

observação da amostra original seja selecionada mais de uma vez para a nova amostra, nomeada de amostra *bootstrap*. Esse processo de amostragem é repetido inúmeras vezes, gerando  $b$  amostras *bootstrap*, com base nas quais se calcula a estatística de interesse. O número de reamostras requeridas depende da aplicação para a qual ela será utilizada, mas tipicamente  $b$  deve ser de pelo menos 1000 quando a reamostragem for aplicada para a construção de intervalos de confiança (Briggs *et al* 1997; Mooney *et al* 1993).

A validade do método *bootstrap* é baseada em duas suposições: 1) à medida que o tamanho original da amostra se aproxima do tamanho da população, a distribuição da amostra tende para a distribuição da população; 2) à medida que o número de replicações de amostras *bootstrap* tende ao infinito, a aproximação de *bootstrap* da distribuição amostral de uma estatística se aproxima da verdadeira distribuição amostral (Mooney *et al* 1993).

A aproximação de *bootstrap* da distribuição amostral da ICER é obtida de forma muito semelhante a do caso de uma única amostra, citado anteriormente. No entanto, uma vez que a ICER é estimada com base em quatro estatísticas (duas para o custo e duas para a efetividade, a partir de duas amostras independentes de custos e efetividades), cuidados devem ser tomados para a aplicação do *bootstrap* em cada uma dessas amostras (Efron e Tibshirani 1994; Chaudhary e Stearns 1996; Willan e O'Brien 1996; Briggs *et al* 1997; J L Severens 1999).

Para que seja possível gerar uma distribuição de *bootstrap* para uma ICER, os seguintes passos devem ser seguidos (assumindo-se que há  $n_A$  pacientes no grupo de tratamento alternativo e  $n_B$  no grupo de tratamento padrão):

1. Gerar uma amostra com reposição de  $n_A$  pares de custo/efetividade a partir dos dados do grupo experimental cuja amostra tem tamanho  $n_A$ ;
2. Gerar uma amostra de  $n_B$  pares de custo/efetividade a partir dos dados do grupo controle com substituição;
3. Calcular a estimativa para a ICER para a reamostra *bootstrap*;
4. Repetir esse procedimento  $b$  vezes.

As  $b$  estimativas de ICER formam a distribuição amostral empírica da ICER.

Com base nesses valores se estima a ICER pontualmente e por intervalo de confiança. A estimativa pontual é a média entre as estimativas das  $b$  amostras. Para a estimativa por intervalo de confiança, diferentes métodos são citados na literatura, sendo os quatro mais comumente utilizados os seguintes: aproximação normal, percentil, percentil-t e método do viés corrigido e acelerado. A escolha do melhor método é, no entanto, específica ao contexto de sua aplicação (Briggs *et al* 1997). Assim, uma vez que o estimador da ICER ( $\hat{ICER}$ ) é um estimador tendencioso, onde o viés e assimetria da distribuição precisam ser ajustados, o método mais indicado é o do viés corrigido e acelerado (BCa - *bias-corrected and accelerated*) criado por Efron, em 1987 (Briggs *et al* 1997). Esse estimador consiste em uma modificação do método do percentil, envolvendo ajustes algébricos para percentis selecionados, com o objetivo de servir como pontos de extremidade nos intervalos de confiança (Briggs *et al* 1997). O método BCa, bem como os outros três métodos citados anteriormente, encontram-se implementados no pacote “boot” do *software* R (Canty e Ripley 2016).

### 3.7 Aplicativo computacional

A idéia da ponderação pela probabilidade inversa tem sido estudada e apresentada sob vários contextos, mas são muito poucos os aplicativos computacionais desenvolvidos e disponibilizados aos interessados em aplicar a técnica. Até onde se tem conhecimento, apenas dois programas computacionais estão disponíveis na literatura, e nenhum deles está aplicado à estimação de custos. Um deles é o pacote nomeado de “*ipw*” e implementado no *software* R. Utiliza a ponderação pela probabilidade inversa para ajustar modelos estruturais marginais, com o objetivo de estimar os efeitos causais em uma análise de dados (Geskus e van der Wal, 2011). O outro é uma macro implementada no *software* estatístico Stata, nomeado de “*teffects ipw*” (Cattaneo 2010). Consiste na estimação dos efeitos do tratamento em dados observacionais, via ponderação pela probabilidade inversa, com o objetivo de corrigir o problema resultante de dados faltantes.



## 4 EXEMPLO DE APLICAÇÃO DO ESTIMADOR IPW

Nesta seção, um exemplo é apresentado com os passos necessários para a estimação do custo médio de um tratamento utilizando o estimador IPW. Na Seção 4.1 o cenário do exemplo e a descrição do banco de dados utilizado são descritos. O banco de dados utilizado é um banco hipotético com dados criados pelo pesquisador de modo a replicar cenários comumente encontrados em bancos de dados reais. Na Seção 4.2, a aplicação do IPW é demonstrada através da construção de uma tabela, que tem como resultado o custo médio estimado do tratamento, bem como os custos ponderados para cada período de tempo. Na Seção 4.3, é mostrado como, a partir desses últimos resultados, os valores de custo dos indivíduos censurados são recuperados e imputados ao banco de dados. Por fim, na Seção 4.4, essa imputação é explicada detalhadamente, através de conjuntos de dados hipotéticos.

### 4.1 Descrição do exemplo

Suponha que um ECR com coleta *piggyback* de dados de custo é conduzido para comparar dois tratamentos: P (tratamento padrão) e A (tratamento alternativo). O evento de interesse é a morte (desfecho binário). Pacientes são aleatorizados em um dos dois tratamentos e acompanhados desde o início do tratamento até a morte ou por um período de meio ano ( $\tau = 6$  meses), o que ocorrer primeiro. Dados de custos são computados mensalmente ( $t_j$ , para  $j=1, 2, \dots, 6$ ). Os dados apresentados na Tabela 1 representam os custos observados de 10 pacientes ( $n=10$ ) submetidos ao tratamento alternativo. Os custos de cada paciente estão dispostos nas linhas da tabela e os períodos de tempo estão representados nas colunas ( $t_1 - t_6$ ). A última coluna apresenta o custo total acumulado ( $CT = M_i(t_6)$ ) de cada paciente durante os 6 meses de estudo. A última linha apresenta a soma e a média dos custos para cada período de tempo e para o custo total.

Note que alguns pacientes foram retirados do estudo a partir de certos períodos de tempo, de modo a simular situações de morte ou de censura. Se antes de seis meses o paciente for a óbito, isso é considerado um evento (e); caso o paciente abandone o estudo, ou não morra até o fim dos seis meses, isso é considerado como censura (c).

Por exemplo, os custos do paciente 2, o qual foi censurado, estão disponíveis apenas por quatro meses ( $t_1$  até  $t_4$ ), tendo uma história de custo acumulada de

R\$10.466,00 reais ( $M_2(t_4) = 10.466,00$ ). Ou seja, sabe-se que o paciente 2 teve um custo total de R\$10.466,00 por quatro meses de tratamento. Os pacientes 6 e 9 também foram censurados. O paciente 3 vem a representar o primeiro caso de morte com apenas um mês de acompanhamento, gerando um custo total acumulado de R\$ 4.253,00 reais ( $M_3(t_1)$ ). Os pacientes 5 e 10 também morreram durante o estudo, mas com tempos de acompanhamento de três e quatro meses, respectivamente.

**Tabela 1-** Dados de custo de um ECR hipotético

Paciente	t1	t2	t3	t4	t5	t6	CT
1	3140	2990	2890	2500	1788	1330	14638
2	2600	2286	2740	2840	c		10466
3	4253	e					4253
4	2436	2040	1788	1411	1314	1099	10088
5	3325	3540	3800	e			10665
6	2070	2870	c				4940
7	2960	2630	2898	2936	2500	1903	15827
8	2360	1200	2700	1630	1460	1780	11130
9	2600	3560	2200	1990	3600	c	13950
10	2870	2414	3790	4005	e		13079
<b>Soma</b>	<b>28614</b>	<b>23530</b>	<b>22806</b>	<b>17312</b>	<b>10662</b>	<b>6112</b>	<b>109036,00</b>
<b>Média</b>	<b>2861,4</b>	<b>2614,4444</b>	<b>2850,75</b>	<b>2473,1429</b>	<b>2132,4</b>	<b>1528</b>	<b>10903,60</b>

Ainda, a soma no tempo  $t_1$  é igual a 28.614,00 e representa o custo total observado até o tempo  $t_1$  (1 mês de tratamento) dos 10 pacientes da amostra. A média igual a 2.861,40 (28.614,00/10) representa o custo médio por paciente por um mês de tratamento. Analogamente, o custo médio do tratamento alternativo por paciente, por um período de seis meses, é de R\$10.903,60 ( $\frac{1}{n} \sum_{i=1}^{10} \sum_{j=1}^6 M_i(t_j) - M_i(t_{j-1}) = 10.903,60$ ).

Note que as médias foram calculadas considerando que todos os pacientes que foram a óbito e também os que foram censurados tiveram custo zero depois que saíram do estudo. Este raciocínio de custo zero é verdadeiro para os pacientes que foram a óbito, mas não para àqueles que foram censurados. Ou seja, o custo médio calculado desta maneira deve subestimar o verdadeiro custo médio, visto que, apesar de não se conhecer a real situação dos pacientes censurados, é provável que esses permaneceram gerando custos por, pelo menos, algum tempo.

## 4.2 Calculando a estimativa IPW para o tratamento alternativo

Nesta seção será explicado como calcular a estimativa do custo médio total de um tratamento quando os dados de custo observados estiverem no formato apresentado na Tabela 1 e for utilizado o estimador BTP. Os resultados gerados por cada passo do cálculo, listados abaixo, são apresentados na Tabela 2.

1. *Passo 1 - Período de tempo:* Primeiramente, deve-se analisar de que forma os dados foram coletados (mensalmente, anualmente, etc) para, então, definir os períodos de tempo do estudo. O primeiro período de tempo compreende o tempo entre  $t_0$  (entrada do paciente no estudo) e  $t_1$  (primeiro tempo de observação de dados de custo do estudo). Na Tabela 1, os dados foram coletados mensalmente por um período de estudo de 6 meses. Então, para fins de cálculos, são considerados sete pontos de tempo ( $t_0$  até  $t_6$ ). Esses tempos irão compor a primeira coluna da Tabela 2 e o número de linhas da tabela será igual a quantidade de tempos (no exemplo, igual a 7).
2. *Passo 2 - Indivíduos em risco:* A segunda coluna da Tabela 2 apresenta o número de indivíduos em risco (de ser censurado ou ter o evento de interesse) no início de cada período de tempo. O número de indivíduos em risco nos tempos  $t_0$  e  $t_1$  serão sempre iguais ao tamanho da amostra, visto que no início do estudo todos os pacientes estão sob o risco de sofrerem o evento ou de serem censurados. O número desses indivíduos em risco, na medida em que os pacientes forem sofrendo o evento ou sendo censurados, diminui gradativamente.
3. *Número de pacientes que sofrem evento ou censura –* O número de pacientes que sofrem o evento ou são censurados compõem a terceira e a quarta coluna da Tabela 2. A quantidade de eventos ou de censuras deve ser registrada no período de tempo anterior a ocorrência dos mesmos. Por exemplo, os pacientes 2 e 10 da Tabela 1 compareceram a todas as consultas até o quarto mês. Logo, seus custos não estão disponíveis no período de tempo  $t_5$  da Tabela 1, onde se observa uma censura e uma morte, respectivamente. Contudo, se observarmos a Tabela 2, a censura e morte desses pacientes estão computadas no tempo  $t_4$  (linha 5) e não no período de tempo  $t_5$ . Ou seja, para efeito de

construção da tabela, considera-se que a censura está vinculada ao seu último registro de custo disponível no banco de dados.

4. *Taxa de risco ( $H(t)$ )* – A taxa de risco representa o número de indivíduos censurados ao longo do período, dividido pelo número de indivíduos em risco no início do período. Esse valor estima a probabilidade de censura no período. Por exemplo, para o tempo  $t_5$ , tem-se:

$$H(t_5) = \frac{\text{n}^\circ \text{ de censuras}(t_5)}{\text{n}^\circ \text{ de indivíduos em risco}(t_5)} = \frac{1}{5} = 0,2$$

Logo, com base nos dados observados, estima-se que a probabilidade de um indivíduo que ainda estava em risco no tempo  $t_5$  ser censurado entre os tempos  $t_5$  e  $t_6$  é de 0,20. Ou seja, estima-se que 20% dos indivíduos que ainda estão sob risco no tempo  $t_5$  vão ser censurados até o tempo  $t_6$ .

5. *Probabilidade de não ser censurado ( $P(t)$ )* - Essa probabilidade é resultado de uma unidade menos a taxa de risco, ou seja,  $1 - H(t)$ . Então, como a taxa de risco para o tempo  $t_5$  é de 20%, a estimativa para a probabilidade de não ser censurado será de 0,8, como se segue:

$$P(t_5) = 1 - H(t_5) = 1 - 0,2 = 0,8$$

Logo, com base nos dados observados, estima-se que a probabilidade de um indivíduo, que ainda estava em risco no tempo  $t_5$ , não ser censurado entre os tempos  $t_5$  e  $t_6$  é de 0,80. Ou seja, estima-se que 80% dos indivíduos que ainda estão sob risco no tempo  $t_5$  continuarão sob risco no tempo  $t_6$ .

6. *Probabilidade total de não ser censurado ( $K(t)$ )* – Essa probabilidade se trata de uma probabilidade acumulada, e sua fórmula é dada por  $K(t_j) = P_i(t_{j-1}) * K_i(t_{j-1})$ . Considerando o tempo  $t_5$  esta probabilidade será:

$$K(t_5) = P(t_4) * K(t_4) = 0,857 * 0,888 = 0,7619$$

Logo, com base nos dados observados, estima-se que a probabilidade total de não ser censurado até o tempo  $t_5$  é de 0,7619. Ou seja, estima-se que 76,19% dos indivíduos do estudo ainda estejam sob risco no tempo  $t_5$ .

7. *Custo total bruto* – o custo total bruto é calculado para cada período de tempo e consiste na soma de todos os custos disponíveis no período. Considerando novamente o período de tempo  $t_5$ , o custo total para esse período é:

$$\begin{aligned}\sum_{i=1}^{10} M_i(t_5) &= M_1(t_5) + M_4(t_5) + M_7(t_5) + M_8(t_5) + M_9(t_5) \\ &= 1788 + 1314 + 2500 + 1460 + 3600 = 10662\end{aligned}$$

Logo, a soma de todos os custos disponíveis no período  $t_5$  é de R\$ 10.662,00, o que representa o total observado de custos no quinto mês de tratamento. Observe que os pacientes que foram a óbito ou foram censurados no quinto mês contribuem com zero nesta soma.

8. *Custo total ponderado* – Corresponde à penúltima coluna da Tabela 2 e é o resultado da divisão do custo total do tempo  $t$  pela probabilidade total de não ser censurado, como se segue

$$IPW_5 = \frac{10662}{0,76190} = 13993,88$$

Então, o custo total para o período de tempo  $t_5$ , ponderado pelo estimador IPW, é de R\$13.993,88 reais. Observe que o custo total ponderado é um pouco maior que o custo total bruto, representando que a ponderação IPW considera custos diferentes de zero para os pacientes que foram censurados no período. Ainda, quanto maior a probabilidade total de não ser censurado, menor será a diferença entre o custo total bruto e o custo total ponderado.

9. *Custo médio final* – Corresponde a última coluna da tabela e é resultado da soma de todos os custos totais ponderados, dividido pelo número de indivíduos que participaram do estudo, ou seja, que estão em risco no período de tempo  $t_0$ , como se segue

$$\overline{IPW}_{\text{final}} = \frac{\sum_{j=1}^6 IPW_j}{n} = \frac{2861,40 + 23530 + 25656,75 + 19476 + 13993,88 + 10027,50}{10} = 12129,81$$

Logo, estima-se que o custo médio final por paciente para o tratamento alternativo, por um período de tempo de seis meses, é de R\$ 12.129,81. Essa é a estimativa pontual, mas para se obter alguma medida de erro dessa

estimativa é necessário calcular o intervalo de confiança para esse custo médio.

**Tabela 2** - Detalhamento dos cálculos para estimação dos custos pelo método IPW

Período	Nº risco	Evento	Censura	H(t)	P(t)	K(t)	Custo total bruto	Custo total ponderado	Custo Médio Final
0	10	0	0	0	1	1	0	0	12129,81
1	10	1	0	0	1	1	28614	28614,00	
2	9	0	1	0,11111	0,88889	1	23530	23530,00	
3	8	1	0	0	1	0,888	22806	25656,75	
4	7	1	1	0,14286	0,85714	0,888	17312	19476,00	
5	5	0	1	0,2	0,8	0,762	10662	13993,88	
6	4	0	4	1	0	0,609	6112	10027,50	

### 4.3 Estimador IPW visto como um método de imputação

Tendo sido calculado o custo ponderado para cada tempo, será possível imputar os valores de custo faltantes para todos os indivíduos censurados do estudo, e assim ser obtido o custo total por indivíduo (soma dos valores das linhas da Tabela 3). Tomando como exemplo novamente o tempo  $t_5$ , observa-se que cinco dos dez indivíduos do estudo não tem os seus custos disponíveis para esse período (Tabela 1), sendo os indivíduos 3, 5 e 10 pela ocorrência de evento e os indivíduos 2, 6, e 9 pela ocorrência de censura. Partindo do pressuposto de que após o evento os indivíduos deixam de gerar custos, os valores imputados no banco de dados para os indivíduos 3, 5 e 10 serão iguais a zero. No caso dos indivíduos censurados, o cálculo para a imputação dos dados no tempo  $t_5$  é realizado da seguinte forma:

1. Calcula-se o custo adicional pela ponderação IPW, sendo esse o resultado da diferença entre o custo total ponderado para o período ( $IPW_5$ ) e o custo total bruto ( $M(t_5)$ ).

$$Custo\ adicional_{t_5} = (IPW_5) - (M(t_5)) = 13993,88 - 10662,00 = 3331,88$$

2. Divide-se o resultado do custo adicional pelo número de indivíduos censurados no período.

$$Valor\ de\ imputação = \frac{Custo\ adicional_{t_5}}{número\ de\ indivíduos\ censurados_{t_5}} = \frac{3331,88}{2} = 1665,938$$

Logo, cada um dos indivíduos censurados (2 e 6) recebe no período de tempo  $t_5$  o valor de custo, a ser imputado, de R\$ 1.665,938, conforme Tabela 3:

**Tabela 3** - Dados de custos imputados a partir da ponderação IPW em um ECR hipotético

Paciente	t1	t2	t3	t4	t5	t6	CT
1	3140	2990	2890	2500	1788	1330	15268
2	2600	2286	2740	2840	1665,938	1305,167	13494,556
3	4253	0	0	0	0	0	4253
4	2436	2040	1788	1411	1314	1099	9325
5	3325	3540	3800	0	0	0	16433,27
6	2070	2870	2850,75	2164	1665,938	1305,167	6127
7	2960	2630	2898	2936	2500	1903	15686
8	2360	1200	2700	1630	1460	1780	9660
9	2600	3560	2200	1990	3600	1305,167	15308,222
10	2870	2414	3790	4005	0	0	13684
<b>Soma</b>	<b>28614</b>	<b>23530</b>	<b>25656,75</b>	<b>19476</b>	<b>13993,875</b>	<b>10027,5</b>	<b>121298,13</b>
<b>Média</b>	<b>2861,4</b>	<b>2353</b>	<b>2565,675</b>	<b>1947,6</b>	<b>1399,3875</b>	<b>1002,75</b>	<b>12129,81</b>

#### 4.4 Entendendo o estimador IPW na prática

Essa seção tem o propósito de auxiliar o leitor a entender como a ponderação IPW se comporta diante de diferentes cenários e como valores de custos podem ser imputados a partir dessa. A ponderação dos custos de cada período de tempo pelo método IPW é calculada ao se dividir o custo total bruto em certo período de tempo pela probabilidade total de não ser censurado nesse mesmo período.

Em posse dos custos ponderados, em cada período de tempo, é possível imputar valores de custos aos pacientes que foram censurados. Os valores a serem imputados, por sua vez, dependem do cenário de estudo em que se dá a ponderação: *a)* cenário apenas com eventos; *b)* cenário apenas com censura e *c)* cenário com ocorrência de eventos e censura.

Assim, segue análise de pequenos conjuntos de dados, os quais facilitam o entendimento do método. Para facilitar a visualização dos custos que são imputados pelo método, atribui-se o custo de R\$100,00 para cada paciente, independentemente do período.

**Cenário 1 – Apenas com a ocorrência de eventos:** A Tabela 4 apresenta dados de um cenário com a ocorrência apenas de eventos. Note que os valores imputados são todos iguais a zero, o que torna as duas últimas linhas da tabela (custo total bruto e custo ponderado) iguais. Ou seja, o estimador IPW considera que, após o evento, o paciente deixa de gerar custos.

**Tabela 4 -** Cenário com apenas a ocorrência de eventos.

Paciente	t1	t2	t3	t4	t5	t6	total	período	em risco	morte	censura	H(t)	P(t)	S(t)
1	100	0	0	0	0	0	100	0	10	0	0	0	1	1
2	100	100	100	0	0	0	300	1	10	1	0	0	1	1
3	100	100	100	100	100	100	600	2	9	0	0	0	1	1
4	100	100	100	100	100	100	600	3	9	1	0	0	1	1
5	100	100	100	100	100	100	600	4	8	0	0	0	1	1
6	100	100	100	100	100	100	600	5	8	0	0	0	1	1
7	100	100	100	100	100	100	600	6	8	0	7	0,88	0,13	1
8	100	100	100	100	100	100	600							
9	100	100	100	100	100	100	600							
10	100	100	100	100	100	100	600							
Total	1000	900	900	800	800	800	5200							
IPW	1000	900	900	800	800	800	5200							

**Cenário 2 – Apenas com a ocorrência de censura:** A Tabela 5 apresenta cenários com a ocorrência apenas de censuras. Nesse tipo de cenário, não importa o tempo de ocorrência da censura e nem em que quantidade ela ocorre. O valor a ser imputado será a média dos custos dos indivíduos que ainda estão presentes no período de tempo. Por exemplo, no tempo  $t_4$ , os indivíduos 1, 2 e 3 tiveram custos imputados no valor de R\$ 100,00 reais, correspondente a média dos nove custos disponíveis para esse tempo. Isso, porque o método considera que o fato de um paciente ter abandonado o estudo, não significa que ele deixou de gerar custos e que, além disso, esses tendem a ser semelhantes aos dos demais pacientes do estudo. Observe que a linha referente ao custo total bruto dos tempos  $t_2$  até  $t_6$  apresenta custos subestimados em comparação com a linha referente aos custos ponderados.



**Tabela 5** – Cenário apenas com a ocorrência de censura

Paciente	t1	t2	t3	t4	t5	t6	total	período	em risco	morte	censura	H(t)	P(t)	S(t)
1	100	100	100	100	100	100	100	0	10	0	0	0	1	1
2	100	100	100	100	100	100	100	1	10	0	4	0,4	0,6	1
3	100	100	100	100	100	100	100	2	6	0	0	0	1	0,6
4	100	100	100	100	100	100	100	3	6	0	0	0	1	0,6
5	100	100	100	100	100	100	600	4	6	0	0	0	1	0,6
6	100	100	100	100	100	100	600	5	6	0	0	0	1	0,6
7	100	100	100	100	100	100	600	6	6	0	6	1	0	0,6
8	100	100	100	100	100	100	600							
9	100	100	100	100	100	100	600							
10	100	100	100	100	100	100	600							
Total	1000	900	800	700	600	500	4000							
IPW	1000	1000	1000	1000	1000	1000	6000							

**Cenário 3 – Com ocorrência de eventos e censura:** Neste cenário, os valores de custo imputados aos indivíduos censurados se modificam bastante e dependem do período de tempo em que ocorrem os eventos. Quando a censura ocorre em período de tempo anterior ao evento, os custos são imputados com a média dos custos dos indivíduos que ainda se encontram no estudo. Então, do mesmo modo que ocorre no Cenário 2, os custos imputados aos indivíduos 1, 3, e 4, no período de tempo  $t_2$  e  $t_3$  da Tabela 6 serão de R\$ 100,00 reais. Porém, quando há ocorrência do evento e censura concomitantemente no mesmo período de tempo, o valor de custo imputado para os indivíduos censurados diminui conforme aumenta o número de eventos. No período de tempo  $t_4$  da Tabela 6, a ocorrência de um evento resultará numa diminuição no valor imputado de R\$ 100 para R\$ 85,7 reais, nos custos dos indivíduos 1, 3 e 4. Por outro lado, o custo imputado a esses mesmos indivíduos para o próximo período ( $t_5$ ) aumenta para R\$ 89,30 devido a ocorrência de uma nova censura. Logo, o aumento do número de censuras, em período de tempo onde não se observa o evento, trará um aumento no valor de custo a ser imputado. Contudo, o aumento do número de eventos resultará em menor valor de custo a ser imputado para os indivíduos que abandonaram o estudo. Essa lógica é bastante coerente, pois se há um aumento de experiências de evento, espera-se também que maior seja a probabilidade de que aqueles que abandonaram o estudo possam também terem tido o evento.

Entretanto, se durante o estudo o evento ocorrer antes da primeira censura, o valor imputado a essa será a média dos custos disponíveis no período de tempo (Tabela 7). Esse tipo de cenário foge à lógica das situações apresentadas anteriormente, e é

motivo de alerta, pois quando há muitos eventos antes da ocorrência da primeira censura, pode se ter superestimação do custo médio do tratamento.

**Tabela 6 - Cenário com evento e censura**

Paciente	t1	t2	t3	t4	t5	t6	total	período	n em risco	morte	censura	H(t)	P(t)	S(t)
1	100	100	100	85,7	89,3	71,4	100	0	10	0	0	0	1	1
2	100	100	100	0	0	0	300	1	10	0	1	0,1	0,9	1
3	100	100	100	85,7	89,3	71,4	200	2	9	0	2	0,2	0,78	0,9
4	100	100	100	85,7	89,3	71,4	200	3	7	1	0	0	1	0,7
5	100	100	100	100	100	0	500	4	6	0	0	0	0,83	0,7
6	100	100	100	100	89,3	71,4	400	5	6	1	0	0	1	0,583
7	100	100	100	100	100	100	600	6	5	0	5	1	0	0,583
8	100	100	100	100	100	100	600							
9	100	100	100	100	100	100	600							
10	100	100	100	100	100	100	600							
Total	1000	900	700	600	500	400	3600							
IPW	1000	1000	100	857,1	857,1	685,7	5400							

**Tabela 7- Cenário com ocorrência de evento antes da primeira censura**

Paciente	t1	t2	t3	t4	t5	t6	total	período	n em risco	morte	censura	H(t)	P(t)	S(t)
1	100	0	0	0	0	0	100	0	10	0	0	0	1	1
2	100	100	100	100	85,71	90,48	200	1	10	0	0	0	1	1
3	100	100	100	100	85,71	90,48	300	2	9	1	1	0,11	0,89	1
4	100	100	100	100	0	0	400	3	8	1	2	0,13	0,88	0,89
5	100	100	100	100	100	90,48	500	4	7	0	2	0	1	0,78
6	100	100	100	100	100	100	600	5	6	1	3	0,166	0,83	0,78
7	100	100	100	100	100	100	600	6	5	5	3	1	0	0,65
8	100	100	100	100	100	100	600							
9	100	100	100	100	100	100	600							
10	100	100	100	100	100	100	600							
Total	1000	900	900	800	600	500	4500							
IPW	1000	900	900	900	771,43	771,43	5243							

## 5 OBJETIVOS

Objetivo geral: Apresentar a construção de ferramentas computacionais em *software* Excel e R, para estimação de custos pelo método IPW conforme abordado por Bang e Tsiatis (2000), com o objetivo de lidar com o problema da censura em dados de custos longitudinais.

Objetivos específicos: Propiciar ao pesquisador um maior entendimento do uso do método IPW, bem como a interpretação dos seus resultados, de modo a facilitar a aplicabilidade desse método na estimação do custo de um tratamento, inclusive como método de imputação de dados faltantes.

## 5 REFERÊNCIAS BIBLIOGRÁFICAS

- Bang H, Tsiatis AA. Estimating medical costs with censored data. *Biometrika*. 6 de janeiro de 2000;87(2):329–43.
- Başer O, Gardiner JC, Bradley CJ, Yüce H, Given C. Longitudinal analysis of censored medical cost data. *Health Econ*. maio de 2006;15(5):513–25.
- Briggs AH, O'Brien BJ. The death of cost-minimization analysis? *Health Econ*. março de 2001;10(2):179–84.
- Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Econ*. agosto de 1997;6(4):327–40.
- Canty A, Ripley B. boot: Bootstrap Functions (Originally by Angelo Canty for S) [Internet]. 2016 [citado 26 de fevereiro de 2016]. Recuperado de: <https://cran.r-project.org/web/packages/boot/index.html>
- Carvalho MS, Andreozzi VL, Codeço CT, Campos DP, Barbosa MTS, Shimakura SE. Análise de sobrevivência: teoria e aplicações em saúde [Internet]. Editora FIOCRUZ; 2011 [citado 26 de janeiro de 2016]. Recuperado de: <http://books.scielo.org/id/xqkmm>
- Cattaneo MD. Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *J Econom*. abril de 2010;155(2):138–54.
- Chaudhary MA, Stearns SC. Estimating confidence intervals for cost-effectiveness ratios: an example from a randomized trial. *Stat Med*. 15 de julho de 1996;15(13):1447–58.
- Crowder M. Longitudinal Data Analysis edited by Garrett Fitzmaurice, Marie Davidian, Geert Verbeke, Geert Molenberghs. *Int Stat Rev*. 1 de abril de 2009;77(1):148–9.
- Detsky AS, Naglie IG. A clinician's guide to cost-effectiveness analysis. *Ann Intern Med*. 15 de julho de 1990;113(2):147–54.
- Drummond MF. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press; 2005.
- Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. CRC Press; 1994.
- Field A. *Descobrimos a estatística usando o SPSS - 2.ed.* Bookman Editora; 2009.
- Geskus RB, van der Wal, WM. ipw: An R Package for Inverse Probability Weighting [Internet]. 2011 [citado 9 de novembro de 2015]. Recuperado de: <http://www.jstatsoft.org/rt/findingReferences/1921/0>
- Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-Effectiveness in Health and Medicine*. Oxford University Press; 1996.
- Gray AM, Clarke PM, Wolstenholme JL, Wordsworth S. *Applied Methods of Cost-effectiveness Analysis in Healthcare*. 1 edition. OUP Oxford; 2010a.

- Gray AM, Clarke PM, Wolstenholme JL, Wordsworth S. Applied Methods of Cost-effectiveness Analysis in Healthcare. Handbooks in Health Economic Evaluation 3; 2010b.
- Griffiths RI, Gleeson ML, Danese MD, O'Hagan A. Inverse probability weighted least squares regression in the analysis of time-censored cost data: an evaluation of the approach using SEER-Medicare. Value Health J Int Soc Pharmacoeconomics Outcomes Res. agosto de 2012;15(5):656–63.
- Horvitz DG, Thompson DJ. A Generalization of Sampling Without Replacement from a Finite Universe. J Am Stat Assoc. 1 de dezembro de 1952;47(260):663–85.
- Jaffa MA, Gebregziabher M, Jaffa AA. A Joint Modeling Approach for Right Censored High Dimensional Multivariate Longitudinal Data. J Biom Biostat [Internet]. agosto de 2014 [citado 14 de janeiro de 2016];5(4). Recuperado de: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4327878/>
- J L Severens TMDb. Uncertainty of incremental cost-effectiveness ratios. A comparison of Fieller and bootstrap confidence intervals. Int J Technol Assess Health Care. 1999;15(3):608–14.
- Lin DY. Linear regression analysis of censored medical costs. Biostat Oxf Engl. março de 2000;1(1):35–47.
- Lin DY. Regression analysis of incomplete medical cost data. Stat Med. 15 de abril de 2003;22(7):1181–200.
- Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. Biometrics. junho de 1997;53(2):419–34.
- Mooney CZ, Duval RD, Duval R. Bootstrapping: A Nonparametric Approach to Statistical Inference. SAGE; 1993.
- O'Hagan A, Stevens JW. On estimators of medical costs with censored data. J Health Econ. maio de 2004;23(3):615–25.
- O'Sullivan AK, Thompson D, Drummond MF. Collection of health-economic data alongside clinical trials: is there a future for piggyback evaluations? Value Health J Int Soc Pharmacoeconomics Outcomes Res. fevereiro de 2005;8(1):67–79.
- Polanczyk CA. Decisão clínica em hipertensão arterial sistêmica baseada em análises econômicas. Rev Bras Hipertens. 2002;29–34.
- Pullenayegum EM, Willan AR. Marginal Models for Censored Longitudinal Cost Data: Appropriate Working Variance Matrices in Inverse-Probability-Weighted GEEs Can Improve Precision. Int J Biostat. 2011;7(1):1–27.
- Raikou M, McGuire A. Estimating medical care costs under conditions of censoring. J Health Econ. maio de 2004;23(3):443–70.
- Tian L, Huang J. A two-part model for censored medical cost data. Stat Med. 15 de outubro de 2007;26(23):4273–92.

- Vanni T, Luz PM, Ribeiro RA, Novaes HMD, Polanczyk CA. Avaliação econômica em saúde: aplicações em doenças infecciosas. *Cad Saúde Pública*. dezembro de 2009;25(12):2543–52.
- Wijeyundera HC, Wang X, Tomlinson G, Ko DT, Krahn MD. Techniques for estimating health care costs with censored data: an overview for the health services researcher. *Clin Outcomes Res CEOR*. 1 de junho de 2012;4:145–55.
- Willan AR, Briggs AH. *Statistical Analysis of Cost-Effectiveness Data*. John Wiley & Sons; 2006.
- Willan AR, Lin DY, Manca A. Regression methods for cost-effectiveness analysis with censored data. *Stat Med*. 15 de janeiro de 2005;24(1):131–45.
- Willan AR, O’Brien BJ. Confidence intervals for cost-effectiveness ratios: an application of Fieller’s theorem. *Health Econ*. agosto de 1996;5(4):297–305.
- Young TA. Estimating mean total costs in the presence of censoring: a comparative assessment of methods. *Pharmacoeconomics*. 2005;23(12):1229–42.
- Zhao H, Bang H, Wang H, Pfeifer PE. On the equivalence of some medical cost estimators with censored data. *Stat Med*. 30 de outubro de 2007;26(24):4520–30.
- Zhao H, Tian L. On estimating medical cost and incremental cost-effectiveness ratios with censored data. *Biometrics*. dezembro de 2001;57(4):1002–8.
- ipw: An R Package for Inverse Probability Weighting | van der Wal | *Journal of Statistical Software* [Internet]. [citado 9 de novembro de 2015]. Recuperado de: <http://www.jstatsoft.org/article/view/v043i13>

## 6 ARTIGO ORIGINAL

Construção de ferramenta computacional para estimação de custos na presença de censura utilizando o método da Ponderação pela Probabilidade Inversa

Paula Marques Sientchkovski

Patrícia KlarmannZiegelmann

Vanessa Bielefeldt LeottiTorman

Programa de Pós-Graduação em Epidemiologia

Universidade Federal do Rio Grande do Sul

[Este artigo será submetido a revista Value In Health]

## RESUMO

Dados de custo necessários na Análise de Custo-Efetividade são, muitas vezes, obtidos de estudos longitudinais, situação na qual é comum a presença de censura. A ideia da Ponderação pela Probabilidade Inversa (IPW- Inverse Probability Weighting ) vem sendo bastante estudada e sugerida na literatura relacionada a esse problema, mas se desconhece a disponibilidade de ferramenta computacional para sua aplicação nesse contexto. O objetivo deste artigo é apresentar ferramentas computacionais de fácil acesso e manipulação, para estimação de custos utilizando o método IPW. Através da criação de planilhas eletrônicas em software Excel e programação em software R, e utilizando-se banco de dados hipotéticos, busca-se propiciar ao pesquisador maior entendimento do uso do estimador IPW, bem como a interpretação dos seus resultados. Como resultado do estudo, observa-se que as ferramentas desenvolvidas permitem a aplicação do método IPW de modo acessível e mostram-se como facilitadoras para a estimação de custos na presença de censura.

**Palavras-chave:** Ferramentas computacionais; Dados de Custos; Censura; Ponderação pela Probabilidade Inversa;

## ABSTRACT

Cost data needed in Cost-Effectiveness Analysis are often obtained from longitudinal studies, a situation in which the presence of censorship of such data is usual. The idea of Inverse Probability Weighting (IPW) has been extensively studied and suggested in the literature related to this problem, but is unaware of the availability of computational tool for your application in this context. The purpose of this paper is to present the development of computational tools of easy access and manipulation, to estimate costs by IPW method. By creating spreadsheets in Excel software and programming in R software, and using hypothetical database, we seek to provide to the researcher most understanding of the use of IPW estimator and the interpretation of its results. As result of the study, it is observed that the developed tools allow the application of IPW in an accessible way, and show themselves as facilitators for the estimation of costs in the presence of censorship.

**Key words:** Computational Tools; Costs Data; Censored; Censorship; Inverse Probability Weighting.



## Introdução

A Análise de Custo-Efetividade (CEA - Cost-Effectiveness Analysis) é uma metodologia utilizada em avaliações econômicas para comparar duas ou mais tecnologias de cuidado com a saúde. A comparação se baseia nos custos (medidos em unidades monetárias) e nas efetividades (medidas em unidades naturais) das tecnologias.

As estimativas de custo de um tratamento são, muitas vezes, obtidas através de estudos longitudinais prospectivos. Nesse caso, a maneira mais usual de se obter informações tem sido combinar estudos de custo com ensaios clínicos, onde as informações coletadas dos pacientes são obtidas a partir de observações diretas. No entanto, isso pode produzir resultados imprecisos, pois é possível que alguns pacientes não sejam acompanhados até o ponto final de interesse, implicando na ocorrência de censura e, em tese, na subestimação do custo.

A censura ocorre quando o indivíduo abandona o estudo antes do final ou quando, ao final do estudo, o indivíduo ainda não experimentou o evento (por exemplo, a morte). Esta censura se apresenta, muitas vezes, de forma informativa, dado que o custo total no tempo de evento é positivamente correlacionado com o tempo de censura. Esta característica dos dados impede que a estimação do custo médio de um tratamento seja feita por meio de métodos tradicionais de análise de sobrevivência[3,4].

Lin *et al* [3] foram os autores da primeira tentativa realizada com o objetivo de incorporar adequadamente o efeito da censura informativa na estimativa de custos médios. Porém, a consistência desses estimadores não foi garantida sob certas condições de censura. Na sequência, Bang e Tsiatis[4] propuseram quatro estimadores consistentes sob qualquer condição de censura. Posteriormente, um estimador foi proposto por Zhao e Tian[5] sob as mesmas condições. Muitos estudos foram dedicados a comparar a eficiência e a consistência desses estimadores. Como resultado destes estudos se tem que, quando a história de custo dos pacientes está disponível, o estimador simples particionado de Bang e Tsiatis (BTp) deve ser o escolhido[4,6–8]. Esse estimador é também denominado, em literatura recente[9], como estimador IPW (*Inverse Probability Weithing*), nomenclatura que adotaremos neste artigo.

A proposta de criar ferramentas computacionais para aplicação do estimador IPW decorre da existência de diversos estudos que abordam a estimação de custos médicos a partir de dados censurados, com ênfase para esse método[4,5,7–15]. Por sua

vez, esse cenário se contrapõe ao desconhecimento de ferramenta computacional disponível para aplicação do método.

Assim, este artigo propõe a construção, baseada em descrição detalhada do método em literatura recente[9], de ferramentas computacionais para lidar com o problema da censura na estimativa de custos pelo estimador IPW, utilizando-se de planilhas eletrônicas criadas no *software* Excel 2016 e programação em *software* R, versão 3.2.2. Cabe destacar que, apesar do método IPW para estimação de custo ser aplicado, nesse estudo, em bancos de dados hipotéticos, as situações criadas replicam cenários comumente encontrados em bancos de dados reais, buscando propiciar ao leitor, além de uma abordagem intuitiva do método, um maior entendimento de seu uso, bem como a interpretação dos seus resultados.

---

### **Exemplo de Estudo Longitudinal com Dados Censurados**

O método IPW será apresentado com base em dados de um exemplo hipotético. Como exemplo, suponha que um Ensaio Clínico Randomizado (ECR), onde a coleta de dados de custos é conduzida para comparar dois tratamentos: P (tratamento padrão) e A (tratamento alternativo). O desfecho primário de interesse é a morte (desfecho binário). Pacientes são aleatorizados em um dos dois tratamentos e acompanhados desde o início do tratamento até a morte, ou por um período de meio ano ( $t = 6$  meses), o que ocorrer primeiro. Dados de custo são observados mensalmente ( $t_j$ , para  $j=1,2,\dots,6$ ) para cada paciente da amostra. Os dados apresentados na Tabela 1 representam os custos observados dos 10 pacientes ( $n = 10$ ) submetidos ao tratamento alternativo. Os custos de cada paciente estão dispostos nas linhas da tabela e os períodos de tempo estão representados nas colunas ( $t_1 - t_6$ ). Os dados de custo para os pacientes submetidos ao tratamento padrão serão analisados de maneira independente entre si.

Note que alguns pacientes foram retirados do estudo antes do final do período de seis meses, de modo a exemplificar situações de morte ou de censura. Se durante o estudo o paciente fora óbito, considera-se a ocorrência de um evento (e). Se o paciente abandonar o estudo, ou não vier a óbito, considera-se a ocorrência de uma censura (c). Por exemplo, o paciente 3 morreu antes da coleta de custos do segundo mês de tratamento, assim teve seus custos computados apenas no primeiro mês e é considerado um evento. O paciente 2 abandonou o estudo entre o quarto e o quinto mês de

tratamento. Ele é considerado uma censura, tendo contribuído com seus dados de custo durante os quatro primeiros meses. Note também que os pacientes apresentam custos mais elevados em períodos de tempo anterior à censura, bem como antes de sua morte, ocasionando assim uma correlação positiva entre os tempos de morte e de censura, o que é ilustrado como um caso de censura informativa.

**Tabela 1 - ECR hipotético com dados de custos censurados.**

Paciente	t1	t2	t3	t4	t5	t6
1	3140	2990	2890	2500	1788	1330
2	2600	2286	2740	2840	c	
3	4253	e				
4	2436	2040	1788	1411	1314	1099
5	3325	3540	3800	e		
6	2070	2870	c			
7	2960	2630	2898	2936	2500	1903
8	2360	1200	2700	1630	1460	1780
9	2600	3560	2200	1990	3600	c
10	2870	2414	3790	4005	e	

### Dados de Custo Censurados e a Ponderação pelo IPW

O estimador IPW consiste em uma versão particionada do primeiro estimador proposto por Bang e Tsiatis, denominado no artigo seminal[4] de estimador simples ponderado de caso completo (BT). A vantagem do método particionado é considerar todos os dados de custo observados, mesmo que esses sejam censurados, tornando mais precisa a estimativa do custo médio de um tratamento.

Por exemplo, considerando-se os dados da Tabela 1, tem-se que os custos dos indivíduos foram observados mês a mês durante os 6 meses de duração do estudo, e os custos dos indivíduos 2, 6 e 9, os quais foram censurados, puderam ser computados por 4, 2 e 5 meses, respectivamente. Assim, o estimador IPW considera o tempo e o número de ocorrência dessas censuras e calcula para cada período de tempo um valor de ponderação que será aplicado a soma dos custos disponíveis nesses períodos. A aplicação do estimador IPW será detalhada nas próximas seções onde são explicadas as ferramentas computacionais criadas para este fim.

## Ferramentas Computacionais

Nesta seção são apresentadas duas ferramentas computacionais que realizam a estimação do custo médio utilizando o estimador IPW. Os arquivos necessários para a utilização destas ferramentas estão disponibilizados em [https://drive.google.com/drive/folders/0BwzKma\\_dVqzCTEVYZkpvVkJCTGs](https://drive.google.com/drive/folders/0BwzKma_dVqzCTEVYZkpvVkJCTGs)

### *Ponderação IPW utilizando o software Excel*

O aplicativo criado no *software* Excel 2016 (arquivo “IPW”) é composto de duas planilhas eletrônicas, nomeadas: Banco de Dados (Figura 1) e Resultados (Figura 2), como descrito a seguir.

**Planilha “Banco de Dados”:** é a planilha na qual o usuário insere os dados de seu estudo. Com base nos dados inseridos nesta planilha, as demais planilhas são preenchidas de modo automático. Os dados a serem inseridos nessa planilha devem estar em formato análogo ao apresentado na Tabela 1 e as células D6 e G6 devem ser preenchidas, respectivamente, com o “número de indivíduos” e com o número de “períodos de tempo” em que o tempo total de acompanhamento do estudo foi dividido. Para os indivíduos que não tiveram os seus custos computados, devido à ocorrência do evento ou censura em algum intervalo, devem ser colocadas as letras “e” ou “c”, por uma única vez, na coluna referente ao período de tempo em que a falta dos dados foi observada. Por exemplo, o paciente 2 teve seu custo observado até o quarto mês, e depois disto ele foi perdido. Então, no tempo 5 ele deve aparecer como censura. O paciente 3 teve seu custo observado no primeiro mês, mas foi a óbito antes do término do segundo mês, devendo aparecer como evento no tempo 2.

C	D	E	F	G	H	I	J
Inserir nº de indivíduos	10	Inserir nº de períodos de estudo		6			
		1	2	3	4	5	6
1	3140	2990	2890	2500	1788	1330	
2	2600	2286	2740	2840	c		
3	4253	e					
4	2436	2040	1788	1411	1314	1099	
5	3325	3540	3800	e			
6	2070	2870	c				
7	2960	2630	2898	2936	2500	1903	
8	2360	1200	2700	1630	1460	1780	
9	2600	3560	2200	1990	3600	c	
10	2870	2414	3790	4005	e		

**Figura 1-** Planilha “Banco de Dados” com dados de custo hipotéticos representando um tratamento alternativo aplicado à 10 pacientes, durante um período de tempo de 6 meses, onde observou-se a ocorrência de 3 eventos (nos períodos de tempo 2, 4 e 5) e 3 censuras (nos períodos de tempo 3, 5 e 6). Essa planilha está programada para colorir de vermelho e verde, respectivamente, os eventos (e) e as censuras (c). Os dados utilizados como forma de exemplo, podem ser obtidos no arquivo de nome B\_IPW\_TA acessando o link do Google Drive disponível neste artigo.

**Planilha “Resultados”:** esta planilha apresenta os resultados dos cálculos necessários para a estimação do custo médio de um tratamento. A sua estrutura, e os cálculos subentendidos na sua formulação, estão baseados em exemplo dado por Gray *et al*[9], bem como em interpretação da exposição do método IPW por Bang e Tsiatis[4]. A planilha construída possui doze colunas e o seu número de linhas depende do número de *períodos de tempo* (primeira coluna) considerados no estudo. Por exemplo, com base nos dados da Tabela 1, coletados mensalmente por um período de 6 meses, o primeiro

período de tempo se inicia no tempo zero e assim temos sete períodos de tempo ( $t_0$  até  $t_6$ ).

O número de indivíduos em risco (segunda coluna) no tempo  $t_0$  é sempre igual ao tamanho da amostra, já que no início do estudo todos os pacientes estão em risco de sofrerem o evento ou de serem censurados. Na medida em que ocorrem as censuras e/ou eventos, o número de pacientes diminui. Assim, em  $t_3$  todos os pacientes estão em risco, já em  $t_3$  temos oito pacientes em risco em razão da ocorrência de morte e censura nos tempos  $t_1$  e  $t_2$ .

O número de eventos e o número de censuras (terceira e quarta colunas) são computados, para ambos os casos, no período de tempo anterior às suas ocorrências, de modo que os indivíduos com dados censurados já não sejam considerados na coluna com os valores correspondentes aos pacientes em risco ( $n$  em risco). Por exemplo, considerando que na Figura 1 o paciente 9 sofre censura em  $t_6$ , para fins de cálculo essa censura é registrada em  $t_5$ .

A taxa de risco (quinta coluna) estima a probabilidade de censura, ou seja, o número de indivíduos censurados dividido pelo número de indivíduos ainda em risco. Assim, em  $t_5$  20% dos indivíduos ainda em risco foram censurados, como se segue:

$$H(t_5) = \frac{\text{n}^\circ \text{ de censuras } (t_5)}{\text{n}^\circ \text{ de indivíduos em risco } (t_5)} = \frac{1}{5} = 0,2$$

A probabilidade de não ser censurado no intervalo de tempo (sexta coluna) resulta de uma unidade menos a taxa de risco. Por exemplo,  $P(t_5) = 1 - H(t_5) = 0,8$ . A probabilidade total de não ser censurado no início do período de tempo  $t$  (sétima coluna) consiste em uma probabilidade acumulada, e sua fórmula é dada por  $K(t_j) = P(t_{j-1}) * K(t_{j-1})$ . Por exemplo,  $K(t_5) = P(t_4) * K(t_4) = 0,857 * 0,888 = 0,7619$ . Então, cada indivíduo  $i$  que teve custo observado no tempo  $t_5$  terá seu custo ponderado por essa probabilidade total.

As colunas 8 a 12 da planilha compreendem os custos. O custo bruto é a soma de todos os custos observados em cada um dos períodos de tempo, como se segue,

$$\begin{aligned} \sum_{i=1}^{10} M_i(t_5) &= M_1(t_5) + M_4(t_5) + M_7(t_5) + M_8(t_5) + M_9(t_5) \\ &= 1788 + 1314 + 2500 + 1460 + 3600 = 10662 \end{aligned}$$

O *custo bruto médio* é o custo bruto dividido pelo número de indivíduos em risco no período de interesse. O *custo ponderado* (IPW), por sua vez, é o resultado da divisão do custo bruto do período  $t$  pela probabilidade total de não ser censurado no mesmo período. Assim, se em  $t_5$  o custo bruto é R\$ 10.662,00, então o custo ponderado para  $t_5$  será

$$IPW_{t_5} = \frac{10662}{0,76190} = 13993,88$$

Esse custo ponderado é usado para calcular o *custo ponderado médio*, dividindo-se o primeiro pelo número de indivíduos em risco no período  $t_j$ . Assim, em  $t_5$ , temos:

$$\overline{IPW}_{t_5} = \frac{13993,88}{5} = 2798,78$$

Por fim, o *custo final médio* estima o custo médio do tratamento e corresponde à última coluna da planilha. Esse resulta da soma de todos os custos ponderados, divididos pelo número de indivíduos que participaram do estudo, ou seja, em risco no período de tempo  $t_0$ , como se segue:

$$\overline{IPW}_{\text{final}} = \frac{\sum_{j=1}^6 IPW_{t_j}}{n} = \frac{2861,40 + 23530 + 25656,75 + 19476 + 13993,88 + 10027,50}{10} = 12129,81$$

Período de tempo	Nº em Risco	Evento	Censura	H(t)	P(t)	K(t)	Custo Bruto	Custo Bruto Médio	Custo Ponderado	Custo Ponderado Médio	Custo Final Médio
0	10	0	0	0	1	1	0	0,00	0	0	12129,81
1	10	1	0	0	1	1	28614	2861,4	28614,00	2861,40	
2	9	0	1	0,11111	0,88889	1	23530	2614,4	23530,00	2614,44	
3	8	1	0	0	1	0,88889	22806	2850,8	25656,75	3207,09	
4	7	1	1	0,14286	0,85714	0,88889	17312	2473,1	19476,00	2782,29	
5	5	0	1	0,2	0,8	0,7619	10662	2132,4	13993,88	2798,78	
6	4	0	4	1	0	0,60952	6112	1528	10027,50	2506,88	

**Figura 2 - Planilha “Resultados” com base nos dados de custo hipotéticos inseridos na planilha “Banco de Dados”. Planilha com 6 linhas representando o número de período de tempo em que foi dividido o estudo, e com 12 colunas, representando, respectivamente: período de tempo, indivíduos em risco, número de eventos, número de censuras, taxa de risco (H(t)), probabilidade de não ser censurado no intervalo de tempo (K(t)), custo bruto, custo bruto médio, custo ponderado, custo ponderado médio e o custo final médio.**

## Ponderação IPW utilizando o *software* R

O R é um *software* livre que pode ser obtido através de sua página oficial (<http://www.r-project.org/>), onde estão disponíveis informações de como instalá-lo, bem como instalar e carregar pacotes. Será necessária para o funcionamento da programação disponível a instalação do pacote "stringr".

A programação em R para estimação do custo médio de um tratamento através do método IPW foi criada como forma alternativa ao uso das duas planilhas eletrônicas em software Excel. Essa programação se utiliza dos mesmos cálculos apresentados na Seção anterior para a formulação do estimador IPW, e está dividida em dois arquivos tipo “.txt” nomeados como IPW\_R\_Funcoes e IPW\_R. O arquivo IPW\_R\_Funcoes é um arquivo que não necessita ser ajustado pelo leitor e é um arquivo somente de leitura, utilizado através do comando “source ("IPW\_R\_Funcoes.txt")”, contido na programação do arquivo IPW\_R. Esse arquivo é composto por duas funções nomeadas de “CONTA\_C”, “TABELA\_RESUMO”.

A função TABELA\_RESUMO realiza todos os cálculos necessários para a ponderação dos custos pelo método IPW. Essa, tem a função CONTA\_C como uma função auxiliar, com a finalidade de localizar e contar o número de censuras e de eventos do banco de dados. O resultado da função TABELA\_RESUMO é uma tabela com a mesma estrutura e conteúdo da planilha “Resultados” do Excel (Figura 2).

O arquivo IPW\_R contém a programação que deve ser inserida no console R para produzir os resultados. Esse arquivo está organizado em seis títulos (Instala Pacotes, Carrega Bibliotecas, Diretório, Carrega Funções, Leitura Dos Dados e Análise Dos Dados) envoltos ao sinal # (quadrado, jogo-da-velha). O R não lê, como parte da programação, o que vem após esse sinal. Abaixo de cada um desses títulos, estão disponibilizadas, como forma de comentários (após o sinal #), as instruções necessárias para o funcionamento da programação.

Para a utilização do arquivo IPW\_R arquivo, sugere-se os passos a seguir:

**Passo 1:** Criar uma pasta de trabalho (neste artigo esta pasta foi nomeada "F:/Ferramentas\_computacionais"), onde deverá ser salvo os arquivos: IPW\_R, IPW\_R\_Funcoes e o banco de dados a ser analisado.

**Passo 2:** Digitar os dados de custo utilizando o *software* Excel com formato análogo ao apresentado na Tabela 1 (ver o banco de dados hipotético do exemplo disponível no



arquivo B\_IPW\_exemplo\_TA). Observe que a primeira linha do banco deve ser composta pela variável "id" (que identifica o indivíduo) e pelas variáveis tempo t1, t2, ..tn (que representam os custos gastos nos n períodos do estudo). Este banco de dados deve ser salvo na pasta de trabalho criada no passo1.

**Passo 3:** Abrir o arquivo IPW\_R disponibilizado e fazer as modificações indicadas abaixo de cada título. Salve o arquivo após as alterações.

**Passo 4:** Abrir o *Software* R e fazer a instalação do pacote "stringr"(a forma como instalar pacotes encontra-se disponível na página oficial do R).

**Passo 5:** Copiar a programação do arquivo IPW\_R que foi modificada e colar no console do R a fim de obter as saídas com os resultados da programação (Figura 4).

```
> dados
  id  t1  t2  t3  t4  t5  t6
1  1 3140 2990 2890 2500 1788 1330
2  2 2600 2286 2740 2840    c
3  3 4253    e
4  4 2436 2040 1788 1411 1314 1099
5  5 3325 3540 3800    e
6  6 2070 2870    c
7  7 2960 2630 2898 2936 2500 1903
8  8 2360 1200 2700 1630 1460 1780
9  9 2600 3560 2200 1990 3600    c
10 10 2870 2414 3790 4005    e

> tabela
  n_em_risco eventos censuras      h_t      p_t      K_t custo_bruto
t0          10         0         0 0.0000000 1.0000000 1.0000000          0
t1          10         1         0 0.0000000 1.0000000 1.0000000      28614
t2           9         0         1 0.1111111 0.8888889 1.0000000      23530
t3           8         1         0 0.0000000 1.0000000 0.8888889      22806
t4           7         1         1 0.1428571 0.8571429 0.8888889      17312
t5           5         0         1 0.2000000 0.8000000 0.7619048      10662
t6           4         0         4 1.0000000 0.0000000 0.6095238       6112
  custo_bruto_medio custo_ponderado custo_ponderado_medio custo_final_medio
t0                0.000             0.00                0.000          12129.8125
t1             2861.400           28614.00             2861.400
t2             2614.444           23530.00             2614.444
t3             2850.750           25656.75             3207.094
t4             2473.143           19476.00             2782.286
t5             2132.400           13993.87             2798.775
t6             1528.000           10027.50             2506.875
```

**Figura 4 - Saída mostrando os resultados da programação R contida no arquivo IPW\_R, utilizado o banco de dados hipotético do arquivo ("B\_IPW\_exemplo\_TA.csv") e a função TABELA\_RESUMO, disponibilizados no endereço já mencionado na Seção: Ferramentas Computacionais.**

## Discussão

A ideia da ponderação pelo inverso da probabilidade tem sido estudada e apresentada sob vários contextos, no entanto são muito poucos os aplicativos computacionais desenvolvidos e disponibilizados aos interessados em aplicar a técnica. Apenas dois programas computacionais utilizando essa ideia foram encontrados em pesquisa na literatura, e nenhum deles está aplicado à estimação de custos. O primeiro programa é o pacote com nome de “*ipw*” implementado no *software* R[16], que utiliza a ponderação pelo inverso da probabilidade para ajustar modelos estruturais marginais, com o objetivo de estimar os efeitos causais em uma análise de dados[17]. O segundo programa está implementado no *software* estatístico Stata, nomeado de “*teffects ipw*”[18], e consiste na estimação dos efeitos do tratamento em dados observacionais, via ponderação pelo inverso da probabilidade, para corrigir o problema resultante de dados faltantes.

Não foi uma preocupação deste estudo realizar simulações e utilizar banco de dados reais para comprovar o bom desempenho do estimador IPW, pois esse já foi um trabalho realizado por outros pesquisadores[5,6,8]. Bang e Tsiatis[4], por exemplo, utilizaram extensos estudos de simulação com dados fortemente censurados, criando uma variedade de circunstâncias onde há ocorrência de censura. Ainda no mesmo artigo, se utilizam de um conjunto de dados reais de um ensaio clínico realizado pelo Centro Médico da Universidade de Duke, onde aplicaram (retiram aleatoriamente indivíduos do estudo) 11%, 31%, 48% e 67% de censura, afim de comparar os dados censurados com a média amostral dos dados completos. Outros trabalhos também se utilizaram de simulações e banco de dados reais para testar e comparar o IPW com outros estimadores. O resultado dessa comparação, como já salientado na Seção 2 deste artigo, é a preferência pelo método IPW quando o interesse do estudo são os dados longitudinais de custo na presença de censura informativa.

Este estudo, então, acabou por atender a dois propósitos: o de desenvolver e disponibilizar um aplicativo computacional para aplicação dessa técnica, mas também o de aplicá-la em um novo contexto, já que, até onde se tem conhecimento, é a primeira vez que uma ferramenta computacional é oferecida para aplicação do IPW na estimação de custos. Sendo assim, propiciou-se ao pesquisador da área da saúde, por meio desse estudo, a possibilidade de utilização de um método já consolidado na literatura para a

estimação do custo na presença de censura, através da criação de ferramentas computacionais de fácil acesso e aplicação.

Com as planilhas desenvolvidas no *software* Excel e a programação no *software* R, ao mesmo tempo em que se oferece ferramentas em plataformas computacionais largamente utilizadas, oferece-se a possibilidade de entendimento desse método através dos comentários relacionados a essas planilhas. Ainda assim, se não for interesse do usuário conhecer os cálculos que compõem o método IPW, essas planilhas podem ser manipuladas apenas para obtenção dos resultados. Assim, entende-se que as ferramentas desenvolvidas se mostraram eficientes para a estimação do custo de um tratamento num cenário de dados censurados, configurando-se como alternativa confiável e acessível para aplicação do método nesse contexto.

## REFERÊNCIAS

---

- [1] Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-Effectiveness in Health and Medicine. Oxford University Press; 1996.
- [2] O'Sullivan AK, Thompson D, Drummond MF. Collection of health-economic data alongside clinical trials: is there a future for piggyback evaluations? *Value Health J Int Soc Pharmacoeconomics Outcomes Res* 2005;8:67–79. doi:10.1111/j.1524-4733.2005.03065.x.
- [3] Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. *Biometrics* 1997;53:419–34.
- [4] Bang H, Tsiatis AA. Estimating medical costs with censored data. *Biometrika* 2000;87:329–43. doi:10.1093/biomet/87.2.329.
- [5] Zhao H, Tian L. On estimating medical cost and incremental cost-effectiveness ratios with censored data. *Biometrics* 2001;57:1002–8.
- [6] Raikou M, McGuire A. Estimating medical care costs under conditions of censoring. *J Health Econ* 2004;23:443–70. doi:10.1016/j.jhealeco.2003.07.002.
- [7] O'Hagan A, Stevens JW. On estimators of medical costs with censored data. *J Health Econ* 2004;23:615–25. doi:10.1016/j.jhealeco.2003.06.006.

- [8] Zhao H, Bang H, Wang H, Pfeifer PE. On the equivalence of some medical cost estimators with censored data. *Stat Med* 2007;26:4520–30. doi:10.1002/sim.2882.
- [9] Gray AM, Clarke PM, Wolstenholme JL, Wordsworth S. *Applied Methods of Cost-effectiveness Analysis in Healthcare*. Handbooks in Health Economic Evaluation 3; 2010.
- [10] Başer O, Gardiner JC, Bradley CJ, Yüce H, Given C. Longitudinal analysis of censored medical cost data. *Health Econ* 2006;15:513–25. doi:10.1002/hec.1087.
- [11] Detsky AS, Naglie IG. A clinician's guide to cost-effectiveness analysis. *Ann Intern Med* 1990;113:147–54.
- [12] Griffiths RI, Gleeson ML, Danese MD, O'Hagan A. Inverse probability weighted least squares regression in the analysis of time-censored cost data: an evaluation of the approach using SEER-Medicare. *Value Health J Int Soc Pharmacoeconomics Outcomes Res* 2012;15:656–63. doi:10.1016/j.jval.2012.03.1388.
- [13] Lin DY. Linear regression analysis of censored medical costs. *Biostat Oxf Engl* 2000;1:35–47. doi:10.1093/biostatistics/1.1.35.
- [14] Wijesundera HC, Wang X, Tomlinson G, Ko DT, Krahn MD. Techniques for estimating health care costs with censored data: an overview for the health services researcher. *Clin Outcomes Res CEOR* 2012;4:145–55. doi:10.2147/CEOR.S31552.
- [15] Young TA. Estimating mean total costs in the presence of censoring: a comparative assessment of methods. *Pharmacoeconomics* 2005;23:1229–42.
- [16] ipw: An R Package for Inverse Probability Weighting | van der Wal | *Journal of Statistical Software* n.d. <http://www.jstatsoft.org/article/view/v043i13> (accessed November 9, 2015).
- [17] Geskus RB, van der Wal, WM. ipw: An R Package for Inverse Probability Weighting 2011. <http://www.jstatsoft.org/rt/findingReferences/1921/0> (accessed November 9, 2015).
- [18] Cattaneo MD. Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *J Econom* 2010;155:138–54. doi:10.1016/j.jeconom.2009.09.023.

## 7 CONCLUSÕES E CONSIDERAÇÕES FINAIS

As planilhas desenvolvidas no *software* Excel e a programação no *software* R para aplicação do método IPW se mostraram eficientes para a estimação do custo de um tratamento num cenário de dados censurados, configurando-se como alternativa confiável e acessível para aplicação do método, além de preencher uma lacuna referente à aplicação desse, já que se desconhece programação com essa finalidade.

Além disso, ao mesmo tempo que foram criadas ferramentas em plataformas computacionais largamente utilizadas, essas foram desenvolvidas com o objetivo de poderem ser manipuladas para obtenção de resultados mesmo sem o conhecimento pelo usuário dos cálculos que compõem o método IPW, facilitando sua aplicação. Não obstante, ofereceu-se também ao leitor a possibilidade de entendimento desse método através dos comentários das planilhas criadas e apresentadas nesse trabalho, já que nesses comentários está expresso a interpretação realizada do IPW e o raciocínio que fundamentou o desenvolvimento dessas planilhas e a programação.

Por fim, com a aplicação do método IPW, e a consequente obtenção dos custos totais a partir da imputação de dados faltantes, mostrou-se também como é possível calcular a ICER com dados de custo mais precisos, em ferramenta computacional complementar, com impactos positivos para a Análise de Custo-Efetividade.

## 8 APÊNDICES

### Rotina R para estimação dos custos pelo método IPW

```
#####  
# Programação em R para estimação do custo pelo método IPW na presença de censura #  
#####  
  
# Esta programação necessita de alguns ajustes para o seu bom funcionamento como:  
# O banco de dados deve estar no mesmo formato do banco disponibilizado no arquivo  
#B_IPW_exemplo;  
# Após os títulos Diretório e Leitura dos dados devem ser inseridos, respectivamente,  
#conforme os comentários,  
# o endereço onde foi salvo o seu banco de dados e o nome do banco.  
#Após os demais títulos não será necessária nenhuma alteração.  
  
#####  
# 1) Instala Pacotes #  
#####  
#A programação disponível necessita da instalação do pacote stringr,  
#essa instalação é feita uma unica vez no seu R.  
  
#install.packages("stringr")  
  
#####  
# 2) Carrega Bibliotecas #  
#####  
# As bibliotecas deverão ser carregadas sempre que um novo script for aberto  
  
library(stringr)  
  
#####  
# 3) Diretório #  
#####  
#Inserir dentro dos parênteses e entreposto as aspas o endereço aonde foi salvo o arquivo.csv  
com o seu banco de dados  
pasta_dos_dados <-("F:/Ferramentas_computacionais") #fazer alterações conforme  
#comentário acima.  
setwd(pasta_dos_dados) #Não alterar
```

```
#####  
# 4) Carrega Funções #  
#####
```

#O comando source traz as funções editadas no arquivo IPW\_R\_Funcoes.txt para dentro do R.  
source("IPW\_R\_Funcoes.txt")

```
#####  
# 5) Leitura Dos Dados #  
#####
```

#Para a leitura dos dados é necessário inserir dentro dos parênteses e entreposto as aspas o nome seu banco de dados.

# O banco deve estar no mesmo formato do banco de dados hipotético criado para exemplificar o método (IPW B\_IPW\_exemplo).

```
dados <- read.csv2("B_IPW_exemplo_TA.csv")#fazer alterações conforme comentário acima.  
dados
```

```
#####  
#6) Analise Dos Dados #  
#####
```

# Para a análise de dados não será necessária nenhuma alteração nas linhas de comando

```
tabela <- TABELA_RESUMO(dados, ipw.imput = TRUE )
```

```
tabela
```

```
#####  
# Funções em R para estimação do custo pelo método IPW na presença de censura #  
#####
```

```
#####
```

```
# conta_c # -Função auxiliar para identificação de strings em data.frame
```

```
#####
```

```
#Argumentos:
```

```
#x- conjunto de dados, data.frame ou matriz.
```

```
#value- valor a ser buscado.
```

```
conta_c <- function(x, value) {  
  nc <- dim(x)[2]  
  ans <- sum( x[,1] == value, na.rm = TRUE )  
  if ( nc > 1 ) {  
    for (i in 2:nc) {  
      ans <- c(ans, sum(x[,i] == value, na.rm = TRUE ))  
    }  
  }  
}
```

```

    }
    return(ans)
  } else {
    return(ans)
  }
}

```

```
#####
```

```

# TABELA_RESUMO # -Função que computa as estatísticas de sobrevivência, custos e valores a
##### serem calculados de acordo com método IPW retornado ao usuário uma
#tabela resumo.

```

```
# Argumentos:
```

```
# x : conjunto de dados, um data.frame.
```

```
# ipw.imput : gerar valores para imputação por IPW em um conjunto de dados.
```

```
TABELA_RESUMO <- function(dados, ipw.imput = FALSE ) {
```

```
  n_lin <- dim(dados)[1]
```

```
  n_col <- dim(dados)[2]
```

```
  # Limpa valores inadequados
```

```
  dados[dados==c("", "NA")] <- NA
```

```
  dados[is.na(dados)] <- NA
```

```
  # Remove a identificação
```

```
  dados <- dados[,2:n_col]
```

```
  dados <- data.frame(t0=rep("0.0",n_lin), dados)
```

```
  # Converte os fatores para string
```

```
  for (i in 1:n_col) {
```

```
    dados[,i] <- as.character(dados[,i])
```

```
    dados[,i] <- str_replace(dados[,i], ",", ".")
```

```
  }
```

```
  # Computa as estatísticas para formar resultados.
```

```
  eventos <- conta_c(dados, value = "e")[2:n_col]
```

```
  eventos <- c(eventos, 0)
```

```
  eventos_acumulados <- cumsum(eventos)
```

```
  censuras <- conta_c(dados, value = "c")[2:n_col]
```

```
  censuras <- c(censuras, n_lin - sum(censuras[1:(n_col - 1)]) - eventos_acumulados[n_col])
```

```
  censura_acumulada <- cumsum(censuras)
```

```
  n_em_risco <- dim(dados)[1] - ( censura_acumulada - censuras ) - ( eventos_acumulados
- eventos )
```

```
  n_em_risco[n_col] <- censuras[n_col]
```

```
  h_t <- censuras / n_em_risco
```

```
  p_t <- 1 - h_t
```

```
  K_t <- 1
```

```
  for (i in 2:length(p_t)) {
```

```
    K_t <- c(K_t, p_t[i-1] * K_t[i-1])
```



```

}

# Converte os fatores para numeric
dados[dados == c("c","e")] <- NA
for (i in 1:dim(dados)[2]) {
  dados[,i] <- as.numeric(as.character(dados[,i]))
}

custo_bruto = apply(dados, 2, sum, na.rm = TRUE)
custo_bruto_medio = apply(dados, 2, mean, na.rm = TRUE)
custo_ponderado <- custo_bruto / K_t
custo_ponderado_medio <- custo_ponderado / n_em_risco
custo_final_medio <- c( as.character( sum(custo_ponderado) / n_em_risco[1] ),
rep(" ", length(custo_ponderado_medio) - 1) )
val_imputacao <- (custo_ponderado - custo_bruto) / ( censura_acumulada - censuras)
val_imputacao[is.nan(val_imputacao)] <- custo_bruto_medio[ is.nan(val_imputacao) ]
# Agrega tudo em um data.frame e devolve ao usuário.
if ( ipw.imput ) {
  ans <- data.frame(
    n_em_risco,
    eventos,
    censuras,
    h_t,
    p_t,
    K_t,
    custo_bruto,
    custo_bruto_medio,
    custo_ponderado,
    custo_ponderado_medio,
    custo_final_medio)
} else {
  ans <- data.frame(
    n_em_risco,
    eventos,
    censuras,
    h_t,
    p_t,
    K_t,
    custo_bruto,
    custo_bruto_medio,
    custo_ponderado,
    custo_ponderado_medio,
    custo_final_medio)
}
return(ans)
}

```