

Estado da publicação: Não informado pelo autor submissor

Extrapolação de dados curvas de sobrevida em saúde: uma abordagem metodológica ajuste direto com dados agregados

Solange Borges, Ivan Ricardo Zimmermann

<https://doi.org/10.1590/SciELOPreprints.5110>

Submetido em: 2022-11-18

Postado em: 2022-11-18 (versão 1)

(AAAA-MM-DD)

Title: Extrapolation of survival curve data in health: a methodological approach for direct fitting to aggregated data

Título: Extrapolação de dados curvas de sobrevida em saúde: uma abordagem metodológica ajuste direto com dados agregados

Autores: Solange Cordeiro (ORCID: [0000-0002-2833-4198](https://orcid.org/0000-0002-2833-4198))¹, Ivan Zimmermann (ORCID: [0000-0001-7757-7519](https://orcid.org/0000-0001-7757-7519))^{1,2*}.

¹ Programa de Pós-Graduação em Avaliação de Tecnologias em Saúde do Instituto Nacional de Cardiologia (INC); ² Departamento de Saúde Coletiva, Universidade de Brasília (UnB);

*Correspondência: ivan.zimmermann@unb.br

Abstract

Introduction: Economic evaluation models often adopt long time horizons, making it necessary to extrapolate data from clinical research for economic evaluation models. The common methodological proposals available are strongly based on individual patient data (IPD), a scenario not always feasible for the daily routine of the Health Technology Assessment (HTA). Thus, the objective of this study was to propose a method for extrapolation with survival curves with direct fitting to aggregated data.

Methods: The case study consisted of the application of parametric models of survival analysis with the main recommended distributions: exponential, Weibull, log-normal, log-logistics, generalized gamma and Gompertz. The models were adjusted to data from a randomized clinical trial testing therapies (anastrozole and fulvestrant) in the context of metastatic breast cancer with 10 years of follow-up on progression-free survival (PFS) and overall survival (OS). After making the adjustments to the individualized data, obtained by contacting the authors, we sought to validate the application of the adjustment to the aggregated data using nonlinear regressions and optimization algorithms. Both methods were compared in terms of visual inspection and quality of fit (Akaike Information Criteria – AIC and Bayesian Information Criteria – BIC). **Results:** In the two treatment arms, the Weibull and generalized gamma distributions were the ones that best fitted the OS data, both in the individualized and in the aggregated approach, according to statistical and visual inspection criteria. For PFS, log-logistic and log-normal curves were used for anastrozole. In the case of fulvestrant, the best choices would be the log-normal and generalized gamma curves for the individualized data and Gompertz and generalized gamma curves for the aggregated data. In terms of visual inspection, the difference was barely perceptible between the use of the individualized and aggregated models. **Conclusion:** Direct fitting data with survival curves to aggregated data is feasible. Despite differences in the choice of some curves, visual inspection suggests that it is unlikely that these differences have an impact on decision making. The algorithm presented here may be useful in situations where access to IPD is not possible.

Keywords: *Health Economic Evaluation; Survival Analysis; Regression Analysis*

Resumo

Introdução: Os modelos de avaliação econômica frequentemente adotam longos horizontes temporais. Contudo, a pesquisa clínica geralmente tem um curto tempo de seguimento dos participantes, tornando-se necessária a extrapolação de dados para alimentar os modelos de avaliação econômica. As propostas metodológicas disponíveis trabalham fortemente com os dados em sua forma individualizada, cenário nem sempre factível ao cotidiano da Avaliação de Tecnologias em Saúde (ATS). Assim, o objetivo deste estudo foi replicar um método para extrapolação com curvas de sobrevida aplicável a dados agregados. **Métodos:** O estudo de caso consistiu nas aplicações de modelos paramétricos de análise de sobrevivência com as principais distribuições recomendadas: Exponencial, Weibull, Log-normal, Log-logística, Gama generalizada e Gompertz. Os modelos foram ajustados aos dados de um ensaio clínico randomizado de duas terapias (anastrozol e fulvestranto) no contexto do câncer de mama metastático com 10 anos de seguimento nos desfechos de sobrevida livre de progressão (SLP) e sobrevida global (SG). Após a condução dos ajustes aos dados individualizados, obtidos mediante contato com os autores, buscou-se validar a aplicação do ajuste aos dados agregados com uso de regressões não lineares e algoritmos de otimização. Ambos os métodos foram comparados em termos de inspeção visual e qualidade do ajuste (Akaike Information Criteria – AIC). **Resultados:** Na verificação da aplicação no estudo de caso nos dados dos dois braços de tratamentos, as distribuições Weibull e gama generalizada foram as que melhor se ajustaram aos dados de SG, seja na abordagem individualizada quanto na agregada, segundo critérios estatísticos e de inspeção visual. Para a SLP foram as curvas log-logística e log-normal para o anastrozol. No caso do tratamento com fulvestranto, as melhores escolhas seriam as curvas log-normal e gama generalizada para os dados individualizados e Gompertz e gama generalizada nos dados agregados. Em termos de inspeção visual, a diferença foi pouco perceptível entre o uso do modelo individualizado e agregado. **Conclusão:** O ajuste de dados com curvas de sobrevida a dados agregados se mostra factível. Apesar de diferenças na escolha algumas curvas, a inspeção visual sugere que é pouco provável que estas diferenças tenham impacto para a tomada de decisão. O algoritmo aqui apresentado pode ser útil nas situações de impossibilidade de acesso aos dados individualizados.

Descritores: Avaliação Econômica em Saúde; Análise de Sobrevida; Análise de Regressão

INTRODUÇÃO

Considerada um dos grandes pilares da Avaliação de Tecnologias em Saúde (ATS), a avaliação econômica é definida como uma análise técnica que identifica, mensura, valora e compara estratégias em relação aos seus custos e desfechos em saúde [1]. As diretrizes metodológicas de avaliação econômica em saúde são consistentes ao recomendar a adoção de longos horizontes temporais, como exemplo o horizonte por toda a vida, em situações crônicas [2]. Na maioria das intervenções, a efetiva adoção de um horizonte de tempo que contemple a expectativa de vida implicará a necessidade de extrapolação de dados provenientes da pesquisa clínica. Isto é particularmente verdade quando as evidências são provenientes de ensaios clínicos randomizados, onde os custos e barreiras pragmáticas de implementação impedem que o seu tempo de seguimento seja longo o suficiente para abarcar toda a expectativa de vida de seus participantes. Nessas situações, os cálculos da expectativa de vida adotada no modelo podem exigir a adoção de extrapolação de dados.

Neste contexto, a aplicação de modelos de curvas de sobrevivência para que o horizonte do modelo considere a incerteza na extrapolação dos dados disponíveis tem sido frequente, sendo tais curvas estimadas utilizadas para comparações entre dois ou mais tratamentos ao longo do tempo [3]. A função de sobrevivência é uma das principais funções probabilísticas, definida como a probabilidade de uma observação não falhar até um certo tempo t , ou seja, é a probabilidade de uma observação sobreviver ao tempo t , dada uma variável aleatória T não negativa, do tipo contínua, que representa o tempo de falha [4]. Os modelos de análise de dados de sobrevivência podem ser estimados de forma não paramétrica, paramétrica ou semiparamétrica. Com grande aplicação na extrapolação de dados em modelos econômicos na ATS, os modelos paramétricos seguem uma distribuição conhecida de probabilidade e os parâmetros são estimados por meio de ajuste aos dados disponíveis. Como exemplo também frequente, cita-se o uso da distribuição exponencial, situação onde a ocorrência de um evento a uma taxa constante ao longo do tempo é mimetizada [5]. Recentemente, com a difusão dos modelos de Análise de Sobrevida Particionada (PartSA), o uso de extrapolação de dados com curvas de

sobrevida tem sido muito comum, sobretudo, na área de ATS em tratamentos na oncologia [6].

Orientações e recomendações sobre os critérios e métodos a serem adotados na extrapolação de dados no contexto da Avaliação de Tecnologias em Saúde ainda são escassos. Segundo a agência inglesa de ATS National Institute for Health and Care Excellence – NICE, no seu documento de suporte técnico (TSD14), que trata de análise de sobrevivência para avaliação econômica, uma revisão de 45 estudos submetidos na área da oncologia, 32 usaram métodos paramétricos para estimar a sobrevivência, sendo os modelos de Weibull e exponencial os mais usados: Weibull, 23 (51%), exponencial, 20 (44%), Gompertz, 6 (13%), log-logística, 9 (20%), log-normal, 6 (13%) e Gama, com 2 (4%) estudos [7]. Atualmente, uma série de ferramentas e pacotes estatísticos estão disponíveis para se trabalhar com a extrapolação de dados ajustados por meio de curvas paramétricas de sobrevida [8,9]. Apesar dessa importância, os métodos disponíveis são orientados frequentemente a dados individualizados. Tal contexto nem sempre é plausível ao cotidiano da ATS, onde frequentemente o único acesso aos dados de sobrevida se dá por meio de dados agregados em curvas de Kaplan-Meier [10]. Dentre os métodos disponíveis aos dados agregados, destaca-se a proposta de Guyoy et al. (2012)[11], onde se propõe uma aproximação dos dados originais individualizados a partir da derivação das curvas de Kaplan Meier. Contudo, a aplicação do método exige alguns passos e cálculos adicionais a serem conduzidos previamente aos ajustes estatísticos, assim como possui algumas limitações, como na ausência do relato do número de pacientes em risco.

Assim, o presente trabalho tem o objetivo de propor um método aplicável à extrapolação de dados com curvas de sobrevida a partir do ajuste direto a dados agregados.

MÉTODOS

Contexto

Como estudo de caso, adotou-se o exemplo da avaliação de um tratamento no câncer de mama metastático, sendo um contexto da oncologia com frequente relato da necessidade de extrapolação de dados em modelos econômicos [12]. Os dados

adotados como estudo de caso foram provenientes de um ensaio clínico randomizado com a participação de 694 mulheres na pós-menopausa com câncer de mama metastático hormônio positivo não tratadas anteriormente [13]. O estudo comparou a efetividade e segurança do tratamento com anastrozol e sua combinação com fulvestranto, adotando como desfechos a sobrevida livre de progressão (SLP) e a sobrevida global (SG), fazendo uso do método de Kaplan-Meier (K-M) [14]. Em relação à sobrevida global, o grupo com o tratamento anastrozol, contou com 345 mulheres elegíveis destas 261 morreram e 84 foram censurados sem morte. Para a combinação com o fulvestranto, foram 349 elegíveis, onde 247 morreram e 102 foram censuradas sem morte. No desfecho de sobrevida livre de progressão, no grupo com tratamento de anastrozol, 329 tiveram progressão da doença ou morte por qualquer causa e 84 foram censurados sem progressão ou morte. Na combinação com o fulvestranto, 247 tiveram progressão da doença ou morte por qualquer causa e 102 censurados sem progressão ou morte.

Coleta de dados

Os dados necessários para a validação do algoritmo foram retirados da publicação original do estudo [13]. Para fins de comparação entre o ajuste com dados agregados e dados individualizados, foram solicitados aos autores do ECR os dados individualizados que permitissem a reprodução das curvas de sobrevida em análise.

Para a obtenção das estimativas agregadas a partir das curvas de Kaplan-Meier publicadas, foi utilizado o software livre *WebPlotDigitizer* versão 4.2. O detalhamento dos dados obtidos pela digitalização encontra-se no material suplementar. Após a digitalização das imagens disponibilizadas nos artigos, foram estimadas as tábuas de vida com os dados agregados na forma de percentuais de sobrevivência ao longo do tempo. Os dados estimados estão disponíveis em material suplementar. A qualidade da extração dos dados das imagens foi realizada por meio de gráficos da interseção das curvas de sobrevivência dos dados individualizados e agregados.

Ajuste dos modelos

Com intuito de analisar o ajuste dos modelos aos dados de sobrevivência, foram tomados como base os seis principais modelos paramétricos citados em

documento de suporte ao NICE: Exponencial, Weibull, Gompertz, Gama Generalizada, Log-normal e Log-logística [7].

A avaliação dos modelos foi feita por meio da inspeção visual e comparação entre as estatísticas de critérios de informação: *Akaike's Information Criterion* (AIC) e *Bayesian Information Criterion* (BIC) [3]. Na inspeção visual, ambas as curvas ajustadas e as curvas de Kaplan-Meier são plotadas em um mesmo gráfico com o intuito de analisar visualmente a qualidade do ajuste e a plausibilidade clínica das estimativas e comportamentos obtidos [15]. Já, em relação ao critério AIC, proposto por Akaike (1974)[16], trata-se de um estimador qualidade relativa de modelos estatísticos para um determinado conjunto de dados, fornecendo um meio para a seleção do modelo. Diferente de estimativas usuais de erro, como o root-mean-square error (RMSE), sua lógica considera o possível impacto do sobreajuste em situações de número maior de parâmetros e assim busca um balanceamento entre a complexidade e acurácia do modelo [17]. O critério BIC, proposto por Schwarz (1978)[18] possui o mesmo racional, contudo, cita-se que sua penalidade é maior do que no AIC para tamanhos de amostra maiores que sete. Para ambos os critérios, a estimativa absoluta obtida não possui interpretação prática, mas apenas sua comparação relativa. Em seu racional quanto menor o valor obtido, esteja ele em uma escala negativa ou positiva, melhor o modelo [17].

Todas as análises foram conduzidas com auxílio da linguagem R. As rotinas de ajuste aos dados individualizados foram implementadas com o auxílio do pacote estatístico “flexsurv”[8], o qual usa funções de sobrevivência previamente validadas e amplamente utilizadas na análise na área da saúde [9]. Para os dados agregados, obtidos da imagem da curva de Kaplan-Meier, buscou-se a aplicação de modelos regressão não lineares (NLS). Apesar de não tão difundidos no campo da análise de sobrevivência em saúde, os modelos NLS são aplicados em um amplo escopo científico. Seu racional é o de, a partir de uma função previamente definida, aproximar o modelo por um ajuste não linear refinando os parâmetros da função por sucessivas iterações [19]. Assim, sua proposta é essencialmente pautada no apoio de algoritmos de otimização. Para o presente trabalho foi adotado o pacote `nls()` do R. Para tanto, foram consultadas e formuladas as equações das funções de sobrevivência descritas na documentação do pacote “flexsurv”[8]. O código completo e os bancos de dados utilizados podem ser disponibilizados mediante solicitação.

RESULTADOS

Dados de sobrevivência observada

Após contato com os autores do estudo original, foram disponibilizados os dados individualizados e desidentificados para fins de comparação das estimativas obtidas com os dados agregados. Conforme observado Figura 1, ao compararmos as estimativas de sobrevida pelos dados agregados e individualizados, apesar de um desvio em alguns pontos na sobrevida livre de progressão do anastrozol, as estimativas extraídas das imagens possuem alta correlação com as estimativas dos dados individualizados.

Ao serem reproduzidas a análise de sobrevivência com um estimador de Kaplan-Meier com os dados individualizados disponibilizados e a digitalização das curvas publicadas, obtiveram-se curvas muito similares às curvas do estudo original (Figura 2). Após a digitalização das imagens disponibilizadas no artigo científico, observa-se uma grande similaridade com os dados originais, conforme disposto na Figura 2. Com os dados disponíveis foi possível avaliar os ajustes com os métodos tradicionais nos dados individualizados e com NLS nos dados agregados. Os parâmetros calculados das distribuições por ambos os métodos estão disponibilizados na Tabela 1.

Inspeção visual

Os braços de tratamentos foram avaliados separadamente para o ajuste das curvas para o anastrozol e anastrozol + fulvestranto com os dados individualizados para sobrevida global e sobrevida livre de progressão (Figure S1). Para os dois braços de tratamentos as curvas estimadas que melhor se ajustaram visualmente aos dados foram as distribuições Weibull e gama generalizada com adição da Gompertz para o tratamento combinado com o fulvestranto, comparado com a curva de Kaplan-Meier para sobrevida global. Para a sobrevida livre de progressão nos dois braços de tratamento quase todas as curvas ficaram bem próximas da curva de Kaplan-Meier, as duas que ficaram menos próximas foram exponencial e Weibull. A seguir são apresentados os resultados da inspeção visual e estatísticas de ajuste baseados nos dados agregados.

Na estimativa com dados agregados, nos dois braços de tratamentos, as curvas estimadas que melhor se ajustaram visualmente aos dados foram as distribuições Weibull e gama generalizada comparado com a curva de Kaplan-Meier para a sobrevida global (Figura S2). Para a sobrevida livre de progressão, visualmente, todas as curvas estão bem próximas sendo assim possível descartar apenas as distribuições exponencial e Weibull para ambos os braços de tratamento.

Estatísticas de Ajuste

Analisando o ajuste aos dados individualizados, ao considerar as estatísticas de ajuste AIC e BIC, onde o menor valor indica a melhor qualidade de ajuste, verificou-se que as distribuições Weibull em primeiro lugar e gama generalizada em segundo possuem o menor valor para os dois critérios de avaliação para o tratamento com o anastrozol (Tabela 2). Para o tratamento com a combinação do fulvestranto, tem-se que as distribuições com menores valores das estatísticas de ajuste são gama generalizada (AIC), Weibull (BIC), para a sobrevida global. Ao analisar a sobrevida livre de progressão verifica-se que para o tratamento anastrozol + fulvestranto as curvas log-normal e gama generalizada se ajustaram aos dados nos dois critérios de informação (AIC e BIC), já para o tratamento com anastrozol foi a curva log-logística e gama generalizada para o critério AIC e log-normal para o critério BIC.

Utilizando os mesmos critérios de informação AIC e BIC no ajuste aos dados agregados, verificou-se que a distribuição gama generalizada possuiu o menor valor para os dois testes e para os dois braços de tratamento, sendo a curva de Weibull a segunda opção no caso da Sobrevida global. Para a sobrevida livre de progressão a curva log-logística foi a primeira opção para o tratamento anastrozol e a segunda opção a log-normal. Para anastrozol+fulvestranto foram gama generalizada e Gompertz, respectivamente, conforme resultados demonstrados na Tabela 2.

Comparação das curvas escolhidas

Analisando os valores dos parâmetros obtidos (Tabela 1), assim como a disposição visual das principais curvas escolhidas, é possível avaliar as diferenças

práticas entre as estimativas de sobrevivência obtidas a partir de dados individuais daquelas obtidas das curvas dos dados agregados (Figure 3).

DISCUSSÃO

O presente estudo teve sucesso em implementar um ajuste direto de dados de sobrevivência a partir de dados extraídos por digitalização de imagens publicadas. Pode-se interpretar dos resultados que houve semelhança entre as curvas plotadas dos dados individuais versus dados agregados, mostrando ser plausível utilizar os dados de imagens digitalizadas na ausência de dados individualizados.

As principais diretrizes em ATS são ainda escassas em respeito à aplicação e condução da extrapolação de dados com curvas de sobrevivência[7]. Dentro dos pontos de consenso, observa-se a necessidade de minimamente escolher mais de uma curva, sendo as principais: exponencial, Weibull, log-logística, log-normal, gama generalizada e Gompertz. É também comum a recomendação de se verificar as curvas por inspeção visual para descartar as distribuições que não se ajustam aos dados e adotar testes de critérios de informação (AIC/BIC) para confirmar quais dentre as curvas estimadas estão mais aderentes aos dados. Em nosso estudo, após a estimativa de curvas de sobrevivência a partir de dados agregados, todos os principais métodos de seleção de curvas adotados nos modelos com dados individualizados puderam ser executados com dados agregados.

Depois de estimadas as curvas paramétricas para os dois conjuntos de dados (individual e agregado), verificou-se em nosso estudo que em ambos os tratamentos, para o desfecho sobrevivência global, adotaríamos as distribuições de Weibull ou gama generalizada, com pouca diferença entre as duas. Já na sobrevivência livre de progressão, as curvas ficaram quase todas bem próximas da curva de Kaplan-Meier para os dois conjuntos de dados e braços de tratamento, ficando difícil a escolha por inspeção visual. Tal situação demonstra a importância do critério estatístico de AIC/BIC para a escolha das curvas, ficando entre log-normal e log-logística para o tratamento com anastrozol, Gompertz e gama generalizada para anastrozol + fulvestranto. Contudo, destaca-se que nenhuma dessas abordagens é garantia de solução de incertezas em relação às projeções futuras, sendo uma escolha orientada, mas pautada no julgamento. Todavia, pesquisas tem buscado analisar a incerteza das extrapolações por meio de sua comparação com dados reais e ressaltado sua importância. Como exemplo, Bullement et al. (2019)[20], após revisar 11 estudos de avaliações de tecnologias concluiu que as extrapolações iniciais empregadas por fabricantes e

pareceristas geralmente previram razoavelmente bem a sobrevida global quando comparadas a dados mais maduros publicados posteriormente.

Embora já bastante difundida, a extrapolação com curvas de sobrevivência ainda tem suas estratégias metodológicas voltadas ao nível de dados individualizados. Essa é uma importante barreira, dado que o acesso a esse tipo de dado ainda é um desafio. Em publicações recentes no tema de acesso a dados individualizados, Scutt (2020) [21] e van Middelkoop (2020) [22] discutem a dificuldade de realizar estudos com dados individualizados de pacientes, já que muitos pesquisadores ainda não compartilham os dados de seus estudos. Ao nosso conhecimento, esse é o primeiro estudo que compara os métodos de extrapolação baseados em ambos os dados individualizados e agregados com um método direto de ajuste estatístico. Com o ponto do acesso limitado aos dados individualizados, Guyoy et al. (2012)[11], propuseram com sucesso um algoritmo para viabilizar o ajuste de curvas de sobrevida a partir das publicações de curvas de Kaplan-Meier. O método tem sido adotado até hoje em publicações e decisões no contexto da ATS[23]. Entretanto, até a efetiva aplicação do ajuste, o método proposto por Guyoy et al. (2012)[11] exige a condução prévia de alguns passos com o intuito de reconstruir a coorte simulada. Da mesma forma, os próprios autores relatam algumas limitações em situações de escassez de dados, sobretudo em relação à ausência do relato dos dados dos números de indivíduos em risco [11]. Recentemente, Rogula et al. (2022)[24], propuseram uma alternativa promissora com a mesma intenção de reconstrução das curvas, contudo, aplicável em situações de censura. Nosso método se apresenta como uma alternativa mais simples, pautada apenas nos dados digitalizados das curvas de Kaplan-Meier.

Como limitações deste estudo, destaca-se que o modelo proposto foi testado em uma única condição e tecnologia na área de oncologia. Todavia, considerando o compartilhamento de dados e códigos do presente estudo, não observamos impedimentos para que os métodos aqui executados possam ser replicados em qualquer outro contexto de análise de curvas. Outro ponto limitante foi o fato de que apesar do consenso nas curvas de sobrevida global, alguma inconsistência foi observada na escolha das curvas de sobrevida livre de progressão ao comparar os métodos pela estimativa agregada versus individual. Contudo, é importante considerar que mesmo nos modelos com dados individualizados havia muita incerteza na escolha dessas curvas, dada a proximidade dos valores de AIC e BIC, assim como em sua

inspeção visual. Da mesma forma, a diferença prática nas curvas obtidas entre os métodos pode ser considerada mínima e com pouco potencial de impacto nos modelos de avaliação econômica. Tais limitações em nossos resultados ressaltam que a extrapolação com base em dados agregados possa não ser uma escolha de preferência na modelagem em ATS, quando disponíveis os dados individualizados.

Dado que a disponibilidade de dados individualizados é ainda uma barreira para a implementação de modelos de extrapolação de dados com curvas de sobrevivência na ATS, o ajuste de curvas a partir de dados agregados é uma ferramenta de grande valor. Nossa proposta, em código reproduzível e acessível publicamente, se apresenta como mais uma alternativa de apoio à tomada de decisão em ATS. Mesmo sendo uma alternativa factível, ressalta-se que as conclusões aqui alcançadas são baseadas em um contexto isolado e específico, sendo recomendável que mais estudos em outros contextos e terapias sejam conduzidos buscando ampliar o escopo de sua aplicação.

Agradecimentos

O presente estudo só foi possível graças ao Dr. William Barlow (Universidade de Washington-EUA), que prontamente disponibilizou os dados individualizados do ensaio clínico *Overall Survival with Fulvestrant plus Anastrozole in Metastatic Breast Cancer*.

Contribuição dos autores

SB e IZ discutiram e construíram o desenho inicial do estudo. SB trabalhou na coleta de dados. SB e IZ implementaram o modelo. SB conduziu as simulações e analisou os principais resultados. Todos os autores contribuíram igualmente para a discussão e avaliação dos resultados finais.

Conflito de interesses

Nenhum declarado

5 REFERÊNCIAS

1. Drummond FM, Mark JS, Karl C, Greg LS, George WT. *Methods for Economic Evaluation of Health Care Programmes*. Oxford University Press; 2015.
2. Brasil. Ministério da Saúde. Secretaria de Ciência T e IEstratégicosD de C e Tecnologia. *Diretrizes metodológicas: Diretriz de Avaliação Econômica*. 2. ed. Brasília; 2014.
3. Colosimo EA, Giolo SR. *Análise De Sobrevivência Aplicada*. 1ª edição. São Paulo: Blucher; 2006.
4. Kleinbaum DG, Klein M. *Survival analysis: A Self-learning text*. 3rd ed. Springer; 2012.
5. Carvalho MS, Andreozzi VL, Codeço, Claudia Torres Campos, Dayse Pereira Barbosa MT, Shimakura SE. *Análise de sobrevivência*. 2nd ed. Rio de Janeiro: Editora Fiocruz; 2011.
6. Woods BS, Sideris E, Palmer S, Latimer N, Soares M. Partitioned Survival and State Transition Models for Healthcare Decision Making in Oncology: Where Are We Now? *Value in Health*. 2020;23:1613–21.
7. Latimer N. NICE DSU technical support document 14: survival analysis for economic evaluations alongside clinical trials-extrapolation with patient-level data. Sheffield: Report by the Decision Support Unit. 2011;
8. Jackson C. **flexsurv** : A Platform for Parametric Survival Modeling in R. *J Stat Softw*. 2016;70.
9. George B, Seals S, Aban I. Survival analysis and regression models. *Journal of Nuclear Cardiology*. 2014;21:686–94.
10. Briggs A, Claxton K, Mark JS. *Decision Modelling for Health Economic Evaluation*. Published in the United States by Oxford University Press Inc., New York ©; 2006.
11. Guyot P, Ades A, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol*. 2012;12:9.

12. Marguet S, Mazouni C, Ramaekers BLT, Dunant A, Kates R, Jacobs VR, et al. European cost-effectiveness study of uPA/PAI-1 biomarkers to guide adjuvant chemotherapy decisions in breast cancer. *Eur J Cancer*. Elsevier Ltd; 2016;63:168–79.

13. Mehta RS, Barlow WE, Albain KS, Vandenberg TA, Dakhil SR, Tirumali NR, et al. Combination anastrozole and fulvestrant in metastatic breast cancer. *New England Journal of Medicine*. 2012;367:435–44.

14. Mehta RS, Barlow WE, Albain KS, Vandenberg TA, Dakhil SR, Tirumali NR, et al. Overall Survival with Fulvestrant plus Anastrozole in Metastatic Breast Cancer. *New England Journal of Medicine*. 2019;380:1226–34.

15. Latimer NR. Survival Analysis for Economic Evaluations Alongside Clinical Trials—Extrapolation with Patient-Level Data. *Medical Decision Making*. 2013;33:743–54.

16. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr*. 1974;19:716–23.

17. Stoica P, Selen Y. Model-order selection. *IEEE Signal Process Mag*. 2004;21:36–47.

18. Schwarz G. Estimating the Dimension of a Model. *The Annals of Statistics*. 1978;6.

19. Baty F, Ritz C, Charles S, Brutsche M, Flandrois J-P, Delignette-Muller M-L. A Toolbox for Nonlinear Regression in *R*: The Package **nlstools**. *J Stat Softw*. 2015;66.

20. Bullement A, Meng Y, Cooper M, Lee D, Harding TL, O'Regan C, et al. A review and validation of overall survival extrapolation in health technology assessments of cancer immunotherapy by the National Institute for Health and Care Excellence: how did the initial best estimate compare to trial data subsequently made available. *J Med Econ*. Taylor & Francis; 2019;22:205–14.

21. Scutt P, Woodhouse LJ, Montgomery AA, Bath PM. Data sharing: experience of accessing individual patient data from completed randomised controlled trials in vascular and cognitive medicine. *BMJ Open*. 2020;10:e038765.

22. van Middelkoop M, Lohmander S, Bierma-Zeinstra SMA. Sharing data—taming the beast: barriers to meta-analyses of individual patient data (IPD) and solutions. *Br J Sports Med.* 2020;54:822–4.

23. Wang T, Scuffham P, Byrnes J, Downes M. Cost-effectiveness analysis of gene-based therapies for patients with spinal muscular atrophy type I in Australia. *J Neurol.* 2022;269:6544–54.

24. Rogula B, Lozano-Ortega G, Johnston KM. A Method for Reconstructing Individual Patient Data From Kaplan-Meier Survival Curves That Incorporate Marked Censoring Times. *MDM Policy Pract.* 2022;7:238146832210776.

Figures

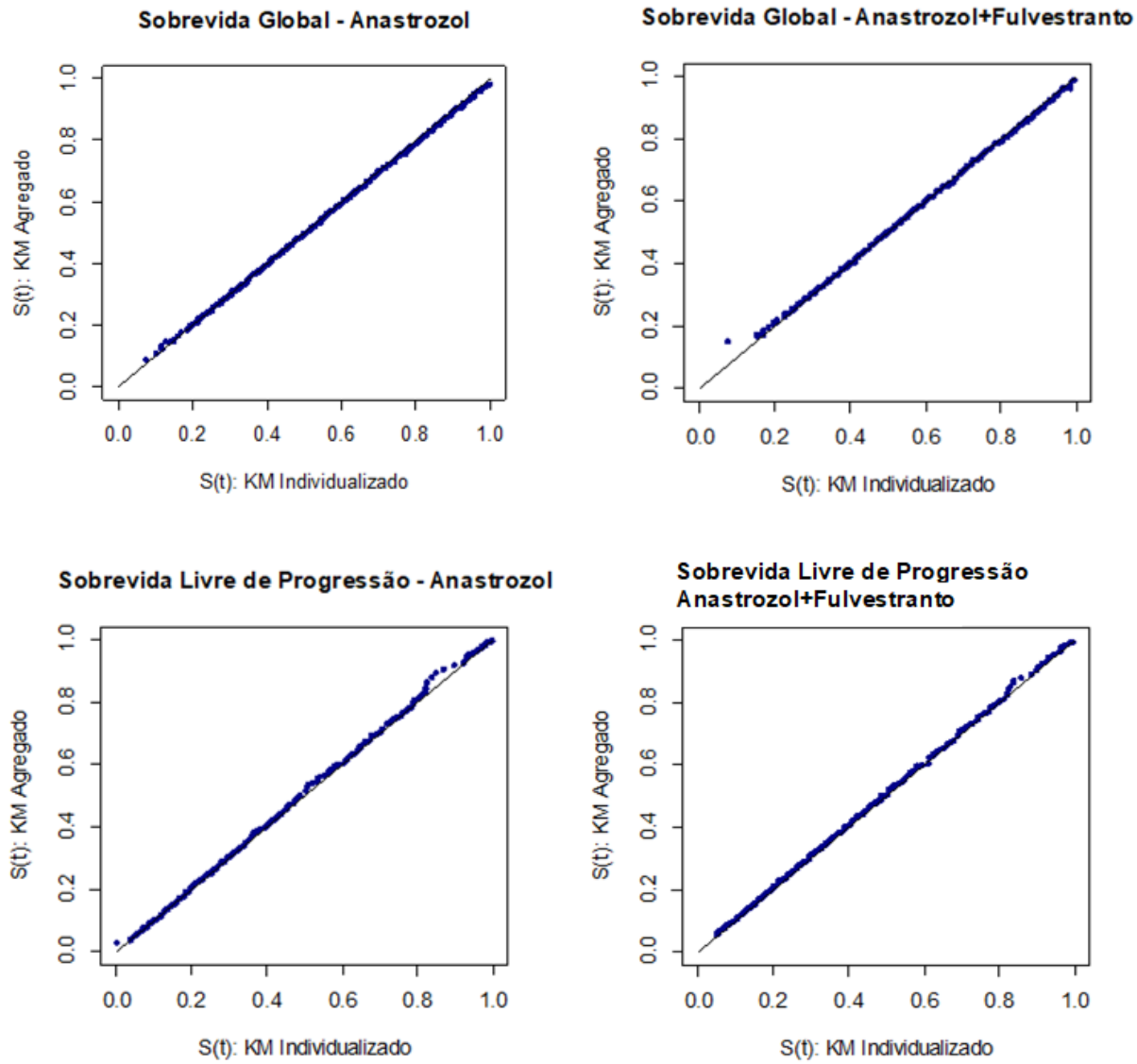


Figura 1 - Curva de sobrevivência de Kaplan-Meier dos dados individualizados versus dados agregados

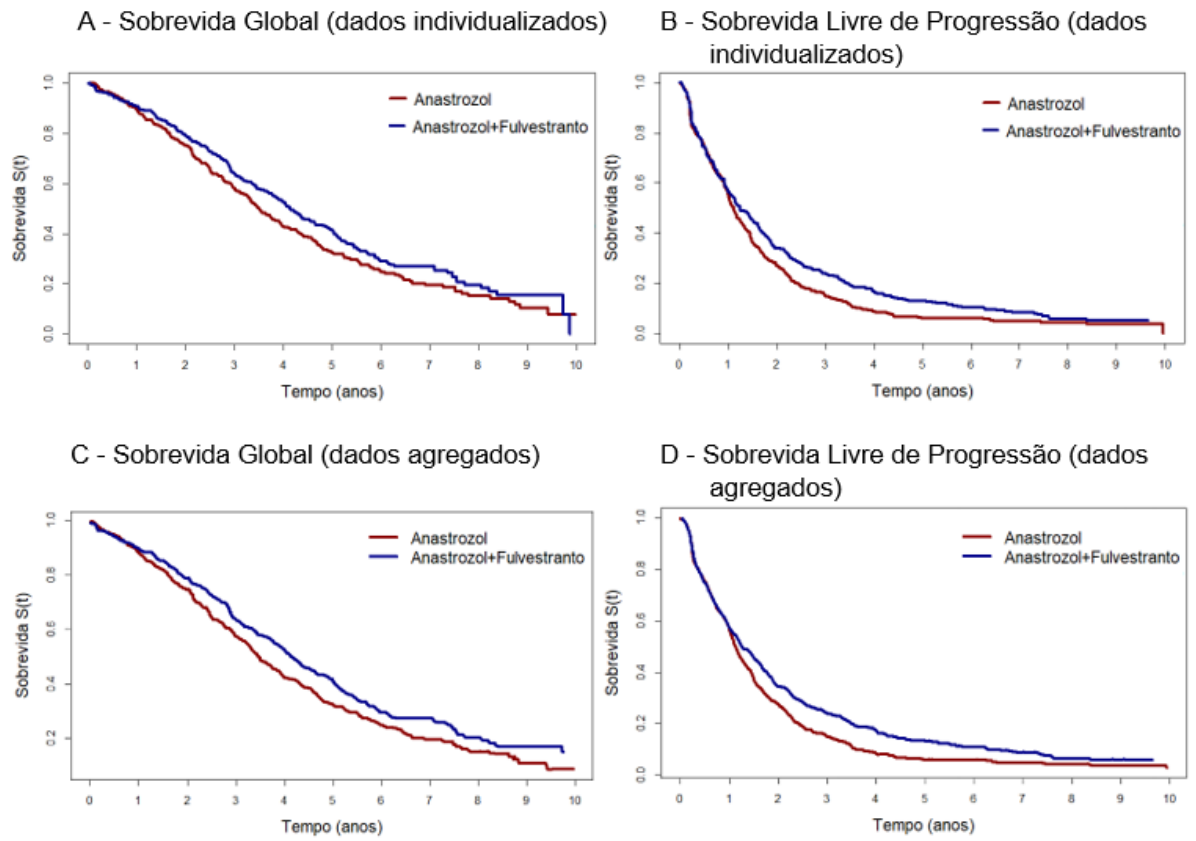


Figura 2 - Curvas obtidas das figuras digitalizadas para os dois braços de tratamento

Fonte: Elaboração a partir dos dados disponibilizados, 2019 [9].

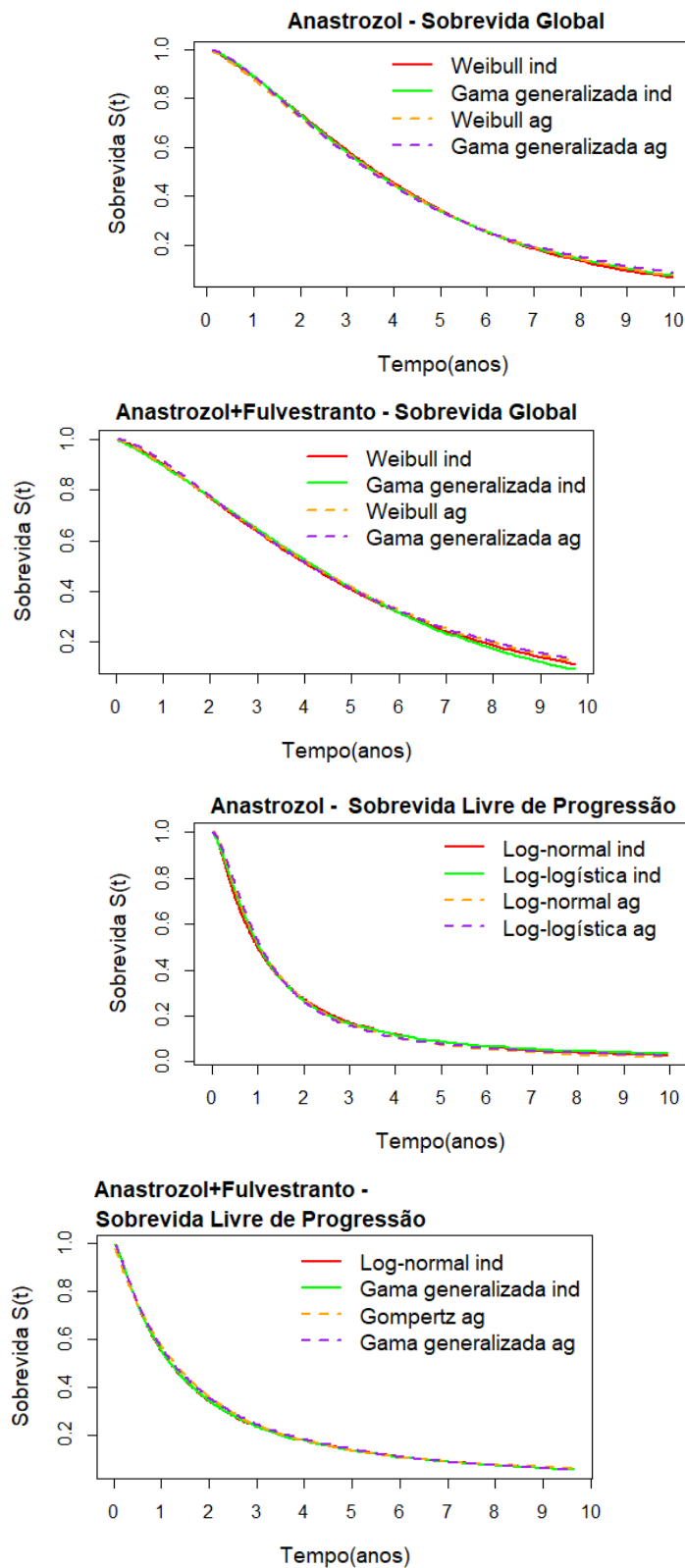


Figura 3 – Comparação das curvas com dados individuais vs agregados, sobrevida global e sobrevida livre de progressão

TABELA 1: Estatísticas do critério de informação AIC e BIC, sobrevida global e livre de progressão, para os tratamentos anastrozol e anastrozol+fulvestranto, para os dados individualizados e agregados

Distribuição		Sobrevida Global				Sobrevida livre de progressão			
		Anastrozol		Anastrozol+ Fulvestranto		Anastrozol		Anastrozol+ Fulvestranto	
		AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Dados Individualizados	Exponencial	1342.930	1346.771	1346.975	1350.822	1018.7919	1022.633	1147.795	1151.641
	Weibull	1313.428	1321.109	1323.797	1331.490	1019.9017	1027.583	1141.275	1148.968
	Log-normal	1334.343	1342.024	1376.212	1383.904	998.9785	1006.660	1112.601	1120.294
	Gama Generalizada	1314.779	1326.301	1323.454	1334.994	996.9148	1008.437	1114.537	1126.077
	Log-logística	1319.823	1327.504	1341.979	1349.672	996.1412	1003.822	1120.047	1127.740
	Gompertz	1325.198	1332.879	1323.652	1331.345	1005.5318	1013.213	1126.659	1134.352
Dados Agregados	Exponencial	-2903.829	-2894.301	-3156.700	-3146.931	-4431.018	-4421.212	-3514.812	-3505.074
	Weibull	-5153.568	-5139.276	-5088.019	-5073.365	-4458.237	-4443.529	-4908.973	-4894.366
	Log-normal	-4531.393	-4517.101	-4664.436	-4649.782	-5319.177	-5304.469	-6501.675	-6487.067
	Gama Generalizada	-5526.168	-5507.113	-5229.058	-5209.520	-5318.292	-5298.681	-6668.986	-6649.510
	Log-logística	-4825.909	-4811.617	-4986.192	-4971.538	-5622.003	-5607.295	-6135.980	-6121.373
	Gompertz	-4137.853	-4123.561	-4379.635	-4364.981	-4837.604	-4822.896	-6633.494	-6618.887

Nota: AIC: *Akaike's Information Criterion* e BIC: *Bayesian Information Criterion*

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença [Creative Commons CC-BY](#).
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.