

# Análise Sequencial usando R para Monitoramento de Vacinas e Drogas Pós-Comercializadas

Ivair R. Silva<sup>1</sup> †

<sup>1</sup> *Departamento de Estatística, Universidade Federal de Ouro Preto, Ouro Preto, MG, Brasil  
Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care  
Institute, Boston, MA, USA*

**Resumo:** *Este artigo descreve alguns dos principais métodos de análise sequencial voltados ao monitoramento de eventos adversos seguidos da pós-comercialização de drogas e vacinas. No contexto usual em que o número de eventos adversos segue uma distribuição de Poisson, o método MaxSPRT é ilustrado pelo uso do pacote SEQUENTIAL do software de programação estatística R.*

**Palavras-chave:** Análise sequencial; Distribuição Poisson; MaxSPRT; Poder; Tempo médio de monitoramento

**Abstract:** *This paper is dedicated to describe some of the main methods for sequential analyses devoted to post-market drug/vaccine safety surveillance. Considering the very practical problem of having adverse events arriving according to a Poisson stochastic process, here we present the use of the Sequential package for applications of the MaxSPRT method. The Sequential package was programed in the R language.*

**Keywords:** Sequential analysis; Poisson distribution; Power; Expected time of surveillance.

## Introdução

A análise estatística sequencial consiste em monitorar, pelo uso de técnicas de inferência estatística, informações amostrais que são coletadas de forma periódica e cumulativa no tempo. Procedimentos clássicos de análise, tais como estimação pontual, estimação intervalar ou testes de hipóteses, são aplicados repetidas vezes conforme os dados vão se acumulando ao longo do tempo. A teoria nesta área ganhou ampla notoriedade após a proposta de e Wald (1945), que desenvolveu o teste sequencial da razão de probabilidades, geralmente referenciado apenas pela sigla SPRT resultado de seu nome em inglês *sequential probability ratio test*. O SPRT tem sido extendido e refinado para aplicações em várias áreas da ciência. Os métodos baseados no SPRT e em suas extensões podem ser classificados como testes generalizados da razão de verossimilhanças sequenciais (SIEGMUND; GREGORY, 1980; LAI, 1991).

---

† Autor correspondente: [ivaiarest@gmail.com](mailto:ivaiarest@gmail.com).

Apesar da aparente similaridade entre a análise sequencial e a teoria para gráficos de controle da qualidade, estes são dois tipos distintos de análise e a diferença reside, principalmente, no objetivo de cada um. Em controle da qualidade, assume-se que o processo estocástico a ser monitorado encontra-se sob condições previamente entendidas, denominadas condições de controle. Assim, o objetivo dos gráficos de controle é monitorar os dados no decorrer do tempo visando detectar futuras mudanças significativas da média e/ou da variância que venham a ocorrer após o início do monitoramento. Diferentemente, em análise sequencial a premissa é que, se o processo possuir características distintas das que se espera, estas já se fazem presentes no processo original mesmo antes do início do monitoramento. O objetivo da análise sequencial, portanto, é identificar se certas hipóteses sobre o processo estocástico original serão confirmadas pelos dados que chegam sequencialmente ao longo do tempo.

A formulação da problemática tratada por técnicas de análise sequencial é bem geral e, por isso, encontra aplicações em diversas áreas da ciência. Entretanto, pode-se dizer que desafios surgidos nas áreas das ciências biológicas, medicina, ensaios clínicos e vigilância da saúde pública em geral, impulsionaram a construção de grande parte dos métodos de análise sequencial. Uma extensa coleção de aplicações de análise sequencial a problemas envolvendo ensaios clínicos é fornecida por Jennison e Turnbull (2000).

Na última década, ocorreu uma intensificação dos esforços dirigidos ao desenvolvimento de métodos de monitoramento de drogas e vacinas pós-comercializadas. Corroborando este fato a criação no ano de 2000, pelo Governo brasileiro, do Sistema de Vigilância Epidemiológica de Eventos Adversos Pós-vacinais. Outro exemplo é o projeto ‘Mini Sentinel Initiative’ (PLATT et al., 2012; SILVA; KULLDORFF, 2014) do ‘U.S.A. Food and Drug Administration (FDA)’. A relevância do tema se justifica porque o efetivo controle das probabilidades dos erros tipo I e tipo II implica proteção à saúde da população exposta aos riscos de medicamentos recém-comercializados. A análise sequencial favorece a rápida detecção de potenciais riscos à saúde da população. Quanto antes um risco aumentado é identificado, maior o número de pessoas que serão protegidas de reações adversas indesejáveis.

O presente artigo se dedica a descrever alguns dos principais métodos de análise sequencial que tem sido usados para o monitoramento da segurança de vacinas e drogas pós-comercializadas. Especificamente, serão descritos: teste de Pocock; teste de O’Brien e Fleming para dados binomiais; SPRT; MaxSPRT; e o método da função gasto de alpha. Naturalmente, o acervo de métodos apresentados pela literatura é extenso, muitos dos quais são obtidos por extensões dos métodos tratados aqui. O presente material não pretende apresentar uma ampla revisão dos métodos existentes, mas tão somente tratar de alguns dos mais utilizados atualmente para monitoramento de eventos adversos. Para uma visão mais geral sobre os métodos de análise sequencial existentes e suas aplicações, sugerimos a leitura do livro de Jennison e Turnbull (2000) e do artigo de Cook et al. (2012). Como uma opção para análise de dados oriundos da distribuição Poisson, apresentamos aqui um exemplo de aplicação do teste MaxSPRT pelo uso do pacote Sequential do software R. Tal aplicação poderá ser de interesse maior àqueles leitores que ainda não possuem muita familiaridade e experiência com análise sequencial e que pretendem fazer suas primeiras análises sob tal abordagem. Assim, este artigo visa contribuir para um primeiro contato com a formulação dos problemas de análise sequencial, apresentar algumas das principais soluções existentes na literatura, e descrever a estruturação dos dados no formato requerido pelo pacote Sequential, bem como a interpretação das saídas apresentadas pelo mesmo.

## Monitoramento do Risco de Eventos Adversos

No monitoramento da segurança de drogas e vacinas pós-comercializadas, quando medicamentos e/ou vacinas são aprovadas para comercialização, o intuito é certificar se produto médico comercializado pode oferecer riscos não esperados à saúde da população. Isto pode ocorrer tanto por questões puramente aleatórias, como também pelo fato de que as experimentações clínicas realizadas no processo de aprovação e certificação de tal produto, devido a questões éticas e por limitações financeiras, podem não contar com amostras representativas de toda a diversidade (idade, etnia, tipos de problema de saúde, gestação, remédios concorrentes e etc.) de que é constituída a população que consumirá o produto. Deste modo, pode ocorrer uma elevação do número médio de ‘eventos adversos’.

Entende-se por ‘evento adverso’ toda situação clínica ocorrida em tempo variável após a utilização de produtos imunobiológicos, respeitando-se um diagnóstico diferencial adequado, o afastamento de situação coincidente e a plausibilidade biológica do evento (MINISTÉRIO DA SAÚDE, 2005). Encontram-se, ainda, outras definições para evento adverso que incluem em sua descrição a área a qual está associado. Por exemplo, se vinculado ao estudo de vacinas, o evento adverso poderia ser conceituado como ‘qualquer sinal ou sintoma grave e/ou inesperado e/ou indesejável que ocorra após a vacinação.’ (MONTEIRO et al., 2011). Ou, se vinculado ao estudo de medicamentos comercializados e distribuídos para a população, poder-se-ia defini-lo como qualquer sinal ou sintoma grave, ou indesejável, após a ingestão/absorção de drogas.

No contexto do monitoramento de eventos adversos seguidos de drogas e vacinas recém aprovadas, a vantagem do teste sequencial sobre o teste clássico (não sequencial) é a possibilidade de se antecipar as conclusões sobre a hipótese de que a droga/vacina é segura, pois os dados são analisados tão logo as primeiras observações amostrais são acessadas, o que possibilita uma rápida tomada de decisão. Em vigilância da segurança de vacinas pós-comercializadas, a rápida detecção de que uma vacina deve ser retirada do mercado, implica evitar danos a saúde de um elevado número de habitantes da população vacinada.

Recentemente, esforços tem sido dirigidos para o desenvolvimento de métodos voltados ao monitoramento de drogas e vacinas pós-comercializadas, para o qual o efetivo controle das probabilidades dos erros Tipo I e Tipo II implica proteção à saúde da população em geral, e mitigação das perdas de capital financeiro e humano associadas a uma retirada indevida de medicamentos/vacinas oriundas de anos de intensas pesquisas investidos em seu desenvolvimento e aprovação. A importância de se controlar os riscos de se retirar produtos indevidamente, ou da não percepção de um problema de segurança, é enfatizado por Silva e Kulldorff (2014).

Atualmente, monitorar a ocorrência de reações adversas provocadas por medicamentos ou vacinas recém-comercializadas é questão de alta relevância. Deve-se citar o empenho do ‘U.S.A. Food and Drug Administration (FDA)’ em intensificar suas ações direcionadas ao desenvolvimento de ferramentas de teste sequencial para vigilância da segurança de drogas e vacinas pós-comercializadas por meio do projeto Sentinela (PLATT et al., 2012; SILVA; KULLDORFF, 2014).

O Brasil também tem se empenhado em aumentar sua eficiência no controle das taxas de reações advesas seguidas de drogas/vacinas pós-comercializadas. No ano 2000, o Programa Nacional de Imunizações implantou oficialmente o Sistema de Vigilância Epidemiológica de Eventos Adversos Pós-vacinais, que permite a coleta de informações de forma abrangente e consistente. Em 15 de julho de 2005, foi publicada a Portaria nº 33 da Secretaria de Vigilância em Saúde do Ministério da Saúde, classificando, assim, os eventos

adversos pós-vacinação como agravo de notificação compulsória. Uma lacuna que se percebe no Brasil é a escassez de artigos científicos que proponham métodos estatísticos de análise sequencial adequados à realidade brasileira. Os trabalhos que tratam das questões epidemiológicas e logísticas estão sendo amplamente discutidas e aprimoradas no Brasil (WALDMAN et al., 2011). Mas deve-se considerar, também, o aspecto sequencial da vigilância de eventos adversos, que, além de poderem responder rapidamente à existência de um risco elevado, também podem garantir um controle rigoroso das probabilidades dos erros Tipo I e Tipo II.

## O Problema dos Testes Múltiplos em Análise Sequencial

Definimos por ‘análise sequencial’ o ato de se realizar inferências estatísticas repetidas vezes com base em um banco de dados que evolui no tempo. Quando a análise consiste especificamente na realização de testes estatísticos de hipóteses, tem-se a situação amplamente conhecida como testes múltiplos. Os testes múltiplos, se não operacionalizados de forma adequada, podem levar a probabilidades de erro Tipo I expressivamente maiores do que o nível de significância nominal adotado pelo analista. Este fato será ilustrado com o exemplo a seguir.

Suponha que um teste sequencial por agrupamento será operacionalizado para verificar se uma vacina, aplicada em certa população, é confiável ou não com respeito a um determinado tipo de evento adverso. Suponha ser de conhecimento médico que os sintomas de tal evento adverso também podem ocorrer na população independentemente da vacinação. Assim, se  $(p \times 100)\%$  desta população foi vacinada, para  $p \in (0, 1)$ , e se de fato a vacina não aumenta a ocorrência dos sintomas associados ao evento adverso, então, para um montante de  $n$  ocorrências de eventos adversos, denote por  $Y_n$  o número de indivíduos que receberam a tal vacina. Assim, se a vacina não altera a probabilidade de ocorrência de sintomas associados ao evento adverso,  $Y_n$  segue uma distribuição binomial de parâmetros  $(n, \theta)$ , com  $\theta = 1/(1 + z)$ , e  $z = (1 - p)/p$ . Mas, se a vacina oferecer um aumento no risco de ocorrências do tal sintoma, então  $\theta = 1/(1 + z/RR)$ , com  $RR > 1$  sendo o risco relativo desconhecido associado à vacinação. A título de ilustração, façamos  $p = 0,02$ , ou seja, 2% da população alvo foi vacinada. Pode-se ter interesse em testar  $H_0 : RR \leq 1$  contra  $H_1 : RR > 0,02$ . Para tanto, vamos usar um nível de confiança  $\alpha = 0,05$ . Suponha também que dois testes foram previamente agendados, sendo que o primeiro deverá ocorrer após observar-se um total de 200 indivíduos apresentando os sintomas monitorados, e um segundo teste deverá ocorrer após observar-se outros 300 indivíduos sintomáticos após o primeiro teste.

Se tratarmos o problema sem levar em conta que dois testes foram planejados, o valor crítico do primeiro teste será  $c_1 = 7$ , pois  $Pr(Y_{200} \geq 7 | RR \leq 1) \leq Pr(Y_{200} \geq 7 | \theta = 0,02) = 0,0493$ . Assim, se o número de vacinados, dentre os 200 primeiros indivíduos, for maior que 6, então se concluirá que a vacina causa um aumento na ocorrência de tais sintomas. Do contrário, o monitoramento será continuado até que se observe os outros 300 indivíduos. O valor crítico não sequencial para o segundo teste será  $c_2 = 16$ , pois  $Pr(Y_{500} \geq 16 | \theta = 0,02) = 0,047$ . Note que a probabilidade de erroneamente rejeitarmos  $H_0$  no segundo teste só será igual a 0,047 se apenas este teste for realizado. Mas, como um teste já teria sido realizado no tempo 200, a máxima probabilidade de se rejeitar a hipótese nula, sob a hipótese nula, será ligeiramente maior:

$$\begin{aligned}
Pr(\text{rej. } H_0 | \theta = 0,02) &= Pr(Y_{200} \geq 7 | \theta = 0,02) + Pr(Y_{200} < 7, Y_{500} \geq 16 | \theta = 0,02) \\
&= 0,0493 + \sum_{y=0}^6 [Pr(Y_{300} > 15 - y | \theta = 0,02) Pr(Y_{200} = y | \theta = 0,02)] \\
&= 0,0792.
\end{aligned} \tag{1}$$

Conclusão: para  $\theta = 0,02$ , a probabilidade do erro Tipo I é próxima de 0,08, o que é indesejavelmente maior do que o nível nominal de 0,05 previamente estipulado.

Se o número de testes aumentasse muito, e é isso que ocorre na prática, a probabilidade do erro Tipo I seria consideravelmente aumentada, o que tornaria o monitoramento complementamente obsoleto, pois levaria a altas probabilidades de se identificar problemas de risco alterado mesmo quando uma vacina fosse confiável. Para contornar este problema, a abordagem correta é ajustar os valores críticos para as múltiplas análises. As formas de se fazer tais ajustes, bem como suas consequências sobre a poder estatístico e o tempo médio de monitoramento, são os temas principais do campo da análise sequencial.

## Métodos para Análise Sequencial

Os métodos de análise sequencial podem ser categorizados em dois grandes grupos, os contínuos e os por agrupamento. O primeiro permite que o investigador realize testes de hipóteses tantas vezes quanto desejar sem que isso comprometa a o nível de significância nominal global de interesse. Procedimentos contínuos são úteis, por exemplo, para monitoramento da segurança de drogas e vacinas pós-comercializadas. Com respeito ao teste sequencial por agrupamento, os dados são analisados em intervalos de tempo, regulares ou irregulares, a medida com que grupos de observações (ou indivíduos) entram no estudo. Os procedimentos por agrupamento são comumente utilizados em testes para experimentos clínicos (JENNISON; TURNBULL, 2000).

Para formalizarmos o conceito de ‘procedimento de análise sequencial contínuo’, seja  $X_t$  um processo estocástico indexado em  $t$ , onde o tempo  $t$  pode ser contínuo ou discreto.

**Definição 1. (Procedimento sequencial contínuo não truncado)** Para  $A(t)$  e  $B(t)$  funções arbitrárias de valor real, um procedimento de análise sequencial contínuo **não truncado** é qualquer procedimento que: i) encerra o monitoramento para a rejeição da hipótese nula no primeiro tempo  $t_1$  tal que  $X_{t_1} \geq B(t_1)$ , para  $0 < t_1 < \infty$ , e  $X_t > A(t) \forall t \leq t_1$  ou; ii) encerra o monitoramento para a não rejeição da hipótese nula no primeiro tempo  $t_1$  tal que  $X_{t_1} \leq A(t_1)$ , para  $0 < t_1 < \infty$ , e  $X_t < B(t) \forall t \leq t_1$ .

$A(t)$  e  $B(t)$  são chamados de barreiras de parada inferior e superior, respectivamente. Observe que o procedimento sequencial não truncado não estipula um limite superior para o tempo de monitoramento, ou seja,  $X_t$  continuará sendo monitorado indefinidamente se não atingir as barreiras inferior ou superior. Portanto, um truncamento sempre será necessário na prática, tal como definido a seguir.

**Definição 2. (Procedimento sequencial contínuo truncado)** Para  $A(t)$  e  $B(t)$  funções de valor nos reais arbitrárias, um procedimento de análise sequencial contínuo **truncado** é qualquer procedimento que: i) encerra o monitoramento para a rejeição da hipótese nula no primeiro tempo  $t_1$  tal que  $X_{t_1} \geq B(t_1)$ , para  $0 < t_1 < L$ , e  $X_t > A(t) \forall t \leq t_1$ , ou; ii) encerra o monitoramento para a não rejeição da hipótese nula se  $X_t < B(t) \forall t \leq L$  e ii.1)  $X_t \leq A(t)$  para algum  $t < L$ , ou ii.2)  $t > L$ .

$L$  é o tempo máximo de monitoramento no procedimento truncado. O valor de  $L$  é fixado de forma arbitrária com base em características desejáveis, tal como um poder estatístico de interesse.

Para os objetivos deste artigo, definamos também o procedimento sequencial unilateral contínuo truncado.

**Definição 3. (Procedimento sequencial unilateral contínuo truncado)** Seja  $B(t)$  uma função de valor nos reais arbitrária. Um procedimento de análise sequencial unilateral contínuo truncado é qualquer procedimento que: i) encerra o monitoramento para a rejeição da hipótese nula no primeiro tempo  $t_1$  tal que  $X_{t_1} \geq B(t_1)$ , para  $0 < t_1 < L$ ,  $\forall t \leq t_1$ , ou; ii) encerra o monitoramento para a não rejeição da hipótese nula se  $t > L$ .

Quando  $X_t$  é discreto, a interpretação do procedimento contínuo (truncado ou não truncado) é que um teste será executado para cada evento individual que for observado. Se, do contrário, para algum intervalo de tempo, for estipulado um valor maior que 1 para o número de casos a serem acumulados entre dois momentos de teste consecutivos, então dizemos que o procedimento é por agrupamento. No caso geral em que  $X_t$  pode ter uma distribuição discreta ou contínua, e para  $t$  discreto ou contínuo, o conceito de teste sequencial por agrupamento é formulado de forma mais geral por estabelecer que testes só serão realizados em momentos de tempo (medidos em tempo calendário ou na escala do processo estocástico) pré-estabelecidos. No caso do procedimento contínuo ambos os casos, bilateral e unilaral superior, foram aqui contemplados com uma definição a parte. Para o caso por agrupamento, apresentaremos a definição apenas da versão unilateral, pois a bilateral fica óbvia por analogia à definição que demos para o caso contínuo.

**Definição 4. (Procedimento sequencial unilateral por agrupamento)** Para um conjunto de constantes  $A_1, \dots, A_G$ , e uma sequência  $\{t_i\}_{i=1}^G$  de tempos, um procedimento sequencial por agrupamento é todo procedimento que rejeita a hipótese nula se  $X_{t_i} \geq A_i$  para algum  $i \in [1, \dots, G]$ .

Observe que também é cabível formular uma versão unilateral com apenas uma barreira inferior. Porém, como as aplicações são geralmente voltadas a reduzir o tempo de monitoramento apenas nos casos em que  $H_0$  deve ser rejeitada, trataremos neste artigo apenas do caso unilateral superior, tal como nas definições 3 e 4.

Um analista interessado pelo procedimento não truncado deve refletir sobre as implicações desta escolha. Para os casos em que  $X_t$  segue uma distribuição binomial( $t, p$ ), por exemplo, Gandy (2009) mostra que o valor esperado do tempo de interrupção do monitoramento é infinito se  $p = \alpha$ , onde  $\alpha$  é a probabilidade do erro Tipo I do procedimento sequencial. Para  $p = \alpha + \delta$ , mesmo quando  $\delta$  for pequeno, como 0.001, por exemplo, o tempo médio de monitoramento será da ordem das centenas de milhares. Ou seja, na prática, a vacina não seria retirada do mercado antes que centenas de milhares de pessoas

tivessem sido expostas a ela. Apesar disto, um dos precursores da análise sequencial, e sem dúvida o mais importante desenvolvimento teórico da área, foi construído com base em uma abordagem não truncada, que é o clássico teste sequencial da razão de probabilidades proposto por Wald (1945). A seção 4.3 fornece uma descrição sucinta do método. Porém, iniciaremos com a descrição de dois testes por grupo, que são os métodos de Pocock e de O'Brien & Fleming, descritos nas próximas duas sub-seções. As três sub-seções que se seguem tratarão de dois métodos da razão de probabilidades e do método do gasto de alpha.

## Teste de Pocock

Sejam  $Y_1, Y_2, \dots$  uma sequência de observações independentes e identicamente distribuídas, todas apresentando a mesma variância  $\sigma^2$  conhecida e seguindo alguma distribuição contínua. O teste proposto por Pocock (1977) é direcionado a testar  $H_0 : E[Y_t] \leq \mu_0$  contra  $H_1 : E[Y_t] > \mu_0$ . Por simplicidade, considere que um teste será realizado a cada vez que um montante de  $m$  observações forem observadas. Para o  $i$ -ésimo teste, o método de Pocock é baseado na estatística padronizada:

$$Z_i = \frac{1}{\sqrt{im\sigma^2}} \left( \sum_{j=1}^{im} Y_j - im\mu_0 \right) = \frac{1}{\sqrt{im\sigma^2}} (X_{im} - im\mu_0), \text{ with } i = 1, \dots, G.$$

O monitoramento é interrompido, e  $H_0$  é rejeitada, para o primeiro  $i$  tal que  $Z_i > c$ , onde  $c$  é um valor crítico constante para todo  $i = 1, \dots, G$ , i.e., interrompa o monitoramento se  $X_{t_i} > A_i = c + \frac{t_i\mu_0}{\sqrt{t_i\sigma^2}}$ , onde  $t_i = im$  e  $X_{t_i} = \sum_{j=1}^{t_i} Y_j$ . Note que, neste exemplo,  $t$  é discreto e representa a ordem de chegada das observações. O valor crítico  $c_i$ ,  $i = 1, \dots, G$ , é obtido por se assumir que  $Y_t$  segue uma distribuição normal.

Se o teste for para a comparação de duas populações, digamos populações  $a$  e  $b$ , para as quais as observações  $Y_{a1}, Y_{a2}, \dots$  e  $Y_{b1}, Y_{b2}, \dots$  são observadas, com hipóteses  $H_0 : E[Y_{at}] \leq E[Y_{bt}]$  contra  $H_1 : E[Y_{at}] > E[Y_{bt}]$ , então a estatística de teste de Pocock é da forma:

$$Z_i = \frac{1}{\sqrt{2im\sigma^2}} \left( \sum_{j=1}^{im} Y_{aj} - \sum_{j=1}^{im} Y_{bj} \right), \text{ with } i = 1, \dots, G,$$

e novamente a suposição de distribuição normal é crucial para obtenção dos valores críticos.

## Teste de O'Brien & Fleming

O'Brien e Fleming (1979) propuseram que o valor crítico não seja uma constante com respeito a  $t$  mas sim uma função decrescente em  $t$ . Sendo  $c$  o valor crítico estabelecido para o teste de Pocock, O'Brien e Fleming (1979) sugeriram o uso do valor crítico  $c_i = c\sqrt{G/i}$  para o  $i$ -ésimo teste.

Veja que, com este valor crítico, comparativamente ao método de Pocock, será mais difícil que  $H_0$  seja rejeitada no início do monitoramento, e isto fará com que uma amostra maior seja acumulada antes que o monitoramento seja interrompido fazendo com que, para um mesmo tamanho máximo amostral, o teste de O'Brien & Fleming apresente um maior poder estatístico. Seguindo a notação da Seção 4.1, o monitoramento é interrompido, e  $H_0$  é rejeitada, para o primeiro  $i$  tal que  $Z_i > c_i$ , onde  $c_i$  é o valor crítico para o  $i$ -ésimo teste, i.e., interrompa o monitoramento se  $X_{t_i} > A_i = c_i + \frac{t_i\mu_0}{\sqrt{t_i\sigma^2}}$ , onde  $t_i = im$  e  $X_{t_i} = \sum_{j=1}^{t_i} Y_j$ .

O procedimento para o caso de comparação de duas populações é análogo ao apresentado na Seção 4.1, exceto que agora o valor crítico a ser usado é  $c_i$ .

Em resumo, o teste de Pocock é mais recomendado para os casos em que a redução do tempo médio de monitoramento é o fator mais importante na análise. Mas, se o poder for a medida de performance mais importante, então o método de O'Brien & Fleming é mais indicado. Para uma discussão mais aprofundada sobre as propriedades do teste de Pocock e do método de O'Brien & Fleming sugerimos o livro de Jennison e Turnbull (2000).

## Teste Sequencial da Razão de Probabilidades de Wald

Os métodos apresentados nas seções 4.1 e 4.2 se enquadram na categoria de procedimentos sequenciais por agrupamento. Falaremos agora de um dos principais precursores dos métodos sequenciais contínuos. O teste sequencial da razão de probabilidades (SPRT) foi introduzido por Wald (1945), e recebeu este nome por ser baseado na razão de verossimilhanças. O SPRT é bastante geral no sentido de poder ser usado para diversas distribuições de probabilidade para  $X_t$ . Apesar disto, por questões didáticas, aqui ele será descrito para o caso particular em que  $X_t$  segue uma distribuição de Poisson. A fundamentação básica é a mesma para qualquer outra distribuição de probabilidade.

Seja  $C_t$  um processo de Poisson homogêneo, a tempo contínuo, e seja  $c_t$  o correspondente valor observado dada uma realização de  $C_t$ . Suponha que, sob a hipótese nula, o valor esperado de  $C_t$  seja fixa e conhecida, denotada por  $\mu_t$ , e que, sob a hipótese alternativa, seu valor esperado seja  $r\mu_t$ , ou seja, a taxa do processo é alterado em  $r$  vezes em relação ao esperado sob  $H_0$ . Portanto,  $r$  é dito ser o efeito de alteração do risco de um evento relativamente ao que se espera sob  $H_0$  ou, simplesmente, risco relativo. Para o caso particular  $t = 0$ , tempo  $C_0 = c_0 = 0$ . Assim, as hipóteses a serem testadas no esquema sequencial são  $H_0 : r = 1$  contra  $H_1 : r = r_1 \neq 1$ , onde  $r_1$  é uma constante fixa e conhecida para a qual deseja-se fazer o teste. A estatística de teste SPRT é definida para cada  $t$  segundo a teoria clássica da razão de verossimilhanças, a estatística de teste é dada por:

$$U_t = \ln \frac{\Pr(C_t = c_t | r = r_1)}{\Pr(C_t = c_t | r = 1)} = \ln \frac{e^{-r_1\mu_t} (r_1\mu_t)^{c_t} / c_t!}{e^{-\mu_t} (\mu_t)^{c_t} / c_t!} = \ln e^{(1-r)\mu_t r^{c_t}} = (1-r)\mu_t + c_t \ln r, \quad (2)$$

se  $t > 0$ , e  $U_t = 0$  para  $t = 0$ .

Para todo  $t > 0$ , a estatística a ser monitorada é  $\Lambda_t = \sum_{j=1}^t U_j$ . As regras de decisão seguidas de cada teste são: continue o monitoramento se  $\psi_{w,t} = 0$ ; encerre o monitoramento e rejeite  $H_0$  se  $\psi_{w,t} = 1$  e; encerre o monitoramento e não rejeite  $H_0$  se  $\psi_{w,t} = 2$ , onde:

$$\psi_{w,t} = \begin{cases} 0, & \text{if } a < \Lambda_t < b, \\ 1, & \text{if } \Lambda_t \geq b, \\ 2, & \text{if } \Lambda_t \leq a. \end{cases} \quad (3)$$

A grande contribuição de Wald (1945) foi o estabelecimento de uma regra exata para a escolha de  $a$  e  $b$  em função das probabilidades dos erros tipo I e tipo II. A solução exata pode levar a algum trabalho numérico dependendo da distribuição de  $X_t$ . Assim, Wald (1945) fornece uma aproximação simplificada fazendo  $a = \ln[\beta/(1-\alpha)]$  e  $b = \ln[(1-\beta)/\alpha]$ ,



onde  $\alpha$  é a probabilidade de erro tipo I global do teste (probabilidade de erroneamente rejeitar  $H_0$ ), e  $\beta$  é a probabilidade de erro tipo II (probabilidade de não rejeitar  $H_0$  quando esta deveria ser rejeitada). Por exemplo, suponha que queiramos  $\alpha = 0.05$  e um poder de 0,99 ( $\beta = 0,01$ ). Assim, as barreiras de monitoramento inferior e superior são aproximadamente iguais a -4.55 e 2,99, respectivamente.

Para ilustrar, suponha que  $C_t$  seja o número de casos de eventos adversos que chegam, após o tempo zero (momento de aplicação da vacina) a um posto de atendimento. Suponha que  $t$  seja contínua e medido em dias e frações de dia. Por simplicidade, façamos  $\mu_t = t$ , e digamos que  $c_{1,2} = 1$ ,  $c_{1,5} = 2$ ,  $c_{3,3} = 3$ ,  $c_4 = 4$ ,  $c_{4,8} = 5$  e  $c_{5,2} = 6$ , ou seja, os tempos de chegada dos seis primeiros casos foram 1,2, 1,5, 3,3, 4, 4,8, 5,2. Observe o fato de que, como o monitoramento é contínuo, precisamos saber os tempos de chegada de cada caso individualmente para que a teoria concernente à escolha de  $a$  e  $b$  em função dos erros do tipo I e II seja aplicável. Suponhamos que, sob  $H_1$ , estamos interessados em testar  $r = 2$ . Após monitorar o processo durante 6 dias, fazendo testes a cada 15 minutos, produziríamos um gráfico como o da Figura 1. As linhas vermelhas horizontais representam as barreiras  $a = -4.55$  e  $b = 2,99$ . Neste monitoramento fictício, como a barreira superior não foi ultrapassada, a hipótese nula não teria sido rejeitada até o tempo 6, mas, como a barreira inferior também não foi tocada no intervalo  $(0, 6]$ , o monitoramento deve prosseguir com a coleta de mais dados para que o nível de significância de 0,05 e o poder de 0,99 sejam atendidos.

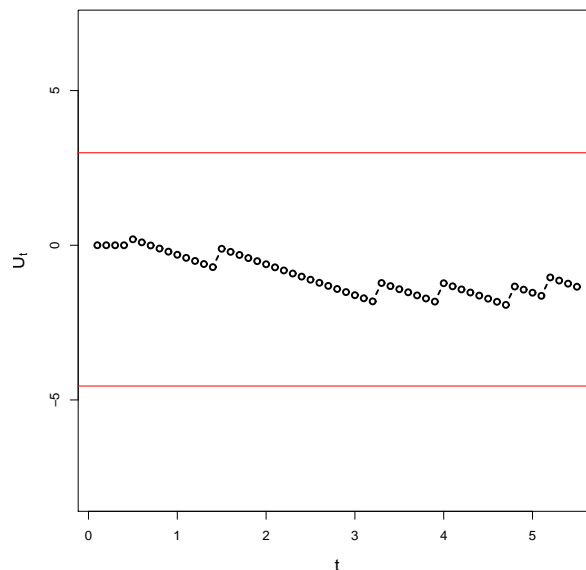


Figura 1: Exemplo ilustrativo para monitoramento de um processo Poisson.

Vários livros foram dedicados a mostrar aplicações e divulgar extensões do SPRT, dentre os quais podemos listar os livros de Ghosh et al. (1997), Govindarajulu (2004), e Mukhopadhyay e de Silva (2009). Alguns livros também expõem abordagem baseadas na ideia de reduzir hipóteses compostas em hipóteses simples para adaptar as condições da razão de verossimilhanças às condições clássicas do SPRT (LAI, 2001; GHOSH, 1992). Vale citar também os casos particulares de extensões, tal como o teste para dados de Poisson de e Abt (1998), o que trata hipóteses que envolvem parâmetros de ruído Bangdiwala (1982), ou um método para o caso corriqueiro de testar proporções de amostras binomiais (HOEL et al., 1976; JOANES, 1972; MEEKER, 1981). Como não poderia faltar, existem propostas exclusivas para dados normalmente distribuídos  $s$  (LACHIN, 1981; VAN DER

TWEEL et al., 1996). Uma versão sequencial ‘stepwise’ foi proposta por Huang (2004). Construções baseadas na definição de função custo também são bastante encontradas na literatura (HOLM, 1985; SCHIPPER et al., 1997).

Uma das extensões mais recentes é o procedimento proposto por Kulldorff et al. (2011). O método fornecido por Kulldorff, ao contrário do SPRT, é truncado, e foi especialmente desenvolvido para tratar de dados binomiais ou Poisson oriundos das contagens de eventos adversos no contexto de vacinas pós-comercializadas. A próxima seção é dedicada a descrever o método.

## SPRT Maximizado

Uma deficiência relevante do SPRT clássico é a possibilidade de se ter poderes bem abaixo do alvo estipulado para cálculo das barreiras  $a$  e  $b$ . Isso ocorre quando a veracidade sobre o risco relativo destoa muito do parâmetro  $r_1$  arbitrado na construção da estatística  $\Lambda_t$ . Na realidade, este é um problema típico dos testes para hipóteses simples mesmo no contexto não sequencial. Kulldorff et al. (2011) fornece exemplos práticos bem elucidativos para destacar como e quando este problema pode comprometer inconvenientemente as conclusões. Assim, no lugar de fixar um valor para  $r_1$ , Kulldorff et al. (2011) propôs o procedimento baseado na maximização da verossimilhança, sob  $H_1$ , com respeito a  $r$ . Com isto, o teste passa a ser para hipóteses compostas, pois  $r$  é estimado por máxima verossimilhança no processo de construção da estatística de teste, e daí seu nome, MaxSPRT. Seguindo a definição de Siegmund e Gregory (1980) e de Lai (2002), o MaxSPRT pode ser classificado como um ‘procedimento de teste sequencial generalizado da razão de verossimilhanças’. Seu desenvolvimento se deu como uma resposta aos esforços do projeto do governo dos Estados Unidos, no âmbito do ‘Disease Control and Prevention sponsored Vaccine Safety Datalink’, para monitoramento eficaz de vacinas pós-comercializadas Lieu et al. (2007).

Seja  $C_t$  a variável aleatória que conta o número de pacientes que receberam a vacina antes do tempo  $t$  e que tiveram um evento adverso de 1 a  $W$  dias após receberem a vacina. Seja  $c_t$  o correspondente valor observado para o número de pacientes que apresentaram eventos adversos. Para eventos adversos raros, é razoável modelar  $C_t$  como um processo Poisson. Sob a hipótese nula,  $C_t$  segue uma distribuição Poisson com média  $\mu_t$ , refletindo um conhecimento a priori sobre a taxa de ocorrência de eventos adversos, ajustada por idade, sexo e outras covariáveis. Sob a hipótese alternativa,  $C_t$  segue uma Poisson com média  $RR\mu_t$ , onde  $RR$  é o risco relativo desconhecido devido à vacina. A estatística MaxSPRT, definida por meio da teoria da razão de verossimilhanças por Kulldorff et al. (2011), é escrita da seguinte forma:

$$\Lambda_t = \ln \max_{H_a} \frac{P(C_t = c_t | H_a)}{P(C_t = c_t | H_0)} = \max_{RR > 1} \ln \frac{e^{-RR\mu_t} (RR\mu_t)^{c_t} / c_t!}{e^{-\mu_t} (\mu_t)^{c_t} / c_t!}. \quad (4)$$

O argumento que resolve o último termo da expressão (4) é  $c_t/\mu_t$ . Assim,

$$\Lambda_t = \ln [e^{\mu_t - c_t} (c_t/\mu_t)^{c_t}] = (\mu_t - c_t) + c_t \ln(c_t/\mu_t), \quad (5)$$

onde  $c_t > \mu_t$ ,  $\Lambda_t = c_t \ln(1 + z)$  se  $c_t = \mu_t$ , e  $\Lambda_t = 0$ , caso contrário.

Na abordagem de vigilância sequencial contínua, a estatística  $\Lambda_t$  é monitorada para todos os valores de  $t > 0$ , e a vigilância é interrompida para a rejeição de  $H_0$  tão logo  $\Lambda_t$  se mostre maior que um valor crítico  $CV$ , ou, quando  $\mu_t = T$ , caso em que a hipótese

nula não é rejeitada.  $T$  é definida a priori como um limite superior para o tamanho amostral, definido na escala do número médio de eventos adversos sob a hipótese nula. Nesta definição, até mesmo um único evento adverso poderia ser responsável pela rejeição da hipótese nula caso o tempo até sua ocorrência fosse extremamente baixo. Uma versão alternativa do MaxSPRT requer um número mínimo de eventos adversos, digamos  $M$ , antes que a hipótese nula possa ser rejeitada, o que pode simultaneamente aumentar o poder estatístico e diminuir o tempo médio de detecção de um risco elevado (SILVA; KULLDORFF, 2012). O valor crítico  $CV$ , o poder estatístico, o tempo médio até sinalização de um risco elevado e o tempo máximo de monitoramento, podem todos serem calculados usando algoritmos para cálculo numérico iterativos via cadeias de Markov (KULLDORFF et al., 2011).

Quando não é possível obter uma estimativa confiável de  $\mu_t$ , uma opção é aplicar o ‘conditional maximized sequential probability ratio test (cMaxSPRT)’, proposto por Li e Kulldorff (2010). O cMaxSPRT não requer a estimação de  $\mu_t$ , e é baseado no teste Monte Carlo.

## Função Gasto de Alpha

Vimos que a estratégia comumente utilizada para interrupção do monitoramento é a adoção de um valor crítico,  $c_t$ , para monitorar  $X_t$ . Alternativamente, ao invés de se definir o limiar  $c_t$ , pode-se definir um limite superior para o montante de probabilidade associada à ocorrência do erro Tipo I em cada teste a ser realizado. Isto é feito pelo uso do conceito de ‘função gasto de alpha’. Para o caso em que o processo  $X_t$  é a tempo discreto, o que se faz é ditar a parcela da probabilidade de erro Tipo I a ser gasta em cada um dos  $G$  testes que serão realizados. O valor total de probabilidade do erro Tipo I, gasto ao final dos  $G$  testes, deverá atender ao nível global de significância,  $\alpha$ , de interesse. Denote o gasto pontual de alpha, no  $t$ -ésimo teste, por  $f_\alpha(t)$ ,  $t = 1, 2, \dots, G$ , e seja  $F_\alpha(t) = \sum_{l=1}^t f_\alpha(l)$  uma função arbitrária tal que  $F_\alpha(G) = \sum_{t=1}^G f_\alpha(t) = \alpha$ .  $F_\alpha(t)$  é chamada de ‘função acúmulo da probabilidade do erro tipo I’. Para o caso em que o processo  $X_t$  é a tempo contínuo, temos que  $\int_0^t f_\alpha(l)dl = F_\alpha(t)$ , e  $\int_0^L f_\alpha(l)dl = \alpha$ , onde  $L$  é o tempo máximo de monitoramento. Um vasto acervo de propostas para a escolha de  $F_\alpha(t)$  pode ser encontrado na literatura. Jennison e Turnbull (2000) apresenta uma rica discussão sobre o tema, bem como alguns exemplos de funções mais usuais, dentre as quais dá destaque à forma do tipo potência:

$$F_\alpha(t) = \min \{ \alpha t^\rho, \alpha \}, \quad (6)$$

onde  $t$  representa uma fração do tempo máximo de monitoramento, e  $\rho$  é um parâmetro de ajuste a ser arbitrado pelo analista. Veja que, para o caso particular  $\rho = 1$ ,  $F_\alpha(t)$  é uma reta com inclinação positiva igual a  $\alpha$ . De acordo com Jennison e Turnbull (2000), estudos anteriores sugerem o uso de valores para  $\rho$  em torno de 1, 1,5 e 2.

## Otimidade dos Procedimentos Sequenciais Contínuos sobre os Desenhos Sequenciais por Grupo

Mostraremos aqui que, para qualquer procedimento sequencial por grupo, sempre existe um procedimento sequencial contínuo que é ‘tão bom quanto’ e que, para processos de Poisson, sempre existe um procedimento contínuo melhor. Este é um importante resultado provado por Silva e Kulldorff (2014), pois mostra que um procedimento contínuo nunca deve ser deliberadamente substituído por um procedimento por grupo. Primeiramente, definiremos a noção de procedimento sequencial uniformemente melhor em termos de quatro métricas que são de importância crucial para mensurar a performance de um teste sequencial. Para tanto, seja  $X_t$  um processo estocástico não negativo com espaço de estados discreto. Para fins didáticos, digamos que o processo  $X_t$  descreva o número de eventos adversos que ocorrem na janela de tempo  $[0, t]$ .

**Definição 5. (Procedimento sequencial uniformemente melhor)** Sejam  $D_1$  e  $D_2$  dois procedimentos para análise sequencial. Para  $D_j$ , o vetor de medidas de sua performance será denotado por  $(\alpha_j, \beta_j, E[S_j], E[L_j])$ , onde  $\alpha_j$  é a probabilidade do erro Tipo I;  $\beta_j$  é a probabilidade do erro Tipo II (o poder estatístico é  $1 - \beta_j$ );  $S_j$  é a variável aleatória que representa o tamanho amostral quando a hipótese nula é rejeitada (tempo esperado de sinalização), e  $L_j$  é o tamanho amostral quando no final do monitoramento quando a hipótese nula não é rejeitada (tamanho amostral máximo).  $L_j$  pode ser definido arbitrariamente como uma variável aleatória ou fixado a priori. O procedimento sequencial  $D_1$  é no mínimo tão bom quanto o procedimento  $D_2$  se  $\alpha_1 \leq \alpha_2$ ,  $\beta_1 \leq \beta_2$ ,  $E[S_1] \leq E[S_2]$ , e  $E[L_1] \leq E[L_2]$ . Se pelo menos uma destas quatro inequações for estrita, então  $D_1$  é dito ser uniformemente melhor que  $D_2$ .

Em palavras, um procedimento sequencial é uniformemente melhor que um outro se ele é tão bom quanto o segundo em todas as quatro medidas de performance e melhor em pelo menos uma das quatro. A definição acima é útil para caracterizar três importantes propriedades envolvendo a comparação entre procedimentos sequenciais por grupo. O próximo teorema introduz a primeira propriedade.

**Teorema 1** *Para qualquer processo estocástico  $X_t$ , não decrescente e indexado por um tempo contínuo ou discreto, e para qualquer procedimento sequencial por grupo, sempre existe um procedimento sequencial contínuo que é uniformemente melhor.*

**Prova 1** *Baseado no procedimento sequencial por grupo, considere o procedimento sequencial contínuo onde  $L = t_G$ ,  $t_0 = 0$  e  $B(t) = A_i$ , para  $t \in (t_{i-1}, t_i]$ . Este procedimento sequencial contínuo rejeita  $H_0$  se, e somente se, o procedimento sequencial por grupo rejeitar. Pelo fato de que esta afirmação é válida independentemente de  $H_0$  ser verdadeira ou não, o procedimento por grupo e o contínuo terão a mesma probabilidade de erro tipo I e o mesmo poder. Ademais, porque  $L = t_G$ , eles também finalizarão o monitoramento no mesmo tempo nos casos em que  $H_0$  não for rejeitada. Agora, seja  $S_c$  a variável aleatória que representa o momento em que o monitoramento contínuo é interrompido para a rejeição da hipótese nula. Isto é, este é o mínimo  $t$  tal que  $X_t \geq B(t)$ . Seja também  $S_g$  a variável aleatória que representa o momento em que o monitoramento por grupo é interrompido para a rejeição da hipótese nula. Ou seja, é o mínimo  $t_i$  tal que  $X_{t_i} \geq A_i$ . Por construção,  $S_c \leq S_g$  para qualquer realização do processo estocástico  $X_t$  e, assim,*

$E[S_c] \leq E[S_g]$ , implicando que sempre existe um procedimento contínuo que é tão bom quanto qualquer procedimento por grupo.

Esta demonstração provê um mecanismo para construção de um procedimento sequencial contínuo que é tão bom quanto um dado procedimento sequencial por grupo.

O Teorema 1 é um dos resultados do artigo de Silva e Kulldorff (2014). A pergunta natural que surge é se podemos obter um procedimento contínuo uniformemente melhor para cada projeto por agrupamento pré-fixado e para qualquer distribuição do processo não-decrescente  $X_t$ . A resposta é: não para qualquer distribuição de  $X_t$ . Mas, para algumas distribuições, um procedimento contínuo uniformemente melhor sempre pode ser obtido. Isto é verdade, por exemplo, para o caso Poisson.

**Teorema 2** *Para qualquer processo estocástico  $X_t$  não decrescente seguindo uma distribuição de Poisson e indexado por tempo contínuo ou discreto, e para um dado procedimento sequencial por agrupamento, sempre existe um procedimento sequencial contínuo uniformemente melhor.*

**Prova 2** *Para um determinado procedimento sequencial por grupo pré-definido, tome um projeto sequencial contínuo por fazer  $L = t_G$ ,  $t_0 = 0$  and  $B(t) = A_i$ , para  $t \in (t_{i-1}, t_i]$ . Considere a notação definida na prova do Teorema 1. Para provar que o tempo máximo de vigilância e as probabilidades dos erros Tipo I e II são as mesmas para o contínuo e o por grupo, aplique exatamente os mesmos argumentos que na prova do Teorema 1. Para derivar a desigualdade estrita em termos do tempo médio de sinalização, acrescentamos o fato de que existe um  $t_i < m < t_{i+1}$  tal que  $E[X_m] - E[X_{t_i}] > 0$ , pois, como  $X_t$  é um processo Poisson, sabemos que  $P(X_m - X_{t_i} > k) > 0$  para qualquer valor de  $k$ . Isto significa que a probabilidade de que a análise contínua venha a rejeitar a hipótese nula no tempo  $m$  é  $P(X_{t_i} < A_i, X_m \geq A_{i+1}) > 0$ , o que implica  $E[S_c] < E[S_g]$ .*

Como uma consequência do Teorema 2, concluímos que, no caso Poisson, é sempre possível considerar testes adicionais em um teste por agrupamento de modo a obter um outro que seja uniformemente melhor. Isto segue do fato de poder-se usar os valores críticos  $A_i$  como as barreiras associadas a um teste adicional arbitrário em um tempo  $t^*$ , onde  $i$  será tal que  $t^* \in (t_i, t_{i+1}]$ .

**Corolário 1.** Se  $X_t$  é um processo Poisson, então, para cada procedimento sequencial por grupo com testes definidos nos tempos  $I = \{t_1, \dots, t_G\}$ , sempre existe um procedimento por agrupamento uniformemente melhor, o qual terá tempos de test  $J = \{\tau_1, \dots, \tau_{K+r}\}$ , onde  $I \subset J$ .

Visando oferecer um resultado mais geral sobre a existência de um procedimento contínuo uniformemente melhor para cada procedimento por grupo, no sentido de que isso seja válido para qualquer distribuição do processo estocástico  $X_t$ , fornecemos o Teorema 3, o qual é válido para um tipo especial de procedimento sequencial por grupo.

**Teorema 3** *Para qualquer processo estocástico  $X_t$  indexado por tempo contínuo ou discreto, e para qualquer procedimento sequencial por grupo tendo  $m$  e  $t_i < m < t_{i+1}$  tal que  $P(X_{t_i} < A_i, X_m \geq A_{i+1}) > 0$ , existe um procedimento sequencial contínuo que é uniformemente melhor.*

A última desigualdade exige que exista pelo menos um momento em que os dados possam chegar entre dois testes consecutivos de modo a indicar com certeza uma futura interrupção do monitoramento.

**Prova 3** *Considere o procedimento contínuo em que  $L = t_G$ ,  $t_0 = 0$  e  $B(t) = A_i$ , para  $t \in (t_{i-1}, t_i]$ . Seguindo os mesmos passos usados na prova do Teorema 1, nós vemos que os procedimentos contínuo e por agrupamento terão a mesma probabilidade de erro tipo I, mesmo poder e mesmo tempo máximo de monitoramento. Sejam  $S_c$  e  $S_g$  os tempos em que os procedimentos contínuo e por grupo irão rejeitar a hipótese nula, respectivamente. Por construção,  $E[S_c] \leq E[S_g]$ . Considere que exista  $m$  e  $t_i < m < t_{i+1}$  tais que  $P(X_{t_i} < A_i, X_m \geq A_{i+1}) > 0$ . Se  $1 \leq l < G$  é tal que  $P(X_{t_i} < A_l, X_m \geq A_{l+1}) > 0$ , seja o novo procedimento por grupo baseado nas barreiras  $A_1, \dots, A_l, A_{l+1}, A_{l+1}, \dots, A_G$  para os tempos  $\{t_1, \dots, t_i, m, t_{l+1}, \dots, t_G\}$ . O tempo médio de sinal deste novo procedimento por grupo será menor do que o do procedimento original e, pelo Teorema 5.1, existe um procedimento contínuo que é tão bom quanto este novo procedimento por grupo. Assim, existe um procedimento sequencial contínuo que é uniformemente melhor que o procedimento original.*

O Teorema 3 requer a existência de pelo menos um intervalo de tempo,  $[t_i, t_{i+1}]$ , em que a probabilidade de se observar um novo evento seja maior que zero e que a barreira para o teste no tempo  $t_{i+1}$  possa ser atingida em um tempo  $m \in (t_i, t_{i+1})$  com probabilidade não nula. Portanto, o intervalo  $[t_i, t_{i+1}]$  é um candidato para receber um teste adicional, o que contribui para uma certa redução do tempo médio de sinalização.

**Corolário 2.** Para qualquer processo estocástico não-decrescente  $X_t$ , a tempo contínuo ou discreto, se existem  $m$  e  $t_i < m < t_{i+1}$  tais que  $P(X_{t_i} < A_i, X_m \geq A_{i+1}) > 0$ , então para cada procedimento sequencial por agrupamento com testes nos tempos  $I = \{t_1, \dots, t_G\}$ , sempre existe um procedimento por agrupamento uniformemente melhor com testes nos tempos  $J = \{\tau_1, \dots, \tau_{K+r}\}$ , onde  $I \subset J$ .

## Sequential Package: MaxSPRT para Dados Poisson

Esta seção traz uma descrição, passo a passo, da utilização do pacote ‘Sequential’, de Silva e Kulldorff (2013), programado em linguagem R, (R CORE TEAM, 2014), e que pode ser baixado livremente acessando ‘<http://cran.r-project.org>’. O pacote Sequential contém funções úteis para o planejamento de análise sequencial via teste MaxSPRT para dados Poisson. Para um dado nível de significância e um tempo máximo  $T$  de monitoramento, o pacote Sequential possibilita a obtenção do valor crítico para operacionalização do MaxSPRT e o cálculo exato das características de performance do teste. Outro material que pode ser consultado para uma explicação mais detalhada sobre outros aspectos do pacote é o manual do usuário, Sequential-manual, por Silva e Kulldorff (2013), que é parte do pacote Sequential. Esta seção visa mostrar alguns exemplos de uso do pacote, que não constam no Sequential-manual, para alguns valores específicos das variáveis de entrada. As funções do pacote Sequential são as seguintes:

- **CV.Poisson(SampleSize,D,M,alpha):** esta função é direcionada para o cálculo do valor crítico a ser usado no teste MaxSPRT contínuo. ‘SampleSize’ é o tamanho máximo da amostra (T), o qual deve ser especificado pelo usuário. As entradas ‘D’ e ‘M’ são parâmetros para indicação de um tempo mínimo de espera para início do monitoramento, onde D é definido em termos do número esperado de eventos sob  $H_0$  e M é definido em termos do número observado de eventos. ‘alpha’ é o nível de significância;
- **Performance.Poisson(SampleSize, D, M, cv, RR):** esta é a função que calcula as três características de performance do teste MaxSPRT contínuo, isto é, poder estatístico, tempo médio de interrupção do monitoramento dado que  $H_0$  foi rejeitada, e tempo total médio de monitoramento. Esta função tem como entradas as informações SampleSize, D, M, e o valor crítico ‘cv’. Para isto, o usuário precisa especificar o verdadeiro risco relativo ‘RR’ de interesse;
- **SampleSize.Poisson(alpha,power,M,D,RR):** esta função serve para que se encontre o valor de T necessário que garanta um ‘power’ de interesse para um dado ‘RR’;
- **CV.G.Poisson(SampleSize,alpha,Looks,M):** esta função serve para que se possa calcular o valor crítico do teste MaxSPRT por agrupamento para um pré-definido número G(‘Looks’) de testes;
- **Performance.G.Poisson(SampleSize,cv,Looks,M,RR):** esta função calcula as três características de performance do teste MaxSPRT por grupo, as quais são o poder estatístico, o tempo médio de interrupção do monitoramento dado que  $H_0$  foi rejeitada, e o tempo total médio de monitoramento.

O quadro a seguir apresenta os comandos para uso das funções do pacote Sequential considerando-se alguns valores específicos para os parâmetros de entrada. Os comandos e símbolos utilizados aqui estão de acordo com a estrutura de programação em R. Assim, este exemplo pode ser replicado por se copiar diretamente o script apresentado neste quadro para o console do R. As sentenças seguidas do símbolo ‘#’ servem para indicar os comentários deste autor ao explicar ou detalhar os objetivos das linhas de comando.

**Comandos para cálculos usando o pacote Sequential**

```
library(Sequential)
# The função 'test' abaixo será usada na busca pelo máximo
# tamanho de monitoramento (T) necessário para que se atinja um determinado
# poder de interesse no
# teste sequencial por agrupamento.
test <- function(Tm,RRm,Looks){
cv<- CV.G.Poisson(SampleSize=Tm,alpha=0.05,Looks,M=1)
x<- Performance.G.Poisson(SampleSize=Tm,cv,Looks,M=1,RR=RRm)
x[[1]]
}
# Ajuste de parâmetros e poder de interesse (target)
G=5
RRr<- 2 # risco relativo verdadeiro fixado
powr<- 0.9 # poder de interesse
# Aqui nós aplicamos p método da bisseção para encontrar T
T1<- 5 # T inicial que fornece um poder menor que o desejado
T2<- 18 # T inicial que fornece um power maior que o desejado
pow<- 0; count<- 1
while(abs(pow-powr)>0.000001&count<100) {
Tm<- (T1+T2)/2
pow<- test(Tm,RRm=RRr,Looks=G)
if(pow>powr){T2<- Tm}else{T1<- Tm}
count<- count+1
}
# Depois de rodarmos os scripts acima, nós temos os seguintes resultados:
pow=0.9000003
# e Tm=14.75176.
```



```

# Agora nós usamos o valor Tm obtido acima para calcular o valor crítico para o
# desenho sequencial por grupo com G=5 testes.
cv<- CV.G.Poisson(SampleSize=Tm,alpha=0.05,Looks=5,M=1)
# Aqui estão as características de performance deste desenho por agrupamento
# sob a hipótese nula (RR=1):
Performance.G.Poisson(SampleSize=Tm,cv,Looks=5,M=1,RR=1)
$Power
[1]0.04573175
$'Expected signal time'
[1]7.671945
$'Expected surveillance time'
[1]14.42799
# Aqui estão as características de performance deste desenho por agrupamento
# sob a hipótese alternativa (RR=2):
Performance.G.Poisson(SampleSize=Tm,cv,Looks=5,M=1,RR=2)
$Power
[1] 0.9000003
$'Expected signal time'
[1] 7.0857
$'Expected surveillance time'
[1] 7.852304
# Agora nós encontraremos o valor de T necessário a um poder de 0.9 (com
RR=2 fixado) no
# teste sequencial contínuo com a mesma probabilidade de erro Tipo I igual a
0,04573175,
# que é o mesmo obtido com o desenho por agrupamento.

```

```

T<- SampleSize.Poisson(alpha= 0,04573175,power=0.9,M=6,D=0,RR=2)[[1]]
# O valor de T resultante é 16,1551.
# O valor crítico para o desenho contínuo é obtido com:
cv1<- CV.Poisson(SampleSize=T,D=0,M=6,alpha=0,04573175)
# Aqui estão as características de performance do procedimento contínuo sob
# a hipótese alternativa (RR=2):
Performance.Poisson(SampleSize=T, D = 0, M = 6, c=cv1, RR = 2)
Power SignalTime SurveillanceTime
[1,] 0.9000002 5.986876 7.003697

```

## Agradecimentos

Sou muito grato à comissão organizadora do XIII MGEst, 2014, que me ofereceu a oportunidade de apresentar o presente trabalho, na forma de um tutorial, naquele evento. Agradeço também à Anandra Santos Ribeiro de Oliveira pela grande ajuda na definição e descrição sobre eventos adversos utilizadas neste artigo.

## References

- ABT, K. Poisson Sequential Sampling Modified Towards Maximal Safety in Adverse Event Monitoring. *Biometrical Journal*, v.40, n.1, p.21-41, 1998.
- BANGDIWALA, S. I. A Sequential Likelihood Ratio Test for General Hypotheses. *Sequential Analysis*, v.1, n.1, p.57-80, 1982.
- COOK, A. J.; TIWARI, R. C.; WELLMAN, R. D.; HECKBERT, S. R.; LI, L.; HEAGERTY, P.; MARSH, T.; NELSON, J. C. Statistical approaches to group sequential monitoring of postmarket safety surveillance data: current state of the art for use in the Mini-Sentinel pilot. *Pharmacoepidemiol Drug Saf.*, Bethesda, v.1, p.72-81, 2012.
- GANDY, A. Sequential Implementation of Monte Carlo Tests With Uniformly Bounded Resampling Risk. *Journal of the American Statistical Association*, Alexandria, v.104, n.488, p.1504-1511, 2009.
- GHOSH, B. K. Introduction to Wald (1945) Sequential Tests of Statistical Hypotheses. *Breakthroughs in Statistics - Foundations and Basic Theory*, New York, v.1, p.249-255, 1992.
- GHOSH, M.; MUKHOPADHYAY, N.; SEN, P. K. *Sequential Estimation*. New York, NY: John Wiley and Sons Inc., 1997.
- GOVINDARAJULU, Z. *Sequential Statistics*. Singapore: World Scientific Publishing Company, 2004.
- HOEL, D. G.; WEISS, G. H.; SIMON, R. Sequential Tests for Composite Hypotheses with Two Binomial Populations. *Journal of Royal Statistical Society B*, v.38, n.3, p.302-308, 1976.
- HOLM, S. On the Optimality of Differentiated SPR Tests of Composite Hypotheses. *Metrika*, v.32, n.1, p.5-33, 1985.
- HUANG, W. Stepwise Likelihood Ratio Statistics in Sequential Studies. *Journal of Royal Statistical Society B*, v.66, p.401-409, 2004.
- JENNISON, C.; TURNBULL, B. W. *Group Sequential Methods with Applications to Clinical Trials*. Boca Raton, London, FL: Chapman and Hall/CRC, 2000.
- JOANES, D. N. Sequential Tests of Composite Hypotheses. *Biometrika*, Oxford, v.59, n.3, p.633-637, 1972.
- KULLDORFF, M.; SILVA, I. R. Continuous Sequential Analysis with a Delayed Start. *Submitted*, 2015.
- KULLDORFF, M.; DAVIS, R. L.; KOLCZAK, M.; LEWIS, E.; LIEU, T.; PLATT, R. A maximized sequential probability ratio test for drug and vaccine safety surveillance.

*Sequential Analysis*, v.30, n.1, p.58-78, 2011.

LACHIN, J. M. Sequential Clinical Trials for Normal Variates Using Interval Composite Hypotheses. *Biometrics*, Washington, v.37, n.1, p.87-101, 1981.

LAI, T. L. Sequential Analysis: Some Classical Problems and New Challenges. *Statistica Sinica*, v.11, p. 303-351, 2001.

LAI, T. L. Asymptotic optimality of generalized sequential likelihood ratio tests in some classical sequential testing problems. *Sequential Analysis*, v.21, n.4, p.121-144, 2002.

LI, L.; KULLDORFF, M. A Conditional maximized sequential probability ratio test for pharmacovigilance. *Statistics in Medicine*, v.29, n.2, p.284-295, 2010.

LIEU, T. A.; KULLDORFF, M.; DAVIS, R. L.; LEWIS, E. M.; WEINTRAUB, E.; YIH, W. K.; YIN, R.; BROWN, J. S.; PLATT, R. Real-Time Vaccine Safety Surveillance for the Early Detection of Adverse Events. *Medical Care*, v.45(S), p.89-95, 2007.

MEEKER, W. Q. A Conditional Sequential Test for the Equality of Two Binomial Proportions. *Applied Statistics*, v.30, n.2, p.109-115, 1981.

MINISTÉRIO DA SAÚDE, Portaria n.2.510, Gabinete do Ministro, de 19 de dezembro de 2005. Institui a Comissão para Elaboração da Política de Gestão Tecnológica no âmbito do Sistema Único de Saúde CPGT. *Ministério da Saúde*, Brasil, 2005.

MONTEIRO, S. A. M. G.; TAKANO, O. A.; WALDMAN, E. A. Avaliação do sistema brasileiro de vigilância de eventos adversos pós-vacinação. *Revista Brasileira de Epidemiologia*, v.14, n.3, p.361-371, 2011.

MUKHOPADHYAY, N.; DE SILVA, B. M. *Sequential Methods and Their Applications*. Boca Raton, London, FL: Chapman and Hall/CRC, 2009.

O'BRIEN, P. C.; FLEMING, T. R. A multiple testing procedure for clinical trials. *Biometrics*, Washington, v.35, n.3, p.549-556, 1979.

PLATT, R. et al. The U.S. Food and Drug Administration Mini-Sentinel program: status and direction. *Pharmacoepidemiology and Drug Safety*, v.21(S1), p.1-8, 2012.

POCOCK, S. J. Group sequential methods in the design and analysis of clinical trials. *Biometrika*, v.64, n.2, p.191-199, 1977.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2014. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

SCHIPPER, M.; DEN HARTOG, J.; MEELIS, E. Sequential Analysis of Environmental Monitoring Data: Optimal SPRTs. *Environmetrics*, v.8, n.1, p.29-41, 1997.

SIEGMUND, D.; GREGORY, P. A Sequential Clinical Trial for Testing  $p_1 = p_2$ . *Annals of Statistics*, v.8, n.6, p.1219-1228, 1980.

SILVA, I. R.; KULLDORFF, M. Sequential-package. *R Foundation for Statistical Computing*, Vienna, 2013. URL <http://www.R-project.org>.

SILVA, I. R.; KULLDORFF, M. Continuous versus Group Sequential Analysis for Post-Market Drug and Vaccine Safety Surveillance. *Biometrics: sob revisão*, 2015.

VAN DER TWEEL, I.; KAAKS, R.; VAN NOORD, P. A. H. Comparison of One-Sample Two-Sided Sequential t-Tests for Application in Epidemiological Studies. *Statistics in Medicine*, Bethesda, v.15, n.24, p.2781-2795, 1996.

WALD, A. Sequential Tests of Statistical Hypotheses. *Annals of Mathematical Statistics*, v.16, p.117-186, 1945.

WALDMAN, E. A.; LUHM, K. R.; MOREIRA, S. A.; MONTEIRO, G.; FREITAS, F. R. M. Vigilância de eventos adversos pós-vacinação e segurança de programas de imunização. *Revista de Saúde Pública*, São Paulo, v.45, n.1, p.173-184, 2011.