

# Contribuições do estatístico para o atendimento ao CONSORT Checklist

*Contributions of the statistician for the accordance with the CONSORT Checklist*

Edson Zangiacomi Martinez

## RESUMO

Este artigo é baseado na palestra com mesmo título apresentada pelo autor no encontro “Planejamento e Condução de Estudos Clínicos de Alta Evidência Científica” organizado pelo Laboratório de Análise da Postura e do Movimento Humano (LAPOMH) no dia 5 de julho de 2008 na Faculdade de Medicina de Ribeirão Preto, USP. O artigo discute o papel do estatístico no delineamento, análise de dados e interpretação de resultados de um ensaio clínico, e sua contribuição ao atendimento do CONSORT checklist.

**Palavras-chave:** CONSORT. Bioestatística. Biometria. Método de Bayes. Teorema de Bayes. Ensaio Clínico.

## Introdução

O CONSORT (*Consolidated Standards of Reporting Trials*) pode ser definido como um conjunto mínimo de recomendações para a apresentação dos resultados de ensaios clínicos controlados aleatorizados<sup>1-3</sup>. Compreende um *checklist* contendo 22 itens essenciais para a apresentação dos resultados do ensaio, além de um fluxograma que explicita o número de indivíduos que compõem as fases de recrutamento, alocação em grupos, seguimento e análise de dados. O *checklist* contém recomendações direcionadas a cinco partes de uma publicação científica: o título e o resumo, a introdução, os métodos, os resultados e a discussão. As contribuições de um estatístico são essenciais não somente na escolha e na aplicação de uma metodologia de análise de dados, mas em ou-

tras etapas do planejamento e da condução de um ensaio clínico. Como mencionado por Pocock<sup>4</sup>, “*It is a common mistake to assume that the statistician need only be concerned with the analysis of results. Of course the statistician plays a major role as data analyst but he should also be involved beforehand in the study’s design and conduct.*”

Pocock<sup>4</sup> ainda acrescenta: “*I think an experienced statistician should be a collaborating scientist in ensuring that both protocol design and the interpretation of trial findings conform to sound principles of scientific investigation. In addition, the statistician is often in a good position to act as a ‘policeman’ in ensuring that satisfactory organizational standards are maintained throughout a trial.*” Por outro lado, o estatístico é muitas vezes requisitado para colaborar em um ensaio clínico ape-

1. Professor Associado. Departamento de Medicina Social da Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo

Correspondência:  
Departamento de Medicina Social  
Faculdade de Medicina de Ribeirão Preto - USP  
Av. Bandeirantes, 3900.  
14049-900 - Ribeirão Preto - SP.  
E-mail: edson@fmrp.usp.br

Artigo recebido em 18/12/2008  
Aprovado em 20/05/2009

nas na etapa de análise de dados. Conforme Pocock <sup>4</sup>, “*Unfortunately, statisticians are often used merely as a technical service for analysing data.*” O estatístico pode ter uma participação bastante abrangente nas etapas prévias à coleta dos dados, como na determinação do tamanho amostral, nas estratégias de alocação dos pacientes em grupos de tratamentos e no planejamento de análises interinas. É atribuída ao célebre estatístico e geneticista inglês Ronald Aylmer Fisher (1890-1962) a frase: “*To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of.*”

O presente artigo tem por objetivo discutir como o estatístico pode contribuir no atendimento dos diferentes itens do CONSORT *checklist*, principalmente no que se refere os itens sete (tamanho amostral) e 12 (análise estatística).

## Tamanho Amostral

O item número sete do CONSORT *checklist* mostra a importância de descrever como o tamanho amostral utilizado no ensaio clínico foi determinado, quando o estudo estava sendo delineado. O número de participantes de um ensaio deve ser necessariamente voltado ao item número cinco do CONSORT *checklist*, que diz respeito aos objetivos e às hipóteses do estudo. Isto significa que o pesquisador deve estimar o número de pessoas que devem ser incluídas no ensaio para que os seus propósitos sejam alcançados de maneira satisfatória. Um ensaio clínico que utiliza um tamanho amostral maior que o necessário para atender aos seus propósitos pode infligir questões éticas, dado que um número desnecessário de indivíduos é neste caso exposto a possíveis riscos, desconfortos e inconvenientes trazidos pelo tratamento ou outra intervenção em estudo. Além disso, implica em desperdício de tempo e de recursos financeiros, materiais e humanos. Por outro lado, um tamanho amostral menor que o necessário traz resultados imprecisos e consequentemente inconclusivos.

Outras questões relevantes em relação ao tamanho amostral são apresentadas por Pocock <sup>4</sup>:

a) *Qual é a principal medida de desfecho?* O desfecho pode ser um evento binário, como morte/sobrevida, cura/não cura, apresentou/não apresentou uma reação adversa ao tratamento, ou uma resposta contínua, como níveis séricos de colesterol ou pressão arterial sanguínea. Pode ser ainda

uma escala, como a escala visual-analógica de dor. Assim, o desfecho é representado por uma variável aleatória, que pode ter natureza quantitativa ou qualitativa.

- b) *Como os dados serão analisados para que uma diferença entre os tratamentos seja detectada?* A escolha de uma ferramenta estatística de análise depende dos objetivos do ensaio, do desenho do estudo e da natureza da variável de desfecho. Por exemplo, os dados de um ensaio clínico com coleta prospectiva de dados, onde se objetiva a comparação entre os tempos de sobrevivência entre dois grupos paralelos de tratamento, podem ser analisados com técnicas de análise de sobrevivência, enquanto os dados de um ensaio clínico onde se deseja comparar as proporções de pacientes que responderam a um tratamento ou a um placebo podem ser analisados por um modelo de regressão logística onde variáveis de confundimento podem ser previstas. A determinação do tamanho amostral deve ser compatível à ferramenta estatística prevista para a análise dos dados.
- c) *Que resultado seria encontrado no grupo de controles?* Para determinar o tamanho amostral, precisamos de uma idéia prévia à coleta dos dados do resultado que seria encontrado para os pacientes que recebem o placebo ou o tratamento padrão.
- d) *Qual seria a menor diferença entre os tratamentos considerada importante do ponto de vista clínico, e com qual grau de incerteza?* Digamos que, com base em estudos prévios, é esperado que 20% dos indivíduos que recebem o placebo apresentam alguma melhora clínica (ver item anterior). Para que o tratamento em estudo seja considerado “útil”, qual deve ser a proporção mínima de indivíduos com melhora clínica no grupo que recebe este tratamento? Se esta proporção for fixada em 30%, isto significa que é estabelecido que a menor diferença entre os tratamentos considerada importante é  $30\% - 20\% = 10\%$ , e qualquer diferença entre estas porcentagens menor que 10% não é considerada importante do ponto de vista clínico.

Em um exemplo, um pesquisador deseja investigar o efeito de um tratamento com baixas doses de aspirina em pacientes com angina crônica estável. Pretende-se conduzir um ensaio clínico controlado aleatorizado, com grupos paralelos, onde metade dos pacientes elegíveis será aleatoriamente alocada a um

grupo que receberá o tratamento com aspirina e a outra metade irá compor um grupo de controle, onde será administrado um placebo. Se a principal medida de desfecho é a ocorrência de infartos do miocárdio, ou seja, um evento binário, podemos utilizar a equação:

$$n = \frac{p_1(100\% - p_1) + p_2(100\% - p_2)}{(p_2 - p_1)^2} f(\alpha, \beta)$$

para determinar o número  $n$  de indivíduos a serem alocados em cada grupo <sup>4</sup>. Esta equação se baseia em um teste de hipóteses, que tem por hipótese nula que pacientes tratados e não tratados com baixas doses de aspirina têm as mesmas chances de um evento de infarto do miocárdio, e por hipótese alternativa que as chances de um evento de infarto são diferentes entre tratados e não tratados. Nesta equação, temos que:

$p_1$  é a frequência relativa esperada de eventos no grupo que receberá o placebo.

$p_2$  é uma frequência relativa de eventos associada ao grupo de pacientes que receberá o tratamento sob investigação, de forma que  $p_2 - p_1$  é a menor diferença entre os tratamentos considerada importante em um ponto de vista clínico.

$\alpha$  é o nível de significância adotado para o estudo, ou seja, a maior probabilidade tolerada pelo pesquisador para um erro tipo I. No contexto deste ensaio clínico controlado, o erro tipo I seria cometido ao concluirmos que o infarto do miocárdio ocorre com diferentes frequências entre pacientes tratados e não tratados com baixas doses de aspirina, enquanto, na verdade, os dois grupos possuem a mesma frequência de infartos. Na área da saúde, o nível de significância é geralmente fixado em 5%.

$\beta$  é a chance tolerada de um erro tipo II, ou seja, o erro que cometemos ao concluirmos que pacientes tratados e não tratados com baixas doses de aspirina possuem chances iguais de um evento de infarto do miocárdio, quando na realidade um dos grupos apresenta chances diferentes (maior ou menor) de apresentar o desfecho em questão. A quantidade  $1 - \beta$  é chamada de poder do teste de hipóteses para detectar uma diferença entre os grupos de magnitude  $p_2 - p_1$ . Geralmente  $\beta$  é fixado em 5%, 10% ou 20%, dependendo da gravidade das consequências de se cometer um erro deste tipo.

$f(\alpha, \beta)$  é uma função de  $\alpha$  e  $\beta$ , dada por  $f(\alpha, \beta) = [\Phi^{-1}(\alpha/2) + \Phi^{-1}(\beta)]^2$ ,

onde  $\Phi$  é a função distribuição acumulada de uma distribuição normal padrão e  $\Phi^{-1}$  é a sua função inversa. Podemos encontrar valores para  $\Phi^{-1}$  facilmente, com a ajuda da função INV.NORMP do programa Microsoft Excel. Por exemplo,  $\Phi^{-1}(0,05)$  é obtido digitando a expressão =INV.NORMP(0,05) em uma célula vazia da planilha do Excel, onde obtemos que  $\Phi^{-1}(0,05) = -1,645$ . Se fixarmos  $\alpha = 0,05$  e  $\beta = 0,20$ , encontramos o valor de  $f(0,05;0,20)$  digitando em uma célula vazia do Excel a expressão

$$=(INV.NORMP(0,05/2)+INV.NORMP(0,20))^2$$

que trará o resultado  $f(0,05;0,20) = 7,85$ .

Digamos que é conhecido que a porcentagem esperada de ocorrências de infarto do miocárdio no grupo que receberá o placebo é de  $p_1 = 13\%$  (a frequência esperada em indivíduos que não recebem o tratamento). É necessário ainda especificar a menor diferença entre os tratamentos considerada importante do ponto de vista clínico. Digamos que se o tratamento trouxer uma redução absoluta do risco de infarto do miocárdio menor que 10% ele não será considerado importante, mas reduções iguais ou superiores a 10% indicam que o tratamento cumpre com o seu propósito. Temos assim  $p_2 = 3\%$  (ou seja,  $p_2 - p_1 = -10\%$ ). Assim, o número necessário de indivíduos em cada um dos grupos, necessário para detectar uma diferença absoluta de riscos de infarto igual a 10% entre estes grupos, é:

$$n = \frac{13\%(100\% - 13\%) + 10\%(100\% - 10\%)}{(-10\%)^2} \times 7,85 = 191$$

considerando que a maior probabilidade tolerada de um erro tipo I é  $\alpha = 5\%$  e a maior probabilidade tolerada de um erro tipo II é  $\beta = 20\%$ .

Este exemplo demonstra que o tamanho amostral necessário a um ensaio clínico só pode ser determinado após o pesquisador definir com clareza o objetivo do estudo, a população em estudo (item 3 do CONSORT checklist) através dos critérios de elegibilidade, especificar a hipótese nula e a hipótese alternativa e determinar as probabilidades de erro tipo I e tipo II. Se o estudo mede o tempo até a ocorrência de um evento (análise de sobrevivência), é preciso especificar o processo de recrutamento e definir o tamanho do período de seguimento <sup>5</sup> (follow-up period).

Outras estratégias para a determinação de tamanhos amostrais em ensaios clínicos onde a medida de desfecho é expressa por uma variável quantitativa ou o desenho prevê medidas pareadas são apresentadas por Lachin<sup>6</sup>. Importantes discussões sobre tamanhos amostrais de ensaios clínicos encontram-se ainda nos artigos de Schulz e Grimes<sup>7</sup>, Brasher e Brant<sup>8</sup> e Sylvester<sup>9</sup>.

O item número sete do CONSORT *checklist* trata ainda a necessidade de explicações sobre as *análises interinas*<sup>10</sup>, quando utilizadas. Alguns ensaios recrutam participantes por um período de tempo relativamente longo, sendo as informações sobre o desempenho do tratamento sucessivamente acumuladas. Nestas situações, as *análises interinas* são feitas sobre os resultados acumulados, enquanto o ensaio está em progresso. De acordo com os resultados destas análises, o ensaio pode ser encerrado antes de atingir o tamanho amostral previamente especificado. Em um estudo onde é esperado que uma nova terapia traga resultados melhores que um placebo, por exemplo, o estudo pode ser interrompido tão logo se percebam evidências suficientes de que a nova terapia seja superiora. Encerrar o estudo antes do prazo previsto traz vantagens econômicas e éticas, dado que não haverá mais sentido tratar indivíduos com um placebo, quando já se sabe que há um tratamento adequado.

Os detalhes de uma análise interina devem ser necessariamente especificados antes da coleta dos dados: quais os momentos previstos para a análise interina, os critérios de interrupção, o tamanho amostral estimado<sup>11</sup>, como os dados serão analisados, qual a precisão especificada para classificar o tratamento como superior ao placebo e como as análises estatísticas serão corrigidas. A prática de múltiplas análises estatísticas ao longo do ensaio, com um acúmulo de informações, exige alguma correção nas inferências, dado que a probabilidade de erro tipo I (nível de significância) também será cumulativa ao longo destas análises, e, ao longo do tempo, maior que o valor nominal fixado antes da coleta dos dados.

## Análise dos Dados

Segundo Pocock<sup>4</sup>, “*The essence of a good clinical trial is that it provides truthful and precise information which is relevant to the treatment of future patients. To this end statistical methods provide a scientific basis for trial design. Methods of greatest value must be simple, reliable and*

*readily understood by non-statisticians.*” Desta maneira, a ferramenta estatística escolhida para a análise dos dados de um ensaio clínico não pode ter uma complexidade maior que os dados; problemas simples demandam ferramentas simples, mas problemas complexos em relação ao desenho ou à distribuição dos dados exigem ferramentas apropriadas. Alguns autores apresentam em livros-texto tabelas ou fluxogramas que orientam pesquisadores a escolher uma ferramenta estatística de acordo apenas com a natureza de suas variáveis. Estes procedimentos, conhecidos por “*cookbook approaches*”, são criticados por Murray<sup>12</sup>: “*The selection of an appropriate statistical test requires experience and insight, and one must be wary of a proscriptive cookbook approach.*” Murray<sup>12</sup> coloca ainda que “*the choice of an appropriate statistical procedure requires both an understanding of the strengths and limitations of the different procedures, and also a deep understanding of the clinical context.*” Este pensamento é compatível com o de Paes<sup>13</sup>: “*Da mesma forma que a prática da Medicina não é completamente objetiva, a da Estatística também não. Não existem ‘receitas prontas’ para tratar doentes, assim como não existem fluxogramas que indiquem as técnicas estatísticas que devem ser utilizadas em cada caso. O que existem são ‘práticas comuns’ que podem ser aplicadas ou não, dependendo das condições do estudo.*”

Assim, a ferramenta de análise dos dados deve ser escolhida conforme os objetivos do estudo, a natureza das variáveis que definem as medidas de desfecho, a medida de tamanho de efeito de tratamento (risco relativo, redução de risco, número necessário para tratar) e o desenho do ensaio, e esta escolha exige um profundo entendimento do estatístico sobre as características do estudo. Os “*cookbook approaches*” devem dar lugar à experiência do estatístico e à sua capacidade de lidar com diferentes métodos de tratamento de dados.

## Descrever, antes de analisar

É importante considerar que mesmo as ferramentas simples, como o teste *t* de Student para comparação de médias populacionais, trazem algumas armadilhas. Por exemplo, seja um ensaio clínico aleatorizado onde são comparados dois tratamentos, A e B, tal que as respostas aos tratamentos são expressas por uma variável de natureza contínua. A Figura 1 descreve a distribuição destas respostas por

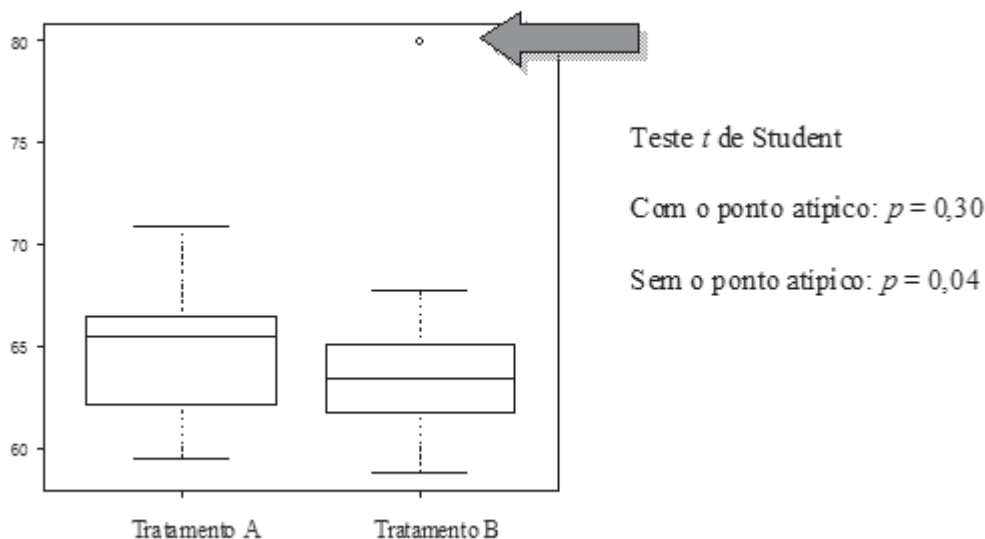
*box-plots*<sup>14</sup>, onde é evidente que um indivíduo submetido ao tratamento B apresentou uma resposta representada por um valor de magnitude demasiadamente grande em relação aos demais. Uma seta destaca na figura este valor atípico. Se aplicarmos aos dados um teste *t* de Student para amostras independentes incluindo este valor atípico, temos um *p* valor de 0,30, que nos leva à não rejeitarmos a hipótese de igualdade de médias populacionais a um nível de significância de 5%. Por outro lado, se retirarmos este ponto atípico do banco de dados, o mesmo teste de hipóteses nos fornecerá um *p* valor de 0,04, agora inferior ao nível de significância de 5%, o que nos leva à rejeitar a hipótese de igualdade de médias. Este exemplo nos mostra que a inclusão ou a exclusão de um valor atípico em relação aos demais em um banco de dados pode nos trazer resultados distintos para o teste de hipóteses.

Se não tomarmos o cuidado de descrevermos detalhadamente os achados de um ensaio clínico, corremos o risco de escolhermos uma ferramenta analítica inadequada para o estudo. Esta etapa de descrição dos achados, que necessariamente antecede a análise dos dados, é chamada *análise exploratória de dados* (AED). A AED foi introduzida por John Wilder Tukey (1915-2000), no seu clássico livro *Exploratory Data Analysis*<sup>15</sup> lançado em 1977. Faz uso de um grande número de técnicas gráficas (como o *box-plot*) e quantitativas, buscando descrever a distribuição das variáveis numéricas, pontos atípicos (*outliers*) que podem trazer influências às medidas descritivas e aos

resultados das análises, e relações e padrões de associações entre variáveis, a princípio ocultas no banco de dados. Embora o pesquisador tenha a possibilidade de prever a ferramenta estatística de análise de dados quando o estudo está sendo delineado, somente na aplicação da AED ele pode definitivamente decidir por um método estatístico ou outro.

### **P valor: uma medida a ser evitada em ensaios clínicos?**

No ano de 2002, o conceituado periódico *The Lancet* publicou uma série de artigos denominada *Epidemiology Series*. No primeiro artigo da série, Grimes e Schulz<sup>16</sup> escreveram: “*Measures of association, such as relative risk or odds ratio, are the preferred way of expressing results of dichotomous outcomes, e.g., sick versus healthy. Confidence intervals around these measures indicate the precision of these results. Measures of association with confidence intervals reveal the strength, direction, and a plausible range of an effect as well as the likelihood of chance occurrence. By contrast, p values address only chance. Testing null hypotheses at a p value of 0.05 has no basis in medicine and should be discouraged.*” Esta postura dos autores em desencorajar do uso dos *p* valores é dirigida às limitações destas medidas, embora elas sejam bastante frequentes nos periódicos da área médica. Segundo Wagner<sup>17</sup>, “*O rótulo ‘estatisticamente significante’*



**Figura 1:** Distribuição das respostas aos tratamentos A e B de um ensaio clínico aleatorizado hipotético ilustradas por *box-plots*, onde percebemos que um ponto atípico influencia o resultado de um teste *t* de Student para amostras independentes.

*é potencialmente enganoso e tem estado sob constantes críticas por parte de vários estatísticos e epidemiologistas durante os últimos anos. Para que se possa avaliar ‘significância’ é igualmente ou mais importante do que o  $p$  valor que se estime o tamanho da associação envolvida. Isso pode ser obtido através de medidas de associação como o risco relativo e seus intervalos de confiança.”*

Os  $p$  valores são derivados dos testes de hipóteses, da teoria clássica estatística. Antes de aplicarmos um teste de hipóteses aos nossos dados, devemos definir o nível de significância, que na área médica é geralmente fixado em 5%. O  $p$  valor é interpretado como o menor valor que o nível de significância deveria ser fixado para que a hipótese nula fosse rejeitada. Em um ensaio clínico aleatorizado controlado, digamos que a hipótese nula estabeleça que o tratamento ou intervenção sob investigação e o tratamento com placebo tragam os mesmos resultados. Um  $p$  valor de 67% (0,67), por exemplo, indica que deveríamos ter especificado um nível de significância de pelo menos 67% para que a hipótese nula fosse rejeitada. Sendo o nível de significância especificado em 5% (menor que 67%), não rejeitamos a hipótese nula. Por outro lado, se o  $p$  valor é menor que o nível de significância adotado, rejeitamos a hipótese nula.

Uma limitação do  $p$  valor é que esta medida não informa o tamanho da diferença entre os resultados da utilização do tratamento sob investigação e do placebo, o que é de grande utilidade na tomada de decisões clínicas. Por sua vez, riscos relativos ou medidas de reduções absolutas ou relativas de risco informam o tamanho destas diferenças, e traduzem a importância clínica da intervenção. A “significância

estatística” não significa “importância clínica”, e a expressão “estatisticamente significativa” pode ser bastante vaga para descrever os resultados de um ensaio clínico<sup>18-21</sup>.

Uma característica importante de um teste de hipóteses é que a sua probabilidade de rejeitar a hipótese nula é sempre relacionada ao tamanho amostral. Tamanhos amostrais pequenos trazem chances pequenas de rejeitar a hipótese de igualdade entre os resultados do tratamento sob investigação e do grupo tratado com placebo, e conseqüentemente, trazem  $p$  valores grandes, ainda que sejam observadas grandes diferenças entre estes grupos. Por outro lado, o uso de amostras grandes traz  $p$  valores pequenos, mesmo que as diferenças entre os dois grupos de tratamentos não sejam relevantes em um ponto de vista clínico.

### Metodologia bayesiana: o futuro dos ensaios clínicos?

A metodologia bayesiana teve uma origem bastante incomum<sup>22</sup>. Ela nasceu de um processo de raciocínio idealizado por Thomas Bayes, um pároco da igreja presbiteriana que viveu em um vilarejo da Inglaterra do século 18, supostamente entre 1701 e 1761. Este raciocínio foi registrado na forma de um ensaio sobre probabilidades condicionais, que Bayes, curiosamente, nunca submeteu para a sua publicação. Dois anos após a morte de Bayes, seu amigo Richard Price (1723-1791) encontrou entre os pertences do pároco o tal manuscrito, que tinha o título “*Na essay towards solving a problem in the doctrine of chances*”. Price encaminhou o ensaio com a demonstração do futuramente famoso Teorema de Bayes à *Royal Society* inglesa, que logo o publicou. Nos dias de hoje, este

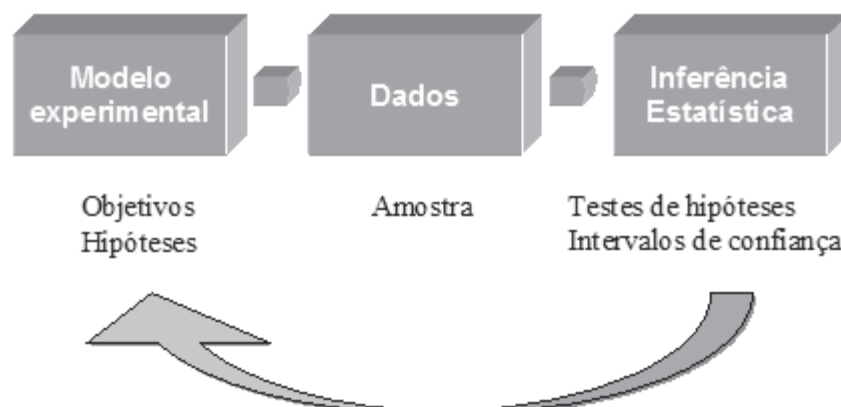


Figura 2: A filosofia dominante na pesquisa: o método frequentista.

ensaio é tido como a base de uma verdadeira revolução em diferentes campos do conhecimento, da genética à teologia <sup>22</sup>.

A filosofia dominante na pesquisa clínica é baseada na chamada *interpretação frequentista*, cujo processo é ilustrado na Figura 2. A partir de um modelo experimental, onde são definidos os objetivos e as hipóteses do ensaio, retira-se aleatoriamente uma amostra da população de interesse (definida pelos critérios de inclusão e exclusão do estudo), sendo os dados assim obtidos submetidos a um processo de inferência estatística, onde testes de hipóteses e intervalos de confiança permitem uma extrapolação dos resultados à população. Os dados, neste processo, trazem uma representação do universo físico, e são descritos matematicamente por uma expressão denominada *função de verossimilhança*.

Na metodologia Bayesiana, estabelece-se que o processo de inferência não é baseado somente no universo físico descrito pelos dados (verossimilhança), mas também no conhecimento prévio do pesquisador sobre o objeto investigado. Este conhecimento prévio traz a subjetividade do investigador, cuja razão é agora integrante do processo. Esta subjetividade, quando expressa por uma função de probabilidades, é denominada *distribuição a priori* do objeto investigado. O teorema de Bayes estabelece que inferências racionais sejam obtidas da chamada *distribuição a posteriori* do objeto (Figura 3), que é proporcional ao produto entre a distribuição a priori (subjetiva) e a função de verossimilhança (universo físico).

Nota-se que a metodologia bayesiana de pesquisa não tem como primeiro passo a observação, mas a especificação do que já se sabe acerca do objeto em investigação. No entanto, como este conhecimento a priori é dependente das crenças e da experiência do observador, os resultados obtidos da distribuição a posteriori não são dependentes apenas dos dados observados como ocorre no método frequentista. Segundo Gill et al <sup>23</sup>, os clínicos são naturalmente bayesianos, ainda que não conheçam formalmente as idéias do bayesianismo. Ao examinar um paciente (universo físico) para um diagnóstico (inferência racional acerca de uma condição), o clínico não se baseia somente em seus dados (o paciente), mas alia a sua experiência prévia e conhecimentos adquiridos em sua formação médica (a razão humana) às condições do paciente (os dados: exames físicos e laboratoriais, sintomas, sinais, queixas), como no teorema de Bayes.

Assim, a estatística bayesiana é baseada no teorema proposto por Bayes (Figura 3). A partir dela, obtemos os *intervalos de credibilidade* para os objetos sob investigação (parâmetros), análogos aos intervalos de confiança resultantes da aplicação do método frequentista. Todos os modelos de regressão do método frequentista (regressão linear e não linear, logística, de Poisson, de Cox para dados de sobrevivência) encontram as suas versões em uma ótica bayesiana. Exemplos de aplicações de métodos bayesianos em ensaios clínicos são encontrados em Greenhouse<sup>24</sup>.

Berry<sup>25</sup> menciona que o uso da metodologia bayesiana para delinear um ensaio clínico é mais va-



Figura 3 – A filosofia bayesiana: o teorema de Bayes (adaptado de Pena <sup>22</sup>).

loroso do que o ato de utilizar ferramentas bayesianas somente para a análise dos resultados. A metodologia bayesiana fornece um mecanismo que permite uma atualização dos resultados acumulados durante o ensaio<sup>26</sup>. Os resultados acumulados podem ser avaliados a qualquer momento, inclusive de forma contínua, com a possibilidade de alterar o desenho do ensaio, por exemplo, desbalanceando a aleatorização para favorecer terapias de melhor desempenho, eliminando ou adicionando braços de tratamentos, e alterando a população de estudo para focar subgrupos de doentes que estão respondendo melhor às terapias experimentais<sup>26,27</sup>.

Alguns autores mencionam que uma vantagem da abordagem bayesiana em ensaios clínicos é a possibilidade de se obter a chamada *distribuição preditiva de probabilidade*, que permite a predição de resultados futuros com base nos resultados atuais do ensaio<sup>24</sup>. Outras vantagens são uma estrutura mais flexível para o monitoramento de ensaios sequenciais<sup>28-31</sup>, a não dependência de tamanhos amostrais grandes (resultados assintóticos) para a validação das inferências e a habilidade de calcular a probabilidade de um tratamento ser mais efetivo que outro<sup>32</sup>.

A Figura 4 ilustra o crescimento do número de artigos publicados em revistas da área médica que fazem algum uso ou menção de ferramentas bayesianas, encontrados em uma busca na base de dados PubMed<sup>33</sup>, utilizando a expressão “*bayesian*”. A figura mostra o número de referências encontradas a cada 10 mil artigos indexados, a cada ano, de 1960 a 2007. Observa-se uma incidência inexpressiva de artigos até meados da década de 80 do último século,

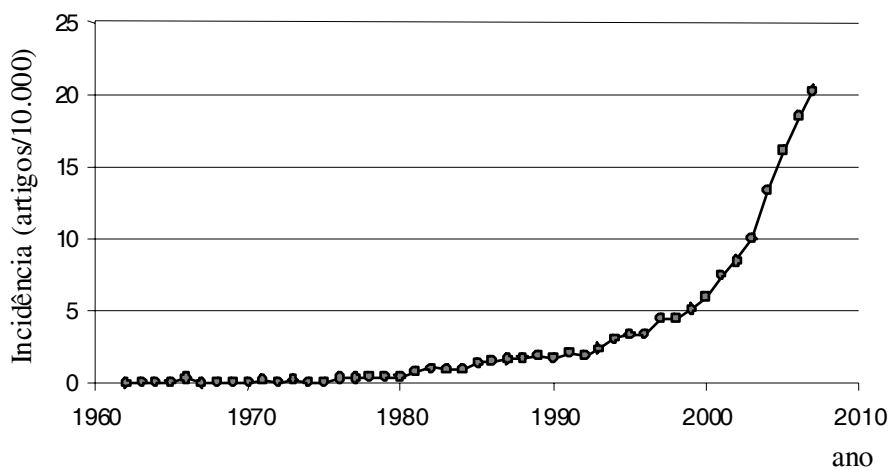
quando se inicia um crescimento propiciado pela popularização dos computadores portáteis, que possibilitaram o uso de algoritmos Monte Carlo em Cadeia de Markov (MCMC), comuns na estimação das distribuições a posteriori. A partir do ano 2000, o uso de técnicas bayesianas na pesquisa médica encontra um novo crescimento, agora propiciado pelo programa de computador WinBugs<sup>34</sup> de domínio público.

## Discussão

O presente artigo mostra que o estatístico tem papel fundamental do delineamento, na análise e na interpretação de dados de estudos clínicos. Sua contribuição ao atendimento dos itens do CONSORT checklist está principalmente na adequada descrição da determinação do tamanho amostral do ensaio, na estratégia de alocação de indivíduos em grupos, na escolha das ferramentas de análise de dados e no auxílio à interpretação dos resultados. Mas, para a sua contribuição ser efetiva, é necessário a ele um amplo conhecimento dos alcances e limitações de cada método estatístico, bem como a habilidade de bem compreender os propósitos do ensaio clínico. Usar inadequadamente uma ferramenta de análise estatística, ainda que simples, pode trazer prejuízos aos resultados no ensaio, como exposto na seção 3.1. Por isso, o estatístico deve ser hábil em descrever os dados do ensaio, destacando as suas características mais relevantes e sendo capaz de tomar decisões sobre as ferramentas analíticas adequadas aos achados desta análise exploratória. É importante que ele conheça as li-

mitações dos *p* valores e as maneiras alternativas e complementares de exibir os resultados do ensaio, descrevendo as magnitudes das discrepâncias entre os grupos de pacientes submetidos ao tratamento investigado e ao placebo por medidas adequadas.

Sendo crescente o uso de técnicas bayesianas na pesquisa em medicina e saúde, é desejável que o estatístico passe por um treinamento formal em estatística bayesiana para tornar-se apto a delinear e analisar dados de ensaios clínicos utilizando tal metodologia.



**Figura 4:** A incidência de artigos encontrados na base de dados. PubMed utilizando a palavra chave “*bayesian*”, de 1960 a 2007.



## ABSTRACT

This article is based on the lecture with the same title presented by the author at the “Planning and Conducting Clinical Trials of High Scientific Evidence” meeting organized by the LAPOMH Laboratory which was held at the Ribeirão Preto Medical School (University of São Paulo) on 5 July 2008. The article discusses the role of the statistician in the design, data analysis and interpretation of the results from a clinical trial, and its contribution to improving the accordance with the CONSORT checklist.

**Keywords:** CONSORT. Biostatistic. Biometry. Bayesian Method. Bayes' Rule. Clinical Trials.

## Referências

1. CONSORT Statement Website. <http://www.consort-statement.org/>
2. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996; 276(8):637-9.
3. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001; 357(9263):1191-4.
4. Pocock SJ. *Clinical trials: a practical approach*. Chichester: John Wiley & Sons, 1983.
5. Wittes J. Sample size calculations for randomized controlled trials. *Epidemiol Rev* 2002; 24(1):39-3.
6. Lachin JM. Introduction to sample size determination and power analysis for clinical trials. *Control Clin Trials* 1981; 2(2): 93-113.
7. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet* 2005; 365 (9467): 1348-53.
8. Brasher PMA, Brant RF. Sample size calculations in randomized trials: common pitfalls. *Can J Anesth* 2007; 54(2): 103-6.
9. Sylvester R. Statistical Principles: Myths or Facts? *Onkologie* 2003; 26:520-1.
10. Posch M, Bauer P. Interim analysis and sample size reassessment. *Biometrics* 2000; 56(4):1170-6.
11. Bristol DR. Sample size determination using an interim analysis. *J Biopharm Stat* 1993; 3(2):159-66.
12. Murray GD. Reply from BJS Statistical Adviser. *Br J Surg* 1996; 83(1): 196.
13. Paes AT. Itens essenciais em bioestatística. *Arq. Bras. Cardiol* 1998; 71(4): 575-580.
14. Williamson DF, Parker RA, Kendrick JS. The box plot: a simple visual method to interpret data. *Ann Intern Med* 1989; 110(11): 916-21.
15. Tukey JW. *Exploratory data analysis*. Reading: Addison-Wesley, 1977.
16. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. *Lancet* 2002; 359(9300):57-61.
17. Wagner M. Significância com confiança? *J pediatr (Rio J)* 1998; 74(4):343-6.
18. Allison GT. Statistical and clinical interpretation of research results. *J Am Pediatr Med Assoc* 2007; 97(2):165-70.
19. Bhardwaj SS, Camacho F, Darrow A, Fleischer AB Jr, Feldman SR. Statistical significance and clinical relevance: the importance of power in clinical trials in dermatology. *Arch Dermatol*. 2004; 140(12):1520-3.
20. Luus HG, Muller FO, Meyer BH. Statistical significance versus clinical relevance. Part I. The essential role of the power of a statistical test. *S Afr Med J*. 1989; 76(10):568-70.
21. Houle TT, Stump DA. Statistical significance versus clinical significance. *Semin Cardiothorac Vasc Anesth* 2008; 12 (1): 5-6.
22. Pena SD. Thomas Bayes: o 'cara'! *Revista Ciência Hoje* 2006; 228: 22-9.
23. Gill CJ, Sabin L, Schmid CH. Why clinicians are natural bayesians. *BMJ* 2005; 330(7499): 1080-3.
24. Greenhouse JB. On some applications of Bayesian methods in cancer clinical trials. *Stat Med* 1992; 11(1):37-53.
25. Berry DA. Introduction to Bayesian methods III: use and interpretation of Bayesian tools in design and analysis. *Clin Trials* 2005; 2(4):295-300.
26. Berry DA. A case for Bayesianism in clinical trials. *Stat Med* 1993; 12(15-16):1377-93.
27. Berry DA. Bayesian clinical trials. *Nat Rev Drug Discov* 2006; 5(1):27-36.
28. Freedman LS, Spiegelhalter DJ, Parmar MK. The what, why and how of Bayesian clinical trials monitoring. *Stat Med*. 1994; 13(13-14):1371-83.
29. Etzioni RD, Kadane JB. Bayesian statistical methods in public health and medicine. *Annu Rev Public Health*. 1995; 16:23-41.
30. Ashby D, Tan SB. Where's the utility in Bayesian data-monitoring of clinical trials? *Clin Trials* 2005; 2(3):197-205.
31. Fayers PM, Ashby D, Parmar MK. Tutorial in biostatistics Bayesian data monitoring in clinical trials. *Stat Med* 1997; 16(12):1413-30.
32. Lewis RJ, Wears RL. An introduction to the Bayesian analysis of clinical trials. *Ann Emerg Med* 1993; 22(8):1328-36.
33. PubMed Website. <http://www.ncbi.nlm.nih.gov/sites/entrez>
34. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000; 10(4):325-37.