

Análise rotineira de dados de vigilância em saúde pública: que procedimentos estatísticos utilizar?

Routine analysis of public health surveillance data: what statistical procedures to use?

Odécio Sanches*

SANCHES, O. Análise rotineira de dados de vigilância em saúde pública: que procedimentos estatísticos utilizar? *Rev. Saúde Pública*, 27: 300-4, 1993. Tem havido no país, nos últimos anos, crescente interesse por questões vinculadas à área tradicionalmente denominada vigilância epidemiológica (que se denominou vigilância em saúde pública). Tem-se defendido, em muitas publicações, o ponto de vista que o fácil acesso a computadores e programas estatísticos deve facilitar a utilização de procedimentos mais sofisticados na análise de dados provenientes de sistemas de vigilância. Considerando que esta colocação, de forma ampla e indiscriminada, pode levar a análises que não tem sustentação teórica formal, apresenta-se posicionamento a respeito. Esclarece-se sobre o que se entende por vigilância em saúde pública e descreve-se, de modo resumido e não-tecnista, as exigências dos procedimentos estatísticos de estimação por intervalo e de testes de hipóteses estatísticas. Comenta-se sobre a origem dos dados de vigilância e o seu caráter não-amostrável e a ausência de aleatoriedade. Discute-se alguns procedimentos descritivos que são e devem ser usados sem prejuízo da qualidade da análise. Com base na literatura, são apresentados alguns procedimentos que têm sido propostos mas que, para sua inclusão rotineira na análise de dados, ainda carecem de mais investigação.

Descritores: Vigilância epidemiológica. Coleta de dados. Interpretação estatística de dados. Inferência.

1- Introdução

A criação e implantação do Sistema Único de Saúde talvez seja um dos fatores que tem estimulado, desde o 1º Congresso Brasileiro de Epidemiologia, realizado na cidade de Campinas, SP, em 1990, um grande e crescente interesse na discussão de questões envolvendo amplo tema convencionalmente denominado vigilância epidemiológica. Esta afirmação é corroborada pelo número de trabalhos apresentados, conferências e oficinas de trabalho realizadas durante o 2º Congresso ocorrido na cidade de Belo Horizonte, MG, no ano de 1992 e, também, publicações em revistas de saúde como, por exemplo, a revista "Saúde e Sociedade" que, em seu lançamento, em 1992, apresenta trabalho relativamente ao tema²⁴.

Nossa preocupação, demonstrada pelo título que decidimos dar ao presente trabalho, surge, principalmente, do entendimento a que podem levar afirmações como:

"... a disponibilidade de programas epidemiológicos e estatísticos de fácil manuseio abriu as portas da análise epidemiológica, até então, restrita a uns poucos iniciados...", Silva²⁴ ou "Métodos estatísticos mais sofisticados, a disponibilidade de computadores e o desenvolvimento de "softwares" estatísticos, para análise, ampliaram o potencial da análise estatística na prática diária da saúde pública..." Thacker & Berkelman²⁵.

Entendemos que afirmações dessa natureza podem induzir os não iniciados, no sentido estabelecido por Silva, ao uso indiscriminado e, conseqüentemente, inapropriado dos procedimentos estatísticos, tornando as análises estatísticas realizadas destituídas de base e sentido ou, no mínimo, sob suspeita.

Nosso questionamento é que não basta a disponibilidade de programas estatísticos de fácil manuseio para a análise de dados provenientes do processo denominado vigilância. É preciso mais: é preciso saber, entre os procedimentos estatísticos de análise de dados, disponíveis em programas

* Professor visitante - Departamento de Epidemiologia e Métodos Quantitativos em Saúde - Convênio FIOCRUZ/CNPq - Escola Nacional de Saúde Pública/FIOCRUZ - Rio de Janeiro, RJ - Brasil.

Separatas/Reprints: O. Sanches - Rua Leopoldo Bulhões, 1480 8º. andar - Manguinhos - 21041-210 - Rio de Janeiro, RJ - Brasil

computacionais de fácil acesso e manuseio, quais podem ser aplicados tendo-se em vista os pressupostos teóricos formais de tais procedimentos e a origem e natureza dos dados produzidos pelos sistemas de vigilância.

Neste trabalho, para estabelecer nossa posição, inicialmente damos o nosso entendimento sobre o processo de vigilância. Em seguida, de forma sucinta e não tecnicista, abordamos as exigências formais para o uso de procedimentos estatísticos inferenciais e, a partir da análise de como são coletados os dados de vigilância, examinamos as restrições para a utilização dos procedimentos de testes de hipóteses estatísticas e de estimação por intervalo, no contexto frequentista introduzido por Neyman e col.¹⁹. Finalmente, comentamos sobre os vários procedimentos que julgamos, pelo menos razoavelmente, compatíveis com os objetivos e dados da vigilância.

2- O Processo de Vigilância em Saúde Pública

Mason¹⁸, na abertura do "Symposium on Statistics in Surveillance", realizado em 1988, no "Center for Disease Control", Atlanta, USA, quando então era seu Diretor, coloca que a vigilância em saúde pública inclui quatro componentes. A primeira é a coleta sistemática e contínua de dados. Análise e interpretação dos dados é a segunda. Isto deve levar, a seguir, à terceira componente: disseminação dos dados e resultados da análise que devem resultar, se necessário, na implementação de ações, baseadas em planejamento, para controlar ou prevenir agravos ou condições adversas à saúde. Uma quarta componente, fechando um ciclo, deve ser introduzida: avaliação periódica do sistema.

Para nós, contínuo é a palavra chave, como também acentuam Thacker & Berkelman²⁵, isto é, entendemos que inquéritos pontuais ou episódicos não constituem vigilância. Como esses autores, também entendemos que um sistema contínuo de coleta e comparação de dados relativos à saúde das pessoas, baseado em um "survey" amostral, também não constitui um sistema de vigilância em saúde pública (como por exemplo os "surveys" do "National Center for Health Statistics", nos EUA, ou uma possível Pesquisa Nacional por Amostragem de Domicílios (PNAD) de saúde, no Brasil), pois para que constitua um sistema de vigilância em saúde pública haveria a necessidade de que as informações estivessem integradas à administração e avaliação de programas específicos de saúde pública, os quais podem, eventualmente, incluir pesquisas epidemiológicas que levam à prevenção.

Enfim, o processo de vigilância em saúde pública pode ser resumido, como o fez Colton¹⁰: "keep a watchfull eye out".

Poderá parecer estranho a muitos leitores da área que, salvo uma única vez, na introdução, não utilizamos o rótulo convencional e formal de vigilância epidemiológica. Ainda que, oficialmente, este seja o termo estabelecido, temos encontrado, na literatura, posições conflitantes a respeito e, às vezes, quase que radicalmente opostas, como por exemplo:

I - "... A vigilância epidemiológica é o ramo da epidemiologia que angaria estas informações. Vigilância epidemiológica não deve ser vista, portanto, como algo separado e distinto da epidemiologia", Silva²⁴.

II - "... Devido ao conteúdo muito mais amplo da epidemiologia, o uso do termo epidemiológico confunde o significado de vigilância no cenário da saúde pública, tendo havido, no passado, a incorporação indevida da pesquisa na definição de vigilância. Por esta razão não nos manteremos fiéis, aqui, ao uso atual do termo epidemiológico para modificar vigilância. Sugerimos que vigilância em saúde pública seria mais apropriado, pois o seu uso conserva as vantagens iniciais do termo epidemiológico, já anteriormente citado, e desfaz um pouco da confusão em torno da prática atual. A vigilância é, mais precisamente, um elemento da saúde pública e, ao lerem o termo, as pessoas devem compreender isto". Tacker & Berkelman²⁵.

Não é nossa intenção, no momento, travar qualquer tipo de polêmica a respeito de conceitos. No entanto, queremos deixar clara nossa posição de concordância com o rótulo vigilância em saúde pública e mais, inclusive, concordância com aqueles que defendem um só sistema de vigilância sob esta denominação.

3- Procedimentos de Inferência Estatística

Um problema, ou uma situação problema, é inerentemente estatística quando envolve variabilidade de natureza aleatória e, em geral, há interesse em se produzir algum tipo de inferência a respeito de um grande número de elementos (população) a partir de experimentos e/ou observações com um pequeno número de tais elementos (amostra) selecionados por algum tipo de procedimento considerado adequado (desenho amostral).

Isto não é, obviamente, uma definição de estatística, como área de conhecimento, já que tal tarefa não parece ser fácil. De fato, basta percorrer a enorme quantidade de livros textos e periódicos específicos na área para detectar mais de uma centena de definições!

O que queremos ressaltar é que se um levantamento qualquer tem objetivos inferenciais, é possível, em muitas circunstâncias, traduzi-los como questões de inferência estatística. E, dois são os grandes problemas de inferência estatística: os de estimação de parâmetros populacionais e os de testes de hipóteses estatísticas a respeito de parâmetros populacionais.

Quer exista uma proposição distribucional a respeito das variáveis aleatórias em questão (inferência paramétrica), quer não haja tal proposição (inferência livre de distribuição, também rotulada não paramétrica), a proposição de seleção aleatória da amostra é uma questão fundamental para a validade dos procedimentos inferenciais estatísticos, quer segundo a escola "fisheriana", quer segundo a escola "neymanpersoniana" - (se é que podemos usar os neologismos). É sempre bom lembrar que esta proposição de seleção aleatória é, também, uma exigência para a construção de modelos de predição tipo regressão, quer seja linear ou logístico.

4- Origem dos Dados de Vigilância e seu Caráter Estritamente Não-Amostral e Não-Aleatório

Dados de vigilância em saúde pública são coletados através de inúmeras fontes: médicos, laboratórios, hospitais e outros serviços de assistência à saúde tem de notificar todos os casos daquelas doenças, ou estados de saúde, especificados por leis (municipais, estaduais ou federais, conforme o caso), que são tidos como notificáveis por essas leis.

Em geral, referem-se a informações a respeito de casos incidentes, assim como dados de ocorrência e diagnóstico. Às vezes, incluem dados censitários relativos à partição da população monitorada, em subgrupos de idade, sexo, local de residência, entre outros.

Tanto os dados coletados quanto as informações censitárias podem estar, e em geral estão, submetidas a erros que, a grosso modo, podem ser classificados como "biases" de informação, erros de não-resposta e erros provocados por extravio ou perda das informações (total ou parcial). Porém, nenhum de tais erros é aleatório no sentido de que, na coleta de dados não existe amostragem aleatória ou alocação aleatória. Em realidade, a população monitorada pode ser imaginada como constituída de dois subgrupos: um com a condição sob vigilância, da qual todos os elementos devem ser capturados pelo sistema, e outro sem a condição. Conseqüentemente, não tem sentido falar-se na presença do erro aleatório, pois não há amostragem aleatória, nem alocação aleatória de unidades de estudo nos dois subgrupos.

Talvez, em algumas condições específicas de monitoramento, possamos tomar o processo de coleta de dados de vigilância como um processo estocástico, com um análogo de amostragem seqüencial, no tempo, com regras de parada bem definidas, a priori. Por exemplo, quando podemos fazer a seguinte analogia: na população monitorada, continuamente, os "não-casos" comportam-se como se fossem produtos não defeituosos num processo de produção industrial contínua e os "casos" comportam-se como produtos defeituosos que surgem na linha de produção. Barnard¹, mostrou que tal processo é estocástico, para o qual Page²⁰ já havia desenvolvido procedimento de análise análogo de um seqüencial. Nessas condições, então, podem ser aplicados procedimentos de inferência estatística, específicos para tais situações.

O que, no entanto, não pode ser feito, por falta de sustentação teórica, é conduzir testes de hipóteses que procuram estabelecer diferenças entre distintos subgrupos em que fica dividida a população monitorada. Estimação por intervalo, dentro de tais subgrupos, também não encontra sustentação já que, em ambas as situações, os procedimentos exigem amostragem aleatória. Poder-se-ia argumentar que, para efeito de estimação, se poderia utilizar estimação baseada em modelo ("model-based estimation") que poderia prescindir da seleção aleatória. Mesmo considerando que esta é uma questão em aberto, na literatura estatística, ainda permaneceria a dificuldade: procedimentos baseados em modelo prescindem da aleatorização mas não prescindem da amostragem, pois usam "purposive sampling", pressupondo um modelo de distribuição para a variável (ou variáveis) em questão. O conhecimento do modelo e a pressuposição de sua validade eliminam a necessidade da aleatorização da amostra. E, no processo de coleta de dados de vigilância, não existe a amostragem no sentido em que é usada com inferência em "superpopulação".

Conseqüentemente, análises estatísticas com dados de vigilância que comparam subgrupos dentro da população monitorada, através de testes de hipóteses, ou que produzem estimações através da construção de intervalos de confiança, para tais subgrupos, devem ser vistos, em nossa opinião, com muita cautela e, até mesmo, com suspeição, no entender de Hall¹⁵.

5- Que tipo de Procedimentos Estatísticos Utilizar?

Tradicionalmente, os dados coletados nos sistemas de vigilância em saúde pública são descritos através de tabelas de frequências (ou dos correspondentes gráficos) das ocorrências de interesse.

Em geral, são descrições estratificadas por grupos de idade, de sexo, de locais de ocorrência e, às vezes, por grupos étnicos.

Há uma concepção errônea, em nossa interpretação, de que procedimentos descritivos, por não trazerem excessiva sofisticação, são pouco eficientes e pouco informativos. Ocorre, a nosso ver, exatamente o contrário. Uma descrição adequada, através de uma tabela de distribuição de freqüências ou, então, uma distribuição por estatísticas de ordem, como por exemplo a distribuição de alguns percentis, pode ser muito mais informativa que qualquer resultado proveniente de um procedimento inferencial estatístico que, além de quase sempre ser inadequado, procura resumir toda a informação em um ou dois parâmetros. A vigilância pode e deve enfatizar taxas e coeficientes com base na população.

Muito embora a melhor forma de exposição de dados simples seja ainda em material textual (pelo menos para os não iniciados), uma exposição gráfica pode fornecer ao leitor uma rápida compreensão acerca de grandes volumes de dados. Os gráficos podem, e devem, servir como instrumentos eficazes para exposição de dados visando tanto a sua simples descrição quanto análises de maior precisão e, ainda, instrumento de comunicação. Há, na literatura internacional, um extenso material a respeito que, a nosso ver, tem sido pouco explorado entre nós, principalmente na análise e disseminação rotineira da informação coletada nos sistemas de vigilância em saúde, como por exemplo, para citar apenas os mais divulgados: Cleveland⁷, Feinberg¹³, Tufte²⁶, Tukey²⁷ e Whittaker²⁹.

A questão do uso das cartas controle na análise rotineira de dados de vigilância é outra questão que merece atenção. Com exceção do esquema de Shewhart²³ pouco se tem visto, ou quase nada, na literatura brasileira, a respeito de outros procedimentos, como por exemplo, a estatística "cusum" (cumulative sums), inicialmente introduzida no controle estatístico de qualidade de produção industrial; Barnard¹, Ewan & Kemp¹² e Page²⁰. Este é um procedimento que se comporta como um processo estocástico, com uma seleção análoga a um sequencial, com regra de parada bem estabelecida, que vem merecendo atenção, na literatura internacional, por inúmeras aplicações a dados de vigilância em saúde, como por exemplo: Bjerkdal & Bakketeig⁵; Kallen & Windberg¹⁶; Levin & Kleine¹⁷; Chen⁶, Frisén¹⁴, Raubertas²¹ e Weatheral & Haskey²⁸.

Da mesma forma, a utilização de séries temporais, como procedimentos de processos estocásticos, após sua utilização inicial nos estudos de morbimortalidade por influenza, de Collins⁹, Clifford e col.⁸ e Serfling²², vêm, nos últimos anos, em nível internacional, recebendo maior atenção.

Embora uma das metas da vigilância em saúde pública sempre tenha sido detectar agrupamentos ou "clusters" temporais e espaciais, poucas vezes a análise estatística formal e específica tem sido aplicada em situações rotineiras, mesmo em nível internacional. Apesar de ser esta uma área que carece, ainda, de mais pesquisas para se decidir a respeito da validade da aplicação rotineira de tais métodos, acreditamos que, entre nós no Brasil, talvez uma das dificuldades de acesso a tais procedimentos é que a grande massa de publicação aparece em periódicos que não são da área médica ou de saúde, embora sua intensão seja atingir os pesquisadores de tais áreas. Citamos, entre tais periódicos: *Statistics in Medicine*, *Biometrics*, *Biometrika*, *Clinical Trials*, *Journal of the American Statistical Association*, *Journal of the Royal Statistical Society*, *Applied Statistics*, entre outros.

Um procedimento que deve ser inserido na rotina de análise de dados provenientes dos sistemas de vigilância em saúde pública, que já vem sendo utilizado com uma certa intensidade em trabalhos de pesquisa em pós-graduação, é o do mapeamento computadorizado que associa computação gráfica, cartografia e distribuição de taxas e/ou coeficientes de morbimortalidade, como por exemplo: Bernadelli & Montomoli², Bithle³, Cook-Mozzafari e col.¹¹, Bithel & Stone⁴, para citar apenas alguns exemplos mais recentes.

6- Conclusão

Como resultado desta reflexão concluímos que procedimentos estatísticos para análise de dados tem limitações naturais que são estabelecidas pelos pressupostos teóricos formais sobre os quais se assenta a construção de tais procedimentos.

Por sua vez, dados oriundos de sistemas de vigilância em saúde pública, como consequência da natureza essencial do processo, são afetados por erros de registros de informação, de perdas e extravios de informação que não são erros aleatórios.

A combinação destes dois fatos exige mais que cautela na utilização de procedimentos de inferência estatística aos dados coletados em sistemas de vigilância em saúde pública.

Outro aspecto que carece mais atenção e exige mais investigação é aquele do tratamento quantitativo, se possível, dos erros não aleatórios que afetam a qualidade dos dados de vigilância.

Os cursos de saúde pública, principalmente em nível de pós-graduação, devem dedicar mais atenção aos aspectos da vigilância em saúde pública e aos possíveis procedimentos quantitativos de análise de dados, com especial ênfase na potencialidade e na limitação dos procedimentos estatísticos.

Agradecimentos

Aos Professores Paulo S. Sabroza e Euclides A. de Castilho, pelas valiosas sugestões.

SANCHES, O. [Routine analysis of public health surveillance data: what statistical procedures to use?]. *Rev. Saúde Pública*, 27: 300-4, 1993. In recent years an increasing interest in epidemiological surveillance (that we prefer to label public health surveillance) has emerged. The viewpoint that the ease of access to computers and statistical software may permit the use of more sophisticated statistical methodologies in the analysis of surveillance data has been profounded in many studies. It is a cause of concern that this attitude, when used indiscriminated by, may lead to analysis lacking in theoretical support. Thus, first, a viewpoint about surveillance in public health is presented and the basic propositions of the theory of statistical hypothesis tests and interval estimation are described briefly and without technicalities. The nature of surveillance data, their nonsampling character and non-random selection are also commented on. Some descriptive procedures that may be used without loss of quality in analysis are described than some procedures that are proposed in the relevant international literature but that need fuller investigation before being introduced in routine data analysis are given.

Keywords: Epidemiologic surveillance. Data collection. Data interpretation, statistical. Inference.

Referências Bibliográficas

- BARNARD, G.A. Control charts and stochastic process. *J. R. Statist. Soc. Ser. B.*, 21: 239-57, 1959.
- BERNARDINELLI, L. & MONTOMOLLI, C. Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Statist. Med.*, 11: 983-1007, 1992.
- BITHEL, J.F. An application of density estimation to geographical epidemiology. *Statist. Med.*, 9: 691-701, 1990.
- BITHEL, J.F. & STONE, R.A. On statistical methods for analysing the geographical distribution of cancer cases near nuclear instalations. *J. Epidemiol. Community Health*, 43: 79-85, 1989.
- BJERKEDAL, T. & BAKKETEIG, L.S. Surveillance of longitudinal malformations and other conditions of newborn. *Int. J. Epidemiol.*, 4: 31-6, 1975.
- CHEN, R. The relative efficiency of the sets and the cusum techniques in monitoring the occurrence of rare events. *Statist. Med.*, 6: 517-25, 1987.
- CLEVELAND, W.S. *The elements of graphing data*. Murray Hill, Bell Telephone Laboratories, 1985.
- CLIFFORD, R.E.; SMITH, J.W.G.; TILLET, H.E.; WHERRY, P.J. Excess mortality associated with influenza in England and Wales. *Int. J. Epidemiol.*, 6: 115-28, 1977.
- COLLINS, S.D. Trends and epidemics of influenza and pneumonia - 1918-1951 *Public. Health Rep.*, 66 (46): 1487-516, 1951.
- COLTON, T. Remarks. *Statist. Med.*, 8: 255-6, 1989.
- COOK-MUZZAFARI, P.J.; DARBY, S.C.; DOLL, R.; FORMAN, D.; HERMON, C.; PIKE, M.C.; VINCENT, T. Geographical variation in mortality from leukaemia and other cancers in England and Wales in relation to proximity to nuclear instalations, 1969-78. *Br. J. Cancer*, 59: 476-85, 1989.
- EWAN, W.D. & KEMP, K.W. Sampling inspection of continuous processes with no autocorrelation between successive results. *Biometrika*, 47: 363-80, 1960.
- FEINBERG, S.E. Graphical methods statistics. *Am. Statistician*, 33: 165-78, 1979.
- FRISÉN, M. Evaluation of methods for statistical surveillance. *Statist. Med.*, 11: 1489-502, 1992.
- HALL, D.B. Analysis of surveillance data: a rationale for statistical tests with comments on confidence intervals and statistical models. *Statist. Med.*, 8: 273-8, 1989.
- KALLEN, B. & WINBERG, J. A Swedish register of longitudinal malformation. *Pediatrics*, 4: 765-76, 1968.
- LEVIN, B. & KLINE, J. The cusum test homogeneity with an application in spontaneous abortion epidemiology. *Statist. Med.*, 4: 469-88, 1986.
- MASON, J.O. Opening remarks. *Statist. Med.*, 8: 253-54, 1989.
- NEYMAN, J. *Probabilidade freqüentista e estatística freqüentista*. Brasília. IMPA-CNPq, 1978. (Monografia de Matemática nº. 31).
- PAGE, E.S. Continuous inspection schemes. *Biometrika*, 41: 100-15, 1954.
- RAUBERTAS, R.F. An analysis of disease surveillance data that uses the geographic locations of the reporting units. *Statist. Med.*, 8: 267-71, 1989.
- SERFLING, R.E. Methods of current statistical analysis of excess pneumonia influenza deaths. *Public Health Rep.*, 78: 494-506, 1963.
- SHEWHART, W.A. *Economic control of quality of manufactured product*. New York, D. VanNostrand Co. Inc., 1931.
- SILVA, L.J. da Vigilância epidemiológica: uma proposta de transformação. *Saúde e Soc.*, 1: 7-14, 1992.
- THACKER, S.B. & BERKELMAN, R.T. Public health surveillance in the United States. *Epidemiol. Rev.*, 10: 164-90, 1988.
- TUFTE, E.R. *The visual display of quantitative information*. Cheshire, Graphic Press, 1983.
- TUKEY, J.W. *Exploratory data analysis*. Reading, Addison-Wesley, 1977.
- WEATHERAL, J.A.C. & HASKEY, J.C. Surveillance of malformations *Br. Med. Bull.*, 32: 39-44, 1976.
- WHITAKER, J. *Graphical models in applied statistics*. New York, Wiley, 1990. (Wiley Series in Probability Statistics).

Recebido para publicação em 8.12.1992

Reapresentado em 16.4.1993

Aprovado para publicação em 7.6.1993