

UMA PROPOSTA PARA DETECÇÃO PRECOCE DE EPIDEMIAS  
COM ENFASE NO CASO DO SARAMPO

Carlos A.C. Pires<sup>1</sup>, Marlo Libel<sup>2</sup>, Ronney B. Panerai<sup>3</sup>

**Resumo** -- A detecção de epidemias de sarampo é importante tanto para melhor aproveitamento de recursos de saúde quanto pelo desenvolvimento de um sistema de detecção passível de aplicação a outros sinais epidemiológicos. Este trabalho apresenta um método objetivo para a detecção precoce de epidemias de sarampo baseado na aplicação da fórmula de Bayes para o cálculo da probabilidade de um dado número de casos caracterizar a existência de um surto epidêmico. Em princípio, esta mesma abordagem pode ser adotada em relação a outras doenças epidêmicas que requerem vacinação urgente como é o caso da meningite. O método foi desenvolvido utilizando os dados de ocorrência de sarampo em 150 cidades de maior número de casos no Rio Grande do Sul, no período de 1971 a 1980. Sendo o método aplicado a este mesmo período, seu desempenho foi avaliado pelo levantamento da Curva de Operação Característica (ROC) : em 347 epidemias foi possível detectar 86% corretamente, com 11% de falsos alarmes e com uma antecedência média de 5 semanas em relação ao pico de epidemia. Finalmente, o método foi então testado com os dados de 1981, simulando assim a implementação do sistema: em 57 epidemias foi possível detectar 70% corretamente, com 12% de falsos alarmes e uma antecedência média de 4 semanas em relação ao pico epidêmico. Os resultados foram considerados satisfatórios, justificando portanto a sua futura implementação em caráter experimental. A possibilidade de utilização do método mesmo sem recursos computacionais, torna viável sua implementação com um mínimo de recursos financeiros.

INTRODUÇÃO

O sarampo é uma doença que embora seja considerada de caráter benigno, ocupa lugar de destaque como causa de mortalidade infantil, especialmente em populações de baixo nível sócio-econômico, onde é uma doença sempre grave devido a desnutrição crônica desta população. No Brasil, na faixa etária de 1 a 4 anos, em 1979 o sarampo foi a terceira causa de morte entre as causas conhecidas, sendo somente precedida pela pneumonia e doenças infecciosas intestinais (Ministério da Saúde, 1982).

São susceptíveis ao sarampo todas as crianças a partir de 6 meses de idade e que não foram vacinadas ou não contraíram a doença, os casos de reinfeção são extremamente raros (Babbot, 1954). O sarampo é uma doença de distribuição

<sup>1</sup> Mestrando do Programa de Engenharia Biomédica - COPPE/UFRJ - Caixa Postal 68.510 - CEP 21.944 - Rio de Janeiro - RJ

<sup>2</sup> Responsável pela Unidade de Vigilância Epidemiológica da Secretaria de Saúde e Meio Ambiente do Rio Grande do Sul

<sup>3</sup> Professor Adjunto do Programa de Engenharia Biomédica da COPPE/UFRJ - Centro de Tecnologia - Bloco H - Sala 327 - Caixa Postal 68.510 - CEP 21.944 - Rio de Janeiro - RJ

universal e que apresenta características epidemiológicas conhecidas, como a periodicidade de seus surtos (2 a 3 anos) e a maior ocorrência de surtos em final de inverno.

A única forma de controle do sarampo atualmente disponível é a vacinação em massa e de rotina da população susceptível. A vacina para sarampo começou a ser efetivamente utilizada na década de 70, porém devido a alta infecciosidade do vírus causador desta doença, o sarampo não foi e dificilmente será uma doença totalmente erradicada.

Para manter o sarampo sob controle, é necessário portanto uma vigilância permanente a fim de detectar epidemias o mais precocemente possível, e tomar medidas visando a proteção principalmente das populações de baixo nível sócio-econômico.

O objetivo deste trabalho é desenvolver um sistema de detecção de epidemias de boa confiabilidade e mínimo atraso na detecção, e cuja implementação seja viável com um mínimo de recursos financeiros. Utilizou-se para isto os dados de ocorrência de sarampo no Rio Grande do Sul, porque além dos mesmos apresentarem boa confiabilidade (Nadal, 1982), são também disponíveis desde 1971. Deve-se salientar também o interesse da Unidade de Vigilância Epidemiológica deste estado, em dispor de um método de detecção passível de implementação dentro do sistema de vigilância já existente.

A filosofia adotada para atingir-se este objetivo foi o desenvolvimento de um método que, mesmo sendo suficientemente simples para ser implementado sem o uso de computador digital, apresente boa confiabilidade de detecção. Considerando-se as séries históricas de 1971 a 1980, o teorema de Bayes foi utilizado para obter-se a curva de "probabilidade de epidemia" versus "número de casos". Esta curva foi aplicada ao período 1971-1980 e também para o ano de 1981, simulando-se assim o funcionamento do sistema com dados inéditos.

## METODOLOGIA

### Aplicação do Teorema de Bayes

A partir do conhecimento da série histórica de ocorrência semanal de sarampo em uma comunidade, é possível calcular-se a probabilidade do número de casos que ocorrem em uma dada semana constituírem uma epidemia ou não. Considera-se "epidemia" o aumento acentuado e rápido do número de casos seguido de um retorno aos níveis anteriores ao surto. Por não existir uma definição quantizada da que seja uma epidemia, assumiu-se para este trabalho que para ocorrer epidemia deveria haver um mínimo de 4 casos semanais de sarampo após o sinal ter sido alisado por um filtro passa-baixa.

Considerando-se a série histórica típica, ilustrada na figura 1, o "sinal epidemiológico" é constituído pela sequência de números que representam o total de casos de sarampo ocorridos durante uma semana em um dado município. São utilizadas 512 semanas do período 1971-1980. Estes dados foram fornecidos pela Unidade de Vigilância Epidemiológica da Secretaria de Saúde do Rio Grande do Sul para os 232 municípios do estado. Somente as 150 cidades com maior número total de casos de 1971 a 1980 foram consideradas.

Na figura 1 podem ser identificados períodos epidêmicos, indicados pela letra "E", além de oscilações no número de casos durante epidemias, observa-se que também em períodos não-epidêmicos ocorrem casos de sarampo, porém com uma

característica aleatória que poderia ser associada a presença de "ruído" em um sistema de comunicações.

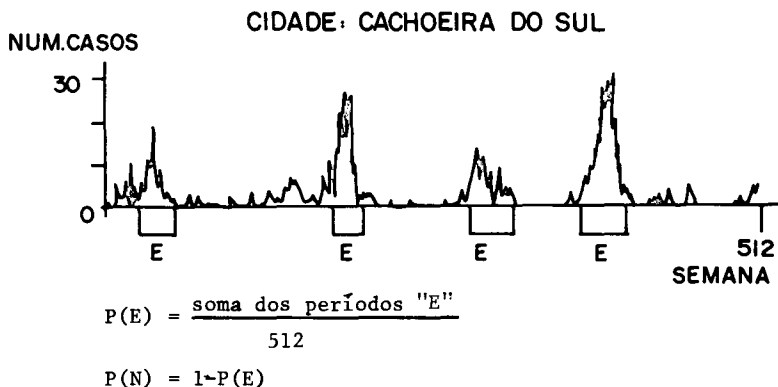


Figura 1. Cálculo da probabilidade "a priori" de ocorrer epidemia em uma cidade, independente do número de casos.

O sistema de detecção de epidemia deve então responder a pergunta: "Dado um certo número semanal de casos de sarampo em um município, qual a probabilidade de que esteja-se em período epidêmico?" Esta probabilidade pode ser calculada pela fórmula de Bayes (Molina, 1931):

$$P(E/C=x) = \frac{P(E) \cdot P(C=x/E)}{P(E) \cdot P(C=x/E) + P(N) \cdot P(C=x/N)} \quad (1)$$

As várias probabilidades que constam da equação (1) seriam:

$P(E)$  : probabilidade "a priori" de ocorrerem epidemias em uma cidade

$P(N) = 1 - P(E)$

$P(C=x/E)$  : probabilidade de ocorrerem  $x$  casos ( $C$ ) durante período epidêmico

$p(C=x/N)$  : probabilidade de ocorrerem  $x$  casos durante período não-epidêmico

$P(E/C=x)$  : probabilidade de existir um período epidêmico dado que ocorreram  $x$  casos semanais de sarampo.

Todas as probabilidades acima podem ser calculadas a partir dos próprios dados para cada uma das 150 cidades em estudo. Na figura 1 está indicado o cálculo de  $P(E)$  e, conseqüentemente,  $P(N)$ . A estimação de  $P(C=x/E)$  e  $P(C=x/N)$  requer a classificação do sinal em "epidemia" e "não-epidemia", a metodologia de desenvolvida para isso é apresentada a seguir.

Considerando-se a presença de flutuações aleatórias no sinal de número de casos, este foi inicialmente alisado com um filtro digital passa-baixa de fase nula, 4ª ordem e frequência de corte de  $0,1\pi$  rad/T (frequência normalizada) (Oppenheim, 1975) a partir de observações sobre o espectro típico do padrão epidêmico (Nadal, 1981). Utilizando-se o sinal filtrado, foi feita a marcação visual do pico de epidemias e, a partir desta referência, sua demarcação com um limiar de 4 casos. Tomando-se os valores correspondentes a "epidemia" e "não-epidemia" é possível construir-se histogramas para cada cidade conforme exemplificado na figura 2A. A partir destes histogramas calcula-se a probabilidade do número de casos ser igual ou maior do que L casos conforme indicado na figura 2B. O fato de se considerar sempre o número de casos como sendo maior ou igual a L, e não em um intervalo  $(L_1, L_2)$  qualquer, garante que estas funções de probabilidade serão monótonas. Isto não contraria a condição de exclusividade dos eventos no teorema de Bayes, já que

$$P(C > X_L) = \sum_{j=L}^n P(C = X_j).$$

Calculadas as probabilidades  $P(X > L/E)$  e  $P(X > L/N)$ , conforme indicado na figura 2B, a fórmula de Bayes (equação (1)) pode ser adaptada para o cálculo correspondente de  $P(E/X > L)$  que é a probabilidade desejada, de ocorrer epidemia da do que o número de casos é igual ou superior a um certo valor L. O cálculo desta probabilidade para todos os valores de L pertinentes, em cada cidade, leva a uma curva de probabilidade do tipo exemplificado pela figura 3 para a cidade de São Leopoldo. Através desta função é então possível obter-se a probabilidade semanal de ocorrência de epidemia. A figura 4 mostra o resultado obtido no caso de Porto Alegre, observando-se uma nítida associação de probabilidades e levadas, concomitantemente com períodos epidêmicos.

#### Avaliação do Sistema detector

Para aplicação das curvas de probabilidade de ocorrência de epidemia (Figura 3) é necessário novamente levar em consideração a variabilidade do sinal de ocorrência de sarampo. Idealmente um filtro casado poderia ser utilizado com o objetivo de se maximizar a relação sinal/ruído (Nadal, 1981). No entanto, devido às características não-causais do filtro casado (Lathi, 1979), sua utilização implicaria necessariamente em um atraso de detecção o que não é absolutamente desejável. Como alternativa empregou-se um filtro com resposta impulsiva exponencial com frequência de corte em  $0,12\pi$  rad/T (frequência normalizada), cujo atraso provocado no sinal é tolerável para a detecção. O filtro proposto também tem a vantagem de permitir a implementação futura do método sem o uso de computador digital. A equação auto-regressiva do filtro é (Hamming, 1977)

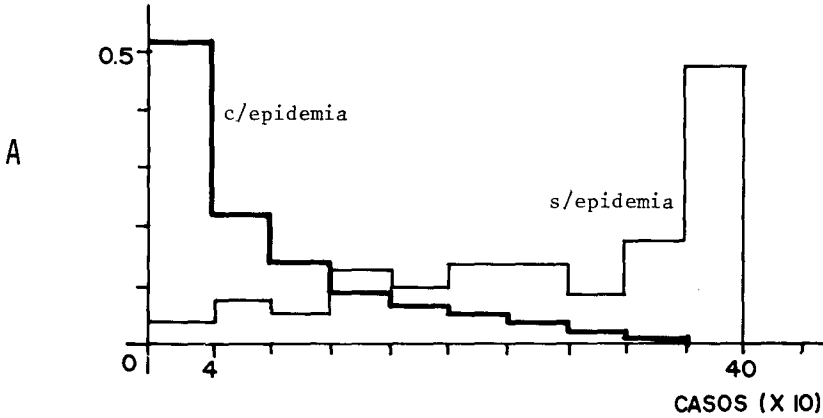
$$Y(n) = 0,7 Y(n-1) + 0,3 X(n) \quad (2)$$

onde  $Y(n)$  corresponde a saída do sinal e  $X(n)$  a entrada.

Para cada cidade o sinal de ocorrência semanal de sarampo foi alisado com o filtro acima e, aplicando-se a curva de probabilidade correspondente, obteve-se o sinal de probabilidade semanal, semelhante ao exemplificado na figura 4. Estipulando-se um determinado nível de probabilidade P, probabilidades acima deste valor representariam detecções de "epidemias" sendo então possível uma comparação com os valores reais determinando-se desta forma a precisão do sistema e a precocidade (em relação ao pico epidêmico) da detecção. Uma vez que o limiar de probabilidade será sempre de escolha do epidemiologista, a avaliação do sistema foi feita para 20 níveis diferentes de probabilidade, não igualmen-

FREQUÊNCIA RELATIVA

CIDADE: SÃO LEOPOLDO



DENSIDADE DE PROBABILIDADE

$$P(X \geq L/E) = (1-A)$$

$$P(X \geq L/N) = B$$

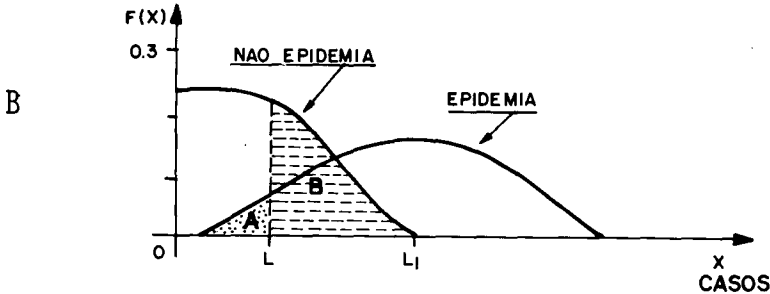


Figura 2. A-Distribuição de frequência relativa ao número de casos no sinal com e sem epidemia. B-Cálculo da probabilidade do número de casos ser igual ou maior que X, dado que é epidemia (E) e dado que não é epidemia (N), a partir do histograma de distribuição de frequência.

te espaçados entre 0.3 e 1.0. Para cada nível determinou-se o número de detecções corretas, falso-positivas e falso-negativas, assim como a precocidade média das detecções corretas. A representação da relação percentual de detecções corretas contra as falso-positivas, para cada um dos 20 níveis de probabilidades considerados, resulta na "Curva de Operação do Receptor" (Rossing e Hatcher, 1980). Esta representação foi adotada para sintetizar o desempenho do sistema. Esta avaliação foi separada em dois períodos distintos: 1971-1980, correspondendo ao próprio conjunto de dados que geraram as curvas de probabilidade, e para o ano de 1981, funcionando como uma simulação real do método enfrentando dados inéditos.

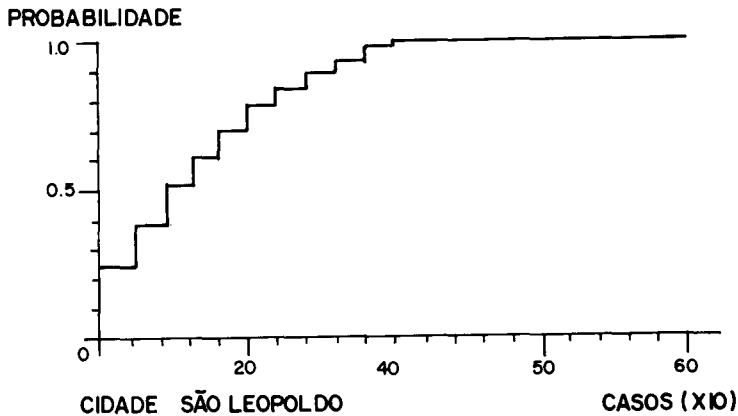


Figura 3. Função de probabilidade do número de casos ser devido a uma epidemia, obtida pela aplicação do Teorema de Bayes.

#### RESULTADOS

A Curva de Operação do Receptor (ROC) para o período 1971-1980 é representada na figura 5 distinguindo-se 3 grupos de cidades de acordo com sua ordem em relação ao número total de casos apresentados. Conforme se reduz o limiar de probabilidade de detecção, aumenta o número de detecções corretas, porém paga-se o preço de ocorrerem detecções falsas. Observa-se na figura 5 que o desempenho do sistema é função do número de casos. Em geral este fator está diretamente relacionado com a população de cada município. Supondo-se que seja tolerável uma ocorrência de 13% de falsos alarmes, a figura 5 mostra que é possível detectar corretamente 92% das epidemias para as grandes cidades, 85% para as médias e 87% para as pequenas cidades.

Uma característica importante na detecção de epidemias de sarampo é a precocidade de decisão. Para o caso presente também verifica-se a influência do número total de casos conforme ilustrado pela figura 6. Utilizando-se novamente como referência a situação de 13% de falso-alarme, resultam precocidades médias de 8,4 e 2 semanas para as três categorias de cidades mencionadas acima. A implicação destes valores na implementação de um sistema de detecção eficaz será discutida posteriormente.

As figuras 7 e 8 apresentam os resultados correspondentes ao ano de 1981 que, apresentando um número bem menor de epidemias, não seria o ideal para uma avaliação de desempenho, mas representa a simulação do detector operando em condições reais. Como seria de se esperar, houve uma queda de desempenho em relação ao período-base (1971-1980), mas ainda mantém-se um nível elevado de detecções corretas contra reduzidos falsos-positivos. No que diz respeito a precocidade, não se observaram alterações significativas entre 1981 e o período 1971-1980.

PROBABILIDADE

CIDADE : PORTO ALEGRE

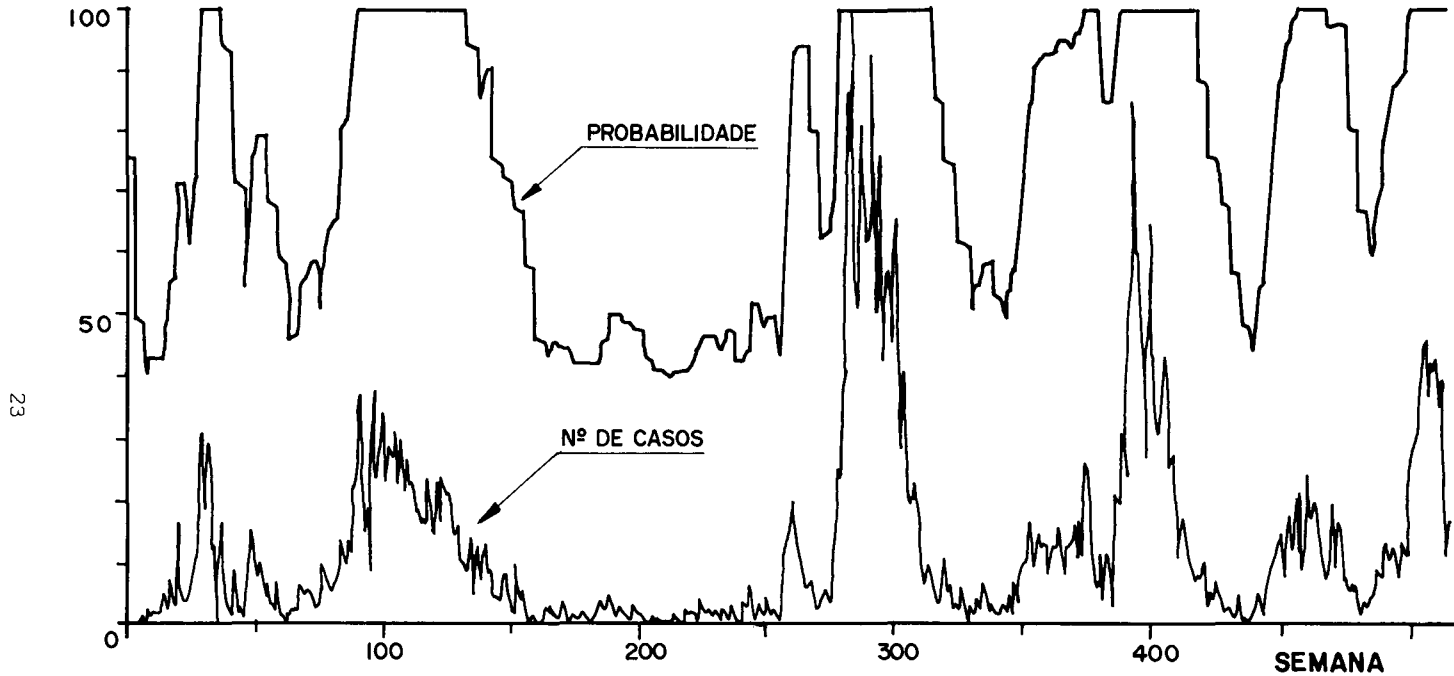


Figura 4. Sinal de probabilidade obtido pela conversão do número de casos em probabilidade de ser este número devido ou não a presença de epidemia (curva superior). A curva inferior corresponde ao sinal original.

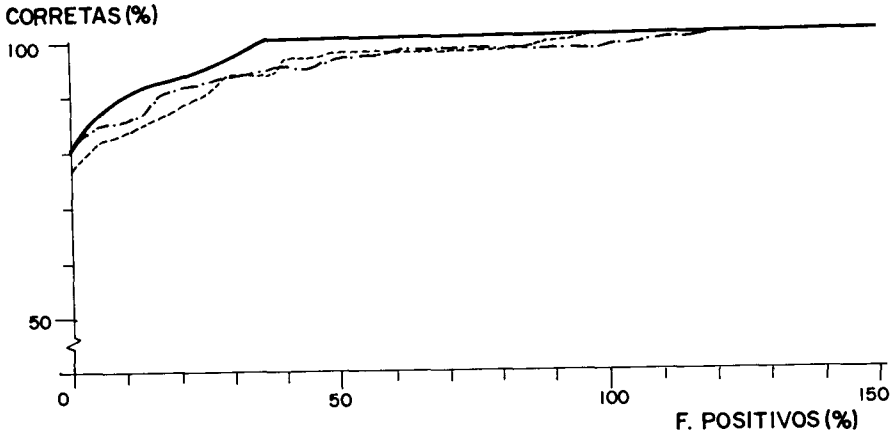


Figura 5. Curva ROC para período 1971-1980. Linha contínua: cidades de ordem 1 a 20 (55 epidemias); linha interrompida: cidades de ordem 21 a 60 (110 epidemias); linha tracejada: cidades de ordem 61 a 150 (122 epidemias). Os percentuais são calculados em relação ao número total de epidemias em cada caso.

### DISCUSSÃO

O método proposto para a detecção precoce de epidemias de sarampo baseia-se unicamente na história de ocorrências em cada município o que resulta em um procedimento de rotina extremamente simples que pode ser realizado manualmente pelos responsáveis pela Vigilância Epidemiológica locais (Pires, 1983). Apesar desta simplicidade, o sistema mostrou um desempenho satisfatório na simulação correspondente ao ano de 1981. Este resultado certamente estimula a sua implementação em caráter probatório permitindo assim uma avaliação com número mais significativo de episódios epidêmicos. Independentemente deste resultado futuro no entanto, vários aspectos do projeto devem ser discutidos e criticados.

A principal limitação do sistema de detecção proposto é a eventual não-estacionariedade da série temporal de ocorrência que afetaria diretamente a relação probabilidade x número de casos alisados (figura 3). Em princípio o período-base considerado (1971-1980) é bastante longo para representar alterações a leitórias no número de casos durante e fora de epidemias. Os resultados obtidos para o ano de 1981 indicam que, caso tenham havido alterações, estas não foram suficientes para prejudicar significativamente o desempenho do detector. Embora os dados referentes a 1982 ainda não tenham sido analisados, existem indicações de uma redução substancial no número de casos em todo o estado do Rio Grande do Sul, como consequência de uma nova estratégia de vacinação que vem sendo adotada. Neste tipo de situação pode-se prever um desempenho pobre no sistema, caso as curvas originais de probabilidade sejam mantidas. No entanto, é óbvio que estas curvas podem ser recalculadas periodicamente fazendo com que o sistema se adapte às novas condições. De forma mais elaborada, seria também possí



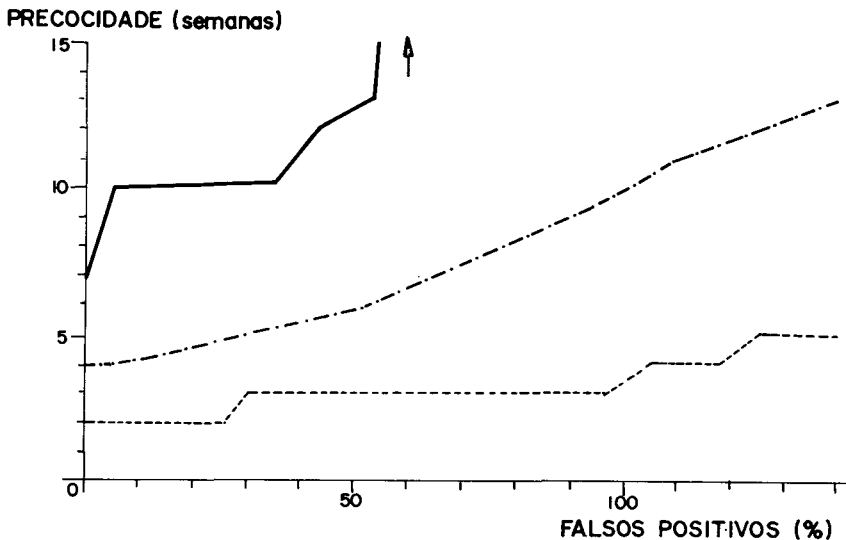


Figura 6. Precocidade média de detecção em relação a semana de pico da epidemia para o período 1971-1980. Linha contínua: cidades de ordem 1 a 20. Linha interrompida: cidades de ordem 21 a 60. Linha tracejada: cidades de ordem 61 a 150.

vel uma atualização das curvas de probabilidade a cada semana, dando-se pêsso diferenciado as amostras mais recentes. Outro caminho seria o emprêgo de filtros adaptativos do tipo "tracking signal" (Trigg e Leach, 1976). Estas evoluções da presente proposta poderão vir a resultar em um sistema mais robusto, mas será imprescindível o uso de um computador para elaboração rotineira dos cálculos, perdendo-se assim uma característica fundamental do detector no estágio em que atualmente se encontra.

Embora para um serviço de Vigilância Epidemiológica seja extremamente importante tomar ciência, em tempo hábil, do início de um processo epidêmico, a utilidade do sistema de detecção desenvolvido ficará sempre dependente da forma como esta informação deverá ser utilizada. Em primeiro lugar caberá ao epidemiologista determinar o nível de probabilidade em que deseja operar, incluindo ajustes deste limiar em função de flutuações sazonais, etc. A determinação do limiar de detecção será função do custo associado a detecções falso-positivas ou falso-negativas, embora a última seja de difícil estimativa. Considerando-se os valores de precocidade obtidos para as cidades maiores (ordem 1 a 60) pode-se inclusive pensar em bloqueio de surtos por revacinação em massa, uma vez que o início de uma epidemia tenha sido detectado. No entanto, os resultados obtidos para as cidades menores, apresentando uma precocidade de apenas 2 semanas indicam que neste caso não seria possível pensar-se em bloqueio. Mesmo assim observa-se que as primeiras 60 cidades contribuem com cerca de 65% do to

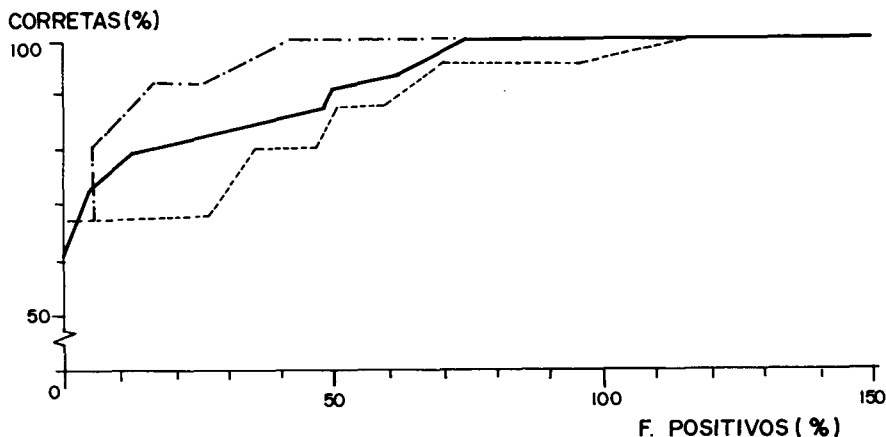


Figura 7. Curva ROC para período de 1981. Linha contínua: cidades de ordem 1 a 20 (16 epidemias). Linha interrompida: cidades de ordem 21 a 60 (16 epidemias). Linha tracejada: cidades de ordem 61 a 150 (25 epidemias). Os percentuais são calculados em relação ao total de epidemias em cada caso.

tal de casos de sarampo no estado.

Uma vez que as notificações de ocorrência de sarampo correspondem somente a cerca de 10% dos casos reais (Libel, 1982), cidades com pequena população apresentam uma variabilidade muito maior no número de casos, não havendo uma distinção muito nítida entre "epidemia" e "não-epidemia". Esta situação, que pode ser definida como de "baixa relação sinal/ruído", explica a menor eficiência do detector para municípios pequenos. Mesmo neste caso, no entanto, pode-se prever uma melhoria no sistema de detecção levando-se em conta não só o número de casos, mas também o padrão temporal do episódio epidêmico. Para o sarampo já foi identificada uma estrutura bastante peculiar em que cada epidemia é constituída por uma sequência de microepidemias com diferentes amplitudes e latências (Sayers e col., 1982). A utilização deste conhecimento, para a elaboração de um sistema de detecção, está sendo testada por Infantosi (1982).

Para finalizar deve-se ressaltar que o método aqui apresentado não depende de qualquer característica própria do sarampo, sendo portanto adaptável para aplicação na vigilância epidemiológica de outras doenças em que a tomada de medidas preventivas seja fundamental.

#### AGRADECIMENTOS

Os autores agradecem a colaboração da Unidade de Vigilância Epidemiológica da Secretaria de Saúde e Meio Ambiente do Rio Grande do Sul no fornecimento dos dados. A equipe de pesquisa envolvida no processamento de sinais epidemiológicos conta também com a participação dos Profs. B. McA. Sayers, do Imperial

College, Antonio Fernando Catelli Infantosi e Flavio Nobre do Programa de Engenharia Biomédica cujas sugestões agradecemos.

Este trabalho foi realizado com apoio financeiro do CNPq e da FINEP.

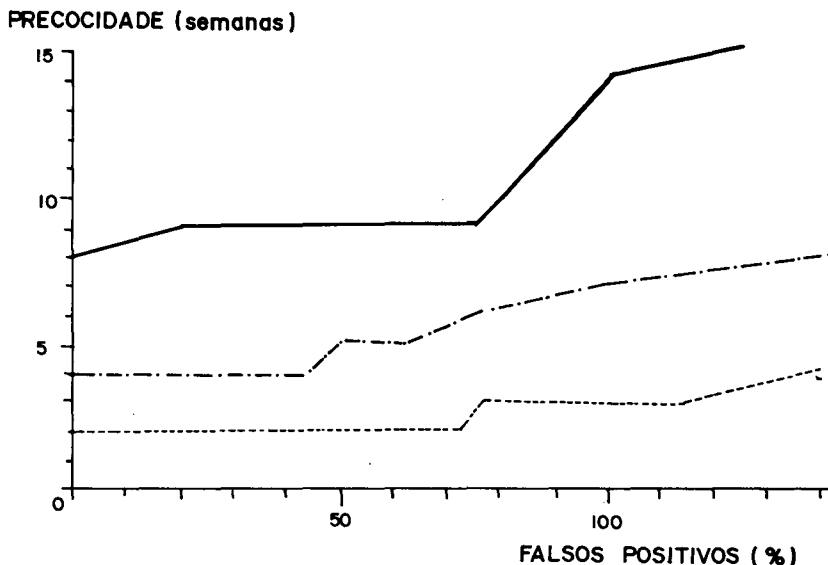


Figura 8. Precocidade média de detecção para o período de 1981. Linha contínua: cidades de ordem de 1 a 20. Linha interrompida: cidades de ordem 21 a 60. Linha tracejada: cidades de ordem 61 a 150.

#### REFERÊNCIAS

1. BABBOT, F.L. (1954), "Modern Measles", American Journal of Medical Science nº 228: 334-361.
2. BRASIL (1982), Estatísticas de Mortalidade 1979. Brasília, Ministério da Saúde.
3. HAMMING, R.W. (1977), "Digital Filters". Englewood, Prentice-Hall.
4. LATHI, B.P. (1979), "Sistemas de Comunicação". Rio de Janeiro, Guanabara Dois.
5. LIBEL, M. (1982), "Avaliação do sistema de Vigilância Epidemiológica do Rio Grande do Sul". Reunião Nacional para Discussão da Estratégia de Vacinação do Sarampo. Fundação SESP, Rio de Janeiro.
6. MOLINA, E.C. (1931), "Bayes' Theorem, an Expository Presentation". The Bell System Journal. Vol. 10, nº 2: 276-283.
7. NADAL, J. (1981), "Emprego de Filtros Casados na Detecção de Surto Epidêmicos de Sarampo". Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro.
8. OPPENHEIN, A.V. e SHAFFNER, R.W. (1975), "Digital Signal Processing". Engle-

wood, Prentice-Hall.

9. PIRES, C.A.C. (1983), "Detecção de Surtos Epidêmicos de Sarampo". Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro.
10. ROSSING, R.G. e HATCHER, W.E. (1980), "A Graphic Method for the Evaluation of Diagnostic Tests". Meth. Inform. Med. Vol. 19 nº 3: 149-155.
11. SAYERS, B.McA., INFANTOSI, A.F.C., PANERAI, R.B., NOBRE, F.F. (1982), "Exploratory pattern-analysis study of a multifocal measles epidemic". IEEE Proceedings, Part A, vol. 129: 693-701.
12. TRIGG, D.W. e LEACH, A.G. (1976), "Exponential smoothing with an adaptative response rate" Operational Research Quartely, vol. 18: 53-59.

## A PROPOSAL FOR THE EARLY DETECTION OF MEASLES EPIDEMICS

ABSTRACT -- This work presents an objective method for the early detection of measles epidemics based on Bayes' Theorem for the computation of the probability of epidemic outbreak associated with the weekly case occurrences of measles. In principle the same approach can be adopted for epidemic surveillance of other diseases, mainly those which require emergency vaccination measures such as meningitis.

Case-occurrences of measles from 150 towns of the state of Rio Grande do Sul, Brazil, in the period Jan/1971 to Dec/1980 have been used as the data set in which this study was based. A total of 347 epidemics were automatically detected yielding the corresponding probability distributions for the number of cases in epidemic and non-epidemic periods. Using this information to apply Bayes' Theorem, results an estimate of the weekly probability of epidemic outbreak. The performance of this detector was evaluated by comparing the original time series with the outcome of several probability threshold levels leading to the receiver operating characteristic (ROC) curve. For the period 1971-1980, it was possible to detect 86% of the epidemics correctly with only 11% of false alarms and an anticipation of 5 weeks in relation to the peak value. The data for 1981 was used for an unbiased test for the method: of a total of 57 epidemics it was possible to detect 70% correctly with 12% of false alarms, in average 4 weeks before the peak value.

The satisfactory results achieved with this technique warrant its implementation for long-term trial under real conditions. For this purpose it is shown how the method can be utilized without a digital computer, therefore not requiring large investments for its implementation.