

# CONTEÚDO

1. Prefácio – H.M. Yang (3 páginas)
2. Circuitos Neurais – C. Timo-Iaria (17 páginas)
3. Modelos Neurais Distribuídos: A Forma no Conteúdo – W.C. Ferreira Jr. (36 páginas)
4. Inferência de Estrutura Funcional em Neurociências – K. Sameshima e L.A. Baccalá (25 páginas)
5. Aspectos do Sistema Imune Fetal e As Doenças Infecciosas – S.D. Passos e H.M. Yang (35 páginas)
6. Uma Abordagem Matemática em Imunologia – S. Barrozo, C.H. Dezotti e H.M. Yang (25 páginas)
7. Sobre A Erradicação de Doenças Infecciosas – Esforço de Vacinação – H.M. Yang e L.K. Hotta (24 páginas)

## Prefácio

Este volume especial de série TEMA traz as publicações de trabalhos apresentados no mini-simpósio durante o XXV Congresso Nacional de Matemática Aplicada e Computacional (CNMAC), realizado em Nova Friburgo, RJ, no período entre 16 e 19 de Setembro de 2002. O mini-simpósio foi organizado pelo Prof. H.M. Yang sob o nome “Matemática aplicada a fisiologia e epidemiologia”, que é o título desta obra.

O mini-simpósio procurou agregar pesquisadores que atuam em fisiologia e em matemática. Os pesquisadores convidados, três médicos e três matemáticos, apresentaram seus trabalhos de pesquisa em neurociência, imunologia e matemática aplicada. O objetivo do mini-simpósio foi justamente, ao congregarem pesquisadores de áreas de conhecimentos distintos, criar um ambiente de discussão e elaboração de idéias a respeito da aplicação da matemática em fisiologia. O interesse tanto dos palestrantes quanto dos congressistas coroou com êxito este encontro.

Um outro objetivo do mini-simpósio era a participação do Prof. Cesar Timo-Iaria, fisiologista de renome, para proferir uma palestra na plenária do Congresso, além de apresentar trabalho no mini-simpósio. Devido a problema de saúde, ele não pôde participar, mas a sua palestra foi lida no mini-simpósio pelo Prof. H.M. Yang.

Portanto, este volume especial é fruto do esforço empenhado pelos palestrantes e, ao mesmo tempo, uma homenagem ao Prof. Cesar Timo-Iaria, que se mostrou extremamente interessado em participar e propor problemas para aplicação da matemática em fisiologia, em especial, na neurociência.

O Prof. Cesar Timo-Iaria é médico formado pela Escola Paulista de Medicina em 1952, e professor de Fisiologia da Universidade de São Paulo há 50 anos. De 1953 a 1964 foi docente do Departamento de Fisiologia da Faculdade de Medicina de Ribeirão Preto da USP e em 1964 foi convidado a se transferir para o Departamento de Fisiologia da Faculdade de Medicina da USP como professor associado. Nesta Faculdade reformou o curso de fisiologia, em que se ministravam 36 aulas práticas diferentes. Em 1973, por força da reforma universitária, transferiu-se para o Instituto de Ciências Biomédicas da USP, na Cidade Universitária, do qual se tornou professor titular em 1972. Em 1984 aposentou-se e retornou à Faculdade de Medicina da USP, onde passou a dirigir um Laboratório de Investigação Médica, função que ainda exerce.

As suas atividades docentes não se restringiram apenas no Brasil. Prof. Cesar foi professor assistente do Departamento de Fisiologia da State University of New York (1958-1959), retornando ao mesmo Departamento como professor associado em 1968. Em 1962 pesquisou no Instituto de Investigaciones Cerebrales, México, onde, com Hernández-Peón, descobriu o centro do sono. Em 1989 trabalhou na Universidade do Texas e em 1998 foi convidado a ministrar uma série de conferências na Faculdade de Medicina da Universidade de Aachen (Alemanha).

Preocupado com as pesquisas de ciências básicas em Medicina, Prof. Cesar inventou a idéia dos Laboratórios de Investigação Médica, os LIMs, que, ao serem criados, foram os responsáveis pela manutenção do nível de pesquisa. Foi o primeiro presidente da Sociedade Brasileira de Fisiologia, criada em 1957, no Rio de Janeiro.

Durante sua carreira, Prof. Cesar fez importantes pesquisas na vasta área da fisiologia. Relaciona-se as suas maiores descobertas.

1. Descobriu em 1963 o centro de sono (em que hoje trabalham vários grupos de pesquisadores americanos).
2. Foi o primeiro a detectar em 1966 o fator natriurético atrial (vários peptídeos da mais alta relevância, produzidos pelo átrio do coração e com muitas funções fundamentais no sistema cardiovascular, no renal, etc.), embora não o houvesse isolado nem caracterizado.
3. Demonstrou em um trabalho ininterrupto que durou 30 anos (1966–1996) a existência de glicceptores no hipotálamo lateral, nos núcleos do trato solitário e no fígado, os quais são os principais responsáveis pela regulação da glicemia; trata-se de sensores (capazes de detectar ínfimas reduções da concentração de glicose no sangue) que medem a glicemia continuamente e, quando esta baixa, eles ativam o fígado para repor o que foi consumido. Em dado instante, sinais oriundos do fígado vão ao sistema nervoso e desencadeiam a sensação de fome e, conseqüentemente, o comportamento alimentar. Esta pesquisa levou-o a aventar a hipótese de que a fome não resulta de hipoglicemia mas do trabalho metabólico do fígado para não deixar a glicemia baixar, tendo usado em suas pesquisas mais de 2.000 gatos, 200 cães, 500 ratos e 40 pessoas.
4. Introduziu em 1977 o rato como animal de escolha para estudar sono e verificou nesta espécie, pela primeira vez em uma espécie não-humana, todas as fases do sono humano, exceto a primeira. Desde 1990 tem feito descobertas interessantíssimas com relação aos sonhos do rato.
5. Descobriu em 1999 o local em que se situa o núcleo neural que regula o sistema imunológico do rato.
6. Descreveu em 2001 no cerebelo todas as fases da vigília e do sono.
7. Participa de relevante pesquisa que demonstra uma relação linear entre a inteligência do rato e a frequência das ondas teta (desde 1999).

O desfecho de uma de suas pesquisas merece ser especialmente relatado: a descoberta do centro do sono. Para isso, ele introduziu o carbacol em um dos núcleos de Gudden, extremo caudal do circuito de Nauta, e em menos de 20 segundos o gato entrou em sono profundo, cursando todas as suas fases. O artigo foi enviado pelos descobridores para a revista *SCIENCE*, que depois de três meses devolveu-o com a escusa de que isto não era prioridade da revista (uma desculpa “esfarrapada” e infantil!). A revista *Experimental Neurology* aceitou-o, mas logo um dos “donos do assunto sono” escreveu, em uma longa revisão, que os gatos (foram mais de 700 experimentos) provavelmente não dormiam, mas estavam em coma. Isto, apesar de a estimulação elétrica da formação reticular do mesencéfalo despertar os animais, procedimento esse que evidentemente jamais poderia despertar um animal do coma! Durante 15 anos este artigo permaneceu no limbo, até que um canadense

(Cordeau) leu-o, repetiu o experimento (uma única vez!) e confirmou aqueles resultados. Imediatamente Hobson, em Harvard, e Siegel, na California, entraram no campo e citaram o artigo apenas uma vez. Desde então ambos são considerados os grandes descobridores do centro do sono...

Enfim, Prof. Cesar Timo-Iaria publicou 82 artigos em revistas internacionais e apresentou 230 comunicações em congressos locais e internacionais. Tendo orientado cerca de 40 mestres e doutores e mais de 100 estudantes, proferiu mais de 200 conferências no Brasil e 16 no exterior.

É com grande admiração que os autores oferecem este trabalho em sua homenagem, esperando que ele assim o considere.

Hyun Mo Yang

## Circuitos Neurais

C. TIMO-IARIA<sup>1</sup>, Laboratório de Neurocirurgia Funcional (LIM 45), Faculdade de Medicina, Universidade de São Paulo, Av. Dr. Arnaldo, 455, CEP: 01246-903, São Paulo, SP, Brasil

**Resumo.** O sistema nervoso, estrutura mais complexa que por ora conhecemos, constitui-se de células especiais, os neurônios, os quais se acoplam segundo vários padrões, com o que geram os circuitos neurais, a maioria sob a forma de microcircuitos. Como em eletrônica, são os circuitos que geram as funções neurais. Em alguns animais inferiores da escala filogenética o sistema nervoso contém poucos milhares de neurônios, suficientes para programar todas as funções desses organismos; na espécie humana, porém, existem cerca de 100 bilhões de neurônios e mais uns 200 bilhões de células da glia (as funções destas últimas como partes de circuitos neurais são ainda pouco conhecidas mas elas também participam da circuitaria neural). Os neurônios constituem provavelmente trilhões de conexões, que permitem gerar os complexos comportamentos que caracterizam os animais superiores, inclusive a atividade mental. A coloração artificial dos neurônios possibilita a identificação de alguns circuitos relativamente simples mas presentemente apenas a microscopia eletrônica permite a identificação de microcircuitos elementares; essa técnica, entretanto, é muito lenta quando se consideram os trilhões de interconexões neurais. Portanto, a descrição e a consequente compreensão de como o acoplamento dos neurônios em circuitos e o dos circuitos em sistemas geram tantas e tão complicadas funções são por ora impossíveis. A eletrofisiologia, estudo dos sinais elétricos gerados pelas células, tem contribuído muito para entendermos como os circuitos funcionam mas tampouco nos possibilita ainda compreender o funcionamento de microcircuitos inteiros. A principal razão dessa limitação é o fato de que os microcircuitos existentes em um volume, por exemplo, de um milímetro cúbico, contêm de poucas centenas a milhares de neurônios. Essa limitação torna impossível registrar os potenciais de cada um dos neurônios de um microcircuito (na realidade, apenas alguns poucos neurônios vizinhos são acessíveis ao registro eletrofisiológico). Conhece-se presentemente a constituição de alguns módulos elementares, que desempenham suas funções na distribuição de informações entre os neurônios e a periferia (canais sensoriais), desta para o sistema nervoso central e deste para a periferia (músculos e os órgãos que efetuam os ajustes hemodinâmicos, ventilatórios, endócrinos etc.).

### 1. Introdução

O nome circuito neural é utilizado para denominar conjuntos de neurônios interligados estruturalmente a fim de gerar padrões funcionais específicos. O termo circuito

---

<sup>1</sup>cetiar@usp.br

significa, literalmente, uma associação fechada mas todos os circuitos neurais têm um ou mais pontos de entrada e de saída de informação, sem o que seriam obviamente inúteis.

O estudo dos circuitos neurais denomina-se *hodologia neural* (*hodologia*, em grego, significa estudo das vias). Esse nome aplica-se ao estudo das vias neurais, das rodoviárias, das telefônicas, das eletrônicas etc. A *hodologia* do sistema nervoso dos animais vertebrados é, provavelmente, o campo mais complexo da ciência, não só pela complexidade intrínseca dos circuitos como pelo elevado número de neurônios que existem no sistema nervoso dos animais de posição filogenética mais elevada na série animal. Na espécie humana calcula-se que há, aproximadamente, 100 bilhões de neurônios e uns 200 bilhões de células gliais, todos direta ou indiretamente associados. Como exemplo de tal associação, é interessante citar as células de Purkiný do cerebelo e as conexões que elas recebem de todo o organismo, para informar esse órgão nervoso sobre como os comportamentos programados são executados; a comparação do programa com a execução permite que o cerebelo efetue as correções à medida que o comportamento se desenvolva. A lesão de certas partes do cerebelo impede essa correção e por isso o comportamento programado é executado sempre com defeitos. Por conseguinte, o cerebelo é essencial a que o programa original seja fielmente produzido. Cada célula de Purkiný recebe entre 300.000 e 400.000 botões sinápticos, que lhe trazem informações de cada parte do organismo e do próprio circuito de que cada uma participa.

Deve-se destacar aqui a analogia entre a circuitaria eletrônica e a circuitaria neural. Para se entender como funciona um aparelho eletrônico é necessário conhecer a constituição e a interação dos elementos que compõem seus circuitos. Para se entender como funciona a circuitaria neural também é necessário conhecer corretamente sua constituição. Entretanto, ao contrário da circuitaria eletrônica, que é inteiramente conhecida, por ser invenção humana, a neural é ainda muito enigmática, não por falta de pesquisa, que é imensa e se iniciou há cerca de 120 anos, porém por causa de sua extraordinária complexidade, à qual temos mínimo acesso experimental.

## 2. O neurônio

Presentemente, apenas a microscopia eletrônica permite saber como se constitui um circuito neural simples, sobretudo do sistema nervoso de animais invertebrados. Acredita-se hoje que entre 80 e 90% dos neurônios que constituem o sistema nervoso são interneurônios, os quais constituem os microcircuitos; os demais estabelecem as conexões entre microcircuitos distantes ou enviam informações aos órgãos periféricos que inervam e controlam (músculos, coração, pulmões, órgãos digestivos), ou que deles recebem, os órgãos sensoriais. Serão necessários muitos milhares de anos de pesquisa intensa para se conhecer adequadamente a constituição dos circuitos neurais por intermédio apenas da microscopia eletrônica. Somente a esperança de que em futuro próximo surjam técnicas suficientemente avançadas para permitir o acesso experimental aos neurônios de cada microcircuito nos alimenta a esperança

de que sua constituição, e portanto seu funcionamento, seja inteiramente esclarecida nos próximos séculos.

Qualquer célula do organismo biológico só pode ser devidamente identificada e estudada quando corada por substâncias (corantes) que reagem com componentes celulares, destacando-os e permitindo por isso identificar cada célula individualmente e estudar sua membrana e sua constituição interna. Poucas estruturas intracelulares podem ser estudadas com luz polarizada ou outras técnicas que não impliquem coloração. No que diz respeito aos neurônios, existem pouquíssimos métodos que os coram por inteiro. Em fins do século XIX o histologista italiano Camillo Golgi inventou uma técnica baseada na reação de componentes dos neurônios com sais de prata, a qual ainda hoje é o principal meio de corar todo um neurônio. Interessante é que, por motivos por ora desconhecidos, a técnica de Golgi cora apenas cerca de 10% dos neurônios em dada região do sistema nervoso. Esse defeito é sua grande qualidade, já que, se corasse todos os neurônios, qualquer região apareceria como uma placa negra, tal é a concentração dessas células na maioria das partes do sistema nervoso. A coloração de Golgi cora poucos neurônios mas os que cora o faz de forma completa; isso permite saber exatamente a forma das células e onde se encontram, fornecendo informações cruciais sobre algumas conexões do neurônio corado. Não estranha, portanto, que por ora seja impossível saber como as células nervosas de um circuito se acoplam entre si e como cada circuito se acopla a outros, a fim de gerar funções específicas. Presentemente conhece-se razoavelmente a constituição de apenas uns poucos tipos de circuitos fundamentais, mas identificados principalmente no sistema nervoso de invertebrados e de alguns vertebrados.

É indubitável que, tal como ocorre em eletrônica, a complexidade crescente do sistema nervoso desde os invertebrados até os vertebrados superiores, inclusive o homem, resulte do número também crescente de circuitos e dos padrões de sua associação. Entretanto, da circuitaria de um radiozinho de bolso à circuitaria de um supercomputador a distância é muitíssimo menor que a da que existe entre o sistema nervoso de uma minhoca e o do sistema nervoso humano. Por isso, desvendar toda a circuitaria constituída pela associação de alto número de neurônios é a chave para a compreensão de como o sistema nervoso gera toda a imensa gama de funções que conhecemos, inclusive a consciência, o raciocínio e outras formas de atividade mental etc.

### 3. Os circuitos

Para se começar a entender a constituição de um circuito é necessário antes de mais nada conhecer a arquitetura de sua unidade fundamental, a célula nervosa ou neurônio.

A Figura 1 mostra em A um neurônio ideal, esquemático. O corpo celular ou pericário emite os dendritos de um lado (por onde entram informações) e o axônio de outro, quase sempre único, por onde saem informações, sob a forma de impulsos elétricos (potenciais de ação). Estes resultam da computação, feita pelo pericário, das informações que entram pelos dendritos. A Figura 1B e a Figura 1C mostram

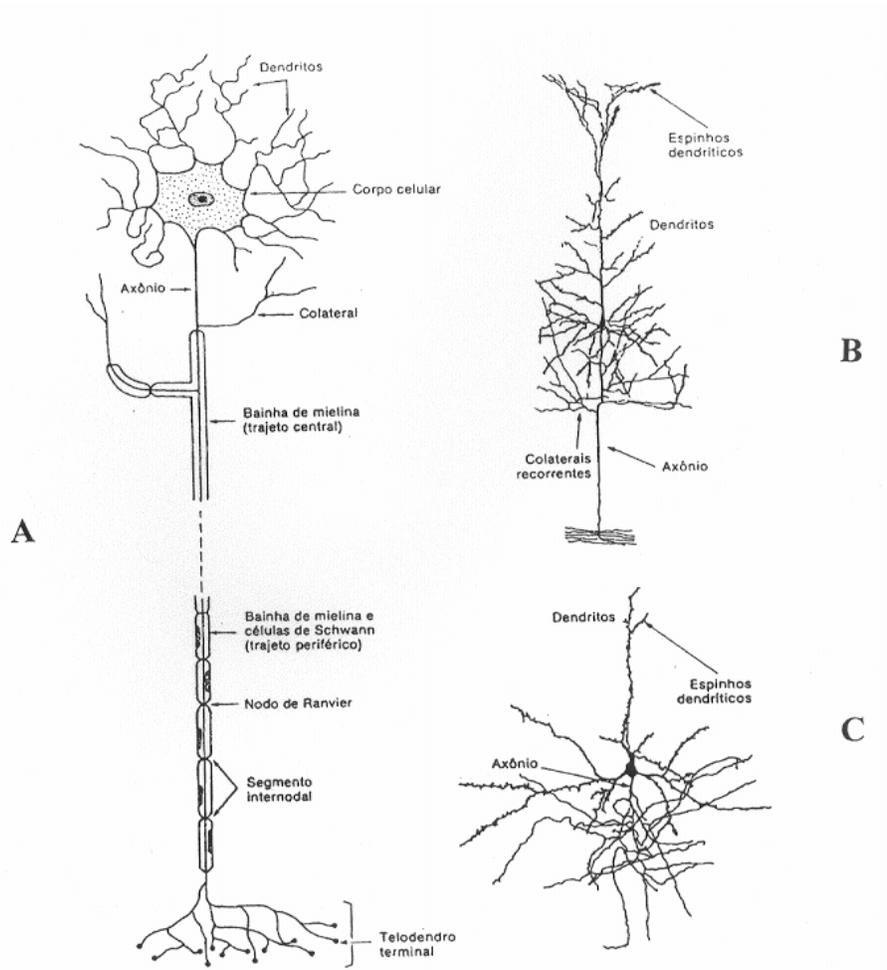


Figura 1: A: esquema de um neurônio, mostrando o corpo celular (pericário), os dendritos (via de entrada de informação) e o axônio (via de saída), este envolvido pela bainha de mielina, constituída de lípidos, por isso isolante, e por células de Schwann, que têm função isolante mas também nutritiva do axônio. Na porção final o axônio se divide em numerosos terminais (telodendro), que transmitem informação às células subseqüentes. O pericário alimenta e mantém os neuritos (dendritos e axônio). B: neurônio do tipo Golgi I do córtex cerebral de coelho, caracterizado por dendritos e axônios longos. O axônio emite numerosos ramos colaterais para o pericário dos neurônios com que se comunica (Cajal, 1911). C: neurônio do tipo II do córtex cerebral de gato, caracterizado por dendritos e axônio curtos, destinados a conexões próximas, que predominam nos microcircuitos.

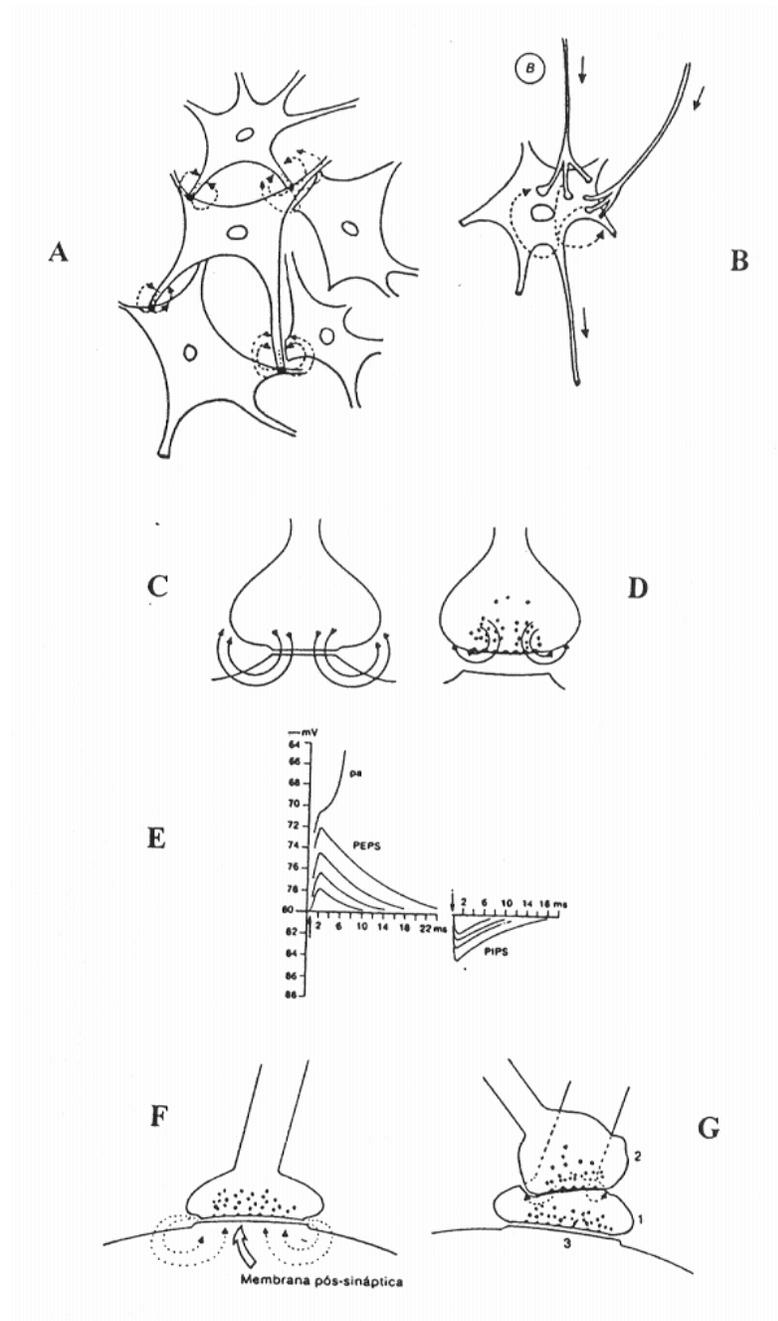


Figura 2: Vide legenda na página 6.

a aparência real de dois neurônios (corados pelo método de Golgi), evidenciando a árvore dendrítica (por onde entram informações) e o axônio (por onde saem informações). Nos dendritos há microscópicas dilatações (espinhos dendríticos), destinadas a aumentar localmente sua superfície, fator importante na transmissão de informações de um neurônio a outro. As conexões entre neurônios se efetuam por intermédio de sinapses, que são contactos diretos ou intermediados por uma fenda onde são liberadas moléculas que efetuam a transmissão (neurotransmissores).

A Figura 2 mostra em A um tipo de conexão curta entre neurônios vizinhos, disposição que deve ser muito comum nos microcircuitos. A Figura 2B esquematiza a forma como um neurônio se acopla a outro distante; pelos dois axônios aferentes descem impulsos (setas) que terminam em sinapses no corpo celular e nos dendritos.

Os axônios da maioria dos neurônios acoplam-se por meio de “botões terminais” (como os denominou Cajal, o descobridor dos circuitos e das sinapses). A Figura 2C ilustra uma sinapse elétrica. O botão terminal acopla-se à membrana de um dendrito ou do pericário de outro neurônio e as correntes que chegam pelo axônio são transferidas diretamente para o neurônio seguinte. Em D vê-se uma sinapse que termina na membrana de outro neurônio mas na qual a transmissão não se realiza por passagem direta da corrente para este. As correntes que fluem pelo botão terminal desencadeiam a liberação, na fenda sináptica, de vesículas que contêm moléculas ativadoras da membrana seguinte (neurotransmissores), por isso essas sinapses denominam-se químicas.

Figura 2 (Legenda). Conexões entre neurônios próximos (A) e distantes (B). C: terminal sináptico (botão terminal) do tipo elétrico, em que a transmissão se faz diretamente entre o botão terminal de um axônio e o ponto do neurônio seguinte ao qual a sinapse se acopla. Note-se que as correntes na sinapse elétrica atravessam diretamente a membrana do neurônio seguinte, já que a fenda sináptica é muito delgada e, portanto, oferece resistência elevadíssima ao fluxo ao longo longitudinal de correntes. D: sinapse do tipo químico; uma vez que a fenda sináptica é relativamente larga, a resistência a correntes ao longo desse espaço é relativamente baixa, o que impede que elas entrem diretamente na membrana do neurônio seguinte. Nesse tipo de sinapse a transmissão de um neurônio para outro é indireta: as moléculas do neurotransmissor liberadas na fenda sináptica acoplam-se a um receptor (radical químico) da membrana, o que despolariza esta última e promove sua ativação. E: à esquerda, curvas de potencial através da membrana, que ilustram convenientemente o que ocorre na membrana durante sua ativação. O potencial de membrana do neurônio pós-sináptico, de cerca de  $-80\text{ mV}$ , reduz-se (despolarização) progressivamente em função da intensidade de sua excitação elétrica. Ao chegar a cerca de  $-72\text{ mV}$  a despolarização atinge o limiar (voltagem crítica da membrana) e desencadeia um potencial de ação (pa na figura) que chega a zero e depois se inverte em algumas dezenas de milivolts, graças à entrada maciça de íons sódio, o quais tornam o interior positivo em relação ao exterior. Enquanto os potenciais parciais, analógicos (PEPS) são localizados, o potencial de ação (pa) propaga-se por toda a membrana do neurônio, chegando aos botões terminais, onde transfere a informação ao neurônio seguinte. À direita, correntes de sentido inverso ao dos

PEPS, denominadas PIPS e que aumentam o potencial de membrana, tornando esta menos excitável. F: fluxo de correntes iônicas entre o botão terminal e a membrana do neurônio seguinte em sinapse de transmissão elétrica. G: o botão terminal de um neurônio (2) acopla-se ao botão terminal de outro (1), aumentando ou diminuindo seu potencial de membrana; no primeiro caso a transmissão pela primeira sinapse (2) bloqueia a transmissão pela segunda (1), ao passo que no segundo caso a transmissão sináptica local pelo neurônio seguinte (3) é facilitada.

Na Figura 2E vê-se como pequenas despolarizações da membrana do neurônio crescem à medida que a liberação de moléculas de neurotransmissores aumenta (PEPS, potenciais excitadores pós-sinápticos), atingindo o limiar crítico. Então, subitamente a membrana se despolariza completamente, originando um impulso nervoso (ou potencial de ação, pa, do qual aqui aparece apenas a parte inicial da fase ascendente). À direita, vê-se o que ocorre com o potencial de membrana em pontos em que deva haver bloqueio da transmissão: o potencial de membrana, graças à liberação de certos neurotransmissores, em vez de diminuir aumenta, o que provoca pequenos potenciais de hipopolarização da membrana (PIPS, potencial inibidor pós-sináptico), tornando muito mais difícil sua ativação por despolarizações parciais.

Na Figura 2F o botão terminal de um neurônio acopla-se à membrana de outro, onde provoca despolarização ou hiperpolarização, o que ativa ou inibe o segundo neurônio, respectivamente. Em G vê-se um botão terminal acoplado ao botão terminal de outro neurônio, onde reduz ou incrementa a transmissão pela segunda sinapse. Esse é um dos arranjos mais poderosos das conexões interneurônicas, visto que é importantíssimo o mecanismo de inativação de neurônios para excluí-los de um circuito, a fim de possibilitar que a informação flua de modo definido em certo instante em certo local de um circuito.

Nos dendritos a informação é quase sempre apenas analógica, ao passo que no axônio a informação é sempre digital. Isso quer dizer que as informações que entram pelos dendritos são potenciais locais e que se manifestam como ondas analógicas positivas ou negativas em muitos pontos de contacto, como acabamos de expor. A soma desses potenciais pelo pericário é que dita se a membrana dessa região do neurônio gera ou não um impulso digital e com que frequência. A voltagem desses impulsos é sempre semelhante mas a frequência varia, e é esta que contém a informação que será transferida para outros neurônios. A Figura 3 ilustra diferentes tipos de associações temporais de impulsos nervosos em axônios. Em 3A, o neurônio emite raros impulsos de frequência baixa e irregular porém quando ativado (estímulo) ele emite imediatamente impulsos (potenciais de ação) de frequência alta no início e depois progressivamente menor, mesmo na vigência do estímulo (que pode ser elétrico artificial ou a partir de impulsos que chegam ao neurônio). Em 3B o neurônio emite salvas de potenciais de ação, separadas por espaços regulares ou diferentes; é óbvio que tanto a frequência de cada salva como os intervalos entre as salvas contêm informação, que será decifrada pelo neurônio que as recebe. Em 3C um estímulo pode provocar bloqueio da gênese de impulsos (à esquerda) ou incrementar a frequência dos impulsos (à direita). Em 3D vê-se o resultado da

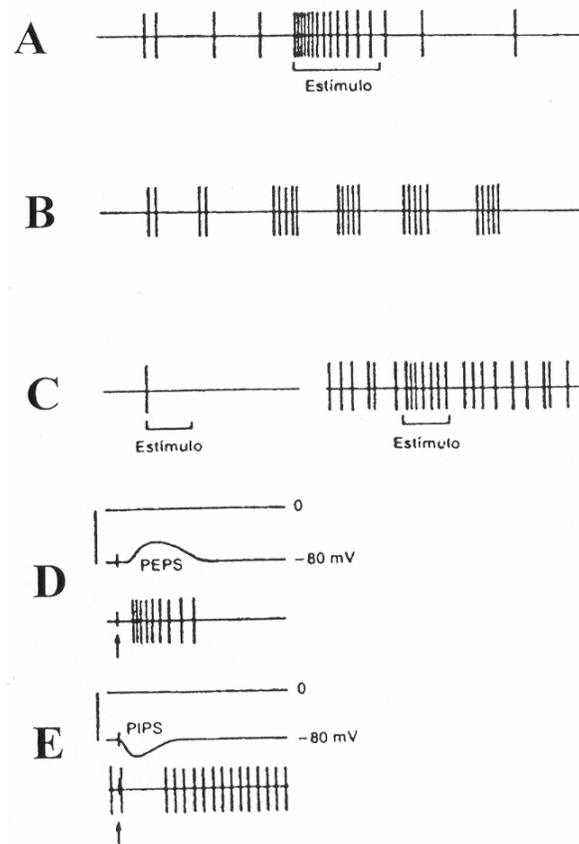


Figura 3: Alguns padrões de informação em um neurônio. A: o neurônio, emitindo potenciais de baixa frequência, é ativado externamente, o que se manifesta como brusco aumento da frequência de impulsos, a qual decai (logaritmicamente) durante a aplicação do pulso de ativação (estímulo). B: o neurônio emite potenciais em surtos de frequências variadas, cada uma delas veiculando uma informação específica aos neurônios seguintes. C: à esquerda, o pulso de ativação provoca apenas um impulso nervoso, seguido de silêncio; à direita, o pulso de ativação aumenta a frequência durante todo o tempo. D: variação do potencial de membrana (que de  $-80\text{ mV}$  baixa a  $-60\text{ mV}$  ou menos) quando o pulso (estímulo) despolariza a membrana (PEPS), o que se manifesta por surto de impulsos de alta frequência. E: variação do potencial de membrana (de  $-80\text{ mV}$  a  $-90\text{ mV}$  ou mais) quando o estímulo provoca aumento do potencial de membrana (PIPS), que se manifesta como bloqueio, enquanto dura o PIPS, da excitabilidade do neurônio; conseqüentemente, os impulsos cessam, retornando quando o PIPS desaparece.

estimulação de um neurônio na membrana de outro, formando-se um potencial estimulador pós-sináptico (PEPS), que reduz o potencial de membrana; enquanto este está reduzido, o neurônio (abaixo) emite uma salva de impulsos. Em 3E vê-se um potencial inibidor pós-sináptico (PIPS) em uma área da membrana de um neurônio e a conseqüente inibição de uma salva de impulsos na saída desse neurônio. Note-se que esta dura enquanto persiste a despolarização da membrana.

Todo circuito neural tem uma ou mais vias de influxo (aférentes ou de entrada) e uma ou várias vias de efluxo (de saída ou eferentes). Entre a entrada e a saída situam-se neurônios que organizam o efluxo de informação em função do influxo. Além disso, em várias etapas do circuito entram informações oriundas de axônios que trazem informações de outros neurônios de fora do circuito, a fim de introduzir impulsos que modificam sua atividade, participando assim de um padrão funcional específico. Em outros pontos do circuito entram outros axônios, que fazem o mesmo. Essas propriedades tornam o circuito a cada instante parte funcional de outros, gerando padrões de circulação de informação extraordinariamente complexos e eficazes. Ao que parece, a maioria das informações em um circuito circula por meio de impulsos propagados.

A Figura 4 mostra um exemplo de circuitos identificados no córtex cerebral (A) por meio da coloração de Golgi, que mostra poucos neurônios mas todos por inteiro. À esquerda vêem-se os números das camadas de células corticais. Note-se que a maioria dos neurônios envia fibras para a camada I, onde se bifurcam; nessa camada há forte interação de informações emitidas por vários neurônios corticais. No córtex occipital, ao qual chegam informações visuais, tal interação pode englobar até 4.000 neurônios. À direita vê-se um pequeno circuito fechado identificado nessa figura por Lorente de Nó, antigo assistente de Cajal. Mais tarde Lorente de Nó identificou circuitos desse tipo (dito reverberantes) em várias regiões do sistema nervoso.

Em B a Figura 4 mostra um tipo comum de circuito, sobretudo ligando vias sensoriais da periferia com músculos. As vias aferentes (que entram em um circuito) nascem, nessa figura, em sensores musculares que medem continuamente o comprimento das células musculares. As vias eferentes nascem em neurônios motores e dirigem-se às células musculares. Esse arco reflexo faz que um músculo se contraia quando estirado; esse fenômeno desempenha funções importantes para controlar a força produzida por um músculo em atividade.

Em C a Figura 4 ilustra um reflexo bem mais complexo, se bem que parecido com o anterior. Note-se que enquanto na Figura B as vias de entrada se conectam diretamente com as de saída, em C há entre entrada e saída outros neurônios (interneurônios) que aumentam a duração da ativação do reflexo.

No esquema B da Figura 4 os impulsos neurais (neste caso provocados por um estímulo, ou pulso elétrico, não por estiramento do músculo) entram na medula pelo ramo aferente e ativam diretamente a via de saída, os motoneurônios, provocando um potencial elétrico simples (com duração em torno de 2 milissegundos) no conjunto de axônios eferentes. No gráfico inferior, C, vê-se que as vias aferentes ativam os motoneurônios por meio de pequenos neurônios (os interneurônios que mencionamos acima), os quais provocam ativação de motoneurônios em cascata, re-



sultando em um potencial simples seguido de outro de menor voltagem porém mais duradouro (resultante da ativação de vários interneurônios). Esse tipo de arranjo hodológico e funcional é comuníssimo em todo o sistema nervoso central.

Figura 4 (Legenda). Algumas formas de circuitos simples. A: circuitos corticais. O córtex cerebral estrutura-se em 6 camadas que trocam complexas conexões entre si. À direita vê-se um circuito reverberante (oscilador), em que um neurônio da camada V ativa interneurônios em série, que o excitam retroativamente, gerando no axônio (eferente) impulsos de frequência variada, segundo o número de interneurônios (Lorente de Nó, 1949). B: esquema de um arco reflexo monossináptico (uma só sinapse em série entre a entrada e a saída do circuito). Um estímulo único aplicado às vias aferentes gera apenas um potencial nas vias eferentes. C: a ativação de aferentes de um arco reflexo intermediado por vários interneurônios provoca um potencial de alta voltagem e curta duração nas vias eferentes, seguido de potenciais de menor voltagem e maior duração, constituídos de impulsos seqüenciais e assíncronos originados pelos interneurônios.

Em gânglios de muitos invertebrados, que constituem partes integradoras de informação complexa no sistema nervoso central desses animais, o número de neurônios é constante e suas posições nos circuitos são invariantes, o que muito facilita seu estudo. Essas características hodológicas permitem o estudo preciso das conexões intraganglionares e da função que cada neurônio desempenha em seu interior. Nos vertebrados, entretanto, o alto número de neurônios e a maior complexidade das conexões são tais que é por ora impossível entender como um microcircuito de fato funciona. É provável que nos vertebrados também haja regularidade de número e de posições de neurônios em um circuito mas devido ao elevado número de neurônios que dele fazem parte é impossível por enquanto saber quantos são e como se articulam esses neurônios.

Uma forma bem conhecida de os neurônios se agruparem é a que caracteriza os núcleos, conjuntos de neurônios que se aglutinam constituindo figuras às vezes circunscritas e reconhecíveis em determinadas regiões (Figura 5). São particularmente reconhecíveis os núcleos que congregam motoneurônios, isto é, neurônios que emitem seus axônios para a periferia e inervam células musculares. Cada motoneurônio envia um axônio a várias células musculares. O conjunto do motoneurônio e das células musculares por ele inervadas constitui uma unidade motora. Quando o motoneurônio emite um impulso todas as fibras por ele inervadas contraem-se concomitantemente. Se o músculo executa movimentos muito precisos um motoneurônio inerva apenas 4 ou 5 células musculares, se executa movimentos grosseiros cada motoneurônio pode inervar centenas de células musculares. Por exemplo: os olhos realizam movimentos sumamente delicados e precisos e suas unidades motoras compõem-se de um motoneurônio para cada 4 ou 5 células musculares. O bíceps braquial, que flete o antebraço sobre o braço e executa movimentos grosseiros, é inervado por um motoneurônio para cada 200 ou mais células musculares. No tórax as unidades motoras podem conter até 600 células musculares por motoneurônio, o que reflete a pouca precisa contração desses músculos,

que em geral se contraem para manter posições (posturas) do tronco como parte de comportamentos múltiplos, não para realizar movimentos finos e precisos.

Cada motoneurônio recebe geralmente milhares de contactos sinápticos pelos quais lhe chegam impulsos elétricos já codificados em centros superiores, a fim de executarem determinada tarefa motora (comportamento). É bem possível que nos vertebrados os motoneurônios ocupem posições mais ou menos fixas no corno anterior da medula, tal como reconhecidamente ocorre nos gânglios de animais invertebrados.

Em várias regiões do sistema nervoso central verifica-se disposição regular dos neurônios em camadas transversais e colunas, o que facilita a participação de cada neurônio em múltiplos circuitos locais, já que formam matrizes estruturais em cada região. A Figura 5A mostra como se dispõem os neurônios no córtex cerebral; é evidente a disposição em camadas horizontais e em colunas verticais. Estudos estruturais e eletrofisiológicos mostram que as colunas verticais constituem módulos funcionais, que são microcircuitos muito complexos acoplados entre si. Desses circuitos emergem vias longas que levam as informações resultantes do processamento nas colunas a outros pontos do sistema nervoso. Por exemplo: quando o córtex cerebral programa certo comportamento (coçar a cabeça, sorrir, falar, caminhar, etc.) tudo o que vai acontecer no comportamento é previamente programado e as informações finais (muito complexas) são enviadas por neurônios específicos a pontos distantes, a fim de haver acoplamento de várias funções. Finalmente, descem impulsos para os motoneurônios em ordem já devidamente programada, os quais ativam músculos específicos e nos fazem executar esses mencionados comportamentos. As posições de motoneurônios em várias regiões do sistema nervoso central aparecem na Figura 5 (v. legenda na pág. 14).

Além da ativação dos motoneurônios na gênese dos comportamentos, descem informações para os circuitos que controlam a pressão arterial, a respiração, a liberação de glicose pelo fígado etc. Esses ajustes vegetativos, como os chamamos, são essenciais para que cada comportamento seja efetuado, visto que os neurônios consomem 20% do oxigênio inspirado pelos pulmões, consumo que equivale a 10 vezes mais do que a média do organismo. Portanto, para que os neurônios funcionem adequadamente é necessário fornecer-lhes oxigênio e glicose em quantidades suficientes para seu funcionamento. Para isso é necessário elevar o fluxo de sangue ao sistema nervoso central, o que se faz por dilatação dos vasos nos locais ativos. Com a vasodilatação a pressão baixa e o sangue então não chega ao sistema nervoso em quantidade adequada; por isso, a pressão também tem que subir. De fato, qualquer que seja o comportamento que estejamos produzindo em dado instante, sempre ocorre aumento da pressão arterial e da respiração.

A Figura 6 representa algumas associações simples entre neurônios. Em 6A um neurônio envia impulsos a outro, por meio de seu axônio, formando sinapses em seus dendritos e pericário. Em 6B dois neurônios associam-se de modo que cada um possa influenciar o outro em serie. Esse tipo de associação permite que um neurônio excitado por outro lhe envie impulsos inibidores, limitando a excitação que está recebendo, e portanto a frequência dos impulsos, ou reforçando-a, mantendo uma atividade elétrica iterativa, oscilante, a qual transmite as oscilações a

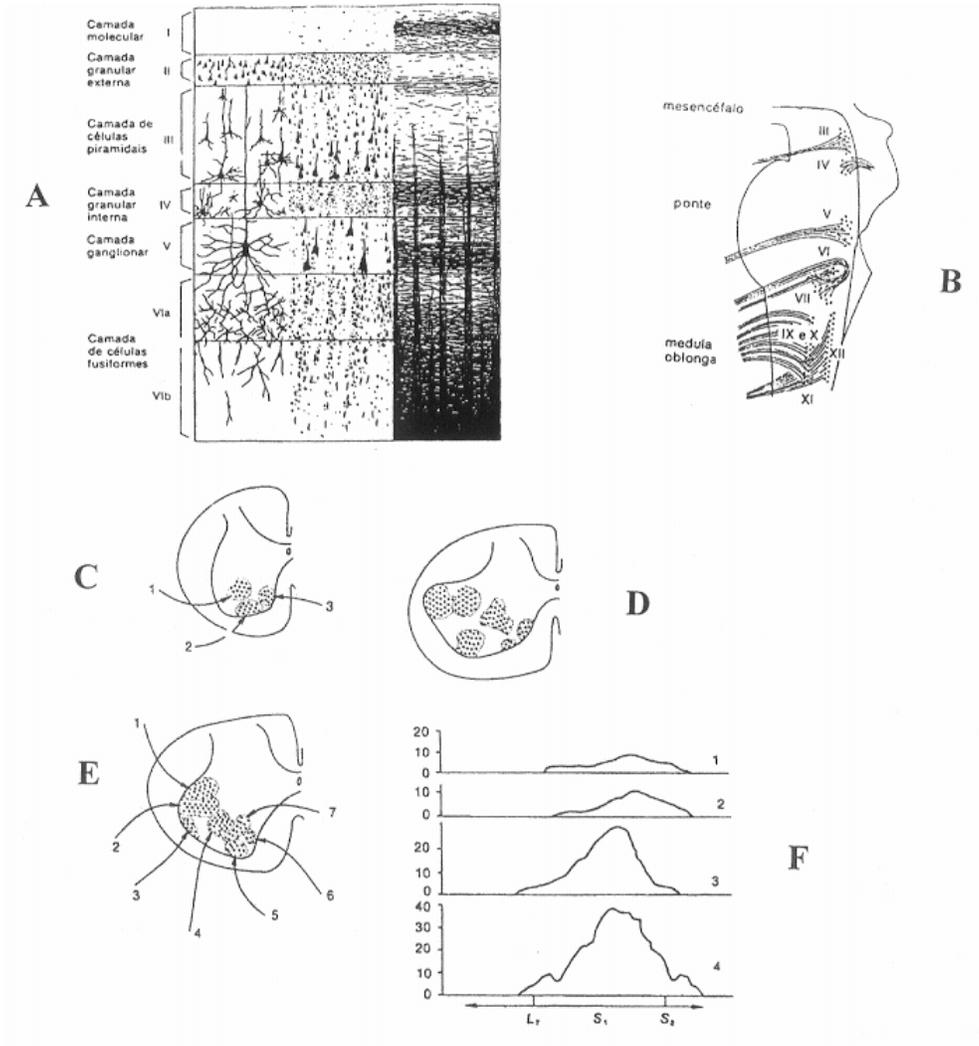


Figura 5: Vide legenda na página 14.

outros circuitos por intermédio de colaterais dos axônios. Em 6C o telodendro de um neurônio se enreda nos dendritos de outro, no qual pode provocar excitação ou inibição. Em 6D está ilustrada uma associação em que dois neurônios paralelos ativam dois outros neurônios, também em paralelo, mas a passagem dos impulsos pode ser regulada para mais ou para menos pelos dois neurônios horizontais, porque estes intervêm na transmissão de informação nas duas cadeias laterais. Em 6E vê-se uma associação que alguns pesquisadores descreveram, mas não muito convincentemente, já que não foi identificada por outros. Nessa associação de neurônios sem dendritos nem axônios a interação implica os pericários dos vários neurônios, em pontos diversos. Se essa associação existe, ela constitui um arranjo ideal entre neurônios, já que permite a gênese de múltiplos circuitos virtuais em função dos locais por onde entram sinais. Em 6F vê-se que um impulso originado no primeiro neurônio excita dois interneurônios, que reexcitam ou inibem o neurônio originador do impulso inicial. Esse tipo de circuito pode manter um processo de reexcitação de um circuito fechado, o qual se completa em cada ciclo com um impulso ou uma salva de impulsos no axônio do primeiro neurônio. Nesse caso o circuito é dito reverberante (v. Figura 4A), que parece executar, entre outras, a função de reter uma informação memorizada nos primeiros instantes do aprendizado, até que ela se consolide em circuitos do mesmo local ou de locais bem distantes no sistema nervoso central.

Figura 5 (Legenda). A: disposição de neurônios em colunas verticais e camadas horizontais no córtex cerebral humano (mapa de Brodmann). Essa disposição é regular no neocórtex de todos os mamíferos mas há diferenças regionais que permitem dividir o córtex cerebral em diversas áreas, algumas funcionalmente já bem caracterizadas (área 17, relacionada com a visão; área 4, relacionada com a mobilização voluntária de músculos durante a emissão de comportamentos, etc.). B: posições de motoneurônios (neurônios que inervam diretamente os músculos) no tronco encefálico. III, IV e VI, motoneurônios que movimentam os olhos. V, motoneurônios que movimentam o maxilar inferior; VII, motoneurônios que movimentam a face (lábios, nariz, pálpebras etc.) e outros. C, D, E: motoneurônios de alguns segmentos da medula espinal, que movimentam a cabeça, o tronco e os membros. Em C, núcleos motores, que movimentam a cabeça do gato. D, motoneurônios que movimentam os membros posteriores. E: motoneurônios que movimentam os membros posteriores. F: representação cartesiana do número de motoneurônios nos núcleos 1 (que flexionam os dedos das patas posteriores), 2 (que estendem os dedos e as pernas posteriores), 3 e 4 (que estendem as pernas), 5 (que estendem os membros posteriores em relação à pelve, 6 (que aproximam os membros posteriores da linha média do corpo) e 7 (idem).

Quando os impulsos recorrentes são excitadores diz-se que a reaferência (ou reatuação, ou retroalimentação, feedback) é positiva. Sendo os impulsos recorrentes inibidores, a circulação de potenciais no circuito tende a reduzir-se ou a cessar. Diz-se, então, que há reaferência negativa.

Na Figura 6G um impulso originado no neurônio 1 aciona o segundo, que emite

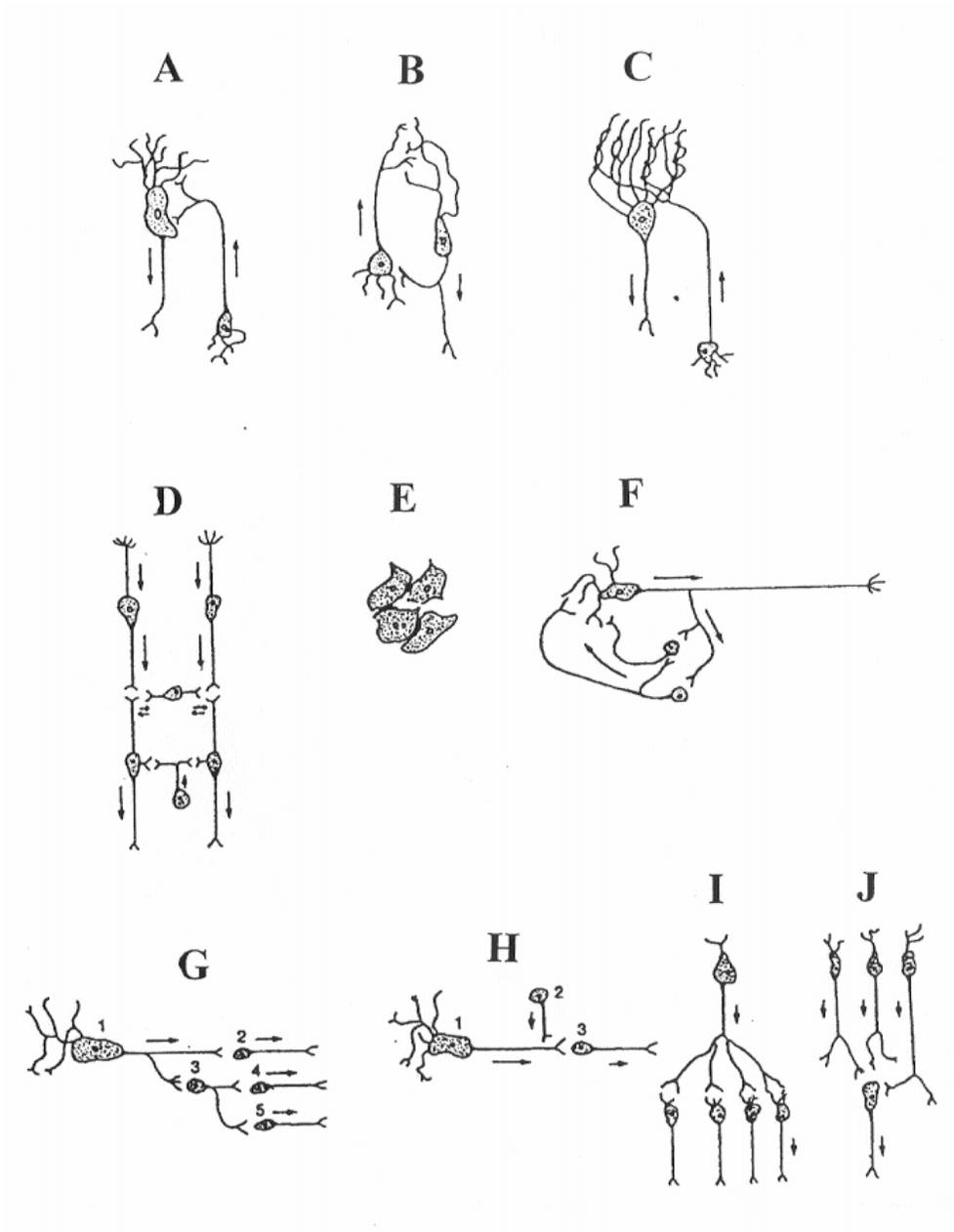


Figura 6: Representação esquemática de alguns tipos de associação de neurônios em circuitos elementares. As flechas indicam o sentido do fluxo de informação. Descrição no texto.

um impulso. Na saída do circuito, pelo ramo colateral do axônio, o neurônio 1 ativa o 3, que mobiliza o 4 e o 5, produzindo impulsos repetidos nos axônios destes dois últimos. Assim, esse circuito é multiplicador de impulsos na saída. Em 6H verifica-se uma associação em que o neurônio 1 ativa o 3 mas, por meio da sinapse no terminal do neurônio 1, o neurônio 2 pode controlar a transmissão de 1 para 3. Em 6I verifica-se divergência de impulsos de um neurônio, que excitam ou inibem vários outros quando entra informação no primeiro; por conexões seriadas tais neurônios podem levar a muitos outros a informação originada no primeiro neurônio. Em 6J, por fim, ocorre o oposto, isto é, convergência de vários neurônios sobre um único, que o ativam como parte de padrões funcionais distintos.

Os circuitos acima descritos, que certamente representam apenas uma fração dos tipos de circuito que existem no sistema nervoso, formam complicadas redes, pelas quais fluem correntes que transferem informação, sob a forma de potenciais locais ou propagados de um ponto a outro. Essas correntes ativam ou inibem pontos bem específicos das redes neurais, fazendo a informação fluir por vias também específicas em dados instantes. Com isso, criam-se momentaneamente vias pelas quais a informação chega aos locais em que provocará seu efeito final. Ao que parece, a maioria dos circuitos contém muitos interneurônios, que funcionando como diminutas válvulas, alteram a direção e o sentido do fluxo da informação e assim determinam a formação de vias específicas em dado instante.

No sistema nervoso humano existem aproximadamente 100 bilhões de neurônios, cada um dos quais recebe milhares de sinapses. A interação desses neurônios por intermédio dessas sinapses cria um número ainda incalculável de conexões, as quais geram elevadíssimo número de padrões funcionais, tais como caminhar, falar, calcular, prestar atenção, sorrir, sonhar, dar uma conferência, ouvir uma conferência etc. etc.

**Abstract.** The nervous system, which is the most complex structure we know so far, is the result of the coupling of specific cells, the neurons. The neurons are coupled according to different patterns, making up the neural circuits, most of which are microcircuits. As in electronics, the circuits are the structures that generate neural functions. In some animals placed in the lowest levels of the phylogenetic scale (such as worms, insects, medusae etc.) the nervous system contains a few thousands of neurons, which are able to generate all the functions necessary for such organisms to survive and reproduce. In humans, however, the nervous system contains nearly 100 billion neurons and 200 billion glial cells (the functions of the latter as parts of neural circuits are still poorly understood). These cells probably make up trillions of connections, which in turn generate the complex behaviors that characterize the phylogenetically upper animals, including mental activity. Artificial staining of the neurons makes it possible to identify some very simple circuits but presently only electronic microscopy allows the detailed study of microcircuits; however, this technique is extremely slow if trillions of neural connections are considered. Therefore, the description of how neurons are coupled to generate microcircuits and, consequently, how such circuits are arranged, a necessary step for the understanding of how complex functions are generated in the nervous system, is not possible as yet. Electrophysiology, the study of the electrical signals generated by living cells, has contributed much to the understanding of how

neural circuits work but it has not led so far to understand the functioning of whole circuits. The main reason for such a limitation is that microcircuits may lodge from hundred to thousands of neurons in a volume that is less than a small fraction of one cubic milliliter. This limitation renders impossible the task of recording the electrical potentials of every neuron in a circuit (in fact, only a few neighboring neurons are accessible to such approach). We know already reasonably well the architecture of only a few elementary neural modules, which exist in invertebrates as well in vertebrates, and that distribute neural impulses to central neurons bringing information from the periphery (sensory pathways) and from the central nervous system to the periphery (motor impulses to muscles and organs that produce blood pressure, respiration, endocrine functions etc.).

## Referências

- [1] T.H. Bullock e G.A. Horridge, “Structure and Function of the Nervous System of Invertebrates”, Vols. 1 e 2., W. H. Freeman and Company, San Francisco and London, 1965.
- [2] G.A. Horridge, “Interneurons”, W. H. Freeman and Company, London and San Francisco, 1968.
- [3] R. Llinás, The intrinsic electrophysiological properties of mammalian neurons; insights into central nervous system function, *Science*, **242** (1988), 1654-1664.
- [4] J. Koester, e S.A. Siegelbaum, Propagated signaling: the action potential, em “Principles of neural science” (E.R. Kandel, J.H. Schwartz e T.M. Jessel, eds.), pp. 151-170, 4<sup>th</sup> edition, McGraw-Hill, New York, 2000.
- [5] R. Lorente de Nó, Cerebral cortex: architecture, intracortical connections, motor projections, em “Physiology of the nervous system” (J.F. Fulton, ed.), pp. 288-330, Oxford University Press, New York, 1949.
- [6] S. Ramon y Cajal, “Histology of the Nervous System in Man and Vertebrates” (N. Swanson e L.W. Swanson, trans.), Vols. 1 e 2, Oxford University Press, New York, 1995.
- [7] J.H. Schwarz e G.L. Westbrook, The cytology of neurons, em “Principles of neural science” (E.R. Kandel, J.H. Schwartz e T.M. Jessell, eds.), pp. 67-87, 4<sup>th</sup> edition, McGraw-Hill, New York, 2000.
- [8] C. Timo-Iaria, Fisiologia do sistema nervoso, em “Fisiologia básica” (M.M. Aires, ed.), pp. 22-300, Editora Guanabara-Koogan, Rio de Janeiro, 1985.

# Modelos Neurais Distribuídos: A Forma no Conteúdo

W.C. FERREIRA JR.<sup>1</sup>, Departamento de Matemática Aplicada, Instituto de Matemática, Estatística e Computação Científica, UNICAMP, CP: 6065, CEP: 13081-970, Campinas, SP, Brasil

**Resumo.** Modelos neurais distribuídos são caracterizados por interações não locais. É apresentada uma discussão sobre interpretações neurobiológicas destes modelos e do fenômeno de morfogênese que ocorre em algumas situações. Focalizaremos em especial os desenvolvimentos que tiveram origem na conjunção dos trabalhos clássicos de H.K.Hartline e J.D.Cowan com o de A.M.Turing relacionados à visão.

## 1. Introdução

*“Os cientistas tentaram, com os seus cérebros, compreender os modelos matemáticos de populações distribuídas, com alguns resultados. Agora, eles tentam entender os seus cérebros com os modelos matemáticos de populações distribuídas!”* sec. Lee A. Segel

A Neurofisiologia moderna, sua abordagem matemática e seu extraordinário desenvolvimento nas últimas décadas são indubitavelmente marcados pelos históricos trabalhos dos fisiologistas ingleses A.L.Hodgkin e A.F.Huxley [19].

O modelo matemático de Hodgkin-Huxley para a dinâmica isolada de um neurônio é uma das mais notáveis e bem sucedidas descrições matemáticas, considerando-se não apenas a Biologia mas qualquer campo do conhecimento. Este fato é particularmente surpreendente se lembrarmos que o tratamento quantitativo em biologia até então restringia-se quase que exclusivamente a abordagens estatísticas e muito raramente se dedicava à formulação de modelos matemáticos derivados de princípios fundamentais.

O histórico artigo de 1952 constituiu-se no fecho de uma série de trabalhos dos mesmos autores que já vinham chamando a atenção de um grande número de fisiologistas com inclinações físico-matemáticas que, a partir daí, foram definitivamente cativados por esta linha de pesquisa. O aparecimento repentino de uma área matematizada de pesquisa em fisiologia foi um acontecimento de grande significação para o desenvolvimento posterior das ciências biológicas muito embora se tratasse

---

<sup>1</sup>wilson@ime.unicamp.br

de uma questão específica de fisiologia celular e aparentemente isolada do universo macroscópico.

Sob uma perspectiva atual é ainda mais notável que um matemático também inglês, A.M.Turing, tenha publicado ao mesmo tempo que Hodgkin e Huxley um artigo sobre modelos matemáticos da morfogênese que se constituiria no outro pilar básico da Biomatemática moderna [62]. Ao contrário, todavia, dos fisiologistas, Turing não publicou qualquer outro artigo anterior (nem posterior!) em áreas correlatas e, talvez por este motivo, o seu trabalho manteve-se incognito da comunidade científica por mais de uma década, e mais ainda da área biológica. Também é interessante frisar que o trabalho de Turing despontava com uma abordagem macroscópica e populacional, totalmente oposta à tendência reducionista e molecular que tomou conta da biologia a partir da segunda metade do século XX. Além disso, a finalidade precípua deste trabalho era apresentar um possível mecanismo para elucidar os processos “mágicos” que resultam na formação das complexas estruturas biológicas e não a dedução de um modelo matemático construído inequivocamente a partir de princípios e hipóteses solidamente firmados.

A abordagem matemática de Turing apresenta assim um novo paradigma na Biomatemática e é essencialmente distinta daquela empregada por Huxley e Hodgkin. Esta última é perfeitamente enquadrada dentro do paradigma clássico da Física (ou, da Engenharia) cujo objetivo patente é sempre descrever, e prever, a evolução dinâmica de um sistema a partir de “leis” fundamentais. O trabalho de Turing não impõe dedutivamente a necessidade do modelo matemático apresentado mas apresenta argumentos que fazem com que a sua explicação se torne plausível, ou melhor, que as hipóteses relacionadas sejam suficientes para a explicação do fenômeno. Poderíamos dizer que esta abordagem tem um caráter hipotético, qualitativo e suficiente, mas não pretende o caráter de uma “lei biológica”, [57]. Ainda como aspecto inovador da metodologia científica, a estratégia de Turing faz uso implícito do princípio de parcimônia de Ockam (“*dentre todas as explicações consistentes para uma observação, aquela mais simples é, provavelmente, também a mais correta*”) no sentido de buscar princípios suficientes, e suficientemente simples, que possam explicar fenômenos complexos na forma como são observados.

Não havendo nenhum outro modelo rival que se compare a ele em simplicidade e na sua capacidade explanatória, e sendo o modelo de Turing baseado em argumentos não apenas simples, mas muito bem conhecidos, sem necessidade da inclusão extemporânea de qualquer hipótese obscura e “mágica”, é natural que despertasse por um lado uma adesão entusiasmada quanto por outro uma oposição cética e até ferrenha.

E, de fato, a existência de quaisquer mecanismos autônomos de formação de estruturas encontrava forte resistência mesmo em se tratando de sistemas simples como as reações químicas inorgânicas, a ponto do, hoje famoso, trabalho experimental de N.Belousov de 1951 ter sido rejeitado “*in limine*” para publicação na União Soviética até a década de 1970. O editor ridicularizou a observação experimental de Belousov, certamente sem tentar reproduzi-la, como sendo impossível frente às leis da termodinâmica. (“*Não me confundam com os fatos!*”). Pode-se imaginar então, qual não seria a reação contra o abstrato mecanismo de Turing para a explicação

de um processo ainda “inobservável”. Entretanto, foi exatamente a química o canal científico que retirou o modelo de Turing da obscuridade, e isto, por obra dos trabalhos desenvolvidos pela escola de físico-química liderada por I. Prigogine em Bruxelas a partir do final da década de 1960, [39]. Até certo ponto é compreensível que o reavivamento da teoria de Turing partisse desta ciência uma vez que o exemplo específico de seu artigo de 1952 se referia a reações químicas. Entretanto, é importante notar que a formação de estruturas em Biologia, ou seja, a morfogênese, era a motivação principal de Turing como se pode ver tanto pelo título do artigo como pela sua frase introdutória reproduzida na citação abaixo.

O desenvolvimento posterior das idéias de Turing em contextos biológicos foi um trabalho liderado principalmente pela escola inglesa de James Murray, [37], assim como por Hans Meinhardt na Alemanha, [34], e Lee Segel nos EUA, [49], [30]. Não seria portanto, surpreendente que, mais cedo ou mais tarde, estas duas vertentes fundamentais da fisiologia matemática, que nasceram incomunicáveis e com escopos e sob paradigmas tão diversos, ainda que próximas no tempo e no espaço, viessem oportunamente se encontrar em alguns pontos da biologia teórica.

No corrente trabalho pretendemos apresentar uma introdução à teoria de Turing sobre possíveis mecanismos de formação de estruturas e, posteriormente, relacioná-las a questões de natureza neurobiológica, especialmente com respeito à visão. A confluência destas vertentes da biologia teórica se deu por conta de um surto de trabalhos surgidos ao final da década de 1970. Estes trabalhos deram início a uma verdadeira revolução no que diz respeito à construção sociológica e psicológica de teorias científicas no decurso da qual antigas barreiras de comunicação e paradigmas de visão científicos tiveram que ser transpostas, [22]. Dentre esses trabalhos, trataremos com maior detalhe os desenvolvimentos que se seguiram aos artigos [16] e [25], [9] e [58]; uma escolha até certo ponto arbitrária que poderia também incluir com igual justificativa vários outros.

O mecanismo de formação de estruturas proposto por Turing em seu modelo matemático é uma das idéias mais profícuas e fundamentais da Biomatemática contemporânea no que esta trata da dinâmica de populações distribuídas continuamente, uma área extremamente ampla que inclui desde populações microscópicas no nível atômico-molecular até o nível de populações humanas. O exemplo original empregado por Turing para descrever a sua teoria de morfogênese baseou-se, possivelmente por motivos pedagógicos, na vetusta teoria da cinética química, onde as populações moleculares são distribuídas segundo a sua posição espacial e as interações ocorrem por interações de reação intermediada pela chamada “lei de ação das massas”. O objetivo central do modelo de Turing é demonstrar a possibilidade de existência de estruturas que seriam caracterizadas como soluções estacionárias e estáveis para as distribuições populacionais de sistemas submetidos a processos de difusão e reação. No exemplo tratado pelo artigo de 1952 as estruturas são representadas por “desenhos”, ou “formas”, no espaço físico tal como também acontece em vários outros casos em neurobiologia [9], [3], [58]. Entretanto, a teoria é facilmente, e vantajosamente, extensível para o estudo de estruturas em espaços de aspectos muito mais gerais, o que é particularmente importante também para a neurobiologia [26], assim como para outras áreas biológicas, como a imunologia, [51].

O modelo de Turing emprega um sistema de equações diferenciais parciais não lineares de difusão cujas propriedades haviam sido muito pouco estudadas até àquela época. A notável exceção a ser feita para este silêncio matemático é um trabalho seminal com motivação também biológica, [28], sobre a existência de ondas viajantes para a equação de difusão com dinâmica vital autônoma, em especial de Verhulst. Esta equação havia sido empregada pelo estatístico e geneticista inglês Ronald Fisher em um famoso artigo de 1935 sobre propagação de genes em populações o que serviu de pretexto para a sua análise matemática. Diz-se, maliciosamente, que a motivação biológica do artigo foi de Fisher e as brilhantes idéias matemáticas nele introduzidas foram inventadas por Kolmogorov, formalizadas por Petrovskii e redigidas por Piskunov.

Este trabalho pode ser considerado com toda a justiça como o terceiro tripé da Biomatemática moderna. A sua influência sobre a neurobiologia teórica / matemática é tão vasta e profunda quanto a de Turing, em grande parte devido também ao fato de que as equações de Hodgkin-Huxley têm essencialmente a mesma natureza da equação de Fisher (KPP-F) e o fenômeno de propagação de ondas, tanto nos axônios [24], [44], como em populações neurais [45], é um tema central da neurobiologia.

A formação e a dinâmica de ondas viajantes e frentes de ondas em sistemas distribuídos e submetidos a processos de difusão e reação é um fenômeno ubíquo na natureza, sendo um exemplo típico disso aquele observado por Belousov. Em Biologia este fenômeno foi observado pela primeira vez no córtex de um cão pelo fisiologista brasileiro A.A.P.Leão em 1944, [29], e posteriormente em diversas outros contextos: [33] e [65].

É interessante observar que a propagação de frentes de ondas também pode ser classificada como um fenômeno de morfogênese de padrões dinâmicos, ao contrário dos padrões estacionários que são o tema central do modelo de Turing e, de certa forma, também o deste trabalho.

Pretendemos apresentar no restante deste artigo, na forma de um esboço, e tanto quanto permitido pelo espaço, algumas idéias básicas da dinâmica populacional no contexto de Turing aplicadas a algumas questões específicas da neurobiologia teórica, com especial atenção para a neurobiologia da visão. A escolha desta área se deve à facilidade com que podemos nela exemplificar de forma prática e concreta alguns processos fundamentais de todo o sistema neurobiológico. Além disso, a visão é o portal por excelência daquilo que, vagamente, reconhecemos ser a “consciência”, este, o atributo mor dos sistemas neurobiológicos e, com absoluta certeza, o tema pinacular de toda a ciência, [4].

Organizaremos a exposição de acordo com a ordem dos seguintes tópicos:

II-A TEORIA DE TURING: PROCESSOS DE ATIVAÇÃO E INIBIÇÃO

III-O MODELO DE TURING PARA INTERAÇÕES NÃO LOCAIS: MODELOS NEURAIIS

IV-MODELO DE SWINDALE PARA O FENÔMENO DE DOMINAÇÃO OCULAR DE HUBEL-WIESEL

V-PADRÕES DE PERCEPÇÃO VISUAL EM ALUCINAÇÕES: O MODELO DE COWAN-ERMENTROUT

## VI-OS MODELOS DE HARTLINE-RATLIFF-KNIGHT-PESKIN: BANDAS DE MACH

## VII-MODELOS COM DISTRIBUIÇÃO CONTÍNUA DE ESTADOS NEURAIIS: OS NOVOS “INSTRUMENTOS ” DE OBSERVAÇÃO

Não tentaremos de forma alguma lecionar sobre os aspectos fisiológicos das questões aqui tratadas por falta eventual de espaço e definitiva de ousadia. Esta dificuldade reverte-se para bem do(a) leitor(a) já que existe uma vasta literatura a respeito escrita com verve e precisão pelos personagens principais do desenvolvimento recente da neurobiologia teórica. Citamos em especial os excelentes textos gerais de biomatemática, [37], [24], [56], e [43], que são particularmente úteis para os leitores provenientes da matemática, e o notável texto de biologia da visão escrito por uma autoridade da área e especialmente apropriado para iniciantes pretensiosos: D.A.Hubel- “Eye, Brain and Vision”, W.H.Freeman, 1988. Uma fonte mais detalhada e acessível sobre fisiologia da visão é encontrada no sítio:

<http://webvision.med.utah.edu> .

## 2. A teoria de Turing: Processos locais de ativação e inibição

*“In this section a mathematical model of the growing embryo will be described. This model will be a simplification and an idealization, and consequently a falsification”*  
Alan Turing.

A estratégia de Turing para explicar a morfogênese biológica em suas manifestações mais elementares consiste na construção de um modelo simples, e portanto falso como ele mesmo enfatiza, que resulte em uma formação “espontânea ” de estruturas , isto é, destituído de qualquer hipótese “maliciosamente arquitetada” ou um mecanismo previamente deliberado para reproduzi-las. Por isso, Turing lançou mão de dois processos corriqueiros que já eram conhecidos e amplamente utilizados na físico-química há mais de um século: a difusão simples (Fick) e a interação de espécies por meio da “lei de ação das massas ”. Tais processos físico-químicos estão presentes em circunstâncias tão variadas e frequentes que parece totalmente implausível a sua suficiência para a formação de estruturas complexas e excepcionais. O que Turing mostrou é que a simples ocorrência destes processos não é, de fato, suficiente para a morfogênese; é necessário que condições de ajustes paramétricos finos sejam também verificados, e é neste ponto que a excepcionalidade e a complexidade do fenômeno serão explicadas.

A lei de ação das massas pressupõe que uma reação efetiva entre duas espécies aconteça com taxa proporcional a uma fração dos encontros entre seus elementos e que, por sua vez, estes encontros aconteçam com uma taxa proporcional às densidades (moles por unidade de volume) das mesmas espécies envolvidas. A taxa (frequência) de encontros depende naturalmente de movimentos microscópicos aleatórios, chamados movimentos brownianos, que são incessantemente efetuados pelas moléculas. Este mesmo movimento browniano resulta em um processo clássico

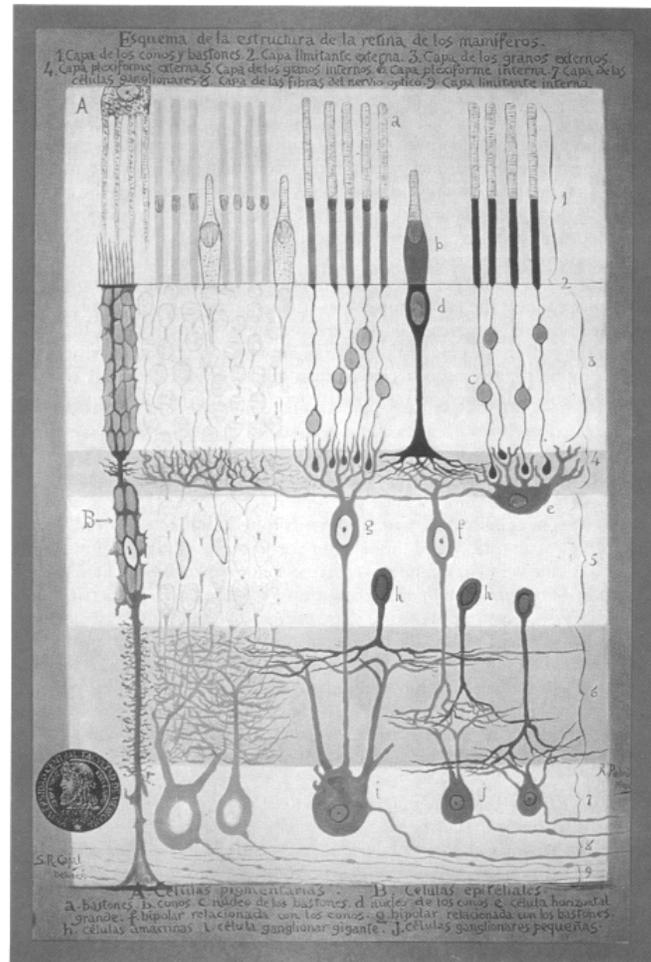


Figura 1: Esboço artístico e preciso do corte da retina de um mamífero desenhado pelo famoso neuroanatomista espanhol Santiago Ramón y Cajal. A luz incide de baixo para cima até atingir as células fotoreceptoras (bastões e cones) na parte superior. O feixe (nervo) ótico sai pelo “ponto cego” da retina, abaixo do desenho. Esta disposição invertida, aparentemente incômoda, deve ter suas razões, mas que a própria razão humana ainda desconhece, [21].

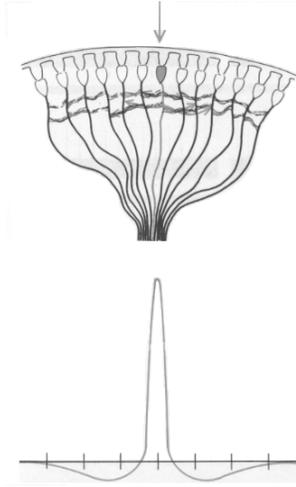


Figura 2: Esboço simplificado do olho “rudimentar” de um *Limulus Polyphemus*. O gráfico abaixo representa a região excitada (positiva) e a região inibida (negativa) como resultado da iluminação pela seta, [47]

de difusão de Fick que já havia sido analisado de maneira simples e notável em alguns dos trabalhos de Einstein no *annus mirabilis* de 1915, [6].

Portanto, podemos dizer que o mecanismo fundamental assumido pelo modelo de Turing foi o processo clássico de difusão que servia de instrumento para possibilitar a intercomunicação e, por conseqüência, a interação entre as moléculas. É importante frisar entretanto que o argumento básico é a possibilidade de interação e não exatamente o mecanismo que o possibilita; este pode ser amplamente generalizado como de fato veremos.

A extensão desta teoria para populações não químicas exigirá a identificação de processos de intercomunicação entre os seus indivíduos que sejam causa e meio de suas modificações mútuas. Para populações de microorganismos, o mecanismo físico-químico browniano que leva a contatos materiais pode ser verificada, [50]. Entretanto, há populações de vários tipos de organismos, micro ou macro, que não efetuam um movimento browniano propriamente dito mas, por outro lado, apresentam comportamentos universais de busca aleatória autônoma cuja formulação matemática é essencialmente idêntica à do exemplo físico-químico. Assim, o modelo matemático e os argumentos de Turing são perfeitamente preservados ainda neste contexto apesar da disparidade da interpretação biológica [49],[50],[23].

Mecanismos muito mais sofisticados de intercomunicação entre organismos e células, fazem uso de sinalizações químicas e são também universalmente encontrados em biologia, [50],[24] e [11]. Nesses casos os indivíduos biológicos interagem pelo contato material com o sinalizador químico através de sensores e não direta-

mente entre si. É interessante observar que o movimento browniano é, em última análise, o processo físico básico que promove a difusão das substâncias sinalizadoras e, portanto, mais uma vez, se encontra no fundamento do mecanismo de interação. Sinalizações sonoras [38] e luminosas [7] são também comuns na sociobiologia de vários outros organismos, pequenos e grandes.

Nos sistemas que nos interessam os indivíduos são os neurônios, e as “espécies” de neurônios a serem consideradas não se constituem necessariamente de células fisiologicamente distintas, mas, por exemplo, de células em estados distintos de excitação [27],[45],[61]. Por outro lado, a intercomunicação entre elas se processa (quase que) materialmente através de sinapses entre axônios e dendritos, isto é, entre partes das próprias células, embora de maneira muito peculiar. Ressalve-se, a bem da verdade, que as sinapses envolvem também uma sinalização química difusiva, mas quase de contato [21]. Os dendritos podem ser amplamente ramificados, atingindo inúmeras outras células, e os axônios podem ser extremamente longos comparados às dimensões do corpo do neurônio propriamente dito. Além disso, a sinapse tem uma eficiência dinâmica e variável com o tempo e uso, e esta característica da intercomunicação neural é um dos aspectos fundamentais e cruciais para o estudo dos comportamentos típicos dos sistemas neurológicos [21],[17].

Apesar da total diversidade de interação sob o ponto de vista fisiológico, é notável que os argumentos de Turing possam permear todas as situações descritas com base em argumentos relativamente simples. A formulação conveniente dos argumentos de Turing em termos de escalas de espaço e tempo que possibilitaram a sua aplicação generalizada deve-se em grande parte a Segel e Levin [49].

Apresentaremos a seguir, de forma quase esquemática, estes argumentos que tornam plausíveis um modelo distribuído interativo para descrição de fenômenos de morfogênese em seus aspectos mais básicos e simples. A nomenclatura, assim como os conceitos básicos utilizados para a construção do modelo matemático serão introduzidos sem maiores formalidades; maiores detalhes podem ser consultados na literatura [12],[14].

É interessante notar inicialmente que os processos difusivos da forma clássica de Fick são comumente relacionados a uma tendência uniformisadora da densidade distribuída. Esta propriedade concorre obviamente na direção oposta à formação de estruturas, ou formas, uma vez que estas são caracterizadas pela manutenção de contornos e fronteiras bem definidas e nítidas. Portanto, de saída, o contexto escolhido por Turing parece não somente implausível como também contaditório. O modelo de Turing mostrou como o comportamento difusivo de duas espécies interativas, desde que em circunstâncias adequadas, pode reverter a tendência uniformizadora da difusão e assim também as nossas expectativas mais ingênuas.

Lembremos que o processo clássico de difusão em uma população descrita pela função densidade  $\rho(x, t)$ , onde  $x \in R^n$  representa o espaço de aspecto da população, é caracterizado por um fluxo da forma  $J = -D \frac{\partial \rho}{\partial x}$ , onde  $\frac{\partial \rho}{\partial x}$ , deve ser entendido como a derivada (gradiente) da função  $\rho(x, t)$ , e  $D$  como o coeficiente de difusão. Para efeito didático, na apresentação do modelo de Turing, o espaço de aspecto será, por enquanto, identificado com o espaço físico e, portanto, a densidade  $\rho(x, t)$  terá a sua interpretação mais usual.

A derivada  $\frac{\partial^2 \rho}{\partial x^2}$  será interpretada, já com vistas à sua generalização, como uma medida (infinitesimal) da heterogeneidade local da distribuição dos indivíduos da população. Para isso, introduzimos o conceito de média no seguinte sentido auto-explicável:  $\langle \rho \rangle_h(x, t) = \int_{B(x, h)} \omega(z) \rho(z, t) dz$ , onde  $B(x, h) = \{z, |x - z| \leq h\}$ , e onde

$\omega$  é uma função de peso positiva, em geral  $\omega = 1$ . É fácil ver que, para pequenos valores de  $h$ , uma medida da heterogeneidade da distribuição nas vizinhanças de  $x$  pode ser definida como  $\delta_h \rho(x, t) = \frac{\rho(x, t) - \langle \rho \rangle_h(x, t)}{\text{vol} B(x, h)}$ , que será analisada segundo uma expansão na ordem de potências de  $h$ . Aproximações de ordem superior em  $h$  podem ser interpretadas como medidas “quase-locais” (não locais) da heterogeneidade e, pela simetria, quando houver, conclui-se que apenas derivadas de ordem par serão retidas nesta expansão. Portanto, em geral, o “fluxo” que promoverá tendências homogeneizadoras “quase-locais” terá a forma genérica  $J_m = \sum_{k=0}^m -D_k \frac{\partial^{2k+1} \rho}{\partial x^{2k+1}}$ ; o fluxo de Fick é de ordem  $m = 0$ . Em modelos neurais a não-localidade é mantida em toda a sua extensão, isto é, para todos as potências de  $h$ , o que implica na utilização de um funcional integral para o fluxo, ou, um operador diferencial de ordem infinita. Estes modelos têm sido relativamente pouco explorados ainda em biomatemática como de resto, em toda a matemática aplicada, [13].

Voltando ao caso clássico, podemos então interpretar o coeficiente de difusão  $D$  como representativo tanto da “motilidade” da espécie como da “transitabilidade” do meio em que se dá a movimentação dos indivíduos. Sob um ponto de vista generalizado podemos interpreta-lo como a capacidade de interação dos indivíduos.

Indivíduos que se movimentam mais rapidamente têm, naturalmente, uma taxa de encontro maior. Este resultado muito conhecido no âmbito molecular (“lei de Arrhenius”) também não tem sido apropriadamente abordado no contexto biológico.

No presente modelo consideraremos este parâmetro como constante mas é facilmente imaginável situações biológicas em que a capacidade de interação entre os indivíduos dependa radicalmente da posição (heterogeneidade do meio), do tempo (influências externas), ou mesmo da densidade local da população.

Considerações simples quanto ao sistema de unidades nos fornecem facilmente a dimensionalidade de  $D$ ,  $[D] = L^2 T^{-1}$ , que independe da dimensão do espaço de aspecto. O significado deste parâmetro como escala intrínseca é simples mas de crucial importância para a compreensão dos argumentos de Turing e, de resto, para qualquer estudo sobre processos difusivos. Para entendermos isto, lembremos que a solução fundamental da equação de difusão nos mostra que o ‘sinal difusivo’ emitido por uma fonte pontual é sentido a uma distância  $x$  depois de um intervalo de tempo proporcional à seguinte expressão  $\tau \propto \frac{x^2}{D}$ . Segel denomina este resultado elementar e fundamental, mas pouco enfatizado, como sendo a ‘regra de ouro’ da difusão, [31], [13].

Uma vez definido o processo de movimentação e estabelecido que a interação se dá “corpo a corpo”, é necessário analisar a dinâmica de reação.

Iniciaremos esta análise com o caso mais simples e mais importante da dinâmica populacional, a dinâmica Malthusiana, cuja representação matemática é dada na forma:  $\frac{du}{dt} = au$ , onde  $a$  uma constante e  $u$  uma concentração homogênea de

população. Mais tarde veremos que esta simplificação é suficiente para os nossos propósitos, se bem entendida, uma vez que consideraremos a dinâmica apenas nas imediações de uma solução onde o modelo Malthusiano fornece uma excelente descrição.

Inicialmente observemos que a expressão acima pode ser escrita como  $\frac{1}{u} \frac{du}{dt} = a$ , o que significa uma taxa de auto influência específica constante. Considerando agora a dimensionalidade dos termos, concluímos que  $[a^{-1}] = T$  e, portanto o parâmetro  $a$  designa uma escala de tempo intrínseca do sistema. Para melhor entender o importante significado deste parâmetro consideremos o caso específico de uma dinâmica auto inibidora, isto é, onde  $a < 0$ , que se inicia com  $N$  indivíduos (em medida dimensionalizada). Observando agora que o número de indivíduos que sobreviveram exatamente até o intervalo  $\{t, t + dt\}$  é  $-du$ , isto é, aqueles que viveram até lá e de lá não passaram ou, ainda, os que morreram no seu transcurso, concluímos que o seu tempo médio de vida é dado por:

$$\tau = \frac{-1}{N} \int_0^N t \, du = a^{-1}. \text{ O mesmo argumento pode ser reproduzido com uma}$$

dinâmica auto ativadora, ou catalisadora, isto é, com  $a > 0$ .

Portanto, reunindo as duas interpretações, concluímos que os indivíduos de uma população submetida simultaneamente a um processo de difusão e a uma dinâmica Malthusiana negativa, percorrem uma distância da ordem de  $\sqrt{\frac{D}{a}}$ . Este resultado simples, mas não trivial, é fundamental para a interpretação da teoria de Turing.

Consideremos agora um sistema de duas espécies distribuídas no espaço segundo as densidades  $\rho_i(x, t)$ ,  $i = 1, 2$ , e submetidas a processos de difusão clássicos com coeficientes de difusibilidade distintos  $D_i \neq 0$ . Suponhamos ainda que as duas espécies interagem pela ‘lei de ação das massas’, de tal forma que o princípio de conservação aplicado a ambas nos fornece o seguinte sistema de equações diferenciais parciais:

$$\frac{\partial \rho_i}{\partial t} = D_i \frac{\partial^2 \rho_i}{\partial x^2} + R_i(\rho_1, \rho_2), \quad i = 1, 2.$$

Uma estrutura neste sistema seria simplesmente uma solução estacionária, isto é, independente do tempo, mas efetivamente dependente do espaço,  $u_i(x)$ , e dinamicamente estável.

Chamaremos estas soluções de “padrões”, ou “formas”. O conceito filosófico de estrutura, padrão ou forma é aparentemente mais rico pois pressupõe alguma “particularidade”. Entretanto, esta “particularidade” depende dos olhos do observador, isto é, se lhe parece comparável a alguma outra estrutura previamente conhecida, ou não! É impossível determinar qual distribuição espacial é um “padrão aceitável” sem uma lista exaustiva ou descritiva do que isto significa. O reconhecimento de padrões é uma questão de relevância central no estudo da cognição visual [55], assim como na teoria de sistemas imunes no que diz respeito ao processo de auto identificação, [51]. Sugirimos ao (à) leitor(a) relacionar este conceito com a sua própria cognição visual atentando para as ilustrações da figura 5 e consultando algumas figuras no sítio, [10].

A teoria de Turing tem por objetivo simplesmente (!) mostrar que a formação de estruturas em um sentido geral pode ocorrer sem a necessidade de um mecanismo

de instruções algorítmicas e teleológicas. A construção de estruturas pré almeçadas como, por exemplo, manchas na pele do leopardo, listas na pele da zebra, [37], ou no cortex visual, [58],[59], atividades neurológicas correspondentes a alucinações [9], e muitos outros, [37], são questões específicas a serem resolvidas nos seus respectivos contextos. A explicação que o mecanismo de Turing de fato oferece a uma vastíssima gama de formação de padrões específicos constata cabalmente que a teoria não é somente é plausível.

A demonstração de existência de soluções padrões para um sistema geral é uma tarefa difícil ainda hoje e seria praticamente impossível há meio século atrás. Além disso, as demonstrações matemáticas rigorosamente técnicas não são exatamente os argumentos mais convincentes em matemática aplicada pois usualmente requerem hipóteses sem significado biológico e frequentemente produzem resultados de difícil, ou inexistente, interpretação.

A estratégia de Turing foi então supor uma situação mais específica em que o sistema homogêneo e destituído de difusão  $\frac{du_i}{dt} = R_i(u_1, u_2)$ , dispõe de uma solução estável, digamos  $u_{i0}$ , e determinar sob que condições (valores dos parâmetros) a introdução de processos difusivos poderiam desestabilizar o sistema, e portanto, leva-lo a uma outra solução, possivelmente estacionária. Enfim, o objetivo era bem mais modesto do que demonstrar diretamente a existência de soluções padrões; era principalmente o de analisar a existência de mecanismos dependentes apenas de processos difusivos clássicos e interações de massa que possam construí-las a partir de estados totalmente sem estruturas (“caldos”), ou seja, uniformes no espaço. Naturalmente, a instabilidade da solução estacionária original, não garante que uma perturbação levaria o sistema a uma solução padrão estacionária; há a possibilidade de que o sistema se aproxime de soluções do tipo frentes de ondas [24], caóticas [24], ou simplesmente sem classificação, que todavia podem ser consideradas como padrões dinâmicos, e o foram pelo próprio Turing. (“*Most of an organism, most of the time, is developing from one pattern into another, rather than from homogeneity into a pattern. One would like to be able to follow this more general processes mathematically also. The difficulties are, however, such that one cannot hope to have any very embracing theory of such processes, beyond the statement of the equations. It might be possible, however, to treat a few particular cases in detail with the aid of a digital computer*”). [62].

A demonstração de que simples processos difusivos poderiam desestabilizar estados uniformes já era uma idéia suficientemente revolucionária para a época.

Os argumentos serão apresentados em uma dimensão, isto é, para  $x \in R^1$  já que em dimensões superiores apenas a manipulação algébrica se torna mais complexa sem qualquer acréscimo conceitual significativo.

Escrevendo o sistema de Turing homogêneo na forma vetorial,  $u = (u_1, u_2)$  e  $\frac{du}{dt} = R(u)$ , a condição de que o estado estacionário uniforme  $u_0 = (u_{10}, u_{20})$  seja assintoticamente estável nos leva à concluir que a matriz jacobiana

$$\frac{\partial R}{\partial u} = \begin{pmatrix} \frac{\partial R_1}{\partial u_1} & \frac{\partial R_1}{\partial u_2} \\ \frac{\partial R_2}{\partial u_1} & \frac{\partial R_2}{\partial u_2} \end{pmatrix}_{(u_0)} = A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

tem apenas autovalores com a parte real negativa, condição equivalente às

seguintes afirmações:

- 1)- $T_r(A) = A_{11} + A_{22} < 0$
- 2)- $\det(A) = A_{11}A_{22} - A_{21}A_{12} > 0$ . [2].

Voltando ao modelo completo de Turing, com difusão, abordaremos a questão sobre a possibilidade de que alguma perturbação, qualquer que seja ela, independente de condições de fronteira, possa desestabilizar a solução uniforme. Observe que  $\rho(x, t) = \rho_0 = (u_{10}, u_{20}) = \text{constante}$ , é uma solução (estacionária, uniforme) do sistema completo. Desestabilizar aqui significará, desestabilizar o sistema linearizado que descreve, aproximadamente, o comportamento dinâmico do modelo matemático quando  $\rho(x, t) \approx \rho_0 = (u_{10}, u_{20})$ , ou seja,  $\rho(x, t) = \rho_0 + \eta(x, t)$ , enquanto a perturbação  $\eta(x, t) \approx 0$ .

O modelo de Turing linearizado nas vizinhanças da solução uniforme é obtido tomando-se apenas os termos lineares em  $\eta$  expandidos na equação

$$\frac{\partial \rho}{\partial t} = \frac{\partial \{\rho_0 + \eta(x, t)\}}{\partial t} = D \frac{\partial^2}{\partial x^2} \{\rho_0 + \eta(x, t)\} + R(\{\rho_0 + \eta(x, t)\}), \text{ ou seja,}$$

$$\frac{\partial \eta}{\partial t} = D \frac{\partial^2 \eta}{\partial x^2} + A\eta = L\eta$$

onde  $D$  é uma matriz diagonal  $D = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}$  e  $D_i$  são os respectivos coeficientes de difusão de cada espécie.

A equação linearizada fornecerá a tendência dinâmica da perturbação  $\rho(x, t) = \rho_0 + \eta(x, t)$  enquanto ela permanecer nas vizinhanças de  $\rho_0$ , ou seja, enquanto  $\eta(x, t) \approx 0$ . Portanto, a questão se resume em analisar a evolução temporal da solução  $\eta(x, t)$  da equação linearizada considerando-se a possibilidade de quaisquer condições iniciais. Para isso, considerando o caráter linear da equação, utilizaremos o princípio de superposição de Fourier que nos leva a dinâmicas na forma  $\eta(x, t) = e^{\lambda t} v_\lambda(x)$ , onde  $v$  é uma solução do problema espectral livre  $Lv = \lambda v$ . Neste caso,  $v$  representará a perturbação inicial  $\eta(x, 0) = v_\lambda(x)$ . Em princípio supomos então que todas as perturbações iniciais possam ser bem representadas como superposição contínua da família  $\{\lambda\}$  de funções espectrais  $\{v_\lambda\}_\lambda$ , o que é a própria essência do Método de Fourier.

Portanto, para desestabilizar o sistema, basta que exista uma destas funções com  $\lambda > 0$ .

Para analisarmos esta questão utilizaremos novamente o princípio de Fourier na representação das funções  $v_\lambda$  como superposição de funções do tipo  $v_{\lambda_k}(x) = e^{ikx} w_k$ , onde  $w_k \in R^2 - \{0\}$  é constante. Observemos que interessa-nos apenas funções limitadas no espaço, razão porque o número de onda  $k$  deve ser real.

Finalmente, as funções básicas  $v_{\lambda_k}(x) = e^{ikx} w_k$  devem ser procuradas como soluções da equação espectral

$$Lv_{\lambda_k}(x) = \lambda_k v_{\lambda_k}(x), \text{ ou seja,}$$

$$\left( D \frac{\partial^2}{\partial x^2} + A \right) e^{ikx} w_k = e^{ikx} (-k^2 D + A) w_k = e^{ikx} \lambda_k w_k,$$

que se reduz ao problema espectral algébrico para

$$(-k^2 D + A) w_k = \lambda_k w_k,$$

cujas soluções, não trivial, somente é possível se

$$\det(-k^2 D + A - \lambda_k) = 0.$$

Enfim, existirá uma perturbação  $e^{ikx}w_k$  inicial não estável para algum  $k$  real, (isto é, cuja evolução

$\eta(x, t) = e^{\lambda_k t} e^{ikx} w_k$  crescerá), somente se a matriz

$$A(k) = -k^2 D + A$$

tiver algum autovalor  $\lambda$  com parte real positiva.

Voltamos pois ao conjunto de condições equivalente à estabilidade de uma matriz  $2 \times 2$  :

$$1- \text{Tr}(A(k)) < 0$$

$$2- \det A(k) > 0,$$

que neste caso “deve ” ser violado.

Já que a primeira condição é automaticamente satisfeita,

$$\text{Tr}(A(k)) = \text{Tr}(A) - k^2(D_1 + D_2) \leq \text{tr}(A) < 0,$$

resta-nos apenas a segunda condição. Portanto, para a existência de perturbações instáveis, é necessário determinar valores dos parâmetros das matrizes  $D$  e  $A$ , (coeficientes de difusão e reação) para que tenhamos  $\det A(k) = D_1 D_2 k^4 - (A_{11} D_2 + A_{22} D_1) k^2 + \det A < 0$  para algum  $k$  real.

Observemos inicialmente que a expressão acima se trata de um polinômio de segundo grau na variável real positiva  $s = k^2$ ,

$$P(s) = D_1 D_2 s^2 - (A_{11} D_2 + A_{22} D_1) s + \det A, \text{ e , como}$$

$P(0) = \det A > 0$ , e  $D_1 D_2 > 0$ , concluímos que a existência de algum  $s > 0$  tal que  $P(s) < 0$ , nos leva à seguinte condição necessária:

$$I) - A_{11} D_2 + A_{22} D_1 > 0$$

que não é todavia suficiente.

Como o valor mínimo da parábola é atingido em  $s_m = k_m^2 = \frac{A_{11} D_2 + A_{22} D_1}{D_1 D_2} (> 0$ , pela condição I) temos  $P(s_m) = \det A - \frac{(A_{11} D_2 + A_{22} D_1)^2}{4 D_1 D_2}$ .

Portanto, a condição final para instabilidade é dada por:

$$II) - \det A - \frac{(A_{11} D_2 + A_{22} D_1)^2}{4 D_1 D_2} < 0.$$

Para que o conjunto de condições matemáticas {I, II} seja de alguma utilidade do ponto de vista do modelo matemático, é importante que analisemos o seu significado e ocorrências dentro do contexto biológico.

Começemos pelo aspecto interativo do sistema.

Como  $\text{Tr}(A) = A_{11} + A_{22} < 0$ , (1), concluímos que pelo menos um dos coeficientes,  $A_{11}, A_{22}$  é negativo. Digamos que  $A_{22} = \frac{\partial R_2}{\partial u_2}(\rho_0) < 0$ . Como, por outro lado, I)- $A_{11} D_2 + A_{22} D_1 > 0$ , concluímos que necessariamente  $A_{11} = \frac{\partial R_1}{\partial u_1}(\rho_0) > 0$ . Portanto, os coeficientes da diagonal de  $A$  devem ter sinais opostos. Estes coeficientes representam as constantes Malthusianas para a dinâmica das respectivas espécies no que se refere à auto influencia que cada uma exerce sobre si nas imediações do estado uniforme, ou seja, devemos ter um elemento auto inibidor e um elemento auto-ativador.

Além disso, como  $\det A = A_{11}A_{22} - A_{21}A_{12} > 0$  (1), concluímos que  $A_{12}A_{21} < 0$ . Portanto, um modelo de Turing para duas espécies é necessariamente constituído por uma de duas formas de reação:

A-  
 $\begin{pmatrix} + & - \\ + & - \end{pmatrix}$ , auto ativador (1) que é também ativador cruzado, e um auto inibidor(2) que também é inibidor cruzado, e

B-  
 $\begin{pmatrix} + & + \\ - & - \end{pmatrix}$ , auto ativador(1) que é inibidor cruzado e um auto inibidor(2) que é um ativador cruzado.

Analisemos agora uma questão mais interessante e que diz respeito à difusão.

Como  $Tr(A) = A_{11} + A_{22} < 0$ , (1), e  $A_{11}D_2 + A_{22}D_1 > 0$ , (I), devemos ter  $D_1 \neq D_2$ , ou seja, no modelo de Turing as espécies auto ativadora e auto inibidora devem ter necessariamente coeficientes de difusão distintos.

Além disso, como  $A_{11}D_2 + A_{22}D_1 > 0$ , (I), ( e  $A_{22} < 0$ ,  $A_{11} > 0$ ), concluímos também que

$$\frac{D_2}{-A_{22}} > \frac{D_1}{A_{11}} \text{ ou,}$$

$$\frac{D_2}{D_1} > \frac{|A_{22}|}{A_{11}}, \text{ já que de } A_{11} + A_{22} < 0 \text{ temos } A_{11} < |A_{22}|.$$

Finalmente temos então a terceira condição:

C-

A difusibilidade do auto inibidor  $D_2$  é maior do que a difusibilidade do auto ativador  $D_1$ .

A primeira “regra de ouro” para a caracterização de um modelo de Turing pode ser então enfatizada pela seguinte frase:

“*O auto-inibidor tem difusão rápida e o auto-ativador tem difusão lenta*”.

Para os nossos objetivos todavia, é interessante desenvolver um pouco mais esta caracterização levando em conta a dimensionalidade das constantes de reação,

$$[A_{11}] = [A_{22}] = T, \text{ das expressões } \left[ \left( \frac{D_2}{-A_{22}} \right)^{\frac{1}{2}} \right] = \left[ \left( \frac{D_1}{A_{11}} \right)^{\frac{1}{2}} \right] = L,$$

e a discussão acima sobre a distância de sinalização de um processo difusivo.

Com isto, podemos interpretar a desigualdade

$$\left( \frac{D_2}{-A_{22}} \right)^{\frac{1}{2}} > \left( \frac{D_1}{A_{11}} \right)^{\frac{1}{2}}$$

como a segunda “regra de ouro” para caracterização do modelo de Turing:

“*O (auto)inibidor é de longo alcance e o (auto)ativador é de curto alcance*”.

Estas regras gerais são importantes para que possamos intuir sobre a ocorrência da formação de padrões em sistemas dinâmicos distribuídos e se constituem em um dos pilares fundamentais da teoria contemporânea da morfogênese.

As aplicações que faremos da teoria de Turing aos sistemas neurobiológicos se baseiam essencialmente nestas interpretações.

Considerando que os processos de ativação e inibição, a curta e longa distancia, são mecanismos inerentes aos sistemas neurobiológicos [21], é natural agora prever que o modelo de Turing deva ser um instrumento de argumentação importante no estudo teórico desta área.(Fig.2)

### 3. O modelo de Turing para interações não locais: Modelos neurais

“*You cannot believe in astronomical observations before they are confirmed by theory*”. Arthur Eddington, *sec.*[53].

Conforme já assinalamos, o conceito clássico do processo de difusão pode ser considerado como uma (auto)transferência homogeneizadora da distribuição de indivíduos no espaço de aspecto que faz uso de um mecanismo com caráter estritamente local, isto é, utiliza apenas informações constantes em uma vizinhança “infinitesimal” da distribuição.

O fluxo difusivo é um funcional diferencial da densidade e, portanto, totalmente determinado pelos valores da função distribuição em uma vizinhança infinitesimal do ponto. Isto faz com que distribuições com características outras completamente diversas, mas com mesma derivada em um ponto  $x$ , determinem o mesmo fluxo de difusão, fixados o meio e a espécie, isto é, o coeficiente de difusão. É importante ressaltar que o termo “infinitesimal” tem um significado matemático preciso mas que, mal interpretado, pode dar a falsa impressão de uma hipótese completamente absurda. O exemplo antológico desta incoerência é o famoso “escândalo” da velocidade infinita dos processos difusivos que ainda é motivo para muito desperdício de tinta. “Paradoxos” desta natureza raramente subsistem a uma análise simples, mas não de todo trivial, sobre as escalas de tempo e espaço utilizadas na formulação do modelo matemático, uma questão central e relativamente pouco analisada. Com isto, economiza-se uma grande quantidade de polêmicas infundáveis e inúteis.

Por exemplo, a derivada com relação à variável tempo em um modelo matemático não pode ser literalmente interpretada como um limite no sentido matemático. Ou seja,  $\frac{d\rho}{dt}(t) = \Delta \rightarrow \lim_{\Delta \rightarrow 0} \frac{\rho(t+\Delta) - \rho(t)}{\Delta}$  não tem sentido “experimental” pois, para  $\Delta$  suficientemente pequeno, como exige o conceito matemático de derivada, a função “experimental”  $\rho$  não se modificaria e o resultado seria sempre nulo. (Qual o tempo gasto por um elétron para realizar uma volta completa no átomo de Bohr? Logo, com  $\Delta$  menor do que este valor, todo o metabolismo estaria congelado!).

Como então resolver esta questão? Na verdade, o modelo matemático contínuo deve ser encarado como uma aproximação matemática de um modelo discreto em que a derivada é um “limite freiado”, isto é,  $\frac{d\rho}{dt}(t) = \frac{\rho(t+\Delta) - \rho(t)}{\Delta}$  onde o  $\Delta$  é pequeno comparado com intervalos onde ocorrem grandes variações de  $\rho(t)$ , mas tomado em uma faixa onde há uma “estabilização” do valor do quociente. Esta faixa caracteriza uma escala de tempo intrínseca do modelo matemático; qualquer efeito exterior que ocorra em um intervalo menor, deve ser considerado como sendo “instantâneo”. Uma das poucas discussões a respeito desta questão pode ser encontrada, em uma forma elementar e clara, em [31], capítulos 6 e 13, ou em [13].

Nos modelos chamados neurais [37] supõe-se que ocorra uma interação “não local” no sentido de que os processos de comunicação, ou de influência, exercidos entre os indivíduos desta população (em geral, neurônios) se dá em um intervalo de tempo muito menor do que a escala de tempo característica para a variação dos estados dos próprios indivíduos. Portanto, “grandes distâncias” podem ser percorridas pela

sinalização em uma escala de tempo muito inferior à escala da dinâmica do sistema e, com isto, para efeito do modelo matemático, podemos considerar estes processos como “instantâneos” e de longa distância.

Isto é o que de fato ocorre nos sistemas neurobiológicos em diversos contextos. O estado de excitação dos neurônios é determinado pela dinâmica de suas cargas iônicas ( $Ca^{++}, K^+, Na^+$ , etc.) que, portanto, dependem do fluxo de matéria (íons) através de membranas, (difusão facilitada), um fenômeno mais lento do que a transmissão de sinais elétricos através dos axônios: 0,1 a 10 metros por segundo. Este último é um processo de reação e difusão *à la* KPP-F. Por outro lado, é importante lembrar que os axônios podem ter um comprimento de ordem muitas vezes maior do que a dimensão do corpo do neurônio (soma) que tem tipicamente de 10 a  $50\mu m$  de diâmetro. (Em casos extremos, os axônios podem atingir até 1m de comprimento!). Já na retina, se considerarmos apenas as imediações da fovea ( $\approx 0,5mm$ ) estamos diante de um sistema de pequena extensão e de interações perfeitamente não locais.

Na verdade, a conexão final, a sinapse, se dá por um processo de difusão de neurotransmissores, mas em um curtíssimo espaço, existente entre dendritos e axônios (“cleft”)  $0,02\mu m$ . Considerando-se que o coeficiente de difusão no caso é da ordem de  $D \approx 10^{-5}cm^2/seg$ , a regra de ouro da difusão nos fornece uma estimativa do tempo gasto para esta etapa:  $\tau_{syn} \approx \frac{(2 \times 10^{-6}m)^2}{10^{-5}(10^{-2}m)^2/seg} = 4miliseg$ , portanto, de ordem muito menor do que o tempo gasto pelo pulso para percorrer o axônio.

Lembremos que o processo clássico de difusão presente no modelo original (e didático) de Turing exerce um papel de “mecanismo de sinalização”, já que a interação entre os indivíduos se dá por contato material e depende naturalmente da sua movimentação (movimento browniano) e da eventualidade do encontro (lei das massas). É importante re-enfatizar, o que raramente é feito, que no modelo de Turing o busilis não está exatamente nos mecanismos básicos (difusão e lei das massas), mas sim na sua finalidade (influência) e na maneira pela qual ela é exercida; ativação pelos vizinhos próximos e inibição pelos distantes. Portanto, não há necessidade da concorrência de duas espécies se os processos básicos são cumpridos.

Consideremos então um protótipo (linear) de um modelo matemático distribuído onde as regras fundamentais de auto interação de Turing sejam representadas (em uma dimensão, para simplificar o argumento) por um processo generalizado de comunicação (“instantânea”) de influências descrito da seguinte maneira :

$$\frac{\partial \rho(x,t)}{\partial t} = \int_{-\infty}^{\infty} w(|x-y|) \rho(y,t) dy = L_w \rho.$$

Observe que este modelo prescinde da dinâmica não linear e das interações cruzadas.

O termo à esquerda denota a taxa de variação do estado do sistema pontualmente em  $x$  (representado pela densidade  $\rho(x,t)$  de apenas uma espécie) e a equação determina que este estado é instantaneamente atualizado por uma ponderação contínua dos estados pontuais em  $y$  de todo o sistema. A função  $w(z)$ , chamada função núcleo de influência, introduz um “peso de influência” que depende unicamente da distância  $z = |x-y|$  do elemento influenciador ao ponto em questão. A forma da

função  $w(z)$  é determinante do tipo de interação que se deseja descrever. Embora o objetivo do modelo seja descrever interações não locais, é razoável que esta função se anule para  $z$  maior do que um certo valor  $d$ . Por mera conveniência matemática, em geral, toma-se uma função não nula mas rapidamente decrescente com  $z \rightarrow \infty$ . Por exemplo, para descrevermos uma interação (auto) ativadora de curto alcance e (auto) inibidora de longo alcance, podemos utilizar funções gaussianas

$G_h(s) = \exp[-hs^2]$ , ( $h > 0$ ), e tomar uma função de influencia da forma (dita de Rodieck) [48]

$$w(z, t) = aG_A(z) - bG_I(z), \text{ onde } a > b > 0, \text{ e } A > I > 0.$$

Observemos agora que o método de Fourier nos mostra que perturbações iniciais  $v(x)$  que sejam auto funções de  $L_w$ , isto é, tais que  $L_w v = \lambda v$ , têm a sua dinâmica determinada por  $\rho(x, t) = e^{\lambda t} v(x)$ .

E, não é difícil determinar que para funções de influência simétricas, as auto-funções do operador integral  $L_w$  são exponenciais  $v(x) = w^{ikx}$ , pois,

$$\int_{-\infty}^{\infty} w(x-y)e^{iky} dy = \int_{-\infty}^{\infty} w(z)e^{ik(x+z)} dz = \left( \int_{-\infty}^{\infty} w(z)e^{ikz} dz \right) e^{ikx}, \text{ de onde}$$

tiramos que

$$\lambda_k = \int_{-\infty}^{\infty} w(z)e^{ikz} dz = \hat{w}(k) \text{ é a transformada de Fourier de } w(z). \text{ No caso}$$

gaussiano (Rodieck) esta transformada pode até ser calculada explicitamente:

$$\int_{-\infty}^{\infty} \exp[-hz^2]e^{ikz} dz = \sqrt{\frac{\pi}{h}} \exp\left\{\frac{-k^2}{4h}\right\}, \text{ ou seja,}$$

$$\lambda_k = a\sqrt{\frac{\pi}{A}} \exp\left\{\frac{-k^2}{4A}\right\} - b\sqrt{\frac{\pi}{I}} \exp\left\{\frac{-k^2}{4I}\right\}.$$

É importante observar como, neste modelo simplificado, o ponto de bifurcação determinado pelo valor máximo de  $\lambda_k$  depende de maneira crucial da relação entre a ativação próxima e a inibição distante.

Completemos agora este modelo com uma dinâmica não linear  $F(\rho)$  de caráter pontual,

$$\frac{\partial \rho(x, t)}{\partial t} = F(\rho(x, t)) + L_w \rho(x, t),$$

ou seja, a taxa de variação  $\frac{\partial \rho(x, t)}{\partial t}$  (no ponto  $x$ ) é influenciada na dinâmica  $F$  apenas pelos indivíduos situados exatamente no ponto  $x$ .

Em outras palavras (matemáticas), o operador  $F$  é pontual, enquanto que  $L_w$  é não local e o operador de difusão  $D \frac{\partial^2}{\partial x^2}$  é local, ou, infinitesimal. É importante enfatizar novamente que os operadores funcionais  $Op$  que determinam a dinâmica de um sistema distribuído,  $\frac{\partial \rho}{\partial t} = Op[\rho]$  podem ter caráter pontual, local ou, não local, dependendo das escalas de espaço e tempo admitidas para o modelo matemático em questão.

O sistema linearizado, digamos, na solução nula, terá a forma

$$\frac{\partial \eta}{\partial t} = F'(0)\eta + L_w \eta$$

e o problema de estabilidade se reduz à análise de existência de valores positivos para a expressão (chamada, relação de dispersão)

$F'(0) + \hat{w}(k) = \lambda(k, \mu)$ , onde  $\mu$  representa (vetorialmente) os parâmetros constitutivos do modelo. O problema de bifurcação tem agora um caráter mera-

mente algébrico e é determinado pelos zeros simples da função  $\theta(\mu) = k \max \lambda(k, \mu)$ . Modelos matemáticos de duas espécies com interação não local, uma ativadora e outra inibidora, podem ser escritos nas formas

$$\begin{aligned} \frac{\partial \rho_A}{\partial t} &= F_A(\rho_A) + T_1 (L_{w_{AA}} \rho_A + L_{w_{AI}} \rho_I) \\ \frac{\partial \rho_I}{\partial t} &= F_I(\rho_I) + T_2 (L_{w_{IA}} \rho_A + L_{w_{II}} \rho_I) \end{aligned}$$

onde as funções  $T_i$  têm por finalidade estabelecer uma saturação nas influências, tipicamente como

$$T(x) = \frac{x}{|x|+1}.$$

Observe que o sinal (ativação ou inibição) e o alcance (próximo, longo) da interação pode ser determinado pelos operadores lineares de influência  $L$ , em cada termo.

Considerações de escala podem também levar à inclusão de uma dinâmica pontual anterior ao ao processo de influência, ou seja,

$$\begin{aligned} \frac{\partial \rho_A}{\partial t} &= F_A(\rho_A) + T_1 (L_{w_{AA}} F_A(\rho_A) + L_{w_{AI}} F_I(\rho_I)) \\ \frac{\partial \rho_I}{\partial t} &= F_I(\rho_I) + T_2 (L_{w_{IA}} F_A(\rho_A) + L_{w_{II}} F_I(\rho_I)). \end{aligned}$$

Esta classe de modelos ocorre no tratamento da dispersão (não local) de populações como, por exemplo, no estudo da dispersão de insetos sociais, [36].

Não prosseguiremos no desenvolvimento matemático destes modelos pois já temos o suficiente para que o(a) leitor(a) possa vislumbrar os argumentos básicos. Se a manipulação algébrica torna-se previsivelmente mais escabrosa com a inclusão de novos termos, por outro lado os argumentos permanecem essencialmente os mesmos, e isto nos basta a esta altura. Maiores detalhes e exemplos podem ser encontrados na bibliografia: [37], [13].

## 4. Modelo de Swindale para o fenômeno de dominação ocular de Hubel-Wiesel

*“Biologists observe things that cannot be explained. Theorists explain things that cannot be observed”.* Aaron Katchalsky

O mapeamento das sinapses no córtex cerebral com respeito à origem dos axônios na distribuição sensorial do corpo é uma preocupação antiga dos neurofisiologistas, [21]. Partes do corpo onde há uma maior concentração de receptores sensoriais, como as pontas dos dedos, especialmente do polegar, e a fovea da retina, ocupam uma parte relativamente grande no mapeamento cortical. Além disso, o mapeamento é dinâmico e competitivo. Por exemplo, em um violinista cujo treinamento for iniciado antes dos 11 anos de idade, há uma quantidade excepcional do cortex destinada aos estímulos sensoriais dos dedos [60]. O conceito que descreve este importante fenômeno é denominado “plasticidade” e deve-se à escola canadense de neuropsicologia, especialmente a Hebb [17].

O estudo do mapeamento dos receptores óticos no córtex tem por objeto de análise um “padrão” fisiológico fundamental para a compreensão de uma das mais importantes funções do sistema neurológico e que é o tema central deste trabalho; a cognição visual.

A retina se constitui na única “parte exterior ” do cérebro e tem por finalidade transformar a luz incidente em sinais que, após um processamento, são enviados a uma parte intermediária, que a grosso modo atua como um relê (nucleo lateral geniculado), para posteriormente atingirem o córtex visual primário ( V1), e daí para outras sucessivas e mais profundas regiões do córtex. A percepção sensorial-cognitiva global da visão provêm de todo este trajeto e de seus respectivos processamentos. Cada “estação de processamento” tem uma tarefa cognitiva específica, progressivamente mais sutil e menos conhecida, conforme nos aprofundamos no córtex [21]. A retina é portanto a estação primeira do cérebro no trajeto de processamento dos estímulos luminosos e nada mais sábio do que iniciar por ela.

Nesta seção trataremos apenas de aspectos relacionados até a primeira estação processadora cortical denominada V1.

Esta arquitetura de processamento têm inúmeras características que instigam a curiosidade do fisiologista quanto às suas funções. Uma das mais notáveis foi descoberta por D.A. Hubel e T.N.Wiesel na década de 1960 e pela qual ganharam o prêmio Nobel em 1981, [21],

[http://apu.sfn.org/images/brainbriefings/visual\\_dev\\_illus\\_large.gif](http://apu.sfn.org/images/brainbriefings/visual_dev_illus_large.gif) .

A transmissão de sinais da retina ao cortex visual se dá pelo chamado nervo ótico que é na verdade um feixe de axônios provenientes da terceira camada de células da retina, os gânglios, de ambos os olhos. Chegando ao cortex, é natural indagar como se arranjarão espacialmente as células receptoras dos axônios de um e de outro olho, já que receberam sinais distintos e que, de alguma forma, deverão ser cognitivamente integrados para a percepção da imagem.

A formação de sinapses no córtex visual primário era reconhecidamente um processo competitivo entre as duas classes de axônios provenientes das duas retinas desde os trabalhos de Hebb sobre plasticidade. Por exemplo, um recém nascido que tivesse um dos olhos oclusos durante algum tempo, apresentaria uma predominância quase completa de sinapses de axônios do olho ativo em detrimento dos relacionados ao olho inativo. Mais do que isso, Hubel e Wiesel observaram que em alguns animais mamíferos, especialmente naqueles que possuem um visão frontal, (felinos e primatas que são predadores, têm; roedores que são presas, não têm), as distribuições das sinapses dos axônios no córtex visual apresentam uma arquitetura notavelmente organizada. Nesses casos há uma alternância de “colunas” ou “faixas” de neurônios do cortex que são predominantemente conectados, respectivamente, a um ou outro olho. Essas colunas formam ainda padrões espaciais que chamam a atenção pela sua semelhança com os da impressão digital, e da pele de animais como a zebra, [21].

Através de experimentos sutis e engenhosos, Hubel e Wiesel demonstraram que no cortex de um primata as sinapses provenientes do olho esquerdo, por exemplo, exerciam uma influência estimuladora na formação de outras sinapses do mesmo olho em uma região interna de raio aproximado de  $200\mu m$ , enquanto que exerciam uma influência inibidora sobre as mesmas na região intermediária entre  $200\mu m$  e  $600\mu m$ , e nenhuma influência apreciável na região externa além de  $600\mu m$ . Por outro lado, as mesmas exerciam uma influência inversa na formação de sinapses do olho direito, isto é, um efeito inibidor na região interna, estimulador na região intermediária, e

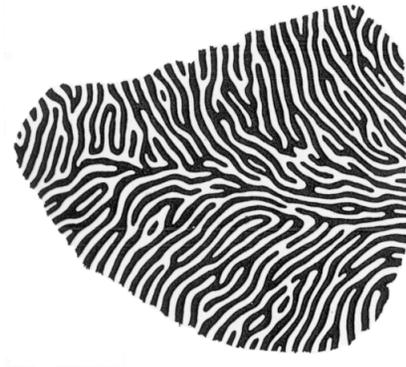


Figura 3: Padrões de faixas em zebras, impressões digitais e/ou colunas de dominação ocular [37].

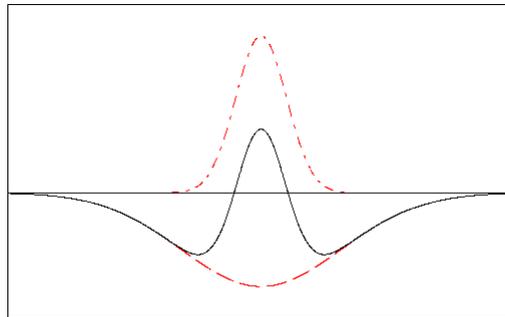
nulo na região externa. Como era de se esperar, observou-se um comportamento simetricamente semelhante das influências das sinapses do olho direito sobre elas mesmas e sobre as do olho esquerdo.

Usando a terminologia do item anterior sobre interações não locais, poderíamos representar matematicamente a influência de uma sinapse do olho esquerdo sobre a formação de sinapses semelhantes por um núcleo de Rodieck do tipo

$$w_s(z, t) = aG_A(z) - bG_I(z) , \text{ onde } a > b > 0, \text{ e } I > A > 0,$$

e sua influência sobre a formação de sinapses rivais do olho direito, com o seguinte núcleo

$$w_r(z, t) = aG_A(z) - bG_I(z) , \text{ onde } b > a > 0, \text{ e } A > I > 0. ([48]).$$



Núcleos de Influência de Rodieck

Figura 4:

Essas influências, como é de se esperar com base no simples bom senso, são cu-

mulativas com respeito à densidade das sinapses influenciadoras, mas estão sujeitas à saturação para uma densidade limiar, ou seja, o efeito Malthusiano é contido após um crescimento inicial.

Naturalmente, apenas com estas informações da fisiologia descritiva, por mais importantes e detalhadas que sejam, não se poderia explicar a formação dos padrões de dominação ocular que é uma questão de caráter essencialmente geométrico e, portanto, matemático.

Coube a Nicholas V. Swindale ([www.swindale.ecc.ubc.ca/gallery/html](http://www.swindale.ecc.ubc.ca/gallery/html)), fisiologista experimental, mas conhecedor também dos argumentos teóricos de Turing, a formulação, em 1979, de um dos primeiros modelos matemáticos para a explicação deste fenômeno de morfogênese [58], [59]. O modelo de Swindale é notável pela parcimônia de suas hipóteses e se constitui em um dos exemplos antológicos da Biomatemática contemporânea.

O modelo faz uso dos operadores  $L_{w_s}$  e  $L_{w_r}$  para representar, respectivamente, os processos de ativação e inibição de populações semelhantes e rivais, e acrescenta uma dinâmica pontual  $F(\rho)$  (por exemplo Verhulstiana,  $F(\rho) = \gamma\rho(1 - \frac{\rho}{K})$ ) para introduzir o processo de saturação da seguinte maneira;

$$\begin{aligned}\frac{\partial \rho_e}{\partial t} &= F(\rho_e)\{L_{w_s}\rho_e + L_{w_r}\rho_d\} \\ \frac{\partial \rho_d}{\partial t} &= F(\rho_d)\{L_{w_r}\rho_e + L_{w_s}\rho_e\}.\end{aligned}$$

Naturalmente,  $\rho_e$  e  $\rho_d$  representam, respectivamente, a distribuição das sinapses do olho esquerdo e do olho direito no córtex visual.

Os estados uniformes  $\rho_e = 0$ ,  $\rho_d = N$ , e  $\rho_e = N$ ,  $\rho_d = 0$  são estáveis, ou seja, não há aparecimento espontâneo de sinapses de nenhum dos dois tipos. Já o estado  $\rho_e = \frac{N}{2}$ ,  $\rho_d = \frac{N}{2}$  pode ser instável para parâmetros adequados e, neste caso, a simulação numérica da dinâmica de pequenas perturbações demonstra claramente a formação estável dos padrões de dominação ocular de Hubel-Wiesel, [58].

O sucesso deste modelo mostra-nos que embora o fenômeno de morfogênese seja extraordinariamente complexo e exasperante quando considerado sob todos os seus aspectos moleculares, celulares e biofísicos, por outro lado, ele pode ser surpreendentemente inteligível através de apenas algumas poucas hipóteses escolhidas com parcimônia e conhecimento de causa. Sem esses “pequenos milagres” o progresso da Biomatemática seria impossível, e sem uma “fé” na sua inesgotabilidade seria desanimador trabalhar nesta área, como de resto, em qualquer parte relevante da ciência.

O desenvolvimento posterior deste e de outros modelos semelhantes para a explicação dos padrões de dominação ocular, assim como de diversos outros fenômenos afins da arquitetura neurológica é revisto em [59] e [35].

## 5. Padrões de percepção em alucinações visuais: O modelo de Cowan-Ermentrout

*“..e, aplicando-lhe saliva aos olhos e impondo-lhe as mãos, perguntou-lhe: Vês alguma coisa? Êste, recobrando a vista, respondeu: Vejo os homens, porém como*

*árvores os vejo. Então novamente lhe pôs as mãos nos olhos, e êle, passando a ver claramente, ficou restabelecido; e tudo distinguia de modo perfeito*". Marcos 8: 23-26.

A visão como cognição não decorre passiva e deterministicamente dos estímulos luminosos que incidem sobre a retina, como se fôsse uma mera impressão fotográfica, mas depende de forma crucial de todo um complexo processamento neurobiológico que apesar de sua robustês, pode ser modificado estruturalmente ou cognitivamente. Nenhum cego de nascença, uma vez curada sua deficiência fisiológica, passa a "ver" o mundo imediatamente; é necessário um tempo de aprendizado cognitivo que fisiologicamente tem a sua correspondência em estruturas e processos, [17]. A modificação destas estruturas fisiológicas em pontos cruciais pode não destruir completamente a cognição mas pode distorcer-la enormemente. Nesta seção trataremos de percepções "anormais", ou alucinações, originadas de modificações estruturais causadas pela supressão de neurotransmissores.

As alucinações visuais são descritas na medicina como sintomáticas de várias patologias. As alucinações decorrentes da ingestão de drogas alucinógenas (LSD e mescalina) foram experimentalmente analisadas na década de 1960, por motivos óbvios, e têm duas características que sugerem um tratamento matemático:

1) A sua origem fisiológica deve estar no córtex visual e independe de estímulos óticos, uma vez que podem ser induzidas em indivíduos cegos por deficiência na retina ou no nervo ótico,

2) Os padrões visuais reportados pelos indivíduos submetidos a estas drogas são geométricos e sempre classificáveis em 5 tipos: i) reticulados ou colmeias, ii) teias de aranha, iii) espirais, iv) túneis (faixas concêntricas), e v) papaventos como na figura 5 abaixo.

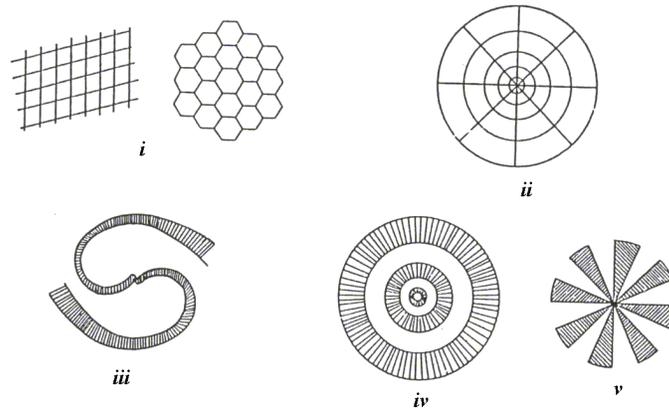


Figura 5: Padrões de percepção visual em alucinações: i) reticulados ou colmeias, ii) teias de aranha, iii) espirais, iv) túneis e faixas concêntricas, e v) papaventos.

Observemos que estes padrões, embora tenham origem intrínseca, são visuais, o que significa que são idênticos aos estímulos óticos provenientes da impressão de imagens reais semelhantes sobre a retina de indivíduos com sua visão intacta. A explicação da ocorrência destes padrões se dará pela análise das atividades do cortex visual, e portanto, dependerá de uma identificação dos mesmos neste contexto, o que somente será possível conhecendo-se o mapeamento entre a retina e o córtex visual. O mapeamento fisiológico entre as superfícies da fovea e do córtex visual foi obtido, de forma relativamente simples, por Cowan em 1977 e consistia basicamente em um logaritmo complexo,  $w = \log z$ , ( $w = (\log r, \theta) \iff z = re^{i\theta} \approx (r, \theta)$ ), onde  $z$  é a coordenada do cortex e  $w$  da fovea. Desta forma, os padrões de colméia na retina se identificam com reticulados no córtex, túneis (faixas concêntricas) com faixas verticais paralelas, papaventos com faixas horizontais e espirais com faixas inclinadas.

A teoria de Cowan-Ermentrout consiste nas seguintes hipóteses que são fundamentadas em observações experimentais:

1-A superfície cortical contém duas famílias, ou camadas, de neurônios distribuídas uniformemente, uma chamada de células excitadoras e outra de células inibidoras.

2-Cada célula pode estar em um estado variável de ativação que será medido pela sua taxa de disparo. Interessa-nos aqui medir o grau de ativação total de cada camada, ou seja a distribuição superficial da taxa de todos os disparos, onde distinguiremos os disparos provenientes das células excitadoras e os das células inibidoras. Denotaremos então por  $E(z, t)$  e  $I(z, t)$  estas respectivas distribuições na coordenada (plana)  $z$  da superfície do cortex visual (V1) no instante  $t$ .

3-Há um decaimento exponencial da atividade do tipo excitador e do tipo inibidor.

4-Há influências não locais: positiva (núcleo gaussiano) da atividade excitadora em  $E$  e  $I$ , e negativa (gaussiana invertida) da atividade inibidora em  $E$  e  $I$ . (A nomenclatura aqui é ligeiramente confusa mas não imprópria e exige um momento de meditação!).

5-As intensidades das influências causadas pelas interações são moduladas por uma função de saturação do tipo  $S(u) = \frac{u}{|u|+1}$ .

6-Uma imagem na retina se traduz por um padrão de células (excitadoras e inibidoras) ativadas no córtex.

Com base em 2),3) e 4), e a experiência dos itens anteriores podemos escrever o modelo de interação não local de Cowan-Ermentrout da seguinte maneira:

$$\frac{\partial E}{\partial t} = -E + S_E (L_{ee}E - L_{ie}I)$$

$$\frac{\partial I}{\partial t} = -I + S_I (L_{ei}E - L_{ii}I)$$

O objetivo de Cowan e Ermentrout era demonstrar que, para um ajuste adequado dos parâmetros do modelo matemático, a solução nula (nenhuma visão!) poderia se tornar instável. Com isto, a própria dinâmica do sistema faria com que perturbações, inevitáveis, se desenvolvessem atingindo estados estacionários estáveis que configuram os padrões de visão alucinatórias no cortex. A modificação dos parâmetros constitutivos do sistema, especialmente na eficiência das sinapses

ativadoras e inibidoras, (motivadas por reações químicas excepcionais dos neurotransmissores), é exatamente o papel desempenhado pelas drogas alucinógenas.

A análise de bifurcação é então feita pelos métodos tradicionais, ainda que um pouco trabalhosa, e o cálculo dos estados estacionários é obtido por simulações numéricas que comprovam a existência dos padrões acima descritos, [37], [3]. Este trabalho comprovou de maneira surpreendente a ubiquidade dos mecanismos de Turing.

O estudo da cognição, e particularmente da consciência, leva-nos imediatamente a questões relacionadas à formação de padrões de formas na dinâmica neurológica. Afinal, esta é a manifestação concreta e objetiva da “consciência”, e portanto a única representação cientificamente analisável dos padrões de percepção. O estudo desta questão, historicamente abordada apenas sob o ponto de vista filosófico e especulativo, tem sido intensamente estudada nos últimos anos sob vários aspectos objetivos [4], [32] onde mais uma vez os argumentos de Turing se mostram como um dos conceitos naturais para sua compreensão.

Apenas como referência passageira de fenômenos analisados nesta área, citamos os conhecidos efeitos denominados “rivalidade ocular ” e instabilidade de percepção visual [32], [1]. Este segundo efeito é facilmente ilustrado nas figuras abaixo onde duas percepções distintas, mas com “estabilidade” precária, são separadamente apreendidas mas facilmente (dinamicamente) intercambiáveis por meio de “pequenas perturbações”. Algumas pinturas de Salvador Dalí exploram de forma magistral a instabilidade da percepção de padrões. Consulte especialmente a pintura “Mercado de escravos com a aparição do busto invisível de Voltaire ” (Dali Museum in Cleveland) reproduzido em [1] e “Velhice, adolescência e infância ” ( St. Petersburg Museum , Florida) reproduzido em [32]. Veja também [10].

Não é difícil aceitar a sugestão de que este fenômeno cognitivo esteja relacionado a dois estados de ativação neurológica estacionários, estáveis, mas passíveis de desestabilização não linear, isto é, instável sob pequenas perturbações não infinitesimais. Este fenômeno é qualitativamente semelhante aos processos de bifurcação, muito comuns em matemática aplicada [31], [2]. Sob esta perspectiva, a caracterização destes “estados cognitivos” seria um problema não somente fisiológico mas de representação de “padrões” no sentido de Turing.

## 6. Os modelos de inibição lateral de Hartline-Ratliff-Knight-Peskin: Bandas de Mach

*“The ultimate goal of many cognitive scientists is to understand intelligence in its highest form: the ability to write poetry, to prove theorems.....Others, myself included, prefer less lofty goals, which seem more likely to be realized. Instead of poets, we wish to understand snails”.* L.A.Segel, [52]

Para o estudo de um sistema tão complexo como o cérebro humano (alguns o consideram como o mais complexo do universo!) é natural que se adote a abor-

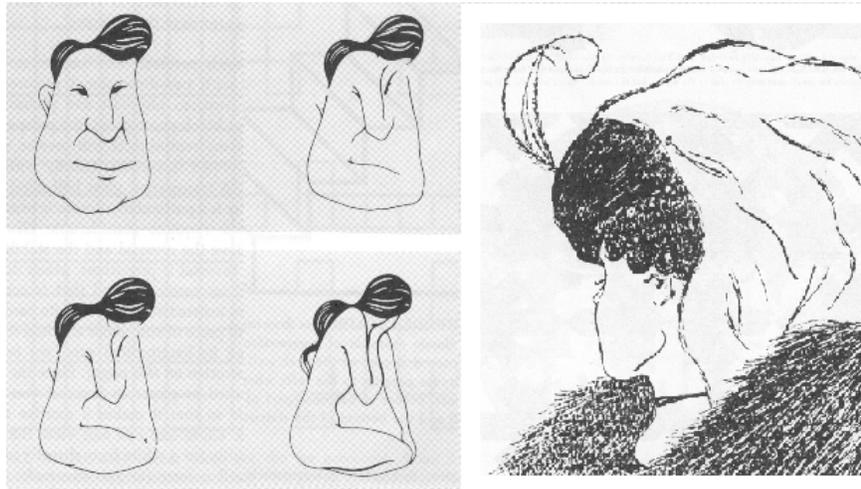


Figura 6: Exemplos clássicos de instabilidade de percepção: “*topetudo*”  $\iff$  *modelo*, de Gerald Fisher 1967, “*damas : jovem*  $\iff$  *velha* ” de W.E.Hill 1915.

dagem mais cautelosa possível ou, como dizem os do sertão mineiro: “mingau quente come-se pelas beiradas”. H.K.Hartline (Nobel de Fisiologia em 1967), F. Ratliff e R.W.Rodieck entre as décadas de 1940 e 1960 adotaram exatamente uma estratégia deste tipo. Iniciando pela parte externa do cérebro, (retina), de um de seus exemplares mais rudimentares (crustáceo *Limulus Polyphemus*), [16], e nas suas manifestações mais simples; o fenômeno de auto inibição e inibição lateral [48].(Fig 1b). Como já tivemos oportunidade de analisar, os processos de inibição a longa distância e ativação a curta distância são os ingredientes de comportamento característicos na construção de um modelo de morfogênese de Turing. Nesta seção todavia não abordaremos especificamente nenhuma formação de padrões e nos limitaremos a questões mais básicas relativas a fenômenos decorrentes do processo de inibição lateral de longa distância.

Uma das funções mais importantes, e simples, do sistema visual é detectar o movimento nas suas imediações, ou seja, codificar a variação de intensidade luminosa que é recebida pela retina. Esta capacidade é crucial especialmente para organismos que são forçados a cumprir o papel de presas em um sistema ecológico.

Os insetos em geral têm um sistema visual rudimentar comparado aos dos mamíferos, formado por uma enorme quantidade de omatídeos que funcionam como pequenos olhos, juxtapostos e independentes mas altamente apropriados para detectar movimentos adjacentes. “*Large eyed ants generally do not respond to prey insects that are standing still, but run after them as soon as they begin to move*”, [64]. As abelhas, mais sofisticadas, utilizam os seus omatídeos como sensores de

“fluxo visual”, com os quais determinam o trajeto de vôo de retorno à colméia, ou à fonte de néctar já encontrada, [63],[15].

A percepção e a cognição de imagens com toda a sutileza que esta tarefa exige é uma capacidade que depende de algumas habilidades elementares que constituem em um sistema orgânico complexo. Fiéis à abordagem da questão no sentido “de baixo para cima”, (“*bottom-up*”), descreveremos rapidamente as observações de Hartline-Ratliff sobre o comportamento da retina do *L.Polyphemos* quando submetida a experiências bem simples, a segunda podendo ser facilmente interpretada na retina do(a) próprio(a) leitor(a).

Os neurônios do sistema em estudo serão classificados quanto ao seu grau de atividade de acordo com a frequência de seus disparos, “spikes”, [21]. A “atividade” de uma região do sistema neural é medida pela taxa de disparos, ou seja, pela concentração de disparos ao longo do tempo.

Experimento 1:

A primeira experiência consiste em iluminar a retina com uma luminosidade constante e, subitamente, diminuir a intensidade desta iluminação. Durante a iluminação inicial a atividade da última camada de neurônios da retina (ganglios) é da ordem de 20-50 por segundo que depende muito pouco da intensidade. Além disso, esta atividade não é deterministicamente previsível pela intensidade da iluminação; este é um processo estocástico e a taxa tem um sentido de média, [21], [53]. A variação temporal de intensidade induz um brusco aumento desta atividade neural que todavia decai rapidamente para uma atividade maior do que a anterior, mas também constante. Sob iluminação modulada a resposta pode ser da ordem de 200 e até 1000 disparos por segundo. Esta observação mostra claramente que o olho responde muito mais forte a variações temporais do que a níveis de intensidade; o patamar de resposta constante posterior é bem abaixo do atingido pelo surto de luminosidade e apenas um pouco acima da resposta ao nível anterior. Usando a nomenclatura de sistemas lineares, podemos dizer que a retina hiper responde ao degrau de intensidade luminosa com um regime transiente curto que rapidamente volta a um outro estado permanente, estacionário e estável.

Ao invés de prosseguirmos na análise deste importante experimento, que testa capacidade adaptativa da retina à variação temporal do estímulo, passaremos ao caso seguinte que trata da resposta neurológica à variação espacial do sinal luminoso.

Experimento 2-Ilumina-se uma região à esquerda com intensidade uniforme e estacionária até uma linha, a partir da qual a intensidade decresce rápida mas com um gradiente constante até atingir a região à direita que é iluminada com uma intensidade também uniforme e estacionária mas com muito menor intensidade.

Neste segundo experimento, observa-se uma resposta de patamar constante ao longo da metade à esquerda, como era de se esperar. Aproximando-se todavia da região de transição verifica-se uma subida íngreme até um pico mergulhando em seguida com grande inclinação até um nível abaixo do patamar à direita para logo em seguida aproximar-se deste patamar.

A explicação discursiva deste fenômeno se baseia no processo de inibição lateral a longa distância. A faixa da Mach à esquerda aparece mais iluminada porque os

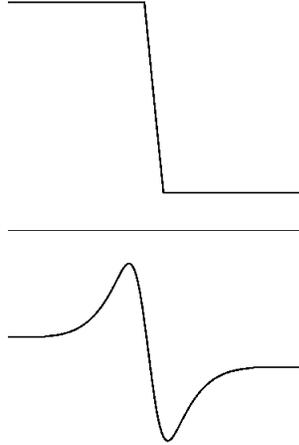


Figura 7: O gráfico acima reproduz a intensidade de iluminação da Fig.8 e o gráfico abaixo descreve a resposta fisiológica da retina ao estímulo que se traduz no efeito Mach.

neurônios estimulados por esta região não recebem uma inibição forte dos neurônios à direita, que não estão estimulados. Um efeito oposto ocorre com a faixa escura da direita por conta da forte inibição dos neurônios (estimulados) à esquerda. Este efeito pode ser melhor observado na figura apresentada no sítio

<http://www.geocities.com.SoHo/Museum/3828/greyspots.html>

A implicação sensorial imediata deste fenômeno é uma constatação experimental simples na visão humana e tem importância fundamental para o padrão cognitivo. Embora este fenômeno tenha sido utilizado pelos pintores através dos tempos [18], a sua descrição científica é relativamente recente ( sec.XIX) e é atribuída ao físico e filósofo austríaco Ernest Mach, pelo que é denominado “efeito Mach”, ou, “bandas de Mach”, [46],[47].

Mostraremos a seguir como o efeito Mach pode ser representado matematicamente como uma simples decorrência em um modelo onde o estímulo visual é submetido à inibição lateral de longa distância. Este modelo foi proposto por Hartline e Ratliff e desenvolvido por Bruce W. Knight [25].

Considerando-se as observações acima, o modelo HR-K pode ser matematicamente escrito da seguinte maneira auto explicativa;

$$R(x, y, t) = E(x, y, t) - \lambda \int_0^t \int_{-\infty}^{\infty} \int_{-\infty}^{+\infty} G(x - \alpha, y - \beta, t - \tau) R(\alpha, \beta, \tau) d\tau d\alpha d\beta$$

onde  $R(x, y, t)$  é a resposta da atividade do sistema no ponto  $(x, y)$  no instante  $t$ ,  $E(x, y, t)$  é o grau de excitação a que o ponto  $(x, y)$  é submetido no instante  $t$  pela luminosidade e  $G(x - \alpha, y - \beta, t - \tau)$  representa a influencia ( inibidora, como é no caso) exercida sobre o ponto  $(x, y)$  no instante  $t$  pela excitação do ponto  $(\alpha, \beta)$

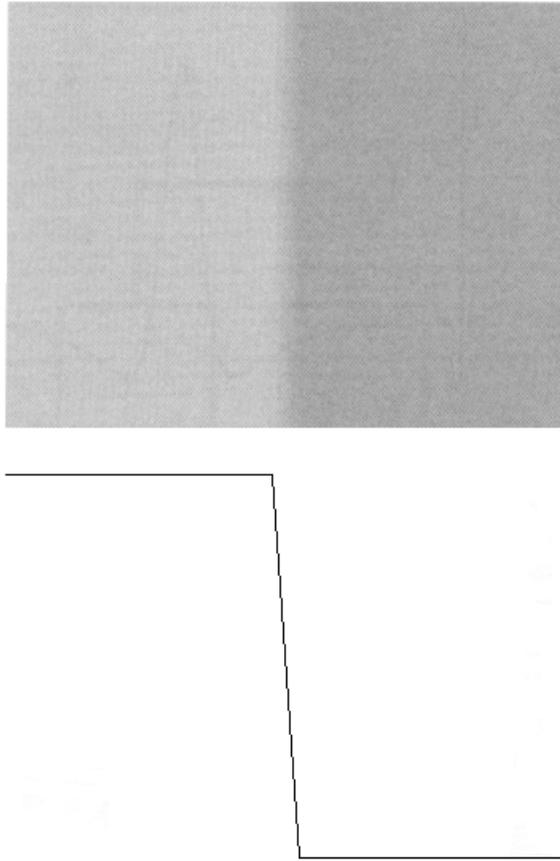


Figura 8: A intensidade de iluminação da figura acima decresce como no gráfico da figura de baixo. Entretanto, o(a) leitor(a) pode atentar para a sua percepção subjetiva de uma (inexistente) faixa mais clara logo à esquerda na região de transição e de uma faixa mais escura à direita da região de transição: as bandas de Mach.

no instante  $\tau < t$ . ( $G(x, y, t) = 0$  para  $t < 0$ ).

Consideremos agora que o efeito inibidor, representado pelo termo integral à direita, se propaga por uma sinalização química difusiva. Então, podemos reescrever este modelo na seguinte forma diferencial proposta por Peskin [43], [24],

$$\begin{aligned} \gamma \frac{\partial E}{\partial t} &= L - E \\ \frac{\partial I}{\partial t} &= \Delta I - I + \lambda R \end{aligned}$$

onde  $E$ , como antes, mede a excitação pontual do sistema,  $L$  é o estímulo luminoso também pontual,  $\gamma$  é uma constante de tempo intrínseca que mede o decaimento livre da excitação,  $I$  é a concentração do “efeito inibidor” (que também decai livremente com a constante normalizada 1), e a resposta pontual  $R$  é dada pela diferença entre a excitação e a inibição:  $R = E - I$ .

É importante frisar que este modelo, relativamente simples, tem por objetivo representar apenas os fenômenos básicos observados naqueles experimentos e não tem obviamente nenhuma pretensão de analisar o sistema visual quanto a questões mais sutis.

A grande vantagem deste tratamento é que podemos prever e analisar os fenômenos que decorrem das propriedades básicas definidas pelos experimentos. Para isto, testamos o sistema com diversos estímulos espaciais e temporais. Por exemplo, se tomamos  $L(x, y) = 1$  para  $x > 0$ , e  $L(x, y) = 0$  para  $x < 0$ , a resposta permanente do sistema (após o período transiente) será estacionário da forma:

$$\Delta I - I - \lambda I + \lambda L = 0$$

$$\text{A sua solução é dada por } I = \left\{ \begin{array}{l} \frac{\lambda}{\lambda+1} \left( 1 - \frac{1}{2} e^{-x\sqrt{\lambda+1}} \right), x > 0 \\ \frac{\lambda}{\lambda+1} \frac{1}{2} e^{x\sqrt{\lambda+1}}, x < 0 \end{array} \right\}, \text{ e}$$

$$R = L - I = \left\{ \begin{array}{l} 1 - \frac{\lambda}{\lambda+1} \left( 1 - \frac{1}{2} e^{-x\sqrt{\lambda+1}} \right), x > 0 \\ -\frac{\lambda}{\lambda+1} \frac{1}{2} e^{x\sqrt{\lambda+1}}, x < 0 \end{array} \right\}$$

cujo gráfico abaixo mostra claramente a formação das “bandas de Mach” de uma maneira semelhante àquela já apresentada na figura 7.

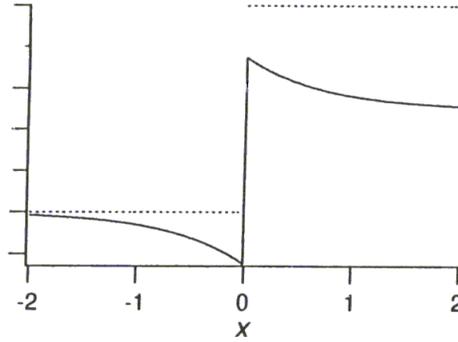


Figura 9:

## 7. Modelos com distribuição contínua de estados neurais: Um novo “instrumento” de observação

“*O sapo não pula por boniteza, ele pula por precisão*” J. Guimarães Rosa

A escolha de alguns aspectos mensuráveis e característicos dos indivíduos de uma população para que a sua distribuição (no “*espaço de aspecto*”) se torne suficiente para descrever uma dinâmica é um dos pontos cruciais na formulação de qualquer modelo matemático de dinâmica populacional [14].

Os modelos matemáticos que apresentamos até o momento simplificam enormemente o comportamento individual dos neurônios. Por exemplo, o “grau de ativação” instantâneo em um ponto do espaço nestes modelos é tomado como uma média temporal (“*integrate-and-fire model*”) que, na escala de tempo, representa de maneira muito rudimentar o comportamento temporal dos neurônios [9]. Com isto estamos desprezando quase completamente a rica dinâmica neural que foi o tema central do trabalho de Hodgkin e Huxley. É claro que em algum momento modelos matemáticos que pretendem uma explicação de fenômenos neurobiológicos cruciais não poderão prescindir de uma descrição mais detalhada desta dinâmica. Ou seja, parafraseando Guimarães Rosa, ou o seu personagem, os neurônios não exibem suas complexas oscilações apenas para nos maravilhar, ou por um mero capricho divino! Ou, como diria Albert Einstein: “*Deus é sutil, mas não malicioso!*”.

É verdade que a grande maioria dos neurônios, de fato, apresenta uma resposta nítida, isto é, extremamente rápida, e essencialmente booleana (sim/não) aos estímulos recebidos. Modelos matemáticos que descrevem fenômenos em uma escala temporal muito maior do que a escala em que ocorre a dinâmica desta resposta são plenamente justificados em considerá-las booleanas. Além disso, pelo lado pragmático da questão, a riqueza da dinâmica de Hodgkin-Huxley [44] não nos deixa opções matematicamente tratáveis no caso de uma população muito grande de neurônios. A propósito, é oportuno lembrar que o tamanho e a complexidade do sistema nervoso real não se mede apenas pelo número de neurônios que o constitui ( $\sim 10^{12}$ ) mas também pela extensa e intensa rede de comunicação ( $\sim 10^{15}$ ) que é uma característica singular e essencial destes sistemas. Um neurônio do cortex pode estar conectado a centenas ou milhares de outros neurônios. Para tornar o problema ainda mais difícil, as escalas de tempo diferem da ordem de  $10^4$ , por exemplo, entre o tempo de percurso de um impulso através de um axônio ( $\mu\text{seg}$ ) ao tempo de processamento de sinais visuais ( $\text{seg}$ ). Também, a capacidade de percepção dos sistemas sensoriais varrem uma extensa gama de escalas como, por exemplo, desde a excitação de uma célula bastão por um único fóton, até à perscrutação de detalhes e cores sob intensa luminosidade.

Entretanto, em alguma circunstância, a descrição de fenômenos neurobiológicos específicos deverá exigir uma representação mais detalhada dos estados das células que compoem o sistema em questão, ainda que tenhamos de lidar com modelos matemáticos mais difíceis de serem analisados.

O estudo destes casos têm sido feito, em geral, por meio de dispendiosas simulações de Monte Carlo e raramente pela análise de um modelo matemático. Ul-

timamente, retomando a estratégia do modelo populacional, uma série de recentes trabalhos da escola de fisiologia matemática da Mount Sinai School of Medicine (N.York) e da Rockefeller University liderada por Bruce W. Knight e Lawrence Sirovich [26],[27],[42], [54] e [41], desenvolveu uma nova abordagem que contempla uma descrição mais detalhada do estado do sistema neural. Este modelo faz uso de “aspectos ” suficientes para a representação analítica de diversos fenômenos ao mesmo tempo que possibilita o seu cálculo em tempo muito menor (aproximadamente da ordem de  $10^{-2}$  ) do que o utilizado nas simulações de Monte Carlo, [40].

O “aspecto”, ou o estado individual, de um neurônio neste modelo, ainda muito simplificado, é descrito pela sua posição no ciclo limite de sua dinâmica que pode ser representada pela variável cíclica  $\theta \in S^1$  (círculo unitário), já que os estados  $\theta = 0$  e  $\theta = 2\pi$  coincidem devido à periodicidade do processo. O estado do sistema, isto é, da população, pode então ser representado por uma função de distribuição  $P(\theta, t)$  que não descreve o aspecto espacial mas apenas a distribuição quanto à excitação. Com isto, o modelo matemático é formulado a partir da aplicação usual de princípios de conservação,

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial \theta} (-D \frac{\partial P}{\partial \theta} + fP)$$

uma vez estabelecidos, neste contexto, as expressões adequadas para os fluxos de transporte e difusão, como é usual em tais modelos, [14] e [27].

É uma característica notável desta abordagem que, embora a dinâmica do neurônio solitário de Hodgkin-Huxley seja completamente não linear, o modelo matemático do sistema neural é representado por uma equação diferencial parcial linear tornando-a assim susceptível ao vetusto e sempre eficiente método de Fourier nas suas diversas formas. Além disso, métodos de projeção para a redução da dimensão do espaço de fase do sistema podem ser utilizados com vantagem, especialmente no cálculo numérico das soluções [55].

A representação do sistema neural distribuído segundo diversos aspectos, não apenas o espacial, proporciona a possibilidade de observação (visual!) de padrões estacionários e dinâmicos que de outra forma se manteriam incógnitos. A estratégia clássica da pesquisa em neurociência, que foi sempre fundamentada na identificação de formas fisiológicas e sua interdependência com as funções do sistema, desenvolveu-se extraordinariamente ao final do século XIX e começo do século XX, como resultado de observações mais acuradas e penetrantes proporcionada pelos instrumentos óticos e técnicas experimentais progressivamente mais modernas. Agora, com a descrição ampliada do sistema neural, será possível detectar padrões de formas não aparentes por uma inspeção à “observação nua ” , isto é, despida dos instrumentos matemáticos. A busca de padrões identificáveis e reproduzíveis do sistema, que é em essência a estratégia científica em qualquer campo de investigação, adquire assim uma importância singular e curiosa no estudo de processos cognitivos.

‘Padrão’ é naturalmente um conceito subjetivo que depende da experiência do observador. Por outro lado, a sua recorrência e estabilidade faz com que ele se torne universal, e objetivo [10]. Com isto, voltamos ao início deste trabalho: usamos o instrumento matemático para a compreensão de alguns aspectos da visão enquanto que, por outro lado, utilizamos a habilidade visual para detectar padrões na sua

representação matemática!

O conteúdo e a essência do sistema neurológico naquilo que ele tem de mais importante podem ser representados enfim como as formas que emergem das diversas dinâmicas em que ele é descrito. Quanto mais os programas de Hodgkin e Huxley e de Turing se entremearrem, tanto mais conheceremos sobre o funcionamento deste sistema.

Se após várias páginas não pudemos convencer o(a) leitor(a) sobre a importância do estudo dos modelos de dinâmica populacional com respeito à emergência de padrões e formas em neurobiologia, só nos resta, como último recurso, apelar a autoridades superiores:

*“Wherever you look in the cortex of the brain, you will find organization....It is far from random....The picture cries out for explanation: to relate structure to function ”* [25], e

*“I have sometimes heard it said that the nervous system consists of huge numbers of random connections. Although its orderliness is indeed not always obvious, I nevertheless suspect that those who speak of random networks in the nervous system are not constrained by any previous exposure to neuroanatomy. Even a glance at a book such as Cajal’s ‘Histologie du Système Nerveux’ (fig.1) should be enough to convince anyone that the enormous complexity of the nervous system is almost always accompanied by a compelling degree of orderliness ”* [21].

**Abstract.** Neural mathematical models are known for their non local interactions. The intertwined development of the classical works by H.K.Hartline and J.D.Cowan with A.M.Turing’s morphogenetic theory is presented with a special focus on vision.

## Referências

- [1] F. Attneave, Multistability in Perception, em [18], pp. 90-99.
- [2] R.C. Bassanezi e W.C. Ferreira Jr., “Equações Diferenciais e Aplicações”, Harbra, São Paulo, 1988.
- [3] P.C. Bressloff e J.D. Cowan, The visual cortex as a crystal, *Physica D*, **173** (2002), 226-258.
- [4] F. Crick, “The Astonishing Hypothesis: The Scientific Search for the Soul”, Scribner’s, New York, 1994.
- [5] P. Dayan e L.F. Abbott, “Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems”, MIT Press, Cambridge, 2001.
- [6] A. Einstein, “Investigations on the Theory of the Brownian motion”, Dover, New York, 1956.
- [7] G.B. Ermentrout, An adaptive model for synchrony in the firefly *Pteroptyx malaccae*, *J.Math.Biol.*, **29** (1991), 571-585.

- [8] G.B. Ermentrout, Neural networks as spatio-temporal pattern-forming system, *Reports on Phys. Progr.*, **61** (1998), 353-430.
- [9] G.B. Ermentrout e J.D. Cowan, A mathematical theory of visual hallucination patterns, *Biol. Cybernetics*, **34** (1979), 137-150.
- [10] M.C. Escher, Optical Illusions and Tessellations, online, <http://www.geocities.com.SoHo/Museum/3828/optical.html>
- [11] C.P. Fall, E.S. Marland, J.M. Wagner e J.J. Tyson, eds., "Computational Cell Biology", Springer, New York, 2002.
- [12] W.C. Ferreira Jr., "Dinâmica de Populações: Modelos Matemáticos, Simulações e Aplicações", Notas de Minicurso, XXI CNMAC, Caxambu, MG, 1998.
- [13] W.C. Ferreira Jr., "Notas de Biomatemática", IMECC- UNICAMP, Campinas, 2000.
- [14] W.C. Ferreira Jr., Comportamento Individual e Dispersão Social, em Anais do X Congresso Latino Americano de Biomatemática, pp.1-17, ALAB 2001.
- [15] K. von Frisch, "The dance language and orientation of bees", Harvard Univ. Press, Cambridge, 1993.
- [16] H.K. Hartline e F. Ratliff, Inhibitory interaction in the retina of limulus (horse-shoe crab), em "Handbook of Sensory Physiology" (H. Autrum, ed.), Vol. VII/2, pp. 381-447, Springer, Berlin, 1972.
- [17] D.O. Hebb, "The Organization of Behavior: A Neuropsychological Theory" , J. Wiley, New York, 1949.
- [18] R. Held, ed., "Image, Object and Illusion, Readings from Scientific American", W.H. Freeman, S. Francisco, 1974.
- [19] A.L. Hodgkin e A.F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve, *Journal of Physiology*, **117** (1952), 500-544. Reproduzido em. *Bull. of Math. Biology* **52**, (1990), 25-71.
- [20] J.J. Hopfield, On theorists and data in computational neuroscience, *Nature Neuroscience Suppl.*, **3** (2000), 1204.
- [21] D.H. Hubel, "Eye, Brain and Vision", W.H. Freeman, S. Francisco, 1988.
- [22] E.R. Kandel e L.R. Squire, Neuroscience: Breaking down scientific barriers to the study of brain and mind, *Science*, **290** (2000), 1113-1120.
- [23] P. Kareiva, Population dynamics in spatially complex environments: Theory and data, *Phil. Trans. R. Soc. Lond. B* , **330** (1990), 175-190.
- [24] J.P. Kener e J. Sneyd, "Mathematical Physiology", Springer, New York, 1998.

- [25] B.W. Knight, The horseshoe crab eye: A little nervous systems that is solvable, em “Lectures in Life Sciences”, pp.113-144, American Math. Society, Providence, 1973.
- [26] B.W. Knight, D. Manin e L. Sirovich, Dynamical models of interacting neuron populations, em “Symposium on Robotics and Cybernetics: Computational Engineering in Systems Applications” (E.C. Gerf, ed.), pp. 1-15, Cite Scientifique, Lille, France, 1996.
- [27] B.W. Knight, Dynamics of encoding in neuron populations: Some general mathematical features, *Neural Computation*, **12** (2000), 473-518.
- [28] A.N. Kolmogoroff, I. Petrovsky e N. Piscounoff, Étude de l'équation de la diffusion avec croissance de la quantité de matière et son application a un problème biologique, *Bull. Univ. Moscow*, **1** (1937),1-25.
- [29] A.A.P. Leão, Spreading depression of activity in the cerebral cortex, *J. Neurophysiol.* **7**(1944), 359-390.
- [30] S.A. Levin e L.A. Segel, Pattern generation in aspect and space, *SIAM Rev.*, **27** (1985), 45-67.
- [31] C.C. Lin e S.A. Segel, “Mathematics Applied to Deterministic Problems in the Natural Sciences”, SIAM, Philadelphia, 1990.
- [32] N.K. Logothetis, Vision: A Window on Consciousness, *Scientific American*, **281** (1999), 44-51.
- [33] H. Martins-Ferreira, G. de Oliveira Castro, C.J. Struchiner e P.S. Rodrigues, Circling spreading depression in isolated chick retina, *J.Neurophysiol.*, **37** (1974), 773-84.
- [34] H. Meinhardt, “Models of Biological Pattern Formation”, Academic Press, New York, 1982.
- [35] K.D. Miller, Ocular dominance and orientation columns, em “The Handbook of Brain Theory and Neural Networks”, (M.A.Arbib, ed.), pp. 660-665, Cambridge, 1998.
- [36] D.C. Mistro, “Modelos para a dispersão de abelhas africanizadas: Um zoom matemático”, Tese de Doutorado, IMECC-UNICAMP, Campinas, São Paulo,1998.
- [37] J.D. Murray, “Mathematical Biology”, Springer, New York, 1993.
- [38] Z. Nédá, E. Ravasz, Y. Brecht, T. Vicsek e A.-L. Barabási, The sound of many hands capping: Tumultuous applause can transform itself into waves of synchronized clapping, *Nature*, **403** (2000), 849-850.

- [39] G. Nicolis e I. Prigogine, “Self-organization in non-equilibrium systems”, J. Wiley, New York, 1977.
- [40] D.Q. Nykamp e D. Tranchina, A population density approach that facilitates large-scale modeling of neural networks: Analysis and application to orientation tuning, *J. of Comput. Neurosc.*, **8** (2000), 19-50.
- [41] D.Q. Nykamp e D. Tranchina, A population density approach that facilitates large-scale modeling of neural networks: Extension to slow inhibitory synapses, *Neural Comput.*, **13** (2001), 511-546.
- [42] A. Omurtag, B.W. Knight e L. Sirovich, On the simulation of large populations of neurons, *J. of Comput. Neurosc.*, **8** (2000), 51-63.
- [43] C.S. Peskin, “Partial Differential Equations in Biology”, Courant Institute Lecture Notes, New York, 1976.
- [44] C.S. Peskin, “Mathematical Aspects of Neurophysiology”, Courant Institute Lecture Notes, New York, 1991.
- [45] D.J. Pinto e G.B. Ermentrout, Spatially structured activity in synaptically coupled neuronal networks: II. Lateral inhibition and standing pulse waves, *SIAM J. Appl. Math.*, **62** (2001), 226-243.
- [46] F. Ratliff, “Mach Bands: Quantitative Studies on Neural Networks in the Retina”, Holden Day, S.Francisco, 1965.
- [47] F. Ratliff, Contour and Contrast, em [18], pp. 16-27.
- [48] R.W. Rodieck, Quantitative analysis of cat retinal ganglion cell responses to visual stimuli, *Vision Research*, **5** (1965), 583-601.
- [49] L.A. Segel e S.A. Levin, Application of nonlinear stability theory to the study of the effects of diffusion on predator prey interactions, em “Topics in Statistical Mechanics and Biophysics” (R.A.Picciorelli ed.), pp. 123-152, Amer.Inst. of Physics, 1976.
- [50] L.A. Segel, Mathematical models for cellular behavior, em “Studies in Mathematical Biology”, (S.A. Levin ed.), pp. 156-190, Mathematical Assoc. of America, 1978.
- [51] L.A. Segel e A.S. Perelson, Computations in shape space: A new approach to immune network theory, em “Theoretical Immunology”, Part Two, (A.S. Perelson,ed.), pp. 321-341, Addison-Wesley, Reading, 1988.
- [52] L.A. Segel, Toward artificial competence, *Springer Lectures in Biomath.*, **100** (1995), 117-121.
- [53] G.M. Shepherd, “Neurobiology”, Oxford Univ. Press, Oxford, 1994.

- [54] L. Sirovich, A. Omurtag e B.W. Knight, Dynamics of neuronal populations: The equilibrium solution, *SIAM J. Appl. Math.*, **60** (2000), 2009-2028.
- [55] L. Sirovich, B.W. Knight e A. Omurtag, Dynamics of neural populations: Low dimensional approximations, *preprint* online: <http://camelot.mssm.edu/people/sirovich> , 2002.
- [56] J. Sneyd, ed., “An Introduction to Mathematical Modeling in Physiology, Cell Biology and Immunology”, Amer. Math. Soc., Providence, 2002.
- [57] C. Stevens, Models are common; good theories are scarce, *Nature Neuroscience Suppl.*, **3** (2000), 1177.
- [58] N.V. Swindale, A model for the formation of ocular dominance stripes, *Proc. R. Soc. Lond. B*, **208** (1980), 243-264.
- [59] N.V. Swindale, The development of topography in the visual cortex: a review of models, *Network*, **7** (1996), 161-247.
- [60] N.V. Swindale, Cortical cartography: what’s in a map?, *Current Biology*, **11** (2001), R730-R733.
- [61] D. Tranchina-Mathematical aspects of vision, em [56], pp. 33-55.
- [62] A. Turing, The chemical basis of morphogenesis, *Phil.Trans. of the Royal Soc. Part B*, **237** (1953), pp.37-72. Reproduzido em *Bull.of Math. Biology*, **52** (1990), pp.153-197.
- [63] R. Wehner-Spatial vision in arthropods, em “Handbook of Sensory Physiology”, vol. VII/6C,(H. Autrum, ed.), pp. 288–661, Springer Verlag, Berlin, 1981.
- [64] E. Wilson, “Ants”, Harvard Univ. Press, Cambridge, 1994.
- [65] A.T. Winfree, “When Time Breaks Down: The Three-Dimensional Dynamics of Electromagnetic Waves and Cardiac Arrhythmias”, Princeton Univ. Press, Princeton, 1987.

## Inferência de Estrutura Funcional em Neurociências

K. SAMESHIMA<sup>1</sup>, Disciplina de Informática Médica, Departamento de Patologia e Laboratório de Neurociência Cognitiva e Processamento de Informação Neural (LIM 45), Faculdade de Medicina, Universidade de São Paulo, Av. Dr. Arnaldo, 455, 01246-903 São Paulo, SP, Brasil

L.A. BACCALÁ<sup>2</sup>, Departamento de Engenharia de Telecomunicações e Controle, Escola Politécnica, Universidade de São Paulo, Av. Prof. Luciano Gualberto, Trav. 3, #158, 05508-900 São Paulo, SP, Brasil.

**Resumo.** O estudo de interações funcionais entre estruturas cerebrais é feito preponderantemente por métodos de correlação. Neste trabalho descrevemos uma abordagem alternativa de análise de séries temporais multivariadas, denominada coerência parcial direcionada (CPD), que torna obsoletos os métodos de correlação ao permitir uma elucidação mais clara da estrutura funcional pela análise simultânea de registros de atividade neural de múltiplas estruturas cerebrais. Discutimos também como a CPD representa o conceito de causalidade de Granger no domínio da frequência. Por fim modelos teóricos e dados experimentais são usados para ilustrar a CPD e compará-la a outras técnicas.

### 1. Introdução

Nas últimas duas décadas tem-se presenciado um progresso vertiginoso na área de neurociências possibilitado pela introdução de novos instrumentos e técnicas. Contudo, apesar dessas modernas técnicas de biologia molecular, neuroanatomia e mapeamento cerebral por imagens funcionais, a monitoração de registros de séries temporais de sinais elétricos cerebrais gerados por despolarização neuronal continua sendo a técnica de escolha para avaliar diretamente os aspectos funcionais, normais ou patológicos, de circuitos neurais em cérebros de mamíferos.

Dentro deste contexto, os métodos de correlação são os preferidos da grande maioria de pesquisadores envolvidos no estudo da função de circuitos neurais pelo registro da atividade cerebral. Estes métodos de correlação são extensamente usados na análise de interações funcionais em eletroencefalografia (EEG), magnetoencefalografia, registros de potenciais de campo local, e mais recentemente nos registros de atividade de neurônios unitários e multi-unitários obtidos em muitas estruturas cerebrais (com registro de dezenas a centenas de neurônios). Esta última abordagem

---

<sup>1</sup>ksameshi@usp.br; apoio financeiro FAPESP, CNPq e PRONEX.

<sup>2</sup>baccala@lcs.poli.usp.br; apoio financeiro FAPESP e PRONEX.

tem recebido uma crescente atenção da comunidade científica mundial pela perspectiva de se poder fazer uma ponte entre as atividades de neurônios unitários isolados e o entendimento da codificação e do processamento de informação por populações neuronais [11,22].

No esforço de se proporcionar instrumentos que possibilitem a análise dessa classe de dados multivariados, surgiram nos últimos anos várias técnicas analíticas: algumas empregando os conceitos da teoria da informação pela medida de informação mútua [8,23,31] ou interdependência entre pares de sinais [1,28]; enquanto outros utilizaram uma extensão das idéias provindas da análise espectral e coerência [9,10,17,24]. Apesar de se dispor de várias técnicas avançadas, uma grande parte dos neurocientistas e os neurofisiologistas clínicos (i.e. os profissionais da eletroencefalografia) relutam em utilizá-las, valendo-se ainda dos métodos de correlação cruzada para análise de atividades de pares de estruturas neurais para inferência de aspectos funcionais do sistema nervoso central.

Todos os métodos que se baseiam no método da correlação restringem-se à consideração da interação simultânea de apenas duas estruturas neurais. Nossa busca por métodos de análise que considerassem mais que duas estruturas simultaneamente surgiu, sobretudo, pela percepção de que para uma inferência estrutural efetiva era necessária a consideração simultânea de muitas estruturas representativas envolvidas na função neural.

Aqui introduz-se o novo conceito de coerência parcial direcionada, que é uma das mais recentes novidades nos esforços de desenvolver métodos e técnicas de determinação de como as estruturas neurais estão funcionalmente interconectadas em pequenos intervalos de tempo [4,26]. Pode-se dizer que sua origem mais remota se deve ao trabalho de Saito e Harashima [25] que introduziram a noção de coerência direcionada entre par de estruturas. Seu método permite a fatoração da função de coerência ordinária de um par de estruturas em duas "coerências direcionadas": uma representando o componente direto e outra o componente de retroalimentação das interações entre duas estruturas neurais. No trabalho original, Saito e Harashima [25] apontam para uma possível lógica por trás de seu método. Este conceito central é hoje conhecido como causalidade de Granger [18].

No passo desses conceitos e idéias, investigamos a generalização de coerência direcionada [2,4,5,26,27] para até dezenas de estruturas processadas simultaneamente e se compara aos testes estatísticos da causalidade de Granger para várias séries temporais simultâneas [20].

O presente trabalho é composto por duas partes: a) inicialmente introduzem-se os fundamentos teóricos do novo método e b) compara-se seu desempenho ao de outros métodos de inferência estrutural usando modelos teóricos e dados experimentais.

## 2. Determinação da estrutura funcional por modelagem autorregressiva multivariada

Na eletroencefalografia quantitativa moderna, a descrição das interações entre diferentes áreas tem se baseado na estimação da matriz de densidade espectral de potência  $\mathbf{S}(f)$ , dada por

$$\mathbf{S}(f) = \begin{bmatrix} S_1(f) & S_{12}(f) & \cdots & S_{1N}(f) \\ S_{21}(f) & S_2(f) & \cdots & S_{2N}(f) \\ \vdots & \vdots & \ddots & \vdots \\ S_{N1}(f) & S_{N2}(f) & \cdots & S_N(f) \end{bmatrix}$$

de registros multicanais de atividade elétrica cerebral  $\mathbf{x}_i(n)$ ,  $1 \leq i \leq N$ .

Em particular, para  $N = 2$ , o método mais utilizado pela fácil compreensão de seus conceitos tem sido o estudo das funções de coerência ordinária dada por

$$C_{ij}(f) = \frac{|S_{ij}(f)|^2}{S_i(f)S_j(f)}$$

que expressa o grau de ativação simultânea de duas áreas  $i$  e  $j$  em função da frequência [4-6].

Saito e Harashima [25] utilizaram argumentos da teoria da informação para introduzir uma extensão da noção de coerência que denominaram coerência direcionada (CD) pela sua habilidade de desvendar o sentido do fluxo de informação ou interação. Na prática sua proposta original estava restrita a  $N = 2$  estruturas, pela necessidade de uma condição *ad hoc* da estrutura do ruído aditivo observado (vide adiante).

A essência dessas idéias pode ser apreendida na forma generalizada para  $N > 2$  com a ajuda do resultado geral de fatoração [14,15] da matriz de densidade espectral de potência dada por

$$\mathbf{S}(f) = \mathbf{H}(f)\mathbf{\Sigma}\mathbf{H}(f)^H \quad (2.1)$$

em que o sobrescrito  $H$  corresponde à transposta hermitiana,

$$\begin{aligned} \mathbf{H}(f) &= \begin{bmatrix} H_{11}(f) & H_{12}(f) & \cdots & H_{1N}(f) \\ H_{21}(f) & H_{22}(f) & \cdots & H_{2N}(f) \\ \vdots & \vdots & \ddots & \vdots \\ H_{N1}(f) & H_{N2}(f) & \cdots & H_{NN}(f) \end{bmatrix} \\ &= [\mathbf{h}_1(f) \quad \mathbf{h}_2(f) \quad \cdots \quad \mathbf{h}_N(f)] \end{aligned}$$

é a matriz de filtros adequados no domínio da frequência e

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22}^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \cdots & \cdots & \sigma_{NN}^2 \end{bmatrix}$$

é matriz de covariância.

Isso leva à seguinte definição generalizada da coerência direcionada de  $j$  para  $i$  expressa por [4,5]

$$\gamma_{ij}(f) = \frac{\sigma_{jj} H_{ij}(f)}{\sqrt{S_i(f)}}, \quad (2.2)$$

em que

$$S_i(f) = \sum_{j=1}^N \sigma_{jj}^2 |H_{ij}(f)|^2,$$

que se reduz à definição de Saito e Harashima [25] quando  $N = 2$ .

Uma importante restrição adicional do método Saito-Harashima é a explícita pressuposição de que  $\Sigma$  é diagonal ( $\sigma_{ij} = 0$  para  $i \neq j$ ). Neste caso,  $|\gamma_{ij}(f)|^2$  pode ser interpretada como a fração da potência na frequência  $f$  originada de  $x_j(n)$  sobre a potência total de  $x_i(n)$ .

Para o caso de  $\Sigma$  diagonal, pode-se mostrar que a coerência ordinária pode ser expressa como

$$C_{ij}(f) = \gamma_i^H(f) \gamma_j(f) = \sum_{k=1}^N \gamma_{ik}^*(f) \gamma_{kj}(f),$$

em que  $\gamma_i(f) = [\gamma_{i1}(f) \dots \gamma_{iN}(f)]^T$ , com \* indicando o complexo conjugado. Isso justifica o nome coerência direcionada uma vez que  $\gamma_{ij}(f)$  pode ser interpretada como uma medida da influência de  $x_j(n)$  sobre  $x_i(n)$  em contraposição a  $\gamma_{ji}(f)$  que descreve a contribuição no sentido oposto.

Para contornar a restrição da necessidade de diagonalidade de  $\Sigma$ , Kaminski e Bli-nowska [19] introduziram uma quantidade alternativa para descrever as interações entre elementos neurais abandonando completamente a formulação com uso de  $\Sigma$ . Eles denominaram este estimador *função de transferência direcionada* (FTD). Este estimador coincide em magnitude com a Eq. 2.2 quando  $\sigma_{jj}$  são normalizados para valor unitário, ou seja

$$FTD_{ij}(f) = \frac{H_{ij}(f)}{\sqrt{\sum_{j=1}^N |H_{ij}(f)|^2}} = \frac{H_{ij}(f)}{\sqrt{\mathbf{h}_i^H(f) \mathbf{h}_i(f)}}.$$

Desta discussão, o problema que se coloca é como calcular  $\mathbf{H}(f)$  e  $\Sigma$  conhecida matriz  $\mathbf{S}(f)$ . Um modo eficiente e prático de fazê-lo é por meio de modelos autorregressivos multivariados. Tais modelos podem ser exemplificados no caso de medida simultânea de  $N = 3$  estruturas:

$$\begin{bmatrix} x_1(n) \\ x_2(n) \\ x_3(n) \end{bmatrix} = \sum_{r=1}^p \begin{bmatrix} a_{11}(r) & a_{12}(r) & a_{13}(r) \\ a_{21}(r) & a_{22}(r) & a_{23}(r) \\ a_{31}(r) & a_{32}(r) & a_{33}(r) \end{bmatrix} \begin{bmatrix} x_1(n-r) \\ x_2(n-r) \\ x_3(n-r) \end{bmatrix} + \begin{bmatrix} w_1(n) \\ w_2(n) \\ w_3(n) \end{bmatrix}.$$

Neste modelo,  $x_1(n)$  depende de seus próprios valores de passado  $x_1(n-r)$ , representado pelos coeficientes  $a_{11}(r)$  e, da mesma forma, também de outras séries, e.g.  $x_2(n-r)$ , nos valores do passado da série  $x_1(n)$  dadas pelos coeficientes  $a_{12}(r)$ .

Este exemplo de modelo autorregressivo multivariado (ARM) pode ser generalizado para  $N$  estruturas da seguinte maneira:

$$\begin{bmatrix} x_1(n) \\ \vdots \\ x_N(n) \end{bmatrix} = \sum_{r=1}^p \mathbf{A}_r \begin{bmatrix} x_1(n-r) \\ \vdots \\ x_N(n-r) \end{bmatrix} + \begin{bmatrix} w_1(n) \\ \vdots \\ w_N(n) \end{bmatrix}. \quad (2.3)$$

De forma geral, o ajuste do modelo conduz às matrizes  $\mathbf{A}_r$  e a  $\Sigma$  na forma de matriz de covariância de  $w_i(n)$ . Pode-se mostrar que

$$\mathbf{H}(f) = \bar{\mathbf{A}}^{-1}(f) = (\mathbf{I} - \mathbf{A}(f))^{-1}$$

em que

$$\mathbf{A}(f) = \sum_{r=1}^p \mathbf{A}_r z^{-r} \Big|_{z=e^{-j2\pi f}}$$

com

$$\mathbf{A}_r = \begin{bmatrix} a_{11}(r) & a_{12}(r) & \dots & \dots & a_{1N}(r) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & a_{ij}(r) & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{N1}(r) & \dots & \dots & \dots & a_{NN}(r) \end{bmatrix},$$

na qual os coeficientes  $a_{ij}(r)$  representam o efeito de interação de  $x_j(n-r)$  sobre  $x_i(n)$ . Neste texto,  $j$  na função exponencial é o valor  $\sqrt{-1}$ .

**Observação 1** *O uso de modelos autorregressivos multivariados (ARM) como os da Eq. 2.3 são comuns a diversos métodos inclusive para a coerência parcial direcionada.*

Agora é fácil mostrar que a simples modificação da Eq. 2.2 para

$$\bar{\gamma}_{ij}(f) = \frac{H_{ij}(f)}{\sqrt{S_i(f)}}$$

resolve o problema de  $\Sigma$  não ser diagonal de modo que

$$C_{ij}(f) = \bar{\gamma}_i^H(f) \Sigma \bar{\gamma}_j(f), \quad (2.4)$$

notando-se que

$$S_i(f) = \mathbf{h}_i^H \Sigma \mathbf{h}_i,$$

com  $\mathbf{h}_i \triangleq [H_{i1}(f), \dots, H_{iN}(f)]^T$ .

## 2.1. Causalidade de Granger e a nova abordagem

Saito e Harashima [25] referem-se a uma possível idéia em que se baseia seu método, atualmente conhecido como causalidade de Granger [18]. Pode-se definir a causalidade de Granger por meio de uma frase: "uma série temporal  $x_j(n)$  Granger-cause outra série  $x_i(n)$ , se o conhecimento do passado de  $x_j(n)$  aumenta significativamente a predição de  $x_i(n)$ ; esta previsibilidade não é recíproca, i.e.  $x_j(n)$  pode Granger-cause  $x_i(n)$  sem que  $x_i(n)$  Granger-cause necessariamente  $x_j(n)$ ".

O conceito de causalidade de Granger é uma idéia advinda da econometria e existem diversos procedimentos de testes estatísticos bem definidos [20]. Sua noção fundamental é que considera **exclusivamente as amostras do passado** para se aumentar a previsibilidade de uma série temporal. Quando se considera a amostra do presente, i.e.  $t = 0$ , de  $x(n)$  contribuindo para o significativo aumento da previsibilidade de  $y(n)$ , na econometria é conhecido como causalidade de Granger instantânea.

A conceituação de causalidade proposta por Granger [18] parece ser bastante interessante também nas neurociências e em outras disciplinas. A avaliação da causalidade de Granger fornece uma medida do grau de interação entre estruturas neurais com o pressuposto de que se a flutuação de ativação de uma série temporal ocorre consistentemente em decorrência de flutuação em outra série por mecanismos de geração ou de influência que estão ligadas a alguma conectividade física como em uma conexão neuroanatômica funcionalmente ativa.

Alguns testes estatísticos da causalidade de Granger (TCG) são baseados no exame direto dos coeficientes  $a_{ij}(r)$  do modelo autorregressivo multivariado, uma vez que este descreve o efeito de predição linear da  $r$ -ésima amostra do passado  $x_j(n-r)$  da séries  $x_j(n)$  na predição de  $x_i(n)$ . Um dos procedimentos para se testar a ausência da causalidade de Granger é mostrar estatisticamente que  $a_{ij}(r) = 0$  para todos os valores de  $r$ , o que rejeitaria a hipótese de que  $x_j(n)$  Granger-cause  $x_i(n)$ . O teste convencional deste tipo aplicável a sinais de atividade neural multivariados foi revisado em [2] e outros testes de Granger estão descritos em [20].

Uma das principais fraquezas da maioria dos testes de causalidade de Granger é que, ao contrário da coerência direcionada, não se pode vislumbrar o cenário de interação entre estruturas neurais no domínio da frequência, exceto quando  $N = 2$  [16].

Adicionalmente, ao verificar a relação entre pares de séries temporais, quando se processa simultaneamente  $N > 2$  estruturas, obtém-se freqüentemente resultados conflitantes entre os achados do CD/FTD e o teste de causalidade de Granger, i.e.  $x_i(n)$  pode apresentar alto valor de CD para  $x_j(n)$  e no entanto se rejeitar a causalidade de Granger de  $x_i(n)$  para  $x_j(n)$  usando-se TCG convencional. Nossa experiência mostra que isso ocorre com freqüência quando são considerados para análise simultânea mais que duas séries temporais, porém para  $N = 2$  (*vide Exemplo a seguir*) CD/FTD e o teste de causalidade de Granger estão sempre em concordância.

Para proporcionar uma figura no domínio da frequência da descrição de causalidade de Granger, propomos um novo conceito, a *coerência parcial direcionada*.

Este é baseado em outro conceito popular entre pares de sinais, a assim chamada *função de coerência parcial*,  $|\kappa_{ij}(f)|^2$ , que descreve a interação entre  $x_i(n)$  e  $x_j(n)$  quando as influências devidas a outras  $(N-2)$  séries temporais são descontadas. O resultado chave é a seguinte fatoração [4]:

$$\kappa_{ij}(f) = \frac{\bar{\mathbf{a}}_i^H(f) \boldsymbol{\Sigma}^{-1} \bar{\mathbf{a}}_j(f)}{\sqrt{[\bar{\mathbf{a}}_i^H(f) \boldsymbol{\Sigma}^{-1} \bar{\mathbf{a}}_i(f)] [\bar{\mathbf{a}}_j^H(f) \boldsymbol{\Sigma}^{-1} \bar{\mathbf{a}}_j(f)]}},$$

em que  $\boldsymbol{\Sigma}$  é a matriz de covariância do erro de predição associado ao modelo na Eq. 2.3 e  $\bar{\mathbf{a}}_k(f)$  é a  $k$ -ésima coluna da matriz

$$\bar{\mathbf{A}}(f) = \mathbf{I} - \mathbf{A}(f) = \begin{bmatrix} \bar{\mathbf{a}}_1(f) & \bar{\mathbf{a}}_2(f) & \dots & \bar{\mathbf{a}}_N(f) \end{bmatrix}.$$

Por analogia com a Eq. 2.4, em que a coerência direcionada reflete os termos para o qual se pode decompor a coerência ordinária, propusemos a seguinte definição generalizada [4].

## 2.2. Definição de coerência parcial direcionada (CPD)

A 'coerência parcial direcionada' de  $j$  para  $i$  é dada por

$$\pi_{ij}(f) = \frac{\bar{a}_{ij}(f)}{\sqrt{\bar{\mathbf{a}}_j^H(f) \boldsymbol{\Sigma}^{-1} \bar{\mathbf{a}}_j(f)}} \quad (2.5)$$

em que  $\bar{a}_{ij}(f)$  é o  $i, j$ -ésimo elemento de  $\bar{\mathbf{A}}(f)$ .

Disto tem-se imediatamente que a coerência parcial entre  $i$  e  $j$  é dada por

$$\kappa_{ij}(f) = \pi_i^H(f) \boldsymbol{\Sigma}^{-1} \pi_j(f),$$

para  $\pi_i(f) = [\pi_{1i}(f), \dots, \pi_{Ni}(f)]^T$ , de onde provém a motivação para o nome de  $\pi_{ij}(f)$ .

Na ocorrência de diagonalidade de  $\boldsymbol{\Sigma}$  com variâncias iguais, a Eq. 2.5 simplifica-se para

$$\bar{\pi}_{ij}(f) = \frac{\bar{a}_{ij}(f)}{\sqrt{\bar{\mathbf{a}}_j^H(f) \bar{\mathbf{a}}_j(f)}}, \quad (2.6)$$

em que

$$\bar{a}_{ij}(f) = \begin{cases} 1 - \sum_{r=1}^p a_{ij}(r) e^{-j2\pi fr}, & \text{se } i = j \\ - \sum_{r=1}^p a_{ij}(r) e^{-j2\pi fr}, & \text{se } i \neq j \end{cases}$$

e

$$\bar{\mathbf{a}}_j(f) = \begin{bmatrix} \bar{a}_{1j}(f) \\ \vdots \\ \bar{a}_{Nj}(f) \end{bmatrix}.$$

A expressão de CPD depende principalmente dos coeficientes  $a_{ij}(r)$  que descreve a relação entre as séries temporais  $x_i(n)$  e os passados de  $x_j(n)$  comparado com os efeitos de  $x_j(n)$  sobre outras séries temporais. Em sua forma simplificada (Eq. 2.6), ela depende exclusivamente de  $a_{ij}(r)$  e, como no caso de FTD, a coerência parcial direcionada proposta,  $\bar{\pi}_{ij}(f)$ , também não depende da formulação baseada em  $\Sigma$ . Da mesma forma como ocorre com CD/FTD [5], as seguintes propriedades de normalização são válidas:

$$0 \leq |\bar{\pi}_{ij}(f)|^2 \leq 1$$

e

$$\sum_{i=1}^N |\bar{\pi}_{ij}(f)|^2 = 1, \quad (2.7)$$

para todos  $1 \leq j \leq N$ .

A igualdade na Eq. 2.7 é um caso especial, pois se se define o vetor de norma quadrática do vetor  $\pi_j(f)$  como  $|\pi_j(f)|^2 = \pi_j^H(f)\Sigma^{-1}\pi_j(f)$ , o resultado geral é  $|\pi_j(f)|^2 = 1$ .

Estas são análogas às propriedades da função de transferência direcionada [19] e coerência direcionada [4,5].

### 3. Análise de modelos teóricos

Para ilustrar o potencial do novo método de análise, serão apresentados diversos exemplos de modelos lineares em que se compara o desempenho da CPD com outros métodos de análise.

Um aspecto importante dos métodos previamente publicados, relacionado à coerência direcionada e suas variantes, é que se pressupunha a matriz  $\Sigma$  diagonal ou ignoravam completamente a sua estrutura, de modo que neste estágio de desenvolvimento está-se utilizando somente a forma simplificada do CPD (Eq. 2.6) nos exemplos que se seguem.

#### 3.1. Equivalência de CPD e FTD para $N = 2$ estruturas

Como demonstrado em [4], quando se considera apenas duas séries temporais,  $N = 2$ , na análise simultânea pode se estudar a interação utilizando a Eq. 2.2 ou Eq. 2.5. Isto é correto uma vez que  $H_{ij}(f)$  é proporcional a  $\bar{\mathbf{A}}_{ij}(f)$ , para  $i \neq j$ , em que a nulidade de um implica também na nulidade de outro termo, indicando ausência

de influência devido a  $x_j(n)$  sobre  $x_i(n)$ . De fato, é fácil mostrar que as resultantes  $|FTD_{ij}(f)|$  e  $|\bar{\pi}_{ij}(f)|$  são idênticas para  $i \neq j$ , uma vez que

$$|FTD_{ij}(f)| = \frac{|H_{ij}(f)|}{\sqrt{|H_{i1}(f)|^2 + |H_{i2}(f)|^2}}$$

e

$$\begin{aligned} \bar{\mathbf{A}}(f) &= \begin{bmatrix} \bar{A}_{11}(f) & \bar{A}_{12}(f) \\ \bar{A}_{21}(f) & \bar{A}_{22}(f) \end{bmatrix} = \begin{bmatrix} H_{11}(f) & H_{12}(f) \\ H_{21}(f) & H_{22}(f) \end{bmatrix} \\ &= \frac{1}{\Delta} \begin{bmatrix} H_{22}(f) & -H_{12}(f) \\ -H_{21}(f) & H_{11}(f) \end{bmatrix}, \end{aligned}$$

com  $\Delta = H_{11}(f)H_{22}(f) - H_{12}(f)H_{21}(f)$ , dando

$$|\bar{\pi}_{ij}(f)| = \frac{|\bar{A}_{ij}(f)|}{\sqrt{|\bar{A}_{1j}(f)|^2 + |\bar{A}_{2j}(f)|^2}}$$

a partir do que, substituindo, obtém-se

$$|\bar{\pi}_{12}(f)| = \frac{|-H_{12}(f)|}{\sqrt{|-H_{12}(f)|^2 + |H_{22}(f)|^2}} = |\gamma_{12}(f)|$$

e

$$|\bar{\pi}_{21}(f)| = \frac{|-H_{21}(f)|}{\sqrt{|H_{22}(f)|^2 + |-H_{21}(f)|^2}} = |\gamma_{21}(f)|.$$

Note-se que a igualdade não ocorre para os índices idênticos, i.e. tem-se  $|\gamma_{11}(f)| = |\bar{\pi}_{22}(f)|$  e  $|\gamma_{22}(f)| = |\bar{\pi}_{11}(f)|$  [4].

### 3.2. Três séries temporais com interações mútuas

Veja o seguinte conjunto de equações que geram três séries temporais:

$$\begin{cases} x_1(n) &= 0, 5x_1(n-1) + 0, 3x_2(n-1) + 0, 4x_3(n-1) + w_1(n) \\ x_2(n) &= -0, 5x_1(n-1) + 0, 3x_2(n-1) + x_3(n-1) + w_2(n) \\ x_3(n) &= -0, 3x_2(n-1) - 0, 2x_3(n-1) + w_3(n) \end{cases} \quad (3.1)$$

em que  $w_i(n)$  são ruídos brancos independentes de média zero e variâncias idênticas. Pode-se observar que a única conexão ausente é a influência explícita do passado de  $x_1(n)$  sobre  $x_3(n)$ . O diagrama estrutural correspondente está na Fig. 1a.

Os correspondentes traçados de  $|FTD_{ij}(f)|$  e  $|\bar{\pi}_{ij}(f)|$  estão mostrados em diagramação matricial, respectivamente, nas Figs. 1b e 1c e a explícita ausência da conectividade ou influência somente é demonstrada pela coerência parcial direcionada, em que se observa claramente que  $|\pi_{31}(f)| = 0$  enquanto  $|FTD_{31}(f)|$  não é nulo (vide Eq. 3.1).

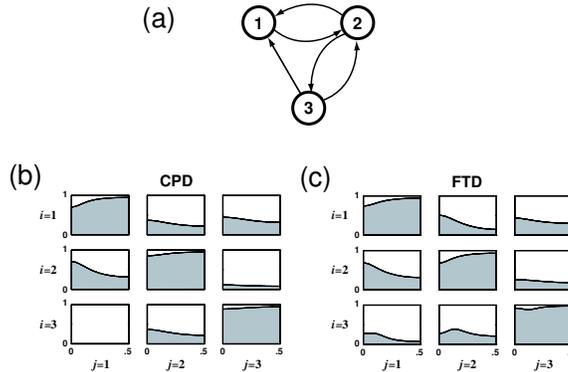


Figura 1: (a) Diagrama de conectividade correspondente à Eq. 3.1. (b) Função de coerência parcial direcionada,  $|\bar{\pi}_{ij}(f)|$ , em diagramação matricial, e (c)  $|FTD_{ij}(f)|$  que descreve o modelo. Note que  $\bar{\pi}_{31}(f) = 0$  indicando inequivocamente a independência da variável  $x_3(n)$  de  $x_1(n)$ . Por outro lado, não se pode obter essa mesma informação a partir de  $FTD_{31}(f)$ .

### 3.3. Variantes de estruturas com cinco variáveis

Introduzindo maior complexidade, três variantes de uma estrutura com cinco variáveis serão analisados a seguir. Considere inicialmente um sistema com a dinâmica expressa por

$$\begin{cases} x_1(n) &= 0,95\sqrt{2}x_1(n-1) - 0,9025x_1(n-2) + w_1(n) \\ x_2(n) &= -0,5x_1(n-1) + w_2(n) \\ x_3(n) &= 0,4x_2(n-2) + w_3(n) \\ x_4(n) &= -0,5x_3(n-1) + 0,25\sqrt{2}x_4(n-1) + 0,25\sqrt{2}x_5(n-1) + w_4(n) \\ x_5(n) &= -0,25\sqrt{2}x_4(n-1) + 0,25\sqrt{2}x_5(n-1) + w_5(n) \end{cases} \quad (3.2)$$

que corresponde ao modelo diagramático da Fig. 2a. Neste modelo, o sinal partindo do oscilador  $x_1(n)$  propaga-se seqüencialmente para  $x_2(n)$ ,  $x_3(n)$ ,  $x_4(n)$  e  $x_5(n)$ . Novamente, as variáveis  $x_4(n)$  e  $x_5(n)$  formam um oscilador independente. Utilizando o modelo ARM específico descrito pela Eq. 3.2, obtém-se  $|\bar{\pi}_{ij}(f)| = 0$  (Fig. 2b) para aquelas conexões diretas inexistentes na Fig. 2a, i.e. não há dependência explícita de  $x_i(n)$  com o passado de outras séries. Por outro lado, na Fig. 2c,  $|FTD_{ij}(f)| \neq 0$  para todo  $i \geq j$ , bem como para  $|FTD_{45}(f)|$  levando a um mapeamento estrutural pelo uso de  $FTD_{ij}(f)$  não tão claro como obtido por meio da informação fornecida pela análise de  $\bar{\pi}_{ij}(f)$ . Atente que, toda vez que  $|\bar{\pi}_{ij}(f)| = 0$ , é um indicativo claro de uma ausência de conectividade direta da estrutura  $j$  para  $i$ . Observe também que o padrão  $|FTD_{ij}(f)| \neq 0$  ocupa todas as diagonais inferiores da matriz.

A partir do modelo anterior, Fig. 2a, se a malha de realimentação de  $x_5(n)$  a

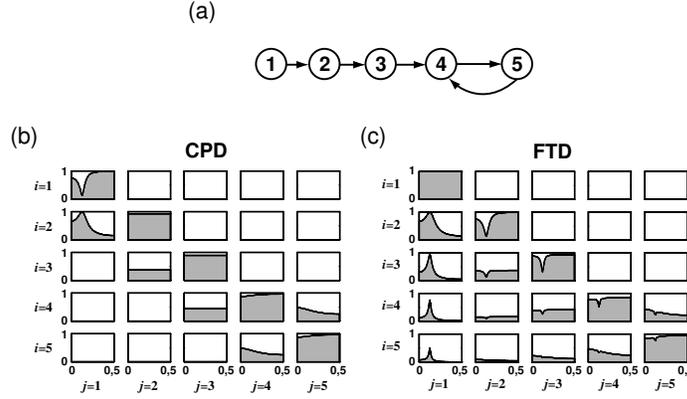


Figura 2: (a) Diagrama correspondente à Eq. 3.2. Em (b) e (c) têm-se  $|\bar{\pi}_{ij}(f)|$  e  $|FTD_{ij}(f)|$ , respectivamente.

$x_1(n)$  fosse fechada modificando-se

$$x_1(n) = 0,95\sqrt{2}x_1(n-1) - 0,9025x_1(n-2) + 0,5x_5(n-2) + w_1(n)$$

que resulta no diagrama de interação dada na Fig. 3a. Como se pode comprovar nessa figura, o gráfico de  $|\bar{\pi}_{ij}(f)|$ , Fig. 3b, continua refletindo fielmente a estrutura funcional correspondente, enquanto, especialmente neste caso particular de malha de realimentação fechada,  $FTD_{ij}(f)$  é não nula em todas as interações possíveis, bidirecionalmente (Fig. 3c).

Como variante final considere uma mudança estrutural imposta à Fig. 2a, em que  $x_1(n)$  se conecta a  $x_4(n)$  por duas vias distintas, por  $x_2(n)$  e  $x_3(n)$ ,

$$\begin{cases} x_1(n) = 0,95\sqrt{2}x_1(n-1) - 0,9025x_1(n-2) + w_1(n) \\ x_2(n) = -0,5x_1(n-2) + w_2(n) \\ x_3(n) = 0,1x_1(n-3) + 0,4x_2(n-2) + w_3(n) \\ x_4(n) = -0,5x_1(n-2) + 0,25\sqrt{2}x_4(n-1) + 0,25\sqrt{2}x_5(n-1) + w_4(n) \\ x_5(n) = -0,25\sqrt{2}x_4(n-1) + 0,25\sqrt{2}x_5(n-1) + w_5(n) \end{cases} \quad (3.3)$$

correspondente ao grafo da Fig. 4a.

Observe que  $FTD_{ij}(f)$ , Fig. 4c, é virtualmente idêntico ao da Fig. 2c, porém a diferença estrutural pode ser facilmente detectado pela análise de  $\bar{\pi}_{ij}(f)$ (Fig. 4b). Mais uma vez  $FTD_{ij}(f)$  falha em mostrar as diferenças estruturais (Fig. 4c comparado a Fig. 2c), que  $\bar{\pi}_{ij}(f)$  revela claramente (Fig. 4b versus Fig. 2b)

É importante notar que em todos esses exemplos  $FTD_{ij}(f)$  é não nula toda vez que existe um caminho que vai da estrutura  $j$  para a estrutura  $i$  no diagrama ou grafo estrutural que descreve o modelo. Portanto, enquanto  $FTD$  revela a existência de um caminho possível para ir de uma estrutura a outra (indicando a

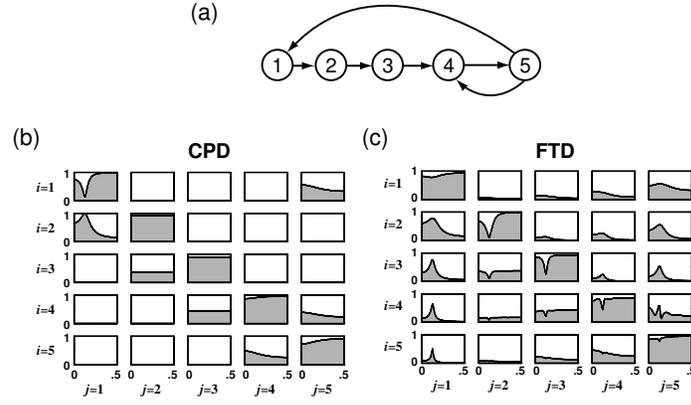


Figura 3: (a) Neste exemplo a malha de fluxo é fechada pela conexão direta de  $x_5(n)$  para  $x_1(n)$ . Em (b) e (c) têm-se novamente  $|\bar{\pi}_{ij}(f)|$  e  $|FTD_{ij}(f)|$ , respectivamente.

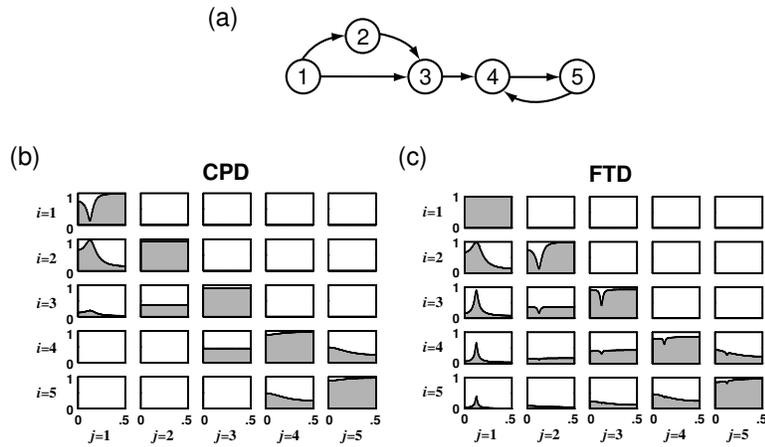


Figura 4: (a) Grafo direcionado do modelo expresso pela Eq. 3.3. Em (b) e (c) têm-se  $|\bar{\pi}_{ij}(f)|$  e  $|FTD_{ij}(f)|$ , respectivamente.

alcançabilidade de uma estrutura por outra) direta ou indiretamente, CPD revela a existência de conexão direta entre pares de estruturas num contexto em que, pelo uso do modelo ARM, as interações entre as várias estruturas consideradas na análise são destrinçadas simultaneamente.

Nos gráficos de CPD mostrados até aqui, os traçados do elemento diagonal da coerência parcial direcionada  $|\bar{\pi}_{ii}(f)|$  estão relacionados com o inverso da densidade espectral de potência pela consideração exclusiva de  $x_i(n)$  consigo mesmo – i.e. quanto de seu próprio passado contribui para o seu estado atual, enquanto o traçado do elemento diagonal da  $|FTD_{ii}(f)|$  é a densidade de potência não originária de nenhuma outra série.

## 4. Análises de potenciais de campo local

Os registros de potenciais de campo local obtidos em preparações crônica de ratos via microelétrodos bipolares são sinais de atividade elétrica neural filtrados na faixa de aproximadamente 0 a 100 Hz. Os potenciais de campo local são sinais mais fidedignos da atividade localizada de grupamento neuronal, porque os microelétrodos são de pequeno diâmetro e posicionados cirurgicamente na estrutura cerebral de interessante. Em um experimento, quando se deseja monitorar objetivamente o estado do rato, um ou mais elétrodos são usados para registros de atividade eletromiográfica, sendo um deles freqüentemente posicionado na musculatura da cabeça, na região posterior do pescoço, para detectar o estado de atividade e tônus muscular.

Inicialmente analisa-se um trecho de registro de atividade elétrica cerebral obtido durante o sono lento, também denominado sono sincronizado, de um rato. É digno de nota que caracterização das diversas fases do sono em ratos foram feitas pela primeira vez por Timo-Iaria e seus colaboradores em 1970 [30].

### 4.1. Análise de registro de sono sincronizado em rato

Fig. 5 contém um trecho de registro de 30 s da fase de sono lento de um rato em que se observa um episódio de fuso no intervalos de 13 a 16 s (segmento **B**) na monitorização de áreas corticais do hemisfério esquerdo, A10, A3 e A17, regiões do hipocampo, CA1 e CA3, e do giro denteado, GD. Os sinais foram amostrados a 256 Hz com 10 bits de resolução usando microelétrodos bipolares com 1 mm de separação. O traçado eletrográfico inferior (OLHOS) mostra a atividade muscular relacionada ao movimento ocular. Neste exemplo, o segmento **A** de 2 s de duração é comparativamente analisado em relação ao segmento **B**.

Inicialmente, estas seis séries temporais foram ajustadas com modelos ARM, a partir do que se estimou CPDs e coerências ordinárias. Uma fotografia de evolução temporal de CPDs e coerência ordinária entre os sinais de A10 e CA1 estimadas considerando conjuntamente todos os seis sinais de neuroatividade da Fig. 5, em representação tempo-freqüência, em escala de cinza, pela colagem de estimativas de segmentos de sinais adjacentes de 2 s de duração (com sobreposição de 50%)

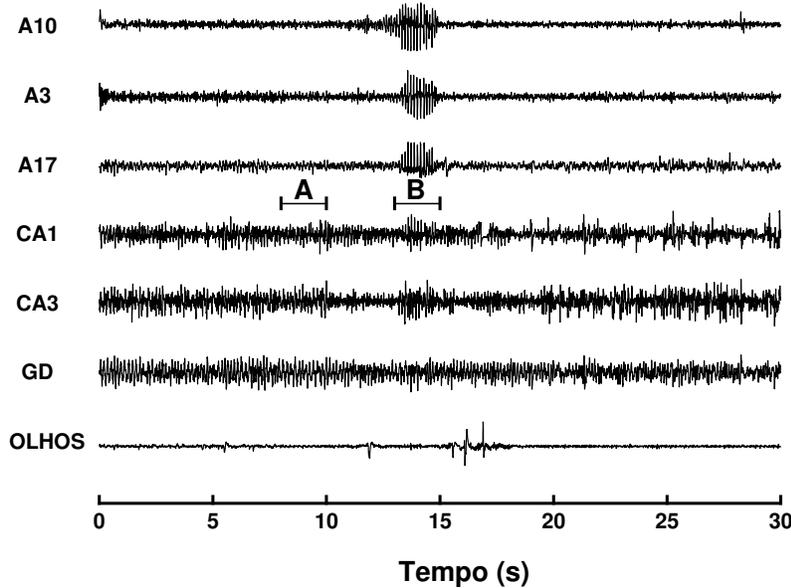


Figura 5: Registro de potencial de campo local de 30 segundos do sono de ondas lentas de rato.

estão na Fig. 6. Nesta figura, os eixos-y representam a frequência na faixa de 0 a 48 Hz. Observa-se na Fig. 6a que  $|\bar{\pi}_{A10 \leftarrow CA1}(f)| \cong 0$  para maioria dos segmentos, i.e. não há fluxo de informação de CA1 para A10, exceto durante o fuso, segmento **B**, quando  $|\bar{\pi}_{A10 \leftarrow CA1}(f)| > 0$ , para  $0 < f < 10Hz$ , enquanto na direção oposta, Fig. 6b,  $|\bar{\pi}_{CA \leftarrow A10}(f)|$  apresenta flutuação contínua ao longo do registro, porém durante o episódio de fuso,  $|\bar{\pi}_{CA \leftarrow A10}(f)| \cong 0$  caracterizando a mudança súbita de direcionalidade de interação durante o episódio do fuso que se inicia ao redor de 12 s e dura aproximadamente 3 s. Note que a interação entre essas áreas é inexistente em nenhuma das direções 8 s antes do fuso, que, por sua vez é precedido imediatamente por um aumento de CPD de A10 para CA1 o que, interessantemente, foi acompanhado de mudança de direção durante o episódio propriamente dito. Observe também o aumento da coerência ordinária, Fig. 6c, 8 s antes e depois do fuso.

Schnider e cols. fizeram uso de CD, para o caso de  $N = 2$ , ao qual escolheram o valor de nível de significância estatística como 0,1 baseado em simulações computacionais exaustivas [29]. Aqui utilizar-se-á, por ora, devido ao fato de a CPD ser também efetivamente uma decomposição de múltiplas séries temporais em pares de interações com exclusão dos efeitos de todas as outras séries, também  $|\bar{\pi}_{ij}(f)|$  significativa acima desse valor. Investigações estão em curso para se estabelecer o nível mais adequado de significância baseado em características estatísticas específicas dos estimadores de CPD utilizados.

Nas Figs. 7a e 7b têm-se os gráficos de CPDs e FTDs obtidos a partir do seg-

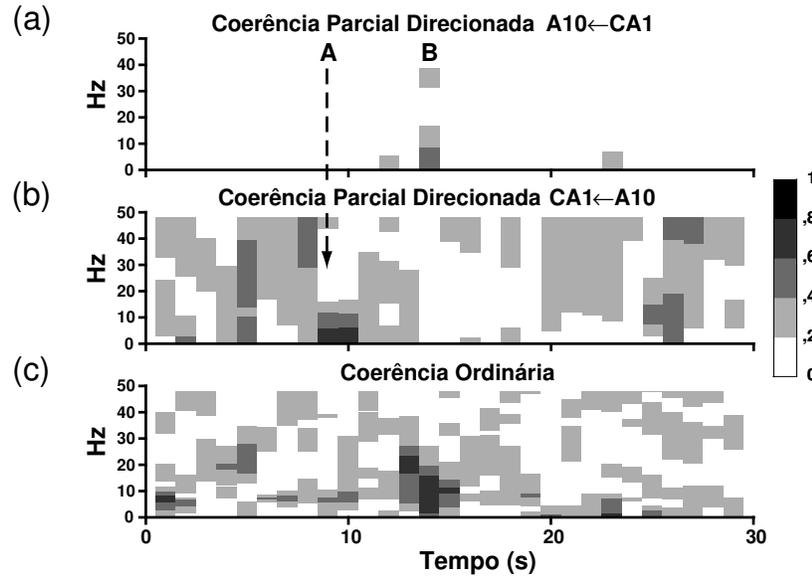


Figura 6: (a) e (b) Coerências parciais direcionadas e (c) coerência ordinária entre as estruturas A10 e CA1.

mento **A** de 2 s de duração, correspondente ao período pré-fuso da Fig. 5, a partir do que se gerou os correspondentes grafos estruturais direcionados mostrados nas Figs. 7c e 7d. A especificação da espessura das setas codificam a força de interação de CPD e FTD:  $0 \leq$  sem seta  $< 0,2 \leq$  pontilhada  $< 0,4 \leq$  linha com espessura com passo de 0,2. Neste período pré-fuso, observa-se ausência de influência do hipocampo sobre áreas corticais estudadas (A3, A10 e A17), que também é corroborada pela FTD e seu corresponde grafo estrutural. Este resultado indica que, pela análise destas seis estruturas, o fluxo de informação no período pré-fuso ocorre exclusivamente das áreas corticais para o hipocampo.

Por sua vez a mesma análise realizada no segmento **B** de ocorrência de fuso da Fig. 5, resultados mostrados na Fig. 8, nota-se significativa retroalimentação do hipocampo para o córtex, principalmente da região CA1 para a área cortical A10 correspondente à área motora do rato, Fig. 8c. A forte conectividade funcional de A17 para o giro denteado presente no período pré-fuso, Fig. 7c, não mais é observada no período de fuso, Fig. 8c.

Em ambas as análises de conectividade funcional, as fotografias fornecidas pela FTD e seus grafos direcionados são bem mais complexos, observando-se um grafo estrutural quase que completamente conectado.

#### 4.2. Análise de registro de rato em atividade exploratória

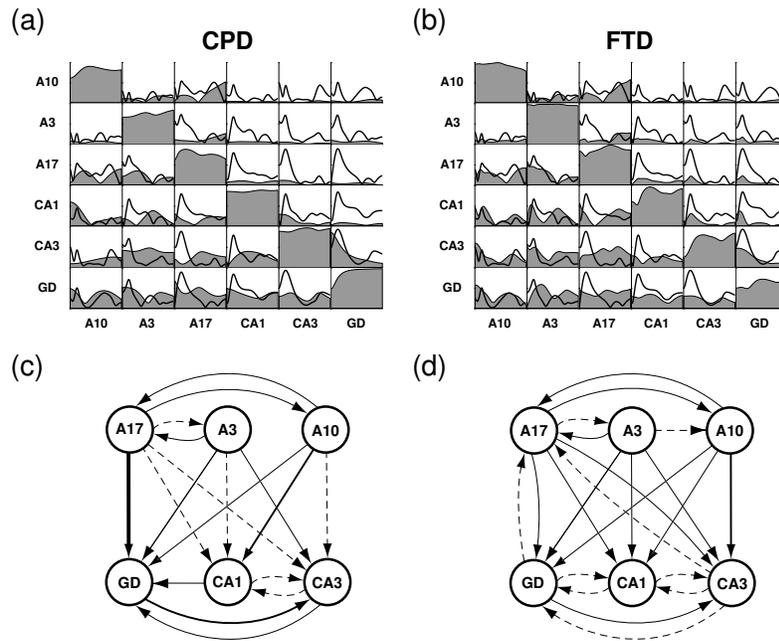


Figura 7: (a) CPDs e (b) FTDs obtidos do segmento **A** da Fig. 5, correspondente ao período pré-fuso, com que se gerou os correspondentes grafos estruturais direcionados mostrados em (c) e (d).

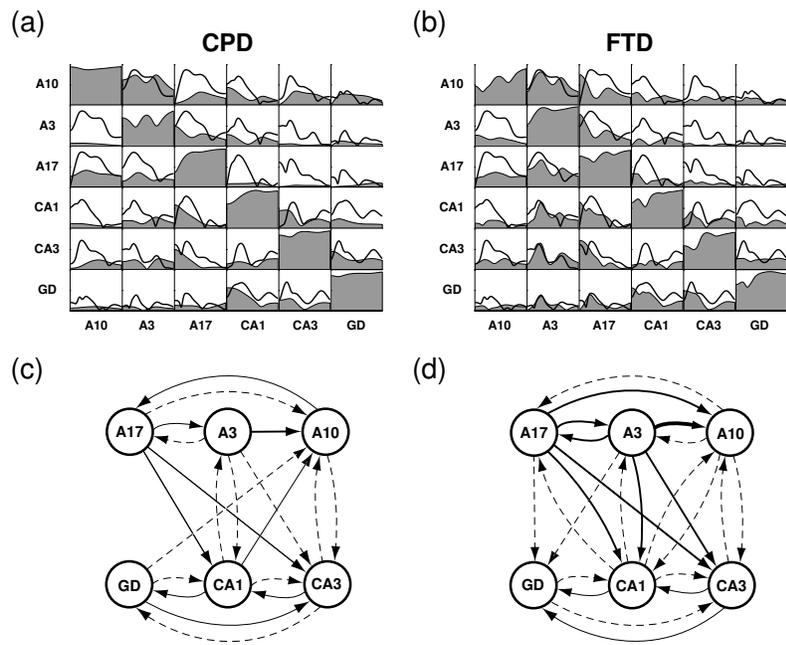


Figura 8: (a) CPDs e (b) FTDs do segmento correspondente ao período de ocorrência do fuso (segmento **B** da Fig. 5) e seus respectivos grafos estruturais direcionados (c) e (d).

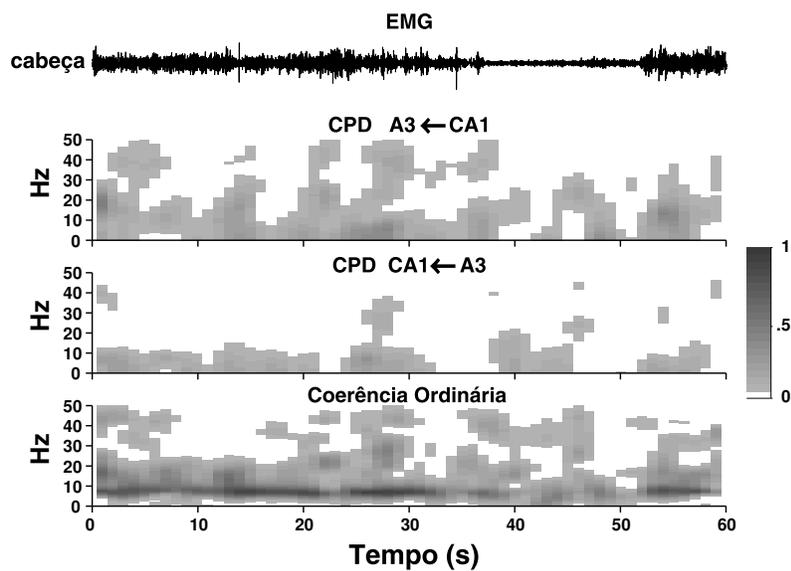


Figura 9: Exemplo de rato em comportamento exploratório em que a evolução temporal da CPD mostra a relação de interação entre a área cortical A3 e CA1 no hipocampo de rato com comportamento estagiado pelo eletromiograma (EMG) dos músculos do pescoço. A escala em cinza foi usada para representar as magnitudes do CPD e da coerência ordinária.

Este último exemplo ilustra a análise de CPD de potenciais de campo local obtidos em rato em comportamento exploratório. Os registros de atividade elétrica cerebral analisados foram obtidos no campo CA1 do hipocampo, áreas corticais somestésica (do tato) (A3) e motora (A10), e núcleo dorsal da rafe (DR), onde são observados ritmos oscilatórios na faixa do teta (ritmo teta) durante o sono dessincronizado e no estado de alerta atenta. A Fig. 9 ilustra a evolução temporal de CPD entre duas estruturas, CA1 e A3, em que se registrou simultaneamente o eletromiograma do músculo da cabeça que é um excelente marcador do estado de atividade, notando-se dois períodos distintos de comportamento. Nos primeiros 30 s deste registro o rato esteve engajado numa atividade exploratória na gaiola com iluminação, então gradualmente diminuiu o seu nível de atividade, como pode ser constatado pela diminuição da amplitude do eletromiograma a partir de 25 s. Ao redor de tempo 52 s, o rato reiniciou a sua atividade exploratória quando a luz da gaiola foi apagada. Como pode ser comprovada pela análise da função de coerência ordinária (também denominada clássica) e os traçados dos registros das Figs. 10 e 11, as oscilações rítmicas são mais proeminentes durante o comportamento exploratório.

Para análise escolheu-se um trecho curto neste episódio, do período entre 18 e 20 s (Fig. 10a) caracterizado por alta amplitude da atividade eletromiográfica dos músculos da cabeça. A análise da CPD e do grafo direcionado (Figs. 10b e c) revela um padrão de retroalimentação (interação bidirecional) e aponta o núcleo dorsal da rafe (DR), talvez, como tendo um importante papel durante o comportamento exploratório no processamento de informação. Um panorama completamente diferente emerge ao analisar o segmento de registro entre 48 e 50 s (Fig. 11a) quando os músculos da cabeça do rato estão com baixa atividade. Ao comparar-se os cálculos de CPD nas Figs. 10b e 10c com os das Figs. 11b e 11c, nota-se que o DR tem inicialmente o papel de fonte de informação e depois torna-se um sorvedouro. Adicionalmente, nota-se que a influência de CA1 sobre DR é essencialmente indireta com o sinal fluindo via A3 e A10, Fig. 11e, em claro contraste com o grafo de conectividade funcional de CPD no período correspondente ao comportamento exploratório (Fig. 10c), em que todas as estruturas recebem substancial influência do núcleo dorsal da rafe (DR). Também do estado de vigília relaxada (Fig. 11c), as informações são passadas para outras estruturas via A10 para DR. Este exemplo evidencia distinto e potencialmente interessante padrão de conectividade funcional que caracterizam os diferentes estados comportamentais que não pode ser mostrado por nenhum outro método de análise de nosso conhecimento.

## 5. Conclusões e Comentários Finais

Com o desenvolvimento da coerência parcial direcionada, uma representação do conceito fundamental de causalidade de Granger no domínio da frequência baseado na análise de processos multivariados, propusemos uma extensão significativa às noções de coerência direcionada e função de transferência direcionada como se pode constatar pelos exemplos apresentados. Em todos os exemplos teóricos apresentados, a CPD foi capaz de indicar a estrutura funcional corretamente, enquanto os

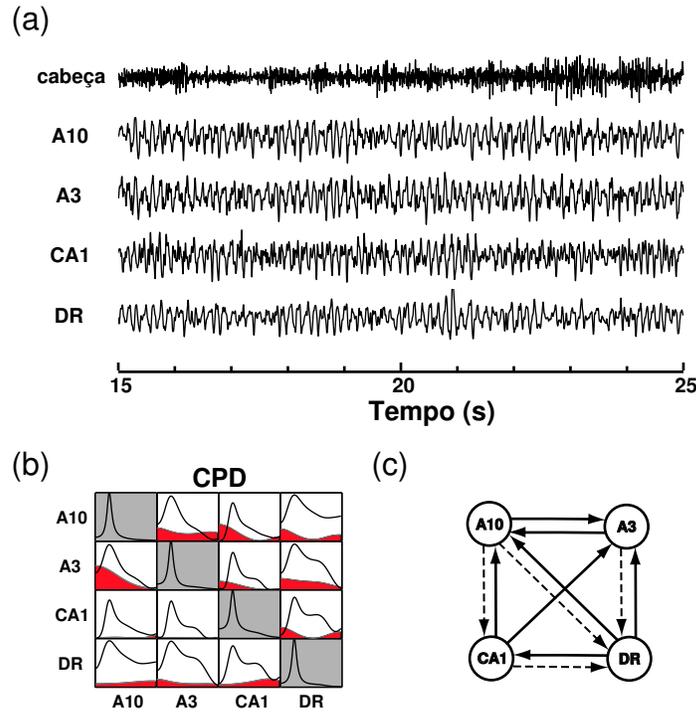


Figura 10: Segmento de registro de 10 s (correspondente ao segmento 15-25 s da Fig. 9, amostrada a 256 Hz) (a) obtido de rato em comportamento de atividade exploratória. O traçado superior (cabeça) corresponde ao eletromiograma de músculos da cabeça; os outros quatro traçados são de potenciais de campo local que mostram oscilações em ritmo teta registradas nas áreas corticais motora (A10) e somestésica (A3), hipocampo (CA1) e rafe dorsal (DR). A análise de CPD (b) determinadas a partir do segmento 18-20 s resulta no grafo de conectividade funcional (c) que indica DR não somente enviando mas também recebendo influências (informações) das outras estruturas. No grafo da CPD, os valores menores são representados por setas tracejadas (pares  $A10 \rightarrow DR$ ,  $A10 \rightarrow CA1$ ,  $A3 \rightarrow DR$ , e  $CA1 \rightarrow DR$ ). Nos gráficos em (b), o eixo-x representa a frequência na faixa de 0 a 32 Hz, enquanto eixo-y está normalizado entre zero e o máximo da densidade espectral de potência, e em todos os outros gráficos os traçados de coerência estão na faixa de 0 e 1.

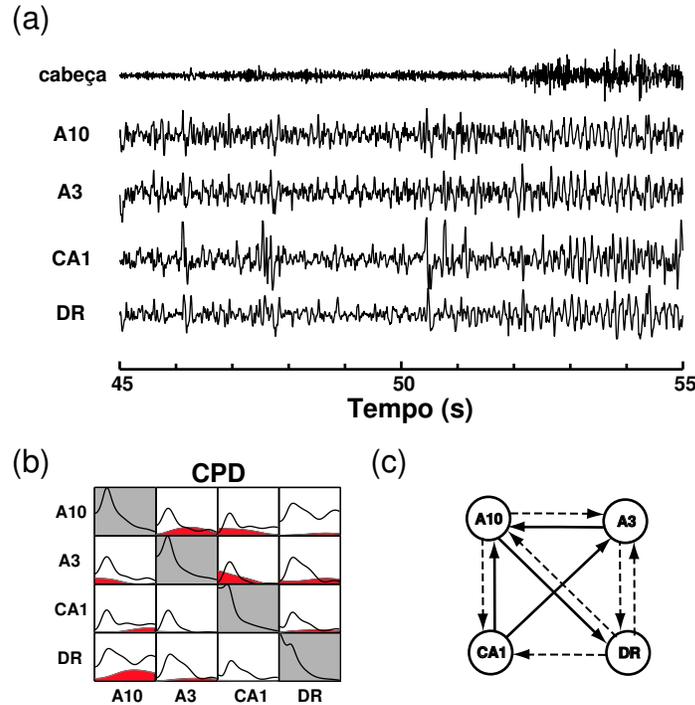


Figura 11: Dez segundos de registro (a), correspondente ao trecho 45-55 s da Fig. 9, amostrado a 256 Hz, mostrando uma transição do estado de repouso para o comportamento exploratório ativo induzida pelo apagar da luz de iluminação da gaiola ao redor de 52 s. Vide Fig 10 para as identificações dos canais de registro. Note-se também que as ondas teta torna-se mais proeminente nas quatro estruturas cerebrais concomitantemente com o aumento da atividade eletromiográfica da musculatura da cabeça. Quando comparada à Fig. 10b, as CPD calculadas para o segmento 48-50 s e seu grafo de conectividade funcional (c) mostram conexões mais fracas partindo de DR para todas as outras estruturas. Note-se também que DR é fundamentalmente um sorvedouro de informação enquanto CA1 é uma fonte de informação para áreas corticais, o que é compatível com a idéia de que neste período o hipocampo, CA1, esteja funcionalmente atuando como fonte de informação memorizada. Nos gráficos em (b), o eixo-x representa a frequência na faixa de 0 a 32 Hz; por sua vez o eixo-y está com escala ajustada para o gráfico de espectro de potência entre 0 e o pico de seu valor, e em todos os outros gráficos os traçados de coerência estão na faixa de 0 e 1.

outros métodos de correlação ou a FTD não permitem uma inferência inequívoca dessas mesmas estruturas.

Nossos exemplos mostraram casos em que CD/FTD e CPD (e por conseguinte GCT) devem concordar, como demonstrado teoricamente para  $N = 2$  no primeiro exemplo. Quando mais que duas estruturas são consideradas simultaneamente, a nulidade de FTD implica necessariamente na nulidade de CPD, e vice-versa, somente se a estrutura da matriz  $\mathbf{H}(f)$  é preservada na sua inversão. Enquanto  $\pi_{ij}(f)$  proporciona uma informação clara sobre a estrutura funcional do sistema estudado pela expressão das interações diretas entre pares de estruturas,  $\gamma_{ij}(f)$  fornece uma informação indireta sobre essas interações, ou melhor, corresponde a uma informação composta obtida a partir das informações de interações diretas. Como discutido em [2,5],  $\gamma_{ij}(f)$  representa a fração de potência em  $x_i(n)$  que originou de cada um dos  $x_j(n)$  de tal sorte que ela descreve a interação entre pares de estruturas, que pode ser direta ou indireta, devida a várias vias de propagação do sinal.

Em outras palavras, a coerência direcionada ou a função de transferência direcionada sumariza as interações estruturais decorrentes de todas as vias de conexões possíveis entre duas estruturas, logo elas serão significantes se existe pelo menos um caminho de propagação da informação.

O cenário proporcionado pela coerência parcial direcionada é muito mais claro pois  $\pi_{ij}(f)$  somente é nula se existe uma interação direta entre duas estruturas quando todas as estruturas relevantes forem consideradas na análise. Outro aspecto de fundamental importância do novo método (que também pode ser inferido por CD/FTD) é a sua capacidade de detectar componentes de retroalimentação entre duas estruturas, ou seja se a interação é uni ou bidirecional.

Com adequada transformação das informações de trens de disparos neuronais por meio de convolução por uma função kernel, é possível aplicar esse método também na análise de dados de potenciais de ação registrados em múltiplas estruturas [12,13,26].

A CPD proporciona uma fotografia mais nítida, ou melhor, permite perscrutar a dinâmica da interação entre estruturas cerebrais pois se pode analisar de segmentos relativamente curtos de séries temporais, alcançando alta resolução temporal comparada a outras técnicas de correlação.

Na prática, a estimação da CPD requer o ajuste adequado de um modelo ARM (assunto este tratado em profundidade em compêndios didáticos [20,21]). Há várias questões teóricas em aberto no processo de estimação de ARM, tais como a estimativa da ordem modelo (i.e. a escolha do valor de  $p$  na Eq. 2.3), os testes diagnósticos quanto ao resíduo do modelo e, talvez mais crucial, a estacionariedade conjunta das séries temporais em que se baseia a estimação dos parâmetros  $a_{ij}(r)$  [7].

Outro aspecto relevante é a importância de se considerar um grande número de estruturas representativas da dinâmica estudada na análise. Isto é pertinente mesmo na análise de CPD pois uma análise de um subconjunto de estruturas pode levar a conclusões errôneas quanto à estrutura funcional verdadeira da interação entre elementos neurais (vide [3]). Por conseguinte, todas as técnicas baseadas na análise restrita a pares de estruturas são ineficientes no estudo da inferência funcional. É, portanto, imperioso realizar análises com inclusão simultânea de um grande número

de estruturas para o entendimento da codificação e do mecanismo de processamento neural.

Embora não tenham sido tratadas neste trabalho, as questões estatísticas relacionadas a estimadores são muito importantes. Existem resultados assintóticos sobre os resultados dos coeficientes  $a_{ij}(r)$  que permitem realizar testes de causalidades de Granger [5,20], porém sua utilidade para os casos de sinais neuroelétricos ainda não é clara devido à natureza praticamente não-estacionária desses sinais.

Todos esses métodos necessitam de um estudo mais amplo baseado em casuística maior. Por fim, o uso de análise de coerência parcial direcionada não se restringe a problemas em neurociências, podendo-se aplicá-la a análise de outras séries temporais multivariadas de sistemas dinâmicos com comportamento predominantemente linear.

**Abstract.** Correlation methods have been the tools of choice in the study of functional interactions among brain structures. In this work we describe a new alternative approach employing multivariate time series (called partial directed coherence) which makes the ordinary correlation tools obsolete as it allows a clearer depiction of the functional structure by taking into account the simultaneous activity of several brain structures. We also discuss how partial directed coherence reflects the the concept of Granger causality in the frequency domain and we employ theoretical models and experimental data to illustrate the new approach, comparing it to other current techniques.

## Referências

- [1] J. Arnhold, P. Grassberger, K. Lehnertz e C.E. Elger, A robust method for detecting interdependences: application to intracranially recorded EEG, *Physica D*, **134** (1999), 419-430.
- [2] L.A. Baccalá e K. Sameshima, Directed Coherence: A tool for exploring functional interactions among brain structures, in "Methods for Neural Ensemble Recordings" (M.A.L. Nicolelis, ed.) pp. 179-192, CRC Press, Boca Raton, 1998.
- [3] L.A. Baccalá e K. Sameshima, Overcoming the limitations of correlation analysis for many simultaneously processed neural structures, *Prog. Brain Res.*, **130** (2001), 33-47.
- [4] L.A. Baccalá e K. Sameshima, Partial directed coherence a new concept in neural structure determination, *Biol. Cybern.*, **84** (2001), 463-474.
- [5] L.A. Baccalá, K. Sameshima, G. Ballester, A.C. Valle e C. Timo-Iaria, Studying the interaction between brain structures via directed coherence and Granger causality, *App. Sig. Process.*, **5** (1998), 40-48.
- [6] J.S. Bendat e A.G. Piersol, "Random Data: Analysis and Measurement Procedures", 2nd ed, John Wiley, New York, 1986.

- [7] C. Bernasconi e P. Konig, On the directionality of cortical interactions studied by structural analysis of electrophysiological recordings, *Biol. Cybern.*, **81** (1999), 199-210.
- [8] N. Brunel e J.P. Nadal, Mutual information, Fisher information, and population coding, *Neural Comput.*, **10** (1998), 1731-1757.
- [9] C.N. Christakos, On the detection and measurement of synchrony in neural populations by coherence analysis, *J. Neurophysiol.*, **78** (1997), 3453-3459.
- [10] R.B. Duckrow e S.S. Spencer, Regional coherence and the transfer of ictal activity during seizure onset in the medial temporal lobe, *Electroenceph. clin. Neurophysiol.*, **82** (1992), 415-422.
- [11] H.B. Eichenbaum e J.L. Davis, "Neuronal Ensembles: Strategies for Recording and Decoding", John Wiley and Sons, New York, 1998.
- [12] E.E. Fanselow, K. Sameshima, L.A. Baccalá e M.A.L. Nicolelis, Thalamic bursting in rats during different awake behavioral states, *Proc. Natl. Acad. Sci. U.S.A.*, **98** (2001), 15330-15335.
- [13] A.S. French e A.V. Holden, Alias-free sampling of neuronal spike trains, *Kybernetik*, **8** (1971), 165-171.
- [14] M.R. Gevers e B.D.O. Anderson, Representation of jointly stationary stochastic feedback processes, *Int. J. Control*, **33** (1981), 777-809.
- [15] M.R. Gevers e B.D.O. Anderson, On jointly stationary feedback-free stochastic processes, *IEEE Trans. Autom. Contr.*, **AC27** (1982), 431-436.
- [16] J. Geweke, R. Meese e W. Dent, Comparative tests of causality in temporal systems, *J. Econometrics*, **21** (1983), 161-194.
- [17] E. Glaser e D. Ruchkin, "Principles of Neurobiological Signal Analysis", Academic Press, New York, 1976.
- [18] C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, **37** (1969), 424-438.
- [19] M.J. Kaminski e K.J. Blinowska, A new method of the description of the information flow in the brain structures, *Biol. Cybern.*, **65** (1991), 203-210.
- [20] H. Lütkepohl, "Introduction to Multiple Time Series Analysis", 2nd ed, Springer-Verlag, Berlin., 1993.
- [21] S.L. Marple Jr, "Digital Spectral Analysis", Prentice Hall, Englewood Cliffs, 1987.
- [22] M.A.L. Nicolelis, ed. *Methods for Neural Ensemble Recordings*. 1998, CRC Press, Boca Raton.

- [23] F. Rieke, D. Warland, R. de Ruyter van Steveninck e W. Bialek, "Spikes: Exploring the Neural Code", MIT Press, Cambridge, 1997.
- [24] J. Rosenberg, D. Halliday, P. Breeze e B. Conway, Identification of patterns of neuronal connectivity - partial spectra, partial coherence, and neuronal interactions, *J. Neurosci. Meth.*, **83** (1998), 57-72.
- [25] Y. Saito e H. Harashima, Tracking of information within multichannel EEG record - causal analysis in EEG, in "Recent Advances in EEG and EMG Data Processing" (N. Yamaguchi e K. Fujisawa, eds.), pp. 133-146, Elsevier, Amsterdam, 1981.
- [26] K. Sameshima e L.A. Baccalá, Using partial directed coherence to describe neuronal ensemble interactions, *J. Neurosci. Meth.*, **94** (1999), 93-103.
- [27] K. Sameshima, L.A. Baccalá, G. Ballester, A.C. Valle e C. Timo-Iaria, Causality analysis of rhythmic activities of desynchronized sleep in the rat, *Soc. Neurosci. Abstr.*, **22** (1996), 27.
- [28] S.J. Schiff, P. So, T. Chang, R.E. Burke e T. Sauer, Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble, *Phys. Rev. E*, **54** (1996), 6708-6724.
- [29] S.M. Schneider, R.H. Kwong, F.A. Lenz e H.C. Kwan, Detection of feedback in the central nervous system using system identification techniques, *Biol. Cybern.*, **60** (1989), 203-212.
- [30] C. Timo-Iaria, N. Negrão, W.R. Schmidek, K. Hoshino, C.E.L. Menezes e T.L. Rocha, Phases and states of sleep in the rat, *Physiol. Behav.*, **5** (1970), 1057-1062.
- [31] S. Yamada, M. Nakashima, K. Matsumoto e S. Shiono, Information theoretic analysis of action potential trains I. Analysis of correlation between 2 neurons, *Biol. Cybern.*, **68** (1993), 215-220.

## Aspectos do Sistema Imune Fetal e As Doenças Infeciosas

S.D. PASSOS<sup>1</sup>, Departamento de Pediatria, Faculdade de Medicina de Jundiaí e Centro de Pesquisas Materno Infantil, CEPEMI-HRFR-DIR IV, Av. dos Coqueiros, 300, CEP: 07780-000, Franco da Rocha, SP, Brasil

H.M. YANG<sup>2</sup>, Departamento de Matemática Aplicada, Instituto de Matemática, Estatística e Computação Científica, UNICAMP, CP: 6065, CEP: 13081-970, Campinas, SP, Brasil

**Resumo.** As curvas sorológicas populacionais fornecem subsídios para programas de vacinação. Realizou-se um estudo prospectivo a fim de descrever o perfil sorológico do rubivírus, através do método do ELISA quantitativo, em 244 crianças de 0 a 2 anos, além de verificar a relação existente entre a concentração de anticorpos de 39 mães (M) e seus recém-nascidos (RN), durante o período de estudo, de 18/11/1990 a 18/01/1991, em Caieiras, São Paulo. Estudou-se, também, a soroconversão em 228 crianças à vacina monovalente da rubéola. A amostra de crianças estudadas foi delimitada de maneira probabilística por conglomerados. A análise estatística empregada nos pares foi a regressão linear, teste dos sinais, observação aos pares e análise de variância uni- e bi-lateral. No decaimento de anticorpos anti-rubivírus, utilizou-se o ajuste das curvas, pelo método da máxima verossimilhança. A análise estatística revelou: a) uma maior concentração de anticorpos no recém-nascido quando os níveis maternos de anticorpos eram mais baixos (menor que 127,5 *U.I./ml*), sugerindo a presença de um mecanismo ativo de transporte transplacentário dos mesmos; b) Mais de 90% das crianças ao nascimento possuem anticorpos anti-rubivírus com rápido decaimento nos primeiros meses. A estimativa para o desaparecimento dos anticorpos recebidos por via transplacentária situa-se entre 3,45 e 6,03 meses; e c) a taxa de soroconversão, após a vacinação, já é elevada (acima de 80%) a partir do quinto mês de vida, aproximando-se de 100% entre 8 e 10 meses de idade.

### 1. Introdução

Desde 1882, quando o biólogo russo Dimitrii Ivanovsky descobriu o primeiro vírus causador da doença do tabaco, outros vírus foram descobertos rapidamente devido a necessidade de se ter uma compreensão dos mecanismos das doenças humanas

---

<sup>1</sup>sauloduarte@uol.com.br; apoio financeiro FAPESP.

<sup>2</sup>hyunyang@ime.unicamp.br; apoio financeiro FAPESP e CNPq.

[57]. Muito embora durante a vida nosso organismo seja constantemente exposto a esses agentes infecciosos, o surgimento da doença depende de inúmeros fatores tais como uma resposta imunológica adequada.

O sistema imune materno-fetal representa o primeiro momento de nossa resposta à equação de sobrevivência humana. Devido à imaturidade do sistema imune fetal, a imunoproteção do feto e do neonato depende da transferência de anticorpos maternos, que é mediada, na vida intra-uterina, pela placenta ao transportar predominantemente imunoglobulinas (IgG) da mãe para o feto [10] [13] [35] [67].

A IgG que atravessa a placenta pode ser detectada precocemente em torno da oitava semana, mas as concentrações fetais de IgG permanecem abaixo de 100 mg/dl até por volta da 17<sup>a</sup> a 20<sup>a</sup> semana [41] quando começam a subir continuamente até cerca da 30<sup>a</sup> semana. Os níveis de IgG nos neonatos a termo podem apresentar concentrações de anticorpos que excedem de 5 a 10% os níveis maternos [40], indicando que a IgG é ativamente transportada para o feto contra gradiente de concentração provavelmente através das vesículas do sincitiotrofoblasto. Tabela 1 demonstra que a razão das concentrações do soro do feto para mãe aumenta de 0,4 na 28<sup>a</sup> semana de gestação para 1,5 na 40<sup>a</sup> semana.

Gestação (em semanas)	Razão entre IgGs Fetal e Materno
28	0,40
30	0,60
32	0,80
34	0,95
36	1,15
38	1,35
40	1,50

Tabela 1 Transferência de anticorpos IgG maternos para feto [60].

É consenso da literatura que há na placenta receptores específicos para diversas proteínas transportadoras. A transferência de proteínas se faz de maneira altamente seletiva, requerendo uma certa disponibilidade de receptores placentários, estando mais relacionada com a configuração molecular do que com o tamanho molecular. A região Fc da molécula de IgG é o elemento funcional de reconhecimento específico neste processo [11] [30] [42]. Todas as subclasses da IgG humana (IgG1, IgG2, IgG3 e IgG4) atravessam a placenta mas envolvem receptores específicos (FcR) na placenta com elevada afinidade por IgG2 e IgG 4 e nenhuma por IgA ou IgM.

Essa passagem é mediada por dois mecanismos principais. O primeiro é de ordem cinética, ou seja, simples transferência passiva na qual o nível de IgG fetal é proporcional às concentrações de IgG materna. O segundo está relacionado com um processo ativo enzimático no qual a IgG é transferida, ativamente, da mãe para o feto. Esta transferência pode ser inibida direta ou indiretamente, pelas altas concentrações de IgG no soro materno. Para baixas concentrações maternas, o sistema é ativado e, proporcionalmente, mais IgG é transferida da mãe para o feto, ocorrendo à inibição desse processo inverso quando ocorre o inverso.

Estes mecanismos passam a ter uma maior importância, durante os processos infecciosos ou imunização, que possam ocorrer durante a gestação [3] [4]. Ambos

determinam um aumento dos anticorpos maternos no período neonatal. Isto possibilita por um lado, aquisição de proteção passiva, pela passagem transplacentária de IgG. Por outro, pode interferir na capacidade de resposta imune a infecções e a vacinação, com possibilidade de ganho de proteção passiva durante a passagem transplacentária de IgG e modular futura capacidade em responder à infecção e e/ou vacina na criança [61]. Por exemplo, Glessen *et al.* [34], em um estudo de imunização de mães vacinadas contra *Haemophilus influenza* tipo b entre 34<sup>a</sup> e 36<sup>a</sup> semana de gestação, demonstraram uma efetiva transferência de IgG anti-hemophilus para o feto, que pareceu proteger o neonato até o quarto mês de vida. Embora a mãe possa estar protegida por anticorpos específicos, nem sempre observamos a mesma proteção ao feto, pois há inúmeros fatores que interferem neste mecanismo de transporte, que precisam ser melhor elucidados.

Considera-se que há infecção hematogênica transplacentária quando um agente presente no espaço interviloso da placenta atravessa o epitélio trofoblástico e localiza-se no córion e, a partir desse local, pode atingir vasos fetais das vilosidades e/ou placa corial e alcançar o feto podendo levar à infecção congênita [12]. A infecção fetal é uma importante causa de morbi-mortalidade apesar da disponibilidade atual de potentes agentes anti-microbianos [25]. O prognóstico do neonato tanto será melhor quando mais precoce for o tratamento pois o início tardio pode levar, dentre outros aspectos, a seqüelas em órgãos com limitado poder de regeneração como o sistema nervoso central. Esse fato pode ser constatado pela resposta imatura resposta imunológica endógena dos neonatos e lactentes jovens.

A prevenção, diagnóstico e tratamento precoce da infecção são objetivos maiores no controle das infecções perinatais. Sempre que possível, a imunização antes da gravidez é desejável e recomendável, pois a imunização seletiva durante a gestação poderia oferecer uma proteção teórica ao feto. O objetivo da imunização seria a produção de imunoglobulinas (IgG) específicas e/ou fatores humorais que podem atravessar a placenta e resultar na proteção passiva da criança, por um período em que persistirem esses anticorpos. Na prevenção de algumas doenças, como tétano neonatal, a vacinação da grávida tem demonstrado algumas evidências de proteção ao neonato, as quais já foram estabelecidas [18] [33] [52], mas ainda há controvérsias em outras, como a rubéola.

Embora exista uma boa passagem seletiva pela placenta de anticorpos maternos para o recém-nascido (RN), principalmente no último trimestre da gravidez, os estudos demonstram uma queda drástica dos anticorpos antivírus da rubéola nos primeiros meses de vida, seguida de um decaimento mensal, conseqüência da perda dos anticorpos recebidos de mães (M) que haviam tido contato com o vírus.

A rubéola em nosso meio constitui um importante problema de saúde pública, pois um grande número de mulheres é infectado durante a idade fértil [66]. Em 1987, um inquérito soropidemiológico em 1.400 amostras de sangue de crianças com idade entre 2 e 14 anos e 329 amostras de sangue de cordão umbilical foi realizado na cidade de São Paulo para verificar a prevalência de anticorpos contra o vírus da rubéola. Um aumento significativo foi encontrado após 6 anos de idade, e 77% dos indivíduos eram soropositivos entre 15 e 19 anos, aumentando para 90,5% no sangue do cordão de crianças cujas mães apresentavam idade entre 20 e 29 anos, alcançando 95,6%

no grupo de 30 a 34 anos. Numerosos estudos têm demonstrado que os anticorpos maternos adquiridos pelos recém-nascidos decrescem de forma exponencial após o nascimento, porém, a maioria destes foram realizados com imunoglobulinas totais e pouca atenção tem sido dada ao declínio de anticorpos específicos [19] [58] [73].

Passos [58] encontrou um total de 5,6% das parturientes não imunes a rubéola, como resultado de uma pesquisa realizada no município de Caieiras, SP, em 1991. Apesar deste resultado estar abaixo quando se compara com os dados da literatura (cerca de 9%), o problema continua tendo uma importante dimensão pois essas mulheres estão sob um risco de terem filhos portadores da síndrome da rubéola congênita (SRC). Dado que a infecção rubeólica fetal pode ocorrer independentemente da infecção materna manifestar-se clinicamente ou ser assintomática, uma gestante infectada irá desenvolver uma viremia. Ocorrendo a viremia, sua placenta será atingida e poderá infectar o feto. Dependendo da idade gestacional em que ocorre a infecção, o vírus poderá acarretar alterações teratogênicas levando a abortos espontâneos, natimortos e crianças com rubéola congênita portadoras de mal-formações evidentes ou não ao nascimento [9] [47].

Dado que a infecção rubeólica fetal pode ocorrer independentemente da infecção materna manifestar-se clinicamente ou ser assintomática, uma gestante infectada irá desenvolver uma viremia. Ocorrendo a viremia, sua placenta será atingida e poderá infectar o feto. Dependendo da idade gestacional em que ocorre a infecção, o vírus poderá acarretar alterações teratogênicas levando a abortos espontâneos, natimortos e crianças com rubéola congênita portadoras de mal-formações evidentes ou não ao nascimento [9]. A preocupação com o aumento do número de mulheres soronegativas, principalmente em nosso meio e como conseqüência, um incremento na possibilidade de virem a ter infecção rubeólica durante a gravidez, podendo acarretar o nascimento de crianças com a SRC, fez com que muitos autores alertassem para a necessidade de um rastreamento sorológico das adolescentes no pré-natal, a despeito de terem tido rubéola ou terem sido vacinadas, bem como o seguimento de mulheres com títulos baixos durante toda a gestação [10] [38] [64] [36].

Um outro recorte de interesse epidemiológico são os profissionais da saúde e educação que atuam com crianças e adolescentes, pois por estarem expostos a vírus que dão quadros de doenças benignas na infância, a mesma evolução pode não ocorrer no adulto. Existe uma recomendação de “American College of Physicians”, o “Advisory Committee on Immunizations Practices”, e vários outros organismos e autores que vacine o pessoal da saúde contra as doenças evitáveis, principalmente sarampo, caxumba, rubéola, hepatite B e varicela [1] [5] [7] [17] [39] [44] [71] [72].

Para se ter uma dimensão do problema, Passos em 1997, realizou-se uma pesquisa entre 165 funcionárias das creches municipais, em idade fértil, no município de Caieiras (SP), a fim de verificar a suscetibilidade e o conhecimento sobre a rubéola [59]. Os resultados sorológicos, pelo método ELISA (método quantitativo imunoenzimático), demonstraram que 4,2% de funcionárias em idade fértil apresentavam-se soronegativas. Em relação ao conhecimento prévio sobre número de relatos de rubéola clínica e vacinação comprovada ao rubivírus, podemos observar que 81,8% (135 em 165) das funcionárias negaram ter tido rubéola e 38,2% (63) referem-se a uma vacinação contra a rubéola nos últimos meses em razão da epidemia de

sarampo de 1997, mas não apresentaram comprovante de vacinação. Quando perguntadas sobre a necessidade de vacinação, 70,9% (117) não tinham conhecimento prévio, 25,4% (42) confirmaram a necessidade e 3,7% (6) não responderam. Conforme pôde ser constatado 135 (81,88%) funcionárias negaram história positiva para rubéola em discordância com a elevada proporção de soropositividade 158 (95,7%), o que produz falta de sensibilidade e especificidade dos antecedentes para prever o estado imunitário.

Considerando elevado custo dos exames sorológicos em relação ao custo da vacina, a abordagem da vacinação parece-nos economicamente viável, eficaz e segura, pois poucas reações vacinais são observadas. A importância do uso da vacinação como forma de proteção contra muitas doenças pode ser fundamentada sobre duas vertentes: a diminuição de casos novos da doença (morbidade) e o baixo custo financeiro da vacina em relação ao tratamento da doença (monetário) [73] ou suas complicações como a SRC. A vacina tríplice viral deve ser preferida à vacina isolada [39].

Sob um enfoque individual, a rubéola poderia oferecer pouca preocupação se não fosse o efeito do vírus sobre o feto, pois numerosos estudos demonstraram os riscos de efeitos teratogênicos nas primeiras 12 semanas de gestação [32] [50]. O objetivo primordial, e universalmente aceita, da vacinação contra a rubéola é a prevenção da infecção em mulheres grávidas e, como consequência, a prevenção da SRC. Entretanto, há uma discordância acerca da melhor maneira de se alcançar este fim, fazendo com que diferentes programas tenham sido propostos por vários países e modificados posteriormente.

Com a finalidade de contribuir na elucidação de alguns aspectos da vasta discussão sobre o efeito dos anticorpos materno contra a rubéola na proteção temporária sobre os recém-nascidos, foram estudadas quantitativamente as amostras de sangue coletadas no campo. Da sorologia contra o vírus da rubéola foram feitos estudos quanto do transporte pela barreira placentária e, também, da soroconversão vacinal. Um outro aspecto importante do estudo foi a obtenção de uma curva sorológica de anticorpos anti-rubéola com os dados locais.

## 2. Metodologia

Foi estudada uma amostra de 244 crianças menores de 2 anos e 39 pares de mães e recém-nascidos, residentes no município de Caieiras, coletada durante o período de 18 de novembro de 1990 a 18 de janeiro de 1991. Este estudo, baseado em uma amostragem da população da cidade de Caieiras, procura avaliar o perfil sorológico em crianças menores de dois anos de idade, antes e após a vacinação contra a rubéola, bem como o nível sérico de anticorpos antivírus da rubéola (AVR) em mães (M) e recém-nascidos (RN) durante o período de estudo [6] [58] [73].

O estudo, além de obter o perfil sorológico de uma comunidade brasileira, procura compreender os fenômenos que medeiam as relações entre os níveis de anticorpos dos fetos pareado com os de suas mães, para entender se há fatores biológicos que contribuem para resultar nas distribuições de concentrações em mães e recém-

nascidos. De igual importância é descobrir se a presença de anticorpos maternos nas crianças afeta a soroconversão vacinal.

## 2.1. Dados de campo

O trabalho de campo desenvolveu-se na Maternidade do Hospital Regional de Caieiras (EMED) e na comunidade urbana. Das crianças recém-nascidas (RN) durante o período de estudo, cujas famílias residiam em Caieiras, foram colhidas amostras de sangue materno (M) no pós-parto imediato e no cordão umbilical. Uma ficha contendo dados de M foi preenchida pela obstetrix e a de RN foi preenchida pelo pediatra do berçário.

Não foram incluídas no estudo as crianças que apresentavam os seguintes problemas:

1. Prematuridade, mal-formação, gemelaridade e anóxia grave;
2. Relato de imunodeficiência congênita ou adquirida;
3. Doenças oncológicas;
4. Processos febris agudos;
5. Tratamento com imunossupressor ou que tivessem recebido imunoglobulinas ou outro derivado sangüíneo no período inferior a três meses; e
6. História de severa alergia a neomicina ou kanamicina ou reação anaflática à proteína do ovo.

Em relação à coleta de sangue, nos recém-nascidos, lactentes e crianças da comunidade cujo acesso venoso era difícil, optou-se pela coleta com seringa, escalpe 21 ou 25, e nas demais crianças pela coleta pelo sistema a vácuo, num total de 1 – 5 *ml*. Uma amostra de sangue do cordão umbilical dos recém-nascidos foi coletada no pós-parto imediato, com sistema seringa e escalpe e, conjuntamente, uma outra amostra de sangue da mãe foi obtida por punção venosa em membro superior, utilizando-se, preferencialmente, o sistema de coleta a vácuo. Todas as amostras de sangue foram imediatamente acondicionadas em geladeira de isopor que continha bolsa de gelo reciclável, e, posteriormente, centrifugadas a 3.500 *r.p.m.* durante 15 minutos (Centrifuga de mesa Bio-Eng), divididas em quatro alíquotas e estocadas a  $-20^{\circ} C$ . O soro congelado era transportado ao laboratório, em gelo seco, para posterior análise.

Além destes casos, por ocasião do estudo, foram coletadas 39 amostras de sangue de M e do cordão umbilical de RN participantes do estudo, sendo que a distribuição por faixa etária materna mais prevalente foi de 20 – 24 anos (48,7%), seguida pela de 25 – 29 anos (18%). Sobre a possibilidade de terem tido rubéola anteriormente, 32 (82%) negaram; 4 (10,25%) disseram ter tido a doença em alguma fase da vida, mas sem comprovação através de testes sorológicos, e 3 (7,75%) não souberam informar.

Em relação às vacinações prévias contra a rubéola, 25 (64,10%) não souberam informar, 13 (33,3%) negaram ter recebido e somente 1 (2,57%) afirma ter recebido, mas sem comprovação vacinal. Em relação ao fato de terem recebido medicamento nos últimos seis meses, 20 (51,3%) informaram ter recebido vitaminas e sintomáticos. Todas as parturientes negaram ter recebido gamaglobulina e sangue no último ano. Na amostra destes 39 recém-nascidos, 24 (61,5%) eram do sexo feminino e 15 (38,5%) do sexo masculino com peso médio ao nascimento de 3.370 g. O parto cesárea foi realizado em 25 (64,1%) das gestantes. O tipo de parto parece não ter influência no estudo das concentrações séricas de imunoglobulinas [48].

Em relação ao método laboratorial, o método quantitativo imunoenzimático (ELISA) utilizado foi ETI-Rubeck-G, Sorin Biomédica, Diagnostici, Vercelli, Itália. O “cut off” de 10 *U.I./ml* foi utilizado para distinguir as crianças soropositivas (assumidas como imunes) das soronegativas. A soroconversão seguida da vacinação foi definida como mudança de soronegativo para soropositivo ou aumento de quatro vezes na concentração de títulos de IgG da primeira para a segunda amostra.

## 2.2. Estudo dos pares mãe–recém-nascido

A análise dos pares das concentrações de anticorpos encontradas na mãe e no recém-nascido foi realizada por métodos estatísticos [22]. Foram utilizados três métodos de análise estatística: teste não-paramétrico, regressão linear e testes paramétricos.

A análise pareada das concentrações de anticorpos encontradas na mãe e no recém-nascido tem o objetivo de verificar se existe alguma relação entre as concentrações de anticorpos nos dois grupos. Os métodos estatísticos, aqui relacionados, permitem avaliar essa possível relação e discutir, também, os fatores geradores das distribuições de concentrações entre mães e recém-nascidos.

### 2.2.1. Teste não-paramétrico

O teste não-paramétrico aplicado à amostra de pares de concentrações de anticorpos de mãe e recém-nascido, para verificar eventuais diferenças, é o teste de sinais. O teste não-paramétrico permite comparar como os dados estão distribuídos (não é preciso conhecer a sua real distribuição probabilística), mesmo que a amostra não obedeça a uma distribuição normal.

O teste de sinais serve para comparar duas amostras que estejam sob a influência de diversas condições. Muitas vezes, algum efeito que se deseja mensurar pode sofrer alterações devido a muitos outros fatores que podem, de alguma forma, “mascarar” o efeito desejado. Portanto, quando se usa teste de sinais (observação em pares), o objetivo é tentar eliminar todos os outros fatores externos. Este teste é aplicável, pois a amostra, observação de pares de concentrações de mãe e recém-nascido, é proveniente de mãe e recém-nascido que são seres distintos, mas a obtenção da concentração dos pares provém de situação similar (sangue e cordão, no ato do nascimento), e cada par é distinto um do outro, pois diferente é a condição de cada mãe e recém-nascido no parto.

O teste de sinais conta os sinais da diferença das concentrações do par mãe e recém-nascido, ou seja,

$$TS = \text{sign}([M] - [RN]), \quad (2.1)$$

onde  $TS$  é o sinal ( $\text{sign}$ ) da diferença. Por exemplo, se  $[M] = 10$  e  $[RN] = 6$ , então o sinal  $\text{sign}$  é  $+$ , portanto  $TS = +$ . O teste de sinais permite estudar se uma das amostras apresenta valores maiores ou menores do que outra. O teste pode ser para um lado (maior ou menor, com nível de significância  $\alpha$ ), ou nos dois lados (diferente, com nível de significância  $2\alpha$ ). Contando o número de sinais positivos, compara-se este número com o número crítico dado pelo nível de significância pré-determinado para uma conclusão estatística a respeito das duas amostras.

### 2.2.2. Regressão linear

A primeira abordagem paramétrica será pelo método de regressão linear. É sabido que a mãe cede ao feto anticorpos que ela carrega. Então os valores da concentração de anticorpos das mães,  $[M]$ , serão considerados variáveis independentes e os valores da concentração de anticorpos dos recém-nascidos,  $[RN]$ , serão considerados variáveis dependentes. Assim, ajusta-se a equação linear

$$[RN] = k_0 + k_1 [M], \quad (2.2)$$

onde  $k_1$  é constante de proporcionalidade (ou coeficiente de regressão) e o coeficiente linear é  $k_0$ . No estudo do par mãe e recém-nascido tem-se  $k_0 = 0$ , pois as mães não podem dar de herança ao seu feto anticorpos que elas não possuem.

O valor da constante  $k_1$ , ajustado pelo método de mínimos quadrados, dá informação sobre o tipo de relação que possa existir entre as concentrações de anticorpos no grupo das mães e no de recém-nascidos.

### 2.2.3. Testes paramétricos

Os testes paramétricos utilizados para estudar eventuais diferenças entre as concentrações de mães e recém-nascidos são testes de observação em pares, análise de variância por classificação simples (“one-way”) e análise de variância com dois fatores (“two-way”).

O teste de observação em pares serve para discriminar duas amostras que estejam sob influência de muitos fatores externos. Este teste baseia-se na diferença da concentração do par mãe e recém-nascido, ou seja,

$$d = [M] - [RN], \quad (2.3)$$

onde  $d$  é o valor da diferença. O teste fundamenta-se na distribuição da diferença  $d$ . A hipótese nula é analisada pelo teste t de “Student”.

A análise de variância permite testar diferença entre duas ou mais amostras de populações simultaneamente. Testa-se a hipótese de que as médias de todos

os grupos sejam iguais através do cálculo da razão das variâncias entre os grupos e dentro dos grupos, resultado normalmente designado de  $F$ . Este valor tem uma distribuição normal, classificada seguindo os graus de liberdade dados pelos números da amostragem e dos grupos observados, e percentil correspondente a uma probabilidade  $1 - \alpha$ . Valores de  $F$  superiores a unidade já sugerem distinção entre os grupos. Valores de  $\alpha$  inferiores a 0,10 são com freqüência utilizados para a rejeição da hipótese nula de igualdade entre os grupos. O fundamento da análise de variância baseia-se no fato de que, se houver grandes diferenças nos valores médios da amostra, então a variância dos grupos combinados será muito maior do que a variância dos grupos separados.

A amostragem composta de concentrações de anticorpos IgG anti-rubéola de mães e recém-nascidos é testada pela análise de variância. Essa amostragem pode ser estudada sob duas hipóteses: a) os dois grupos de concentração, mãe e recém-nascido, são as únicas possibilidades de classificação, ou seja, dentro de cada um dos dois grupos não há nenhuma diferença. Este é o teste de análise de variância com um único fator; e b) além dos dois grupos de concentração, mãe e recém-nascido, existe uma outra categoria, a diferença entre os pares de concentração. Este é o teste de análise de variância com dois fatores. A hipótese nula do teste de variância é analisada pelo teste  $F$  de Fisher.

### 2.3. Estudo do decaimento

O estudo de decaimento da concentração de anticorpos nas crianças de pouca idade tem a finalidade de estimar a meia vida do decaimento, isto é, o tempo necessário para que uma dada concentração diminua pela metade. Ao mesmo tempo, é preciso estimar a idade em que a concentração de anticorpos herdados da mãe seja considerada nula. Este estudo é feito pelo ajuste de curvas.

#### 2.3.1. Regressão linear

O decaimento de anticorpos maternos será estudado considerando duas formas (hipóteses) de diminuição. A primeira forma de decaimento será estudada por meio de uma curva exponencial simples dada por

$$C(a) = C_0 e^{-Ka}, \quad (2.4)$$

onde  $C_0$  é a concentração ao nascimento ( $a = 0$ ) e  $K$  é a taxa de decaimento. Esta abordagem supõe que a taxa de decaimento é constante em todas as idades e é, também, a constante de proporcionalidade entre a taxa de variação da concentração e a concentração em dada idade.

Esta função não-linear pode ser linearizada, se tomar logaritmo neperiano nos dois lados, tomando, então, os parâmetros a serem ajustados as formas

$$\begin{cases} k_0 = \ln(C_0) \\ k_1 = -K. \end{cases} \quad (2.5)$$

Assim, ajusta-se a mesma equação linear dada pela equação (2.2) pelo método de mínimos quadrados. Para evitar problema com valor nulo, aos dados observados, ao serem transformados em logaritmos, devem ser acrescentados uma unidade. Após o ajuste linear, pode-se tomar o anti-logaritmo para obter os valores reais. Porém, como foi acrescentada uma unidade aos valores observados, a curva ajustada tem a forma

$$C(a) = e^{k_0} e^{k_1 a} - 1. \quad (2.6)$$

Da curva ajustada pode-se extrair algumas informações a respeito do decaimento de anticorpos maternos. O tempo necessário para que uma dada concentração diminua pela metade é facilmente obtida da equação (2.4), resultando em

$$a_{1/2} = \frac{\ln(2)}{K}, \quad (2.7)$$

onde  $a_{1/2}$  é a meia vida de decaimento de anticorpos herdados da mãe. A idade em que o anticorpo herdado da mãe pode ser considerado totalmente eliminado  $a_T$  é obtida da equação (2.4), considerando para isso uma concentração limiar arbitrária, dada, por exemplo, por

$$C(a_T) = \frac{C_0}{10}. \quad (2.8)$$

Em outras palavras, quando a concentração inicial (na idade zero) cai para um valor dez vezes menor, supõe-se que a concentração remanescente não mais protege a criança.

### 2.3.2. Ajuste por máxima verossimilhança

O método de ajuste anterior supôs a forma mais simples para o decaimento e, por isso, não requereu nenhuma transformação especial de dados da amostra (o uso de logaritmo foi no sentido de facilitar o ajuste, usando-se um método linear). Quando se supõe que a taxa de decaimento tem dependência com idade e, portanto, a taxa de variação da concentração não é proporcional ao valor da concentração em dada idade, a lei que rege o decaimento de anticorpos maternos não obedece a equação exponencial (2.4).

Para levar em consideração um certo desvio da homogeneidade etária do decaimento, ajusta-se uma logística com um polinômio de terceiro grau dado por

$$C(a) = \frac{1}{1 + e^{c_0 + c_1 a + c_2 a^2 + c_3 a^3}}, \quad (2.9)$$

onde  $c_i$ , com  $i = 0, 1, 2$  e  $3$ , são constantes relacionadas ao grau da potência dado pelo índice  $i$ .

Esta função será ajustada pelo método de máxima verossimilhança. Como a função logística (2.9) está limitada entre os valores 0 e 1, exige-se um tratamento especial dos dados da amostra antes de usá-los. Todas as concentrações serão normalizadas, isto é, todos os dados da amostra serão divididos pelo maior valor da

concentração dentro da amostra. A máxima verossimilhança pode ser definida em termos de seu valor logarítmico ( $l$ ), ou seja,

$$l(\theta) = \sum_{i=1}^n \{g_i \ln [C(a_i)] + (1 - g_i) \ln [1 - C(a_i)]\} \quad (2.10)$$

onde  $\theta$  refere-se aos parâmetros  $c_i$ ,  $i = 0, 1, 2$  e  $3$ , a serem ajustados,  $g_i$  é o valor da concentração normalizada,  $a_i$  é a idade considerada na  $i$ -ésima observação, correspondendo ao valor da concentração  $C(a_i)$ , e  $n$  é o número total da amostragem. A parte constante da expressão de máxima verossimilhança não foi considerada.

A meia vida do decaimento é facilmente extraído a partir da curva ajustada. A idade em que o anticorpo herdado da mãe pode ser considerado totalmente eliminado é obtida, também, a partir do gráfico, considerando a equação (2.8).

## 2.4. Estudo de soroconversão vacinal

O estudo de soroconversão nas crianças de pouca idade tem a finalidade de estudar a capacidade de resposta do sistema imunológico. Este estudo é feito pelo ajuste de máxima verossimilhança, usando-se a equação (2.9), porém com apenas dois valores, isto é: 1 se soroconverteu e 0, no caso contrário. Assim, o parâmetro  $g$  da equação (2.10) assume o valor 1 ou 0.

## 3. Resultados

Os dados de sorologia são analisados da seguinte forma. As 39 amostras de sangue do par mãe e recém-nascido são usadas para detectar se há diferença entre as concentrações de anticorpos anti-rubéola entre os grupos de mães e recém-nascidos. As amostras de 244 crianças menores de 2 anos são usadas para estudar o decaimento de anticorpos maternos anti-rubéola. Finalmente, as 196 amostras de crianças menores de 2 anos são usadas para estudo da soroconversão vacinal.

### 3.1. Estudo do par mãe e recém-nascido

Usa-se métodos estatísticos e ajuste de curvas para estudar as concentrações de anticorpos anti-rubéola encontradas em dois grupos: mãe e recém-nascido. A amostra é composta de 39 pares de concentrações de anticorpos.

#### 3.1.1. Teste não-paramétrico – Teste de sinais

Inicialmente, aplica-se o teste de sinais. Este teste não-paramétrico, em que a distribuição da amostra não depende da distribuição populacional (e nem precisa ser normal), faz a comparação entre as distribuições, sem utilizar nenhum parâmetro. Por isso, pode ser aplicado para verificar se as concentrações de anticorpos no grupo das mães é maior ou menor do que no grupo de recém-nascidos. Esta amostragem é formada por pares de observações em dois seres distintos, as observações de cada par

são feitas quando do momento do nascimento e os diferentes pares de observações são completamente diferentes. Justamente por causa desta última condição é que o teste de sinais é mais adequado, uma vez que o teste  $t$  não se aplica.

O teste de sinais é baseado na contagem de sinais das diferenças entre os pares de observação, e define-se por  $r$  o número de sinais menos freqüente. A hipótese nula ( $H_0$ ) é que cada diferença dos pares de anticorpos anti-rubéola dos grupos de mães e recém-nascidos tem uma distribuição de probabilidade com média 0. Portanto, rejeita-se a hipótese nula se o valor de  $r$  for pequeno. Com respeito ao valor pequeno de  $r$ , pode-se determinar o valor crítico,  $r_c$ , a partir de percentiles para a distribuição da amostra do número de sinais positivos quando a hipótese é verdadeira. No caso em estudo, há  $N = 39$  amostras de pares, e para o teste uni-direcional com um nível de significância estatística  $\alpha = 0,054$ , o valor crítico para o número de sinais positivos é  $r_c = 14$ . Em outras palavras, no teste uni-direcional a hipótese de que a população consiste de 50% ou mais sinais positivos, com nível de significância  $\alpha = 0,054$ , é rejeitada se 14 ou menos sinais positivos são observados na amostra. Semelhantemente, no teste uni-direcional a hipótese de que a população consiste de 50% ou menos sinais positivos, com nível de significância  $\alpha = 0,054$ , é rejeitada se 14 ou menos sinais menos (mais que 24 sinais positivos) são observados na amostra.

No nosso estudo, para a diferença entre as concentrações de mães e recém-nascidos, encontrou-se 10 sinais positivos ( $r$ ) e 27 sinais negativos. Portanto, rejeita-se a hipótese de que, com nível de significância  $\alpha = 0,054$ , 50% ou mais da população tem concentrações de anticorpos de mães maiores que os recém-nascidos. Porém, não se rejeita a hipótese de que, com nível de significância  $\alpha = 0,054$ , 50% ou menos da população tem concentrações de anticorpos de mães menores que os recém-nascidos. Portanto, como o valor de  $r$  é pequeno (pelo menos é menor do que o seu valor crítico), rejeita-se a hipótese nula  $H_0$ .

O teste de sinais permite fazer modificações, e uma delas é sobre a questão de quanto as concentrações de anticorpos de mães são menores que as de recém-nascidos. Para tanto, adiciona-se um valor igual a todos os valores das concentrações de mães. Neste caso, o valor mínimo que se adiciona ao grupo das mães para que a diferença não seja estatisticamente significante é  $C_m = 6,7 \text{ U.I./ml}$ . Com esta modificação nos dados amostrais, tem-se 15 sinais mais, 23 sinais menos, o que não rejeita a hipótese nula  $H_0$ . Este acréscimo gerou um par com concentrações iguais entre mãe e recém-nascido. Portanto, em média, as concentrações de mães são  $6,7 \text{ U.I./ml}$  menores do que de recém-nascidos.

### 3.1.2. Regressão linear

Apresenta-se a amostra com 39 pares de concentrações dos grupos de mães e recém-nascidos em forma de gráfico. Esta amostra é caracterizada por uma diferença de concentração entre os pares considerados. Mostra-se, agora, o ajuste linear dos dados, de acordo com a equação (2.2), para um primeiro estudo paramétrico.

A regressão linear, em que se ajusta apenas o coeficiente angular (ou de regressão) da reta, é mostrada na Figura 1. O valor ajustado para o coeficiente de regressão  $k_1$  é  $1,08 \pm 0,04$  (adimensional), sendo, assim, a equação entre as concentrações dada por

$[RN] = 1,08 \times [M]$ . Com o valor do parâmetro obtido do ajuste determina-se o intervalo de confiança para o coeficiente de regressão linear. Com 95% de significância, o coeficiente de regressão  $k_1$  encontra-se no intervalo entre os valores 1,001 e 1,167. Como o valor do extremo inferior é superior a unidade, conclue-se (com 95% de confiança) que a concentração de mães é menor do que de recém-nascidos, mesmo que a fonte de anticorpos dos recém-nascidos seja a mãe.

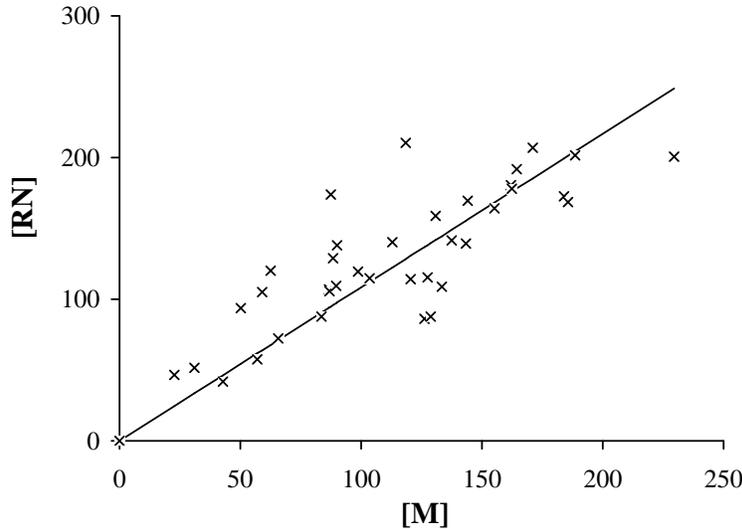


Figura 1: Regressão linear da concentração de IgG anti-rubéola em recém-nascidos  $[RN]$  em função da concentração em mães  $[M]$ . A marca  $\times$  refere-se aos dados observados.

Figura 1 mostra duas tendências (melhor vista na curva de distribuição de resíduos, não apresentada aqui): para valores baixos de concentrações de anticorpos em mães, os valores de recém-nascidos encontram-se acima da curva ajustada, enquanto para valores elevados, as concentrações de anticorpos em recém-nascidos ficam distribuídas em torno da curva. O teste de sinais (não-paramétrico) apenas revelou que há diferença entre os pares. O ajuste de curvas, um método paramétrico, mostrou não somente esta diferença, mas também em que região das concentrações esta diferença ocorre.

Esta concentração elevada entre os recém-nascidos em comparação com as mães, em regiões de baixa concentração de anticorpos nas mães, pode ser vista de outra forma. Ajusta-se a equação

$$[M] = k'_0 - k'_1 \frac{[RN]}{[M]},$$

onde  $k'_0$  e  $k'_1$  são valores positivos. Os parâmetros ajustados são  $k'_0 = 209,47 \pm 23,65$  U.I./ml e  $k'_1 = 77,72 \pm 18,56$  (adimensional). Figura 2 mostra este ajuste. O

coeficiente de regressão situa-se, com 95% de confiança, entre os valores 161,44 e 257,49.

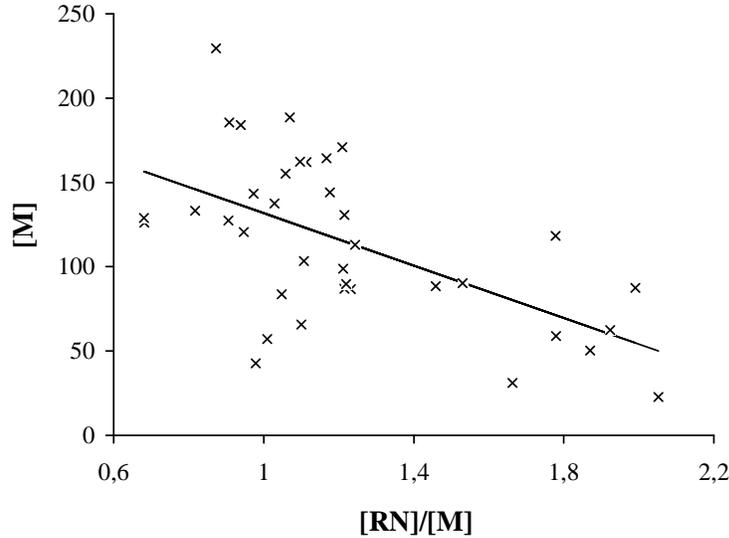


Figura 2: Regressão linear da concentração de IgG anti-rubéola em mães  $[M]$  em função da razão entre as concentrações de recém-nascidos e mães  $[RN]/[M]$ . A marca  $\times$  refere-se aos dados observados.

Figura 2 mostra que, em torno de  $[RN] = [M]$ , ocorre uma maior concentração de pares observados e, também, que o valor da correspondente concentração de anticorpos da mãe situa-se em torno de 125  $U.I./ml$ . A figura mostra que a maioria das mães que participaram do estudo repassaram pelo cordão umbilical um pouco mais ou um pouco menos da sua concentração de anticorpos para os recém-nascidos. A reta ajustada é monotonicamente decrescente, pois concentrações maiores de anticorpos em recém-nascidos ocorrem para baixas concentrações em mães. Por isso, usou-se o sinal menos na equação linear.

### 3.1.3. Testes paramétricos

Os dados amostrais são considerados observações pareadas, para detectar se existe um fator externo que possa causar diferença entre os grupos compostos por mães e recém-nascidos. Por outro lado, muitos outros fatores externos podem mascarar ou obscurecer a real diferença. Para detectar se há de fato um fator externo preponderante, usa-se as observações em pares, e calcula-se a diferença entre cada par. Ao se estudar levando em consideração esta diferença (por exemplo, a concentração de mãe menos a de recém-nascido), assume-se que todos os fatores externos afetam de forma idêntica os dois grupos, sendo assim, removidos pela diferença. O teste

de observações pareadas estuda a hipótese da igualdade de médias entre os dois grupos, isto é, as médias das concentrações de anticorpos nos grupos de mães e recém-nascidos são iguais.

O teste de sinais revelou que há diferença entre os pares estudados. O ajuste de curvas, da mesma maneira, mostrou que esta diferença existe, mas ocorre preferencialmente para baixas concentrações no grupo das mães. Fundamentado nisso, usa-se, agora, o teste paramétrico das diferenças, mesmo sabendo que os diferentes pares de observações são completamente diferentes. Dessa forma, estuda-se se há algum fator comum atuando em todos os pares, focalizando-se para a barreira placentária comum a todos os pares de mães e recém-nascidos, em vez de concentrar-se na diferença de seres.

Como os fatores externos podem ser removidos pela diferença entre os pares observados, testa-se a hipótese nula  $H_0$  de que a distribuição da diferença entre as duas populações tem média 0. Para este fim, usa-se o teste  $t$  com 38 graus de liberdade, pois tem-se 39 pares de observações. Para aplicar o teste  $t$ , calcula-se

$$t = \frac{d - 0}{s_d / \sqrt{39}},$$

com

$$\begin{cases} d = \frac{\sum_{i=1}^{39} (X_i - Y_i)}{39} \\ s_d^2 = \frac{\sum_{i=1}^{39} (X_i - Y_i)^2 - \frac{\left[ \sum_{i=1}^{39} (X_i - Y_i) \right]^2}{39}}{38}, \end{cases}$$

onde  $X_i$  e  $Y_i$  são, respectivamente, as concentrações de mãe e de recém-nascido na  $i$ -ésima observação pareada. Com os 39 pares observados, o valor obtido foi  $t = -3,20$ . Este valor é usado para rejeitar ou não a hipótese nula. Para nível de significância  $\alpha = 0,01$ , o valor de  $t$  de "Student" crítico com 38 graus de liberdade é  $t_{0,99}(38) \approx 2,429$  (entre os valores  $t_{0,99}(30) = 2,457$  e  $t_{0,99}(40) = 2,423$ ). A hipótese de igualdade de médias é rejeitada se  $t < -2,429$  ou  $t > 2,429$ . Como o valor calculado foi  $t = -3,20$ , rejeita-se a hipótese nula e aceita-se a hipótese de que as médias das concentrações entre os grupos de mães e recém-nascidos são diferentes.

Como se detectou uma diferença significativa entre as concentrações dos grupos de mães e recém-nascidos pelo método das observações pareadas, pode-se supor que existem fatores externos que de fato estejam causando esta diferença. Deve-se, então, tentar caracterizar quais são os possíveis fatores que estejam atuando na passagem de anticorpos maternos para o feto (considerando a barreira placentária, assumida para os pares a fim de usar o teste de observações pareadas), de tal forma que, ao nascer, este apresente em média uma concentração maior do que a da mãe, de quem herda os anticorpos. Pela regressão linear já constatamos que a diferença ocorre em baixas concentrações de anticorpos em mães.

De modo semelhante ao caso de teste de sinais, se acrescentarmos  $C_m = 8,8$  U.I./ml em todas as concentrações observadas entre as mães, o valor calculado

para este novo conjunto alterado de dados é  $t = -1,30$ . Para nível de significância  $\alpha = 0,10$ , o valor de  $t$  de “Student” crítico com 38 graus de liberdade é  $t_{0,90}(38) \approx 1,305$  (entre os valores  $t_{0,90}(30) = 1,310$  e  $t_{0,90}(40) = 1,303$ ). Portanto, não se rejeita mais a hipótese nula. O valor médio  $C_m$  encontrado neste teste paramétrico é ligeiramente maior do que o obtido no teste não-paramétrico ( $6,7 U.I./ml$ ). Este valor acrescentado ao grupo das mães, assim como o do teste de sinais, é escolhido dentro da amostra, de forma que ocorra um par com concentrações iguais e, com este valor acrescido, a diferença estatística entre os pares observados dos dois grupos deixa de existir.

Todos os testes já estudados mostraram que há diferença significativa entre as concentrações de anticorpos nos grupos de mães e recém-nascidos. Vamos estudar mais um teste paramétrico, o teste de análise de variância (ANOVA). Este teste fundamenta-se em uma separação de variâncias de todas as observações em partes, cada uma medindo variabilidade atribuída para alguma fonte específica, como variação interna das várias populações ou variações de uma população para outra. Contudo, o teste refere-se à comparação das médias entre várias populações, assumindo que as variâncias destas populações são iguais.

O teste de ANOVA compara o valor de  $F$  (teste de Fisher), obtido da razão entre as variâncias consideradas, com o valor crítico com nível de significância  $\alpha$ , designado  $F_{1-\alpha}$ . Se o valor de  $F$  for maior que o valor correspondente crítico  $F_{1-\alpha}$ , então rejeita-se a hipótese da igualdade de médias entre as populações.

No caso de classificação com 2 populações (mães e recém-nascidos), o teste de ANOVA é semelhante ao do teste de observações pareadas (teste paramétrico). No entanto, este teste considera os valores absolutos observados, e não a diferença, como no teste pareado, por exemplo. Portanto, faz-se o teste de ANOVA com 2 populações. Tabela 2 mostra a análise de variância. Para  $\alpha = 0,05$  nível de significância, os valores críticos são  $F_{0,95}(1,38) = 4,08$  e  $F_{0,95}(38,38) = 1,51$ . Portanto, há diferença entre as médias dos grupos de mães e recém-nascidos, pois  $F(1,38) > F_{0,95}(1,38)$ . Da mesma forma há diferença entre os pares de observações, pois  $F(38,38) > F_{0,95}(38,38)$ .

	Soma de quadrados	Gr. de liber.	Variância	Razão $F$
Colunas	4.290,5	1	4.290,5	$F(1,38) = 10,23$
Linhas	206.728,1	38	5.440,2	$F(38,38) = 12,97$
Entre	15.938,4	38	419,4	–
Total	226.957,0	77	–	–

Tabela 2. Análise de variância com 2 populações.

Mostra-se a motivação para aplicar a análise de variância, mesmo que, para duas populações, este teste fornece o mesmo resultado do teste  $t$  de “Student”. Os resultados anteriores mostraram que a diferença era grande para baixas concentrações de anticorpos em mães, mas praticamente o mesmo para concentrações elevadas. Vamos explorar este fato, realizando vários testes de ANOVA, e determinar o nível de concentração materna em que a diferença entre os pares deixa de existir. Para este fim, considera-se que mães e recém-nascidos pertençam ao mesmo grupo, usando-se, então ANOVA com uma única população. Esta classificação na mesma categoria

dos dois seres distintos fundamenta-se no fato de que, na hora do parto, o sangue nas mães e recém-nascidos é separado apenas pela barreira placentária (motivação para se usar teste de observação pareada).

Tabela 3 mostra a análise de variância considerando todos os dados amostrais em uma única categoria. Para  $\alpha = 0,05$  nível de significância, o valor crítico é  $F_{0,95}(1,76) \approx 3,95$  (entre os valores críticos  $F_{0,95}(1,60) = 4,00$  e  $F_{0,95}(1,120) = 3,92$ ). Portanto, não há diferença entre as médias dos grupos formados por mães e recém-nascidos, pois  $F(1,76) < F_{0,95}(1,76)$ . Quando os grupos (mãe e recém-nascido) e pares (diferentes mães e recém-nascidos) são agrupados, a diferença que existia entre os dois grupos deixa de existir.

	Soma de quadrados	Gr. de liber.	Variância	Razão $F$
Colunas	4.290,5	1	4.290,5	$F(1,76) = 1,4644$
Amostras	222.666,5	76	2.929,8	–
Total	226.957,0	77	–	–

Tabela 3. Análise de variância com 1 população.

Vamos, então, separar este grupo formado por mães e recém-nascidos em dois sub-grupos: Grupo I e Grupo II. No Grupo I são selecionados todos os dados amostrais até um certo valor da concentração de anticorpos da mãe. Conseqüentemente, no Grupo II estão os demais dados, cujas concentrações estão acima do valor considerado. A regressão linear forneceu um valor aproximado para determinar a concentração limiar em que ocorre a diferença entre os grupos de mães e recém-nascidos, que é  $125 \text{ U.I./ml}$ .

Na primeira simulação de sub-grupagem, escolhe-se para a concentração da mãe o valor  $128,8 \text{ U.I./ml}$  como o limite superior para pertencer ao Grupo I. Assim, o Grupo I é formado pelos pares de observações cujas concentrações de anticorpos nas mães  $[M]$  situam-se entre os valores 0 e  $128,8 \text{ U.I./ml}$ , enquanto o Grupo II é formado por mães com concentrações de anticorpos entre  $130,8$  e  $229,5 \text{ U.I./ml}$ . O Grupo I tem 25 pares de amostras, enquanto o Grupo II tem 14 pares. Note que  $130,8 \text{ U.I./ml}$  é o valor imediatamente superior na escala de observação de concentração materna.

Faz-se o teste de ANOVA com 1 população, porém dividida em dois grupos. Tabela 4 mostra a análise de variância para o Grupo I, e Tabela 5 mostra a análise de variância para o Grupo II. Para  $\alpha = 0,10$  nível de significância, os valores críticos são  $F_{0,90}(1,48) \approx 2,81$  (entre os valores críticos  $F_{0,90}(1,40) = 2,84$  e  $F_{0,90}(1,60) = 2,79$ ) e  $F_{0,90}(1,26) \approx 2,90$  (entre os valores críticos  $F_{0,90}(1,24) = 2,93$  e  $F_{0,90}(1,30) = 2,88$ ). Nesta sub-grupagem, tem-se  $F(1,48) < F_{0,90}(1,48)$  e  $F(1,26) < F_{0,90}(1,26)$ . Portanto, não há diferença significativa na média de concentrações de mães e recém-nascidos tanto no Grupo I quanto no Grupo II.

	Soma de quadrados	Gr. de liber.	Variância	Razão $F$
Colunas	4.751,31	1	4.751,31	$F(1,48) = 2,559$
Amostras	89.120,63	48	1.856,68	–
Total	93.871,94	49	–	–

Tabela 4. Análise de variância do Grupo I:  $[M]$  entre 0 e  $128,8 \text{ U.I./ml}$ .

	Soma de quadrados	Gr. de liber.	Variância	Razão $F$
Colunas	296,34	1	296,34	$F(1, 26) = 0,40$
Amostras	19.078,78	26	733,8	–
Total	19.375,12	27	–	–

Tabela 5. Análise de variância do Grupo II:  $[M]$  entre 130,8 e 229,5  $U.I./ml$ .

Como não há diferença, então deve-se proceder a uma nova sub-grupagem. Nesta simulação de sub-grupagem, o par observado de maior concentração materna é retirado do Grupo I e colocado no Grupo II. Assim, escolhe-se para a concentração da mãe o valor 127,5  $U.I./ml$  como o limite superior para pertencer ao Grupo I. Note que 127,5  $U.I./ml$  é o valor imediatamente inferior ao valor 128,8  $U.I./ml$  na escala de observação de concentração materna. Assim, o Grupo I é formado pelos pares de observações cujas concentrações de anticorpos nas mães  $[M]$  situam-se entre os valores 0 e 127,5  $U.I./ml$ , enquanto o Grupo II é formado por mães com concentrações de anticorpos entre 128,8 e 229,5  $U.I./ml$ . O Grupo I tem 24 pares de amostras, enquanto o Grupo II tem 15 pares.

Faz-se o teste de ANOVA com 1 população, porém dividida em dois novos grupos. Tabela 6 mostra a análise de variância para o Grupo I, e Tabela 7 mostra a análise de variância para o Grupo II. Para  $\alpha = 0,10$  nível de significância, os valores críticos são  $F_{0,90}(1, 46) \approx 2,82$  (entre os valores críticos  $F_{0,90}(1, 40) = 2,84$  e  $F_{0,90}(1, 60) = 2,79$ ) e  $F_{0,90}(1, 28) \approx 2,89$  (entre os valores críticos  $F_{0,90}(1, 24) = 2,93$  e  $F_{0,90}(1, 30) = 2,88$ ). Nesta nova sub-grupagem, tem-se  $F(1, 46) > F_{0,90}(1, 46)$  e  $F(1, 28) < F_{0,90}(1, 28)$ . Portanto, há diferença significativa na média de concentrações de mães e recém-nascidos no Grupo I, enquanto para o Grupo II aceita-se a igualdade de médias (não há diferença estatística).

	Soma de quadrados	Gr. de liber.	Variância	Razão $F$
Colunas	5.819,96	1	5.819,96	$F(1, 46) = 3,10$
Amostras	86.295,51	46	1.875,99	–
Total	92.115,47	47	–	–

Tabela 6. Análise de variância do Grupo I:  $[M]$  entre 0 e 127,5  $U.I./ml$ .

	Soma de quadrados	Gr. de liber.	Variância	Razão $F$
Colunas	83,18	1	83,18	$F(1, 28) = 0,09$
Amostras	26.567,95	28	948,86	–
Total	26.651,13	29	–	–

Tabela 7. Análise de variância do Grupo II:  $[M]$  entre 128,8 e 229,5  $U.I./ml$ .

O teste ANOVA permitiu determinar, estatisticamente, o valor da concentração em que a diferença entre os grupos formados por mães e recém-nascidos ocorre. Quando se divide os dois grupos em função do valor de concentração  $C^* = 127,5$   $U.I./ml$ , o sub-grupo com concentrações de mães menores que este valor apresenta diferença nas médias entre mães e recém-nascidos. Por isso, todo sub-grupo (Grupo I) que formar, tendo como limite este valor  $C^*$ , porém para valores máximos menores que  $C^*$ , terá valor de  $F$  tanto maior quanto menor for o valor máximo considerado. Em outras palavras, quanto menor for as concentrações das mães, maior será a

diferença entre as médias. Pode-se dizer que o fator externo que afeta a passagem de anticorpos pela barreira placentária não influencia mais quando a concentração de anticorpos na mãe estiver em torno de  $C^* = 127,5 \text{ U.I./ml}$ .

Levando em consideração todos os resultados acima, pode se assumir que um provável fator externo biológico é a passagem ativa dos anticorpos pela barreira placentária. A passagem de anticorpos pela barreira placentária é um fenômeno biológico. Como tal, há leis físicas e químicas que regulam esta passagem. A importância da análise estatística é detectar se há diferença na amostra, e em que condições esta diferença existe ou deixa de ocorrer. Baseado nos resultados estatísticos, pode-se fazer algumas conjecturas. Uma delas é que o valor de saturação para a passagem ativa de anticorpos pela barreira placentária situa-se em torno de  $C^* = 127,5 \text{ U.I./ml}$ . Assim, biologicamente, conjectura-se que a passagem ativa (com gasto energético) de anticorpos é um fenômeno físico preponderante para baixas concentrações nas mães, até o valor de sua saturação, sendo que acima deste valor, predomina a passagem passiva (difusão). Dessa forma, a passagem ativa modifica ligeiramente as concentrações no par mãe e recém-nascido, aumentando nos recém-nascidos para valores acima daqueles que seriam o esperado apenas pela passagem passiva.

Este estudo estatístico feito a partir de 39 dados amostrais fornece uma base teórica para construir um modelo matemático para estudar a passagem passiva e ativa de anticorpos pela barreira placentária. Um possível modelo matemático, levando em consideração difusão e reação, pode considerar esta barreira como uma membrana porosa e seletiva, e quantificar as concentrações na mãe e no recém-nascido ao longo da gravidez. Uma possível aplicação seria a determinação do período favorável para efetuar a imunização passiva do feto.

### 3.2. Estudo do decaimento do anticorpo materno

Analisou-se o processo de passagem de anticorpos maternos ao feto, em especial com os dados coletados do par no momento do parto. Estes anticorpos têm a finalidade de proteger o recém-nascido nos primeiros dias de sua vida, período em que é incapaz de produzi-los, pela imaturidade do sistema imunológico. Por isso, a proteção é maximizada pela passagem ativa de anticorpos maternos para o feto durante a gestação, sendo que este processo ativo com gasto energético deixa de ser notado para elevadas concentrações de anticorpos em mães, quando ocorre grande fluxo de anticorpos pela passagem passiva (difusiva). Entretanto, sendo proteínas séricas, os anticorpos são degradados, por isso conferem uma proteção passageira, atuando durante um curto período de tempo. Assim, para finalizar o estudo da passagem de anticorpos maternos ao feto, determina-se o período de tempo que estes anticorpos maternos herdados pelo feto podem proteger o recém-nascido.

A questão relativa ao período de tempo que estes anticorpos podem aferir proteção à criança recém-nascida é estudada baseada em uma amostra de 244 crianças menores de 2 anos. Os dados são resumidos na Tabela 8. Esta tabela mostra, por exemplo, que as 8 crianças entre 0 e 1 mês de idade apresentaram anticorpos (maternos) contra o vírus da rubéola.

Primeiro, estes dados são ajustados considerando-se a função dada pela equação (2.4). Mas, quando se aplica o logaritmo, esta função, através das mudanças dadas pela equação (2.5), torna-se linear, igual àquela dada pela equação (2.2). Assim, ajusta-se os dados observados transformados em logaritmo, de acordo com  $\ln [C(a) + 1]$ , pelo método de regressão linear. Figura 3 mostra este ajuste, com os parâmetros ajustados sendo  $k_0 = 1,76 \pm 0,10 \ln (U.I./ml)$  e  $k_1 = -0,42 \pm 0,01 \text{ meses}^{-1}$ .

Idade	Total	Positivos	Proporção	Intervalo de Confiança
0-1	8	8	1	0,63-1
1-2	8	7	0,875	0,47-1
2-3	11	7	0,636	0,31-0,89
3-4	11	6	0,545	0,23-0,83
4-5	11	5	0,455	0,16-0,75
5-6	13	2	0,154	0-0,38
6-7	15	1	0,067	0-0,19
7-8	26	2	0,077	0-0,17
8-9	22	0	0	0-0,15
9-10	14	0	0	0-0,23
10-11	23	0	0	0-0,15
11-12	23	2	0,087	0-0,20
12-24	59	3	0,051	0-0,11

Tabela 8. Soroprevalência da rubéola em crianças menores de 2 anos, estratificada por faixa etária, em um total de 244 crianças, em Caieiras, 1990.

Figura 4 mostra o ajuste da Figura 3 tomando-se os anti-logaritmos. Os parâmetros ajustados são transformados (inversamente) de acordo com a equação (2.5), ou seja,  $C_0 = e^{k_0}$  e  $K = -k_1$ , retomando, assim, a função exponencial dada pela equação (2.4). Deste ajuste, obtém-se a meia vida  $a_{1/2} = 1,64 \text{ meses}$  e a idade de decaimento total de anticorpos maternos  $a_T = 6,03 \text{ meses}$ , calculadas, respectivamente, pelas equações (2.7) e (2.8).

Contudo, os dados da Tabela 8 podem ser ajustados usando a equação (2.9), pelo método de máxima verossimilhança, dada pela equação (2.10). Figura 5 mostra este ajuste, com os parâmetros ajustados sendo  $c_0 = -0,46 \pm 0,10$ ,  $c_1 = 1,06 \pm 0,41 \text{ meses}^{-1}$ ,  $c_2 = -0,04 \pm 0,08 \text{ meses}^{-2}$  e  $c_3 = 0,001 \pm 0,003 \text{ meses}^{-3}$ . A curva ajustada é semelhante à da Figura 4, pois os dois últimos parâmetros podem ser considerados nulos, devido aos erros serem maiores que os valores dos parâmetros. Deste ajuste, obtém-se a meia vida  $a_{1/2} = 1,25 \text{ meses}$  (da curva) e a idade de decaimento total de anticorpos maternos  $a_T = 3,45 \text{ meses}$ , calculadas, respectivamente, pelas equações (2.7) e (2.8).

Os dois ajustes dos dados de soroprevalência em crianças menores de 2 anos mostraram dois aspectos. Primeiro, o decaimento de anticorpos maternos segue um padrão exponencial, isto é, a queda é proporcional à concentração, cuja constante de proporcionalidade não se altera com a idade. Por isso, a curva logística, que considerou heterogeneidade, quando ajustada aos dados, resultou em uma função exponencial (coeficientes  $c_2$  e  $c_3$  nulos). Segundo, a perda de anticorpos maternos

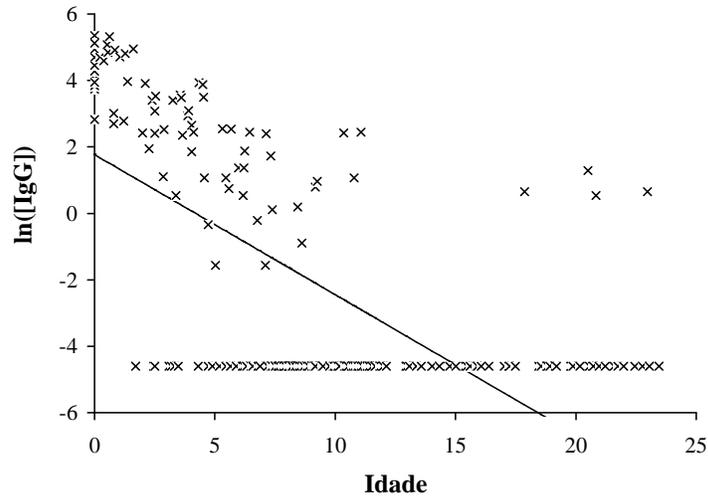


Figura 3: Ajuste por mínimos quadrados do logaritmo da concentração de IgG,  $\ln([IgG])$ , anti-rubéola em crianças em função de idade (em meses). A marca  $\times$  refere-se aos dados observados.

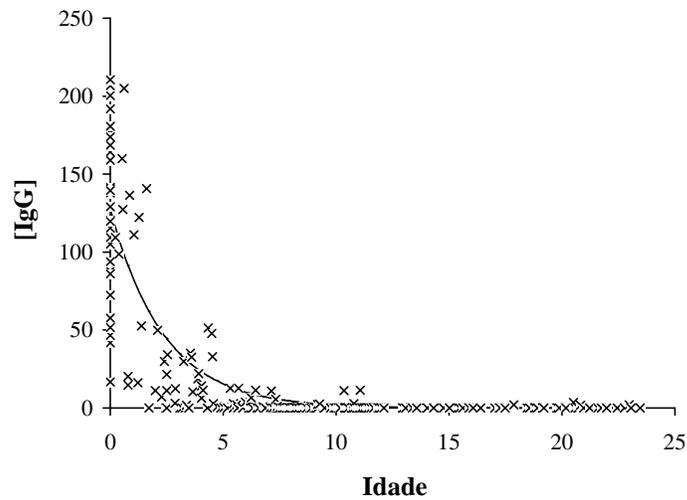


Figura 4: Concentração de IgG anti-rubéola em crianças  $[IgG]$  em função de idade (em meses). A marca  $\times$  refere-se aos dados observados.

ocorre por volta de 6 meses e, assim, quando baseado nestes dados, percebe-se que a comunidade estudada pode e deve ser vacinada contra a rubéola em torno de 6 meses.

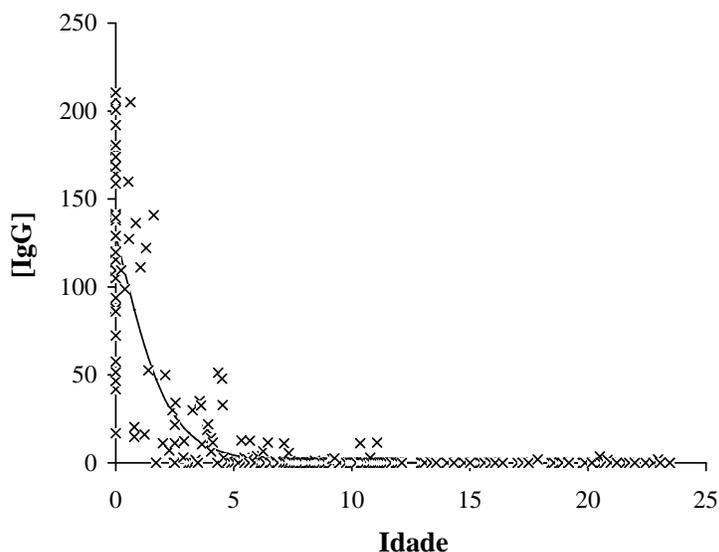


Figura 5: Ajuste por máxima verossimilhança da concentração de IgG anti-rubéola em crianças  $[IgG]$  em função de idade (em meses). A marca  $\times$  refere-se aos dados observados.

Mas, como a presença de anticorpos maternos pode afetar na real estimulação do sistema imunológico da criança pela introdução de vírus atenuado, seja pela imaturidade, seja pelos anticorpos maternos que imediatamente ligam-se aos antígenos, falta estudar a soroconversão vacinal para que a vacinação aos 6 meses de idade seja eficiente e plenamente justificável.

Portanto, para determinar a questão da idade mínima para vacinação em crianças contra a rubéola, faz-se estudo da soroconversão vacinal. Os dados de soroconversão vacinal estão resumidos na Tabela 9. Esta tabela mostra que, por exemplo, as 7 crianças vacinadas contra o vírus da rubéola entre 0 e 1 mês de idade não apresentaram soroconversão, ou seja, a vacina não teve eficácia entre elas. À medida que a idade aumenta, aumentou-se crianças que soroconverteram após receberem a vacina contra a rubéola, sendo que acima de 6 meses quase todas as crianças responderam positivamente à vacinação, soroconvertendo-se.

Os dados da Tabela 9 são ajustados usando a equação (2.9), pelo método de máxima verossimilhança, dada pela equação (2.10). Figura 6, uma curva sigmóide, mostra este ajuste, com os parâmetros ajustados sendo  $c_0 = 3,46$ ,  $c_1 = -1,23 \text{ meses}^{-1}$ ,  $c_2 = 0,07 \text{ meses}^{-2}$  e  $c_3 = -0,01 \text{ meses}^{-3}$ . Desta figura observa-se que 50% das crianças soroconverteram em até 3,53 meses, 93,1% soroconverteram em

até 9 meses e 95,3%, em até os 15 meses.

Idade	Total	Soroconversão	Proporção
0-1	7	0	0
1-2	7	2	0,43
2-3	7	3	0,5
3-4	9	3	0,33
4-5	7	3	0,62
5-6	9	7	0,82
6-7	13	13	1
7-8	22	19	0,91
8-9	21	21	1
9-10	8	7	0,89
10-11	20	20	1
11-12	17	17	1
12-24	49	49	1

Tabela 9. Soroconversão pós-vacinal contra rubéola em crianças menores de 2 anos, estratificada por faixas etárias, em um total de 196 crianças, em Caieiras, 1991.

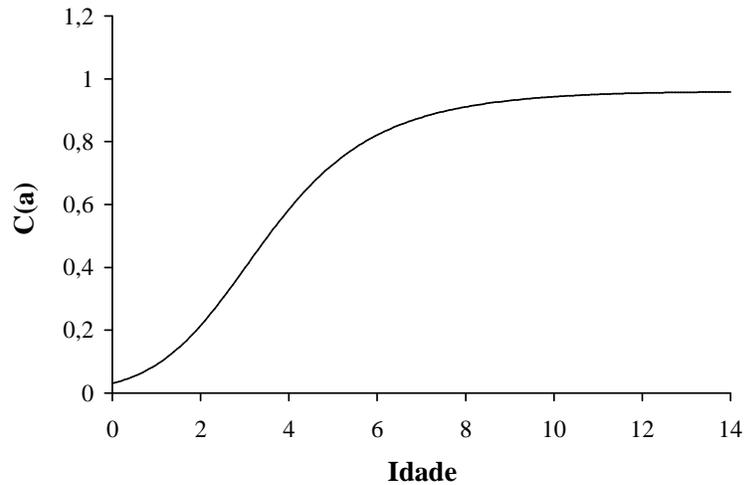


Figura 6: Ajuste por máxima verossimilhança de soroconversão pós-vacinal  $C(a)$  em função de idade (em meses). A marca  $\times$  refere-se aos dados observados.

A curva de soroconversão foi ajustada por uma função logística. Diferentemente da curva de decaimento de anticorpos maternos, observa-se uma dependência com outros fatores que desvia de um comportamento Poissônico (distribuição exponencial). O decaimento de anticorpos maternos é um processo biológico em que uma

dada concentração de imunoglobulinas herdadas decaem pela degradação, que ocorre proporcionalmente à sua concentração. No entanto, a soroconversão não é um fenômeno linear, mas o resultante da interação do binômio anticorpo materno e estimulação antigênica do sistema imunológico. Por isso, a curva de soroconversão foge do padrão exponencial e segue uma forma sigmóide.

A presença de anticorpos maternos protege o recém-nascido, pois este não apresenta ainda o sistema imunológico já amadurecido. No entanto, a presença de anticorpos maternos pode, também, prejudicar o estímulo do sistema imunológico para produção de anticorpos pela criança. Por isso, a relação entre período de decaimento e a idade de soroconversão vacinal deve ser tomada com todo o cuidado. Baseado neste estudo, a vacinação aos 9 meses de idade, atualmente praticada, é bastante razoável, pois nesta idade a maioria das crianças perderam o anticorpo materno e quase todas elas soroconverteram pela vacina. Porém, os resultados mostram que esta idade pode ser abaxada. Mas, para propor este novo esquema de vacinação (6 meses), novos estudos epidemiológicos devem ser feitos para dar mais peso a esta nova prática.

## 4. Discussão

Em nosso estudo, o resultado da soronegatividade ao rubivírus encontrada nas 39 parturientes foi de 5,1%. Os trabalhos de soronegatividade ao rubivírus mostram variabilidade dos resultados nas mulheres em idade fértil, cujo risco médio estimado é de 20% em adquirir a rubéola [15] [16] [20] [23].

Em 1964, Sever *et al.* [65] realizaram um estudo soroepidemiológico multicêntrico anti-rubivírus em 12 hospitais universitários nos EUA (era pré-vacinal). O resultado da soronegatividade encontrada em 600 amostras de soro de mulheres grávidas foi de 17,5%. A faixa de variação da soronegatividade foi de 30% em New Orleans a 8% em Mineapolis, sendo que 25,2% das mulheres soronegativas encontravam-se na faixa de 14 a 19 anos. Cockburn [20], segundo estudo realizado em 40 países, encontrou, nas ilhas Jamaica, Trindade, Havai e Tailândia, a soropositividade entre 34 a 57% para a faixa etária entre 23 e 29 anos. Gil [33] investigou, em 1983 pelo método da inibição da hemaglutinação (I.H.), um total de 1.600 alunas escolares cubanas da rede pública, cuja faixa etária compreendia garotas de 12 a 14 anos, e encontrou 36% de soronegatividade.

Noah *et al.* [54], em um estudo prospectivo de sorovigilância contra o vírus da rubéola em mulheres na faixa de 15 – 44 anos em 1988, encontraram, em uma amostra de 1.031.115 soros de mulheres grávidas, 21,8% de soronegatividade. A faixa etária das mulheres grávidas que apresentou a menor suscetibilidade (3,1%) ficou entre 25 – 29 anos, devendo estas baixas taxas à implementação de vacinas no período pós-natal para as mulheres que não tinham sido vacinadas no período escolar. Durante o período de estudo foram, também, realizados testes sorológicos AVR em 237.974 soros de mulheres não grávidas, que iriam se submeter à vacinação, encontrando-se 18,1% de suscetibilidade.

A preocupação com o aumento do número de mulheres soronegativas e, como

conseqüência, um incremento na possibilidade do nascimento de crianças com a SRC, fez com que muitos autores alertassem para a necessidade de um rastreamento sorológico das adolescentes no pré-natal, a despeito de terem tido rubéola ou terem sido vacinadas, bem como o seguimento de mulheres com títulos baixos durante toda a gestação [36] [46] [74].

Em estudo semelhante ao nosso, Massoud [45], em 1991 na Arábia Saudita, estudou 182 amostras do sangue da mãe e do recém-nascido contra o rubivírus, pelo método ELISA; encontrou 15,4% de soronegatividade ao IgG. O autor faz referências a outros trabalhos realizados em seu país, cuja faixa de soronegatividade em mulheres variou de 4,2 a 29%. Nakao [53], utilizando-se do método da inibição da hemaglutinação em 1993, encontrou, na cidade de Bauru, São Paulo, durante o ano de 1987, média de 9,7% (7,4 a 12%) de soronegatividade entre 689 gestantes estudadas, sendo que 84,8% das mulheres suscetíveis encontravam-se entre 15 – 29 anos, a mesma faixa das mulheres suscetíveis que nós encontramos.

Os baixos valores de soronegatividade encontrados nas parturientes da amostra estudada talvez possam indicar uma alta circulação do vírus selvagem na comunidade, onde os casos de rubéola apareceriam em faixas etárias mais jovens. A imunidade após a rubéola é duradoura, e se tem demonstrado a persistência de anticorpos, em período acima de trinta anos, produzidos pela doença. Azevedo Neto *et al.* [6] determinaram, na comunidade em estudo, a idade média de aquisição da rubéola, que se situa em torno de 7 anos e 4 meses, cujo valor está de acordo com os trabalhos epidemiológicos em populações não vacinadas.

Atualmente, é consenso que a aquisição da imunidade passiva do feto é mediada no útero pela placenta, que transporta IgG da mãe para o feto. Os estudos do intercâmbio materno-fetal em relação aos anticorpos específicos são escassos na literatura médica, principalmente em relação à rubéola, ocorrendo maior número de publicações com outras doenças como o sarampo e malária [10] [21] [26] [38] [42] [51]. Portanto, os dados obtidos e analisados aqui são de grande importância.

O nosso estudo, em uma amostra de IgG-AVR em 39 pares M-RN, mostra que existe uma diferença estatística entre as concentrações da mãe e do recém-nascido, havendo uma maior concentração neste último. Sugere-se, assim, a existência de um transporte ativo pela placenta de IgG contra o vírus da rubéola. Pelo método da análise de variância univariada com um fator, determinamos que o valor crítico máximo para ocorrer transporte ativo situa-se em torno de 128 *U.I./ml*, sendo que, abaixo deste valor, haveria um incremento nos recém-nascidos devido ao transporte ativo. Entretanto, acima daquele valor, o transporte ativo não afetaria na concentração entre recém-nascidos, pois o fluxo pelo transporte passivo seria grande, como evidenciado pela ausência de diferença estatística nas concentrações entre pares M-RN. Neste sentido, Gitlin [31], em 1974, estabeleceu que, quando os níveis maternos de imunoglobulinas totais (Ig) não-específicas para a rubéola aumentassem até cerca de 200 *mg%*, haveria um rápido incremento no seu transporte transplacentário; quando aumentassem até 800 *mg%*, decresceriam os valores transportados e, quando as cifras aumentassem acima de 800 *mg%*, os níveis de Ig eram proporcionais aos níveis maternos.

Martelli e Stangalini [43], em um estudo sorológico de AVR em 1971, pelo método

da imunodifusão simples radial, encontraram maiores títulos de concentrações de anticorpos no sangue do cordão que no sangue materno. Sato *et al.* [63] estudaram 28 amostras de pares M–RN em 1979, encontrando títulos de anticorpos ao vírus do sarampo, caxumba e rubeóla maiores no sangue do cordão do recém-nascido que na mãe, sendo que a proporção média no sangue do cordão do RN–M para rubeóla era de  $1,2 \pm 1,0$ .

A correlação negativa significativa existente em nossa amostragem, de 39 pares M–RN, entre a IgG cordão–mãe para o IgG materno ( $y = 209,47 - 77,72x$ , com intercepto em  $x = 2,70$ ) mostra que, para as taxas elevadas de IgG materno, a taxa de IgG do cordão–mãe é inferior às das mães, apontando para a validade da hipótese de Brambell sobre a passagem de IgG. A hipótese postula que os números de receptores placentários para IgG são os limitantes para a transferência pela placenta [14]. Michaux [49], estudando uma amostra de pares RN–M da população de mães suíças em 1966, encontrou concentrações médias de IgG total mais baixas ( $10,67 \pm 0,28 \text{ mg/ml}$ ) no sangue destas do que nas mães negras bantus ( $19,86 \pm 0,41 \text{ mg/ml}$ ). Os níveis séricos no sangue do cordão dos recém-nascidos mostraram-se mais elevados nas mães suíças.

É consenso da literatura que há na placenta receptores específicos para diversas proteínas transportadas, e a transferência de proteínas faz-se de maneira altamente seletiva, requerendo uma certa disponibilidade de receptores placentários, estando mais relacionada com a configuração molecular do que com o tamanho molecular. A região Fc da molécula de IgG é o elemento funcional de reconhecimento específico neste processo [11] [13] [29] [30] [42]. A IgG encontrada no cordão do recém-nascido é praticamente, na sua totalidade, de origem materna, embora haja pequenos traços endógenos de síntese de IgG. A IgG pode ser identificada como materna pela estrutura antigênica e ausência de marcadores alotípicos paternos (tipo Gm).

A transferência seletiva de IgG da circulação materna para o feto tem sido demonstrada “in vivo”, usando proteínas marcadas com radioisótopos [27]; “in vitro”, usando perfusão da placenta [56] e níveis de Ig no sangue do cordão [28]. Embora exista uma boa passagem seletiva pela placenta de anticorpos maternos do rubivírus (IgG1 e IgG3) para o recém-nascido, principalmente no último trimestre da gravidez, o seu rápido decréscimo durante o primeiro mês de vida, os mecanismos de como se processa esta passagem e a relação quantitativa da IgG entre mãe e recém-nascido são assuntos complexos e controversos [11] [40] [62].

Gitlin [29], em 1971, refere-se à existência de dois mecanismos que medeiam a passagem transplacentária de IgG. O primeiro mecanismo é de ordem cinética, ou seja, a simples transferência passiva na qual o nível fetal é proporcional às concentrações de IgG materna. O segundo mecanismo está relacionado com um processo ativo enzimático no qual a IgG é transferida, ativamente, da mãe para o feto, que, direta ou indiretamente, pode ser inibido pelas altas concentrações de IgG no soro materno. Para baixas concentrações maternas, o sistema é ativado e, proporcionalmente, mais IgG é transferida da mãe para o feto. No caso inverso, haveria uma inibição do processo com conseqüente decréscimo da passagem de IgG.

Barr *et al.* [8], em 1949, encontraram títulos séricos mais elevados de antitoxina diftérica no cordão umbilical de recém-nascido que no soro materno. Wiener

[70], em 1951 estudando títulos séricos de recém-nascido com eritroblastose fetal, encontrou valores próximos aos títulos maternos de anticorpos anti-Rh de mães sensibilizadas. Kohler *et al.* [40], em 1966, encontraram títulos em média 1,2 vezes maior de IgG no sangue do cordão que na mãe, apontando, assim, para um transporte ativo transplacentário. Toivanen *et al.* [68], em 1968, encontraram títulos de anticorpos no soro de cordão de recém-nascidos significativamente mais elevados (ocasionalmente de quatro a oito vezes) que nos títulos maternos em seis dos nove vírus estudados. Os anticorpos que tiveram um aumento nas concentrações dos recém-nascidos incluíram sarampo, influenza A, adenovírus, influenza 1 e 2 e herpesvírus. O anticorpo contra a rubéola não fez parte deste estudo. Logie [42], em um estudo realizado em três países da África em 1973, encontrou resultados semelhantes, reforçando a hipótese do transporte ativo proposto por ele e por Gitlin [29].

Black *et al.* [10], em 1986 em um estudo multicêntrico envolvendo cinco comunidades, África do Sul e New Haven – EUA (nestes dois locais, os nascidos antes e no ano de 1957 e nascidos após 1958), Índia e Taiwan, sobre a variação geográfica da perda de anticorpos maternos anti-sarampo e a prevalência de anticorpos anti-rubéola, demonstraram que os títulos de anticorpos anti-sarampo nos pares M–RN apresentaram uma diferença dos títulos maternos em relação ao cordão de 0,26 a 0,74 vezes a favor da concentração do recém-nascido nas quatro primeiras comunidades, enquanto em Taiwan houve uma diferença negativa de 0,52. Atribuíram esta diferença devido a um eficiente transporte de IgG. Não foi estudada nestes pares a correlação de IgG–AVR entre mãe e recém-nascido.

Gendrel *et al.* [26], em 1988, encontraram resultados semelhantes em amostras pareadas M–RN de IgG anti-sarampo em três tipos diferentes de populações: mães parisienses, mães africanas do Gabão e mães deste país que residiam em Paris. Encontraram uma correlação negativa significativa entre a relação IgG cordão–mãe versus IgG materna nas mães africanas ( $y = 40,2 - 20,5x$ , com intercepto em  $x = 1,96$ ) como nas mães parisienses ( $y = 14,2 - 3,6x$ , com intercepto em  $x = 3,94$ ). Os valores médios de IgG no sangue do cordão dos recém-nascidos das mães parisienses eram 1,34 vezes maiores que nas das mães africanas, lançando a hipótese de que esta diferença se deve ao elevado estímulo antigênico a que estão submetidas as mães africanas. Por viverem em área de alta endemicidade para doenças infecto-contagiosas, como a malária, as mães africanas têm os seus receptores placentários ocupados por outros antígenos, o que poderia diminuir a passagem dos anticorpos anti-sarampo para o recém-nascido. Os autores acreditam que se trata de um fenômeno relacionado às características do meio-ambiente e não étnico, pois as mães do Gabão que residem em Paris têm o mesmo comportamento sorológico que as mães parisienses.

Apesar de não termos encontrado dados na literatura quanto à passagem transplacentária de anticorpos contra a rubéola nos últimos 25 anos, os estudos realizados com outros tipos de vírus reforçam nossos resultados observados, fortalecendo a hipótese da existência de um mecanismo ativo de transporte de anticorpos entre a M–RN, pelo menos no que se refere à população estudada.

A meia vida da concentração sérica IgG–AVR foi estabelecida, levando-se em

consideração as distribuições etárias mensais, calculando-se a idade decimal de cada criança em função do dia da coleta do material. O decaimento de anticorpos AVR pelo método da regressão linear (dados transformados por logaritmo) estimou uma meia vida de 1,64 meses (49 dias), o que determinou o desaparecimento praticamente total dos anticorpos em torno dos 6,03 (181 dias) meses. Pelo ajuste de máxima verossimilhança (dados transformados em frações menores que um), demonstrou-se uma meia vida de 1,25 meses (37,5 dias), e correspondente desaparecimento quase total aos 3,45 meses (103 dias).

As curvas observadas por ambos os métodos mostram uma queda drástica dos AVR nos primeiros meses, seguida de um decaimento mensal, conseqüente à perda dos anticorpos recebidos de mães que haviam tido contato com o vírus da rubéola. Estes dados permitem concluir que o desaparecimento dos AVR, pelo método imunoenzimático, esteja situado entre 4 e 6 meses, idade a partir da qual poderíamos aplicar a vacina contra a rubéola, o que não conflita com os dados da literatura [4] [55]. Por motivo de maior segurança, o melhor momento seria o valor superior deste intervalo, 6 meses, o que possibilitaria uma imunização quase total das crianças. Os dados de soroconversão mostram que a partir do 6<sup>o</sup> mês ocorrem soroconversão vacinal na quase totalidade das crianças. Informações precisas de como ocorre o declínio pós-natal dos AVR são importantes, pois os anticorpos maternos na infância promovem proteção contra as infecções que, por sua vez, podem afetar a eficácia da vacinação, especialmente aquelas referentes às vacinas de vírus vivos e ser um fator na patogênese de certas infecções virais, segundo as publicações disponíveis na literatura [2] [4] [11] [10] [19] [69].

O desaparecimento da imunidade é reflexo do catabolismo de IgG, estando a vida média (valor superior que a meia vida, usada neste trabalho) influenciada por diferentes fatores como a nutrição materna, níveis séricos iniciais da concentração IgG materna, números de receptores placentários livres, idade gestacional do recém-nascido, peso ao nascimento, renda per capita, diferenças geográficas e fatores do meio ambiente [10] [13] [14] [19] [26] [42]. Embora muitos estudos da cinética do catabolismo de anticorpos tenham sido realizados em adultos, é sabido que a vida média dos anticorpos em crianças normais é de 20 – 40 dias, enquanto crianças portadoras de agamaglobulinemia, que receberam, passivamente, gamaglobulina, tiveram uma vida média de 50 – 60 dias [19] [70].

Enders-Ruckle [24], em 1985 utilizando-se da I.H., encontrou um decréscimo de anticorpo AVR de origem materna durante o primeiro ano, caindo de 90% na idade de 0 a 2 meses para 16% na de 9 a 11 meses. Em 1989, Hossain [37] realizou um estudo soropidemiológico contra o vírus da rubéola na Arábia Saudita, através de método ELISA, em uma população estratificada por idade de 0 a 35 anos, e encontrou 100% de soropositividade nas crianças de 0 a 6 meses, seguido por uma queda de 39,3% no período de 7 – 12 meses, demonstrando a perda gradual de AVR, sendo que as mais baixas concentrações foram obtidas na faixa de 13 – 18 meses.

## 5. Conclusão

Os estudos estatísticos e os valores estimados da amostra de sangue coletada no Brasil permitem concluir que:

1. Há uma diferença significativa entre as concentrações de anticorpos antivírus da rubéola entre as mães e os recém-nascidos. Este achado corrobora com a existência de um transporte ativo transplacentário, que favoreceria uma concentração de anticorpos proporcionalmente maior nos recém-nascidos de mães com título de anticorpos abaixo do nível crítico, que se situa entre os valores 127,5 e 128,8 *U.I./ml*.
2. Mais de 90% das crianças, no nascimento, possuem anticorpos anti-rubivírus. Entretanto, observa-se um decaimento rápido da concentração de anticorpos maternos logo nos primeiros meses de vida. Considerando, também, a soroconversão vacinal, conjectura-se que a vacina contra a rubéola pode ser aplicada em crianças na idade de 6 meses.
3. Novos estudos devem ser realizados para podermos responder questões práticas sobre a prevenção por vacina em mulheres grávidas a fim de reduzir as sérias infecções neonatais.

**Abstract.** Rubella is a very important public health problem in Brazil. Serological surveys are the unique available information for the development of immunization programs. In order to establish the serological anti-rubivirus profile, the age dependent seroconversion rates and the relationship between the antibodies concentration of the pair mother  $[M]$  and newborn  $[RN]$ , a prospective study, were carried out in Caieiras, a small town in the outskirts of São Paulo City, Brazil, from November 18, 1990 to January 18, 1991. A quantitative ELISA was used to determine the antibodies concentration against the rubivirus in 244 children comprised on 0 – 2 years of age, as well as in 39 pairs M and RN. The seroconversion study was done with 196 children from the sample. More than 90% of children in this community born with antibodies against rubivirus, with a fast decaying in the first semester of life (complete decaying from 3.45 to 6.03 months). Age dependent maternal antibody decaying study was done by fitting data using maximum likelihood method. The vaccine seroconversion was found to be 50% at 3.53 months and 93.1% at 9 months of age, which is greater than 80% from the fifth month onwards. The M and RN pairs were analyzed by simple linear regression, non-parametric sign test, one and two way analysis of variance and analysis in pairs. With respect to the analysis of the pairs M and RN, linear regression showed the relation  $[RN] = 1.084 \times [M]$  and the signs test indicated a significant difference between pairs. The analysis of variance showed that a critical maximum concentration for M,  $[M]$ , is located on the range  $127.5 < [M] < 128.8 \text{ I.U./ml}$ . The observation of a significant difference between pairs M and RN suggests an active antibody transplacental transport when  $[M]$  is below  $127.5 \text{ I.U./ml}$ .

## Referências

- [1] A.C.P. Task Force on Adult Immunization and Infectious Diseases Society of America, Immunization for special group of patients, em “Guide for Adult Immunization” (American College of Physicians, ed.), 2<sup>nd</sup> ed., Philadelphia, Pa., 1990.
- [2] P.A. Albrecht, F.A. Ennis, E.J. Saltzman e S. Krugman, Persistence of maternal antibody in infants beyond 12 months: mechanism of measles vaccine failure, *J. Pediatr.*, **91** (5) (1977), 715-718.
- [3] M.S. Amstey, R. Insel, J. Muñoz e M. Pichiero, Fetal-neonatal passive immunization against Hemophilus influenza type b, *Am. J. Obstet. Gynecol.*, **153** (1985), 607-611.
- [4] R.M. Anderson e R.M. May, “Infectious Diseases of Human: Dynamics and Control”, Oxford University Press, London, 1991.
- [5] A.S.H.P. Annual Meeting 98, CDC identifies immunizations needed by health care workers, *Am. J. Health Syst. Pharm.*, **55** (1998), 761-764.
- [6] R.S. Azevedo Neto, A.S.B. Silveira, D.J. Nokes, H.M. Yang, S.D. Passos, *et al.*, Rubella seroepidemiology in a non-immunized population of São Paulo, Brazil, *Epidemiol. Infect.*, **113** (1994), 161-173.
- [7] K.J. Bart, W.A. Orenstein, S.R. Preblud e A.R. Hinman, Universal immunization to interrupt rubella, *Rev. Infect. Dis.*, **7** (1985), 77-84.
- [8] M. Barr, A.T. Glenney e K.J. Randall, Concentration of diphtheria antitoxin in cord blood, *Lancet*, **2** (1949), 324-6.
- [9] S.S. Baski e L.Z. Cooper, Rubella and mumps vaccines, *Ped. Clin. N. Am.*, **37** (1990), 651-68.
- [10] F.L. Black, L.L. Berman, J.M. Borogño, *et al.*, Geographic variation in infant loss of maternal measles antibody and in prevalence of rubella antibody, *Am. J. Epidemiol.*; **124** (1986), 442-52.
- [11] J.A. Bellanti, Basic immunologic principles underlying vaccination procedures, *Ped. Clin. North. Am.*, **37** (1990), 513-30.
- [12] A.L. Bittencourt, “Infecções Congênitas Transplacentárias”, Ed. Revinter, Rio de Janeiro, 1995.
- [13] F.W.R. Brambell, W.A. Hemmings e I.G. Morris, A theoretical model of g-Globulin catabolism, *Nature*, **203** (1964), 1352-1355.
- [14] R.W.R. Brambell, The transmission of passive immunity from mother to young and catabolism of immunoglobulins, *Lancet*, **2** (473) (1966), 1087-1093.

- [15] CDC, Rubella vaccine, *Morbid. Mortal. Week Rep.*, **27** (1978),451-454.
- [16] CDC, Immunization of health-care workers: recommendations of the Advisory Committee on Immunization Practices (ACIP) and the Hospital Infection Control Practices Advisory Committee (HICPAC), *Morbid. Mortal. Week Rep.*, **46** (RR-18) (1997), 1-42.
- [17] B. Christenson e Bottiger, Long-term follow-up study of rubella antibodies in naturally immune and vaccinated young adults, *Vaccine*, **12** (1) (1994), 41-45.
- [18] H. Claman, "The Immunology of Human Pregnancy", Humana Pressa, New Jersey, 1993.
- [19] M.J. Cloonan, R.A. Hawker e L.H. Stevens, Postnatal decline of maternally acquired rubella antibodies, *J. Hyg. Camb.*, **68** (1970), 461-468.
- [20] W.C. Cockburn, World aspects of the epidemiology of rubella, *Am. J. Dis.*, **118** (1) (1969), 112-122.
- [21] F. Dabis, R. Waldman, G. Mann, D. Commenges, G. Maadzou e S. Jones, Loss of maternal measles antibody during infancy in a African city, *Int. J. Epidemiol.*, **18** (1989), 264.
- [22] W.J. Dixon e F.J. Massei Jr., "Introduction to Statistical Analysis" 4<sup>th</sup> ed., McGraw-Hill Inc., New York, 1983.
- [23] W.R. Dowdle, W. Ferreira, L.F. de Salles Gomes, *et al.*, WHO collaborative study on the sero-epidemiology of rubella in Caribbean and Middle and South American population in 1968, *Bull. WHO*, **42** (1970), 419-422.
- [24] G. Enders-Ruckle, Rubella antibodies titers in vaccinated and non vaccinated women and results of vaccination during pregnancy, *Rev. Infect. Dis.*, **7** (1985), 103-107.
- [25] R.G. Faix, Maternal immunization to prevent fetal and neonatal infection, *Clin. Obstet. Gynecol.*, **34** (2) (1991), 277-287.
- [26] D. Gendrel, D. Richard-Lenoble e A. Fribourg-Blanc, Transfert des immunoglobulines et des anticorps antirougeoleux de la mère à l'enfant en Afrique et en Europe, *Press méd.*, **32** (17) (1988), 1633-1636.
- [27] D. Gitlin, J. Kumate, J. Urrusti e C. Moraes, The selectivity of human placenta in the transfer of plasma proteins from mothers to fetus, *J. Clin. Inv.*, **43** (1964), 1938-1951.
- [28] D. Gitlin e A. Biasucci, Development of gG, gA, gM, b1A/b1C, C'1 esterase inhibitor, ceruloplasmin, transferrin, hemopexin, haptoglobin, fibrinogen, plasminogen,  $\mu$ 1-antitrypsin, orosomucoid, b-lipoprotein,  $\mu$ 2- macroglobulin and prealbumin in the human conceptus, *J. Clin. Inv.*, **43** (1969), 1433-1438.

- [29] D. Gitlin, Development and metabolism of the immune globulins, em “Immunologic incompetence” (B.M. Kagan e R. Stiehm, eds.), pp. 3-13, Year Book Medical Publishers, Chicago, 1971.
- [30] D.J. Gitlin e D. Gitlin, Cell receptors and selective transfer of proteins from mother to young across tissue barriers, *Ped. Research*, **7** (1973), 290.
- [31] D.J. Gitlin e D. Gitlin, Protein binding by specific receptors on human placenta, murine placenta and sucking murine intestine in relation to protein transport across these tissues, *J. Clin. Invest.*, **54** (1974), 1155-1166.
- [32] N.M. Gregg, Congenital cataract following German measles in the mother, *Trans. Ophthalmol. Soc. Aust.*, **3** (1941), 35-46.
- [33] T.J. Gill, C.F. Repertti, L.A. Metlay, *et al.*, Transplacental immunization of the human fetus to tetanus by immunization of the mother, *J. Clin. Invest.*, **72** (1983), 987-996.
- [34] W.P. Glessen, J.A. Englund, G.R. Siber, *et al.*, Maternal immunization with the capsular polysaccharide vaccine for Haemophilus influenza type B, *J. Infect. Dis.*, **165** (Suppl. 1) (1992),: 134-136.
- [35] J.R. Hobbs e J.A. Davis, Serum gamma globulin levels and gestacional age in premature babies, *Lancet*, **3** (1967), 757-759.
- [36] L. Hornstein, U. Levy e A. Fogel, Clinical rubella with vírus transmission to the fetus in a pregnant woman considered to be immunne, *New. England J. Med.*, **319** (21) (1988), 1415-1416.
- [37] A. Hossain, Seroepidemiology of rubella in Saudi Arabia, *J. Trop. Ped.*, **35** (1989), 169-170.
- [38] P.M. Johnson e P.J. Brown, Review article: Fcg receptors in the human placenta, *Placenta*, **2** (1981), 355-370.
- [39] J.P. Krause, G. Gross, T. Barret, *et al.*, Quality standard for assurance of measles immunity among helath care workers, *Clin. Infect. Dis.*, **18** (1994), 431-436.
- [40] P.F. Kohler e R.S. Farr, Elevation of cord over maternal IgG immunoglobulin: evidence for an active placental IgG transport, *Nature*, **210** (1966), 1070-1071.
- [41] D.B. Lewis e C.B. Wilson, Developmental immunology and roles of defenses in neonatal susceptibility to infection, em “Remington and Klein Infectious Diseases of the Fetus and Newborn Infant”, 4<sup>th</sup> ed., W.B. Saunders Company, 1995.
- [42] D.E. Logie, I.A. McGregor, D.S. Rows e W.Z. Billewicz, Plasma immunoglobulin concentration in mothers and newborn children with special referece to placental malaria, *Bull. Wld. Hlth. Org.*, **49** (1973), 547-554.

- [43] A. Martelli e A. Stangalini, Anticorpi antirusolia e immunoglobuline nel sangue materno e fetale, *G. Bacteriol. Virol. Immuno.*, **64** (9) (1971), 391-396.
- [44] E. Massad, R.S. Azevedo neto, M.N. Burattini, D.M.T. Zanetta, F.A. Coutinho, *et al.*, Assessing the efficacy of a mixed vaccination strategy against rubella in São Paulo, Brazil, *Intern. J. Epidemiol.*, **24** (4) (1995), 121-132.
- [45] M. Massosud, M. El Tagui e W. Saleh, Specific rubella virus IgG and IgM in umbilical cord sera in Saudi Arabia, *J. Egyptian Public Health Association*, **66** (3 e 4) (1991), 387-395.
- [46] A. Mathur, R. Tripathi, V.C. Chaturvedi e P. Mehra, Congenital rubella following inapparent rubella infection, *Indian J. Med. Res.*, **75** (1982), 469-473.
- [47] M.A. Menser, D.C. Dorman, K.G. Kendrick, G.P. Stuart, *et al.*, Congenital rubella: Long-term follow-up study, *Am. J. Dis. Child.*, **118** (1969), 32-34.
- [48] H.W. Mendenhall, Serum protein concentration in pregnancy, *Amer. J. Obstet. Gynec.*, **106** (5) (1970), 718-720.
- [49] J.L. Michaux, J.F. Heremans e W.H. Hitzig, Immunoglobulin levels in cord-blood serum of negroes e caucasians, *Trop. Geog. Med.*, **18** (1966), 10-18.
- [50] E. Miller, J.E. Cradock-Watson e T.M. Pollack, Consequences of confirmed rubella at successive stages of pregnancy, *Lancet*, **1** (1982), 871-874.
- [51] H. McFarlane e A.U.O. Williams, Distribution of malarial antibody in maternal and cord sera, *Arch. Dis. Child.*, **44** (1969), 511-514.
- [52] M.I. Moraes-Pinto, R.S. Oruamabo, F.P. Igbagiri, *et al.*, Antibodies, *J. I. D.*, **171** (1995), 1076-1077.
- [53] N. Nakao, "A Suscetibilidade à Rubéola das Gestantes em Bauru, 1987", Tese de Doutorado, FSP, USP, São Paulo, SP, 1993.
- [54] N.D. Noah e S.E. Fowle, Immunity to rubella in women of childbearing age in the United Kingdom, *B. M. J.*, **297** (1988), 1301-1304.
- [55] D.J. Nokes, R.M. Anderson e M.J. Anderson, Rubella epidemiology in South East England, *J. Hyg.*, **86** (1986), 291-304.
- [56] H. Nöschel, *et al.*, Plazentapassage von natriumsalizylat, em "Fetal and neonatal physiology" (R. Polin e W.W. Fox, eds.), Vol. 1, pp. 130-141, W.B. Saunders Company, 1992.
- [57] M.A. Nowak e R.M. May, "Virus Dynamics. Mathematical Principles of Immunology and Virology," Oxford University Press, Oxford, 2000.
- [58] S.D. Passos, "Rubéola: Perfil Sorológico em Crianças Menores de Dois Anos no Município de Caieiras, SP," Tese de Doutorado, FM, USP, São Paulo, SP, 1996.

- [59] S.D. Passos, Diagnóstico de rubéola em pacientes com exantema, *Pediatrics S.P.*, **22 (3)** (2000), 271-271.
- [60] R.W. Pitcher-Wilmott, P. Hindocha e C.B.S. Wood, The placental transfer of immunoglobulin and subclasses in human pregnancy, *Clin. Exp. Immunol.*, **41** (1980), 303-308.
- [61] F.N. Rasheed, Maternal infections influence infection susceptibility in childhood, *Med. Hypoth.*, **42** (1993), 76-80.
- [62] W.E. Regelman, E.L. Mills e P.G. Quie, Immunology of the newborn, em "Textbook of pediatric infectious diseases" (R.D. Feigin e J.D. Cherry, eds.), pp. 921-939, 2<sup>nd</sup> ed., W.B. Saunders Company, 1987.
- [63] H. Sato, P. Albrechet, D.W. Reynolds, *et al.*, Transfer of measles, mumps, and rubella antibodies from mother to infant, *Am. J. Dis. Child.*, **133** (1979), 1240-1243.
- [64] H.G. Schatzmayr, Aspects of rubella infections in Brazil, *Rev. Infect. Dis.*, **7** (Suppl. 1) (1985), 53-55.
- [65] J.H. Sever, G.M. Schiff e R. Huelmer, Frequency of rubella antibody among pregnant woman and other human and animal population, *Obst. Gynecol.*, **23** (1964), 153-159.
- [66] V.A.A.U.F. Souza, J.C. Moraes, L.M. Sumita, *et al.*, Prevalence of rubella antibodies in a non-immunized urban population, São Paulo, *Rev. Inst. Med. Trop. São Paulo*, **34 (4)** (1994), 373-376.
- [67] R.E. Stiehm e H.H. Fundenberg, Serum levels of immune globulins in health and disease: a survey, *Pediatrics*, **37** (1966), 715-726.
- [68] P. Toivanen, Mantyljarvio e T. Hurvonen, Maternal antibodies in human foetal sera at different stages of gestation, *Immunology*, **15** (1968), 395-303.
- [69] B.O. Vahlquist, The transfer of antibodies from mother to offspring, *Adv. Pediatrics*, **10** (1958), 305-338.
- [70] A.S. Wiener, The half-life of passively acquired antibody globulin molecules in infants, *J. Exp. Med.*, **94** (1951), 213-221.
- [71] W.W. Williams, Guidelines for infection control in hospital personnel, *Infect. Control*, **4** (1983), 326-348.
- [72] W.W. Williams, S.R. Preblud, P.O.S. Reichelderfer, *et al.*, Vaccine of importance in the hospital setting. Problems and developments, *Infect. Dis. Clin. North Am.*, **4** (1990), 701-722.
- [73] H.M. Yang, "Epidemiologia Matemática: Estudos dos Efeitos da Vacinação em Doenças de Transmissão Direta", EDUNICAMP e FAPESP, Campinas e São Paulo, 2001.

- [74] M.J. Yelland, Rubella immune status in general practice, *Aust. Family Physician*, **18 (10)** (1989), 1279-1283.

## Uma Abordagem Matemática em Imunologia

S. BARROZO<sup>1</sup>, Departamento de Matemática, Faculdade de Ciências, UNESP, Av. Luiz Edmundo Carrijo Coube s/n, CEP: 17033-360, Bauru, SP, Brasil

H.M. YANG<sup>2</sup>, Departamento de Matemática Aplicada, Instituto de Matemática, Estatística e Computação Científica, UNICAMP, CP: 6065, CEP: 13081-970, Campinas, SP, Brasil

C.H. DEZZOTI<sup>3</sup>, Departamento de Matemática, Centro de Ciências Exatas e da Terra, UFRN, Rua Frei Jaboatão, 280, ap. 601 – bloco S, CEP: 50710-030, Recife, PE, Brasil

**Resumo.** O estudo e a compreensão do sistema imunológico foram fortemente impulsionados nas duas últimas décadas como parte essencial do esforço de pesquisa sobre a AIDS. O conhecimento resultante, todavia, deverá ter um impacto muito mais abrangente, em particular e de modo especial, no aperfeiçoamento das estratégias de tratamento e prevenção das doenças infecciosas classicamente estudadas pela epidemiologia. Uma pergunta é naturalmente posta: pode a matemática colaborar de alguma forma para este avanço? As pesquisas em imunologia poderão tirar muitas vantagens desta interação com a matemática, ainda que a matemática não possa resolver diretamente todos os problemas da imunologia. Modelos e simulações de aspectos e comportamentos particulares do sistema imunológico poderão reduzir a grande quantidade de experimentos terapêuticos e darem suporte para o progresso da imunologia teórica. Assim, este trabalho tem por objetivo promover uma ampla discussão acerca da contribuição da matemática para o avanço da imunologia, através da elaboração de uma síntese dos modelos matemáticos descritos na literatura, acompanhada de uma análise de aspectos matemáticos e fisiológicos de cada um deles, e um modelo simples que propomos neste sentido.

### 1. Introdução

A Imunologia é definida, no dicionário Houaiss da Língua Portuguesa, como especialidade biomédica que estuda o conjunto dos mecanismos de defesa do organismo contra antígenos (qualquer substância capaz de induzir uma resposta imunológica). Além de estudar a natureza e a prevenção de doenças, esta modalidade biomédica

---

<sup>1</sup>sbarrozo@fc.unesp.br

<sup>2</sup>hyunyang@ime.unicamp.br; apoio financeiro FAPESP e CNPq.

<sup>3</sup>dezotti@ccet.ufrn.br

também busca tratar o câncer e outras doenças que resultam da falha ou mau funcionamento da resposta imunológica (RI), as imunopatologias.

O conjunto das células e substâncias responsáveis pela defesa do organismo, bem como suas respectivas ações, é denominado sistema imunológico (SI). O sistema imunológico dos seres mais complexos, onde está incluída a classe dos mamíferos, caracteriza-se por ser um sistema especializado, descentralizado e inteligente. A especialização diz respeito ao fato de que durante toda a nossa existência somos submetidos à invasão dos mais diversos tipos de patógenos, cada um deles com sua forma específica de invasão, despistamento e reprodução. Além dos ataques externos, nosso organismo está constantemente produzindo células cancerígenas, as quais também devem ser eliminadas. O SI humano deve ser capaz de debelar todas essas ações, possuindo elementos e ações adequadas para responder diferentes tipos de agressões. Outra característica é ser um sistema de ação descentralizado. Na RI não se observa a ação de um gerenciador central que determina as suas várias etapas, ou seja, ela se encontra efetivada em diversos pontos do organismo, como nos órgãos linfáticos periféricos estrategicamente alocados nas várias portas de entrada utilizadas pelos invasores, onde cada um promovendo uma resposta imunológica completa consegue uma resposta rápida à invasão. Muitas vezes, a diferença entre a sobrevivência e a morte encontra-se na rapidez com que um invasor é debelado. A memória imunológica permite que o SI responda com maior rapidez a um invasor, uma vez que já se tenha tido contato com o mesmo. Essa peculiaridade baseia-se no fato de que, uma vez devidamente exposto a um dado patógeno, o SI tem a capacidade de guardar entre seus milhares de linfócitos (divididos em dois grupos, células B, produzidas na medula óssea, e T, produzidas no timo) alguma adaptação direcionada àquele invasor. Temos então o que se convencionou denominar sistema inteligente.

Um outro aspecto relevante da RI diz respeito às doenças conhecidas como auto-imunes, ou seja, doenças onde o SI responde a antígenos do próprio organismo. Sucintamente, o auto-reconhecimento não é geneticamente determinado, mas resulta de um aprendizado das células do SI durante o estágio embrionário do organismo. Influências genéticas e/ou gatilhos ambientais podem quebrar a regra imunológica usual, e então as células B e T passam a responder contra antígenos do próprio corpo. Temos uma série de doenças com esta característica: anemia hemolítica (anticorpos atacam as células vermelhas do sangue), miastenia grave (anticorpos atacam uma proteína vital para que as células musculares recebam sinais dos nervos), diabetes dependente de insulina (as células T destroem a insulina produzida pelo pâncreas) e esclerose múltipla (as células T atacam o isolamento das fibras nervosas no cérebro e na medula dorsal). Na maioria das vezes, os tratamentos de doenças auto-imunes fazem uso de drogas imunodepressoras. O problema de tal tratamento é que ele suprime também a resposta imunológica desejada (por exemplo, contra patógenos), causando ao paciente uma série de infecções oportunistas que podem mesmo levá-lo à morte.

No caso do câncer, o ataque do SI a células do próprio corpo pode ter um efeito benéfico. Uma teoria controversa, enunciada por Lewis Thomas, pleiteava que o dever primário do constante patrulhamento dos linfócitos seria eliminar células pré-

cancerosas. É preciso lembrar que cânceres droga-associados resultam de outros mecanismos que não a imunossupressão. Contudo, em experimentos com ratos foi notado que embora cânceres provocados por viroses ou substâncias cancerígenas provoquem forte resposta imunológica, cânceres espontâneos provocam pouca ou nenhuma resposta imunológica.

Outro caso diverso, em que a ação do sistema imunológico é essencial, diz respeito à rejeição aos transplantes. Nesse caso, o SI, ao reconhecer um tecido e/ou órgão estranho, desencadeia uma série de reações, que muitas vezes resulta em um mau desfecho da intervenção. Neste caso temos também o uso de drogas imunodepressoras com seus efeitos colaterais.

A composição e o funcionamento do sistema imunológico envolvem grande complexidade e sofisticação, integrando continuamente vários tipos de células com um objetivo específico e comum. O sistema imunológico é dividido em dois sistemas principais: o primeiro é o sistema imunológico inato – inato no sentido de que o corpo nasce com a habilidade de reconhecer certos micróbios imediatamente e destruí-los –, e o segundo é o sistema imunológico adaptativo, no qual os anticorpos ocupam um papel fundamental. Nosso sistema imunológico inato pode destruir muitos patógenos no primeiro encontro, porém a imunidade inata não pode proteger contra todas as infecções. Aqueles micróbios que se desenvolvem muito rapidamente não podem ser combatidas unicamente pelas defesas inatas do organismo, as quais, comparativamente, desenvolvem-se lentamente. O sistema imunológico adaptativo habilita o organismo a reconhecer e a responder contra qualquer tipo de micróbio, mesmo que nunca tenha tido contato prévio. Este sistema compreende dois tipos de respostas: uma, dita imunidade humoral e a outra, imunidade celular. A primeira ocorre através da ação de anticorpos, os quais são produzidos por linfócitos denominados células B, as quais são formadas na medula óssea (*bone marrow*) e é particularmente importante no combate a patógenos circulantes na corrente sanguínea. A segunda envolve linfócitos que se originam no timo e por isso são chamados de células T. Estas são divididas em dois tipos, as células T-CD4 e T-CD8, estas últimas sendo chamadas de células T citotóxicas por destruírem células infectadas. As células T citotóxicas são fundamentais para combater infecções virais.

Vemos que, desde a sua formação, constituição e especialização, até a sua maneira de atuar, o SI possui diferentes peculiaridades e muitas delas passíveis de serem modeladas matematicamente. O estudo de modelos matemáticos que procuram explicar algum aspecto do funcionamento do SI, ou ressaltar algum comportamento, o qual pode não ser observado claramente em experimentos, pode trazer grande benefício à própria imunologia. Além disso, tal estudo pode ser interessante quando pensamos na sua influência em outros ramos da ciência médica, como por exemplo, a Epidemiologia (que estuda os diferentes fatores que intervêm na difusão e propagação de doenças, sua freqüência, seu modo de distribuição, sua evolução e a colocação dos meios necessários para a sua prevenção).

Com o intuito de contribuir na investigação deste complexo sistema e das possibilidades de sua compreensão através da modelagem, apresentaremos uma introdução biológica da composição do sistema imunológico a fim de facilitar sua compreensão. Em seguida, apresentamos uma breve revisão de modelos matemáticos

encontrados na literatura que tratam do estudo do sistema imunológico, em especial, adaptativo. Posteriormente, apresentamos e discutimos um modelo que trata a resposta imunológica mediada por células, apresentando a dinâmica de produção de anticorpos mediante a presença de um determinado antígeno. Finalmente, apresentamos algumas conclusões.

## 2. Biologia do sistema imunológico

A resposta imunológica se divide em dois tipos: a resposta inata e a adaptativa.

A resposta imunológica inata (RII) é basicamente efetivada pelos fagócitos e pelo sistema de complementos, que é uma classe de proteínas constituindo em um importante componente desta resposta. Seu nome vem da habilidade que possui em assistir ou complementar a atividade dos anticorpos no combate às infecções. A atividade dos complementos pode se dar de três maneiras: (a) um tipo de complemento, chamado C3, pode-se ligar a quaisquer proteínas, tais como aquelas das bactérias. Ocorrendo isto, a molécula C3 estimula outras moléculas de complementos para se ligarem à bactéria e destruí-la; (b) algumas bactérias são protegidas por uma cápsula de açúcar, os polissacarídeos. Estas cápsulas impedem a ação direta dos complementos sobre a bactéria. Os macrófagos, ao encontrarem estas bactérias, podem produzir interleucina-6, que é carregado pela corrente sanguínea até o fígado. O interleucina-6 instrui o fígado a produzir uma nova proteína que se liga a resíduos de açúcares chamada manose. A seguir, estas proteínas, ligadas às manoses, ligam-se à bactéria, resultando na mudança de forma para poder ativar os complementos, os quais se ligam à bactéria; e (c) as células B são ativadas quando encontram uma bactéria e são estimuladas, pelas células T-CD4 auxiliaadoras, a se proliferarem e produzirem anticorpos. Estes anticorpos se ligam à bactéria e ativam uma proteína de complemento chamada C1q, que ativa outras moléculas de complemento, as quais se ligam à bactéria. Em todos estes casos, os complementos podem matar as bactérias ou recrutar outras células do sistema imunológico, tais como macrófagos, para destruí-las. Ao contrário dos micróbios, nossas células são equipadas com proteínas que desativam os complementos, evitando, assim, o auto-ataque. A imunidade inata é inespecífica no sentido em que todos os componentes da mesma categoria são capazes de reconhecer o mesmo antígeno, e não produzem memória imunológica.

A resposta imunológica adaptativa (RIA) é efetivada pelos linfócitos B e T, e é específica no sentido de que cada célula possui um receptor capaz de reconhecer um único antígeno. Além disso, uma vez convenientemente estimulada, produz memória imunológica. Esta resposta divide-se em dois tipos: resposta imunológica humoral e resposta imunológica célula-mediada ou celular. A resposta imunológica humoral tem por base a ação das células B na produção de anticorpos e tem atuação eficaz contra invasores que vivem fora das células, no sangue ou em fluidos circundando as células. Os anticorpos são os produtos antígeno-específicos das células B, e sua produção na resposta contra infecções é a principal contribuição dos linfócitos B para a RIA. Coletivamente, os anticorpos formam uma família de proteínas plasmáticas,

conhecidas como imunoglobulinas, e sua ação efetora ocorre quando os mesmos atam-se ao patógeno, impedindo que ele, por exemplo, seja capaz de invadir uma célula, ou servindo como um sinalizador para a ação do sistema de complementos na sua eliminação. Também podem atar-se às toxinas produzidas pelo patógeno, impedindo que elas afetem o hospedeiro. No entanto, são ineficazes contra patógenos intra-celulares, uma vez que seus componentes não são lipossolúveis, estando assim impedidos de penetrar nas células.

A resposta imunológica célula-mediada (qualquer RIA na qual as células T antígeno-específicas desempenham um papel principal) age basicamente contra invasores intra-celulares e ocorre em duas etapas. Na primeira etapa faz-se necessário que o sistema imunológico reconheça a presença da infecção em uma determinada célula. Na segunda etapa, ocorre a mobilização para a contenção e posterior eliminação do agente patogênico. Os principais agentes deste tipo de resposta são as células T auxiliares (ou T-CD4) e as células T citotóxicas (ou T-CD8). Entre as auxiliares, dois tipos se distinguem, as puramente auxiliares, ou Th2, e as inflamatórias, ou Th1. As células do tipo Th1 são eficazes contra alguns tipos de bactérias (*Mycobacterium tuberculosis*, agente etiológico da tuberculose, *M. leprae*, causador da lepra, etc.), as quais crescem nas vesículas dos macrófagos e não podem ser eliminadas pelos lisossomas do mesmo. Neste caso, as Th1 funcionam como um disparador da ação efetora, ativando mecanismos de eliminação da própria célula infectada. Os linfócitos T citotóxicos, ou T-CD8, são eficientes contra os vírus. Tais linfócitos reconhecem as células do organismo infectadas pelo vírus e eliminam essas células antes que a replicação viral tenha se completado. O controle de infecções virais através da eliminação da célula infectada é altamente vantajoso. Se a intervenção for suficientemente rápida, as células infectadas podem ser destruídas antes mesmo que as partículas virais dentro delas sejam capazes de se reorganizarem completamente. Nessa fase, qualquer partícula do vírus que porventura seja liberada, não será capaz de infectar novas células, cessando a infecção antes mesmo que ela se propague.

Finalmente, as células Th2 atuam na eliminação de agentes extracelulares mediante a ativação dos linfócitos B. É importante ressaltar que apenas alguns antígenos dotados de propriedades particulares são capazes de ativar diretamente os linfócitos B. Na maioria das vezes, faz-se necessário um segundo sinal emitido pelas células Th2 para que as células B multipliquem e se diferenciem em células capazes de secretar anticorpos. Sendo assim, a ação das células Th2 é quase sempre essencial para que se desencadeie a resposta imunológica efetivada pelos linfócitos B. Passemos então a descrever uma ação completa do SI envolvendo a interação entre linfócitos B e Th2.

Um organismo ao invadir o corpo do hospedeiro pode-se deparar inicialmente com a primeira linha de defesa, o SI inato. Se seu antígeno é de alguma forma atado pelos receptores de um macrófago, este o engolfa e elimina. Alternativamente, o invasor pode estimular o sistema de complementos, e essa ligação dá início a uma série de liberação de substâncias que perfuram a membrana lipídica do patógeno levando-o a morte, ou então sinalizar o alvo que deve ser eliminado pelos fagócitos. Por fim, o patógeno pode possuir uma membrana que neutraliza a ação do sistema de

complementos e impossibilita os fagócitos de reconhecê-lo como estranho ao corpo. Quando isso ocorre, a ação da linha de defesa constituída pelas células B é iniciada.

Um antígeno circulando pela corrente sangüínea é capturado por células do SI chamadas apresentadoras de antígeno, cujas principais são as dentríticas (as células B também conseguem capturá-lo, porém em menor escala). Estas células apresentam o antígeno à uma célula T-CD4, que possui um mecanismo de checagem para verificar se o antígeno é mesmo um corpo estranho. Uma vez confirmada, a célula T se diferencia em Th1 ou Th2, conforme as características do antígeno. No caso de antígenos que necessitam de anticorpos para sua destruição, a célula T diferencia-se em Th2, e se torna ativada. Concomitantemente, uma célula B que depara também com este antígeno, vai capturá-lo. Esse antígeno é internalizado e sofre processos químicos de modo a ser cindido em fragmentos peptídicos. Esses fragmentos peptídicos são atados pelas moléculas transportadoras MHC-II e exibidos na superfície celular de B, que passa a ser chamada B apresentadora de antígeno. Neste tempo é produzida no interior da célula B que internalizou o antígeno, a proteína B7, que também é exibida na superfície da membrana celular de B. Uma célula Th2 ativada pelo mesmo antígeno liga-se então ao complexo MHC-II – fragmento peptídico e seu receptor CD28 ligado a B7. Essa dupla sinalização faz com que a célula Th2 libere substâncias que estimulam a célula B antígeno-específica a se dividir e produzir anticorpos. Seus clones não necessitam de estimulação e produzem o mesmo tipo de anticorpo que aquele da célula B que deu origem a eles. Todas essas células são efetoras, com exceção de algumas delas que são chamadas de células-memória (constituindo, após isso, em uma parte da memória do sistema adaptativo, e na possibilidade de uma posterior invasão pelo mesmo patógeno, a resposta imunológica dar-se-á de maneira mais eficiente e rápida). As células-memória não são efetoras. Uma vez que os anticorpos específicos livres sejam produzidos, estes atam-se ao antígeno do patógeno sinalizando aos sistemas de complementos e fagocitário qual o alvo que deve ser eliminado.

Um aspecto interessante da resposta imunológica adaptativa diz respeito à necessidade de um duplo sinal para que ela se efetive. Como uma resposta errada pode causar danos ao próprio hospedeiro (no caso, a ação pode ser contra alguma célula sadia ou substância própria, também sadia), sempre se faz necessária a ação conjunta de dois sinais para que a resposta imunológica adaptativa entre em ação.

### 3. Estado-de-arte da modelagem matemática

Muitos modelos matemáticos têm sido propostos e estudados com o intuito de contribuir na investigação do funcionamento do sistema imunológico. Será dado aqui um enfoque especial àqueles que descrevem o sistema imunológico adaptativo, com ênfase na interação entre os linfócitos B e T e a produção de anticorpos. Dentre estes, terão prioridades aqueles que usam equações diferenciais ordinárias (EDO) para descrevê-los, embora uma grande quantidade de modelos possam ser encontrados utilizando equações diferenciais parciais, processos estocásticos, autômatos celulares, redes, etc. Alan S. Perelson e colaboradores têm estudado uma grande

quantidade de modelos utilizando EDO, que vêm sendo publicados desde a década de 70 até os dias de hoje. Outros pesquisadores também têm se dedicado a este assunto. Assim, descreveremos aqui aqueles que julgamos mais relevantes para subsidiar os trabalhos futuros que venham a ser desenvolvidos nesta direção.

Em 1976, Perelson, Mirmirani e Oster desenvolveram um trabalho sobre estratégias ótimas em imunologia, que foi dividido em duas partes: o primeiro trata-se da diferenciação e proliferação das células B [17], enquanto o segundo, da produção das células B de memória [18]. A questão central colocada no primeiro trabalho era: qual a estratégia ótima disponível ao sistema imunológico para responder a um ataque antigênico? Para tentar responder a esta pergunta, elaboraram um modelo baseado na presença de determinado antígeno que estimula uma sub-população de linfócitos B em repouso a se diferenciar em linfócitos grandes, os quais se proliferam e começam a produzir anticorpos. Enquanto estão proliferando, os linfócitos podem produzir anticorpos, porém a uma taxa pequena, sendo que a maioria do aparato celular é dedicado à replicação. Todavia, um linfócito grande pode também se diferenciar em uma célula plasma, que é capaz de produzir anticorpos em uma taxa substancialmente grande, embora se divida em uma taxa muito pequena. A pergunta que surge é: qual a melhor estratégia para eliminar o antígeno o mais rápido possível? Este trabalho procurou encontrar um tempo ótimo de proliferação, seguido de diferenciação. Para isso, valeu-se de métodos da teoria de controle ótimo, em particular, o princípio do máximo de Pontryagin, para encontrar a estratégia ótima. No segundo trabalho estudaram a introdução das células de memória dos linfócitos B ao modelo anteriormente analisado. Foi assumido que os antígenos acionam uma sub-população de linfócitos B e estes linfócitos podem, ou se proliferarem e produzirem uma quantidade modesta de anticorpos, ou se diferenciarem em células plasma não divisíveis, as quais produzem uma grande quantidade de anticorpos, ou se diferenciarem em células de memória, as quais não produzem anticorpos.

Em 1993, Fishman e Perelson [9] apresentaram um modelo para a interação das células T, em repouso, com células apresentadoras de antígenos (APC), que resulta na ativação das células T. O modelo foi utilizado para examinar reações de hipersensibilidade tipo retardadas (DTH), que são típicas de imunidade celular. Foi mostrado que as respostas podem ser de dois tipos: ou os antígenos são completamente eliminados ou o sistema imunológico falha e o resultado é uma infecção crônica. Compararam dois tipos de células apresentadoras de antígeno e concluíram que as células dentríticas são mais adequadas para estimular as células T que os macrófagos. O modelo considerou que os antígenos, uma vez detectados e processados pelas APC, são apresentados às células T (neste caso, as auxiliadoras CD4). Estas são ativadas e se expandem por um processo clonal. Nesta fase recrutam, através de linfocinas, células fagocitárias e inflamatórias, que eliminam ou isolam a fonte de estimulação antigênica. Para atacar o problema de escalas presente em modelos como este (as reações de ativação e proliferação são da ordem de 8–16 horas, enquanto que a vida média dos linfócitos T é da ordem de várias semanas), o modelo foi adimensionalizado, tomando como unidade de tempo a vida média dos linfócitos T. Com esta nova escala, as interações entre as células ativadas

atingem um estado quase-estacionário, e esta suposição reduziu o modelo a um sistema de EDO envolvendo apenas três variáveis adimensionais, que representam os receptores sobre as APC, as células T e os antígenos. A quantificação da resposta imunológica, representada pelas células T ativadas, foi obtida explicitamente em função das variáveis do modelo. O sistema foi estudado em equilíbrio, possuindo dois estados estacionários: um trivial (ausência de antígenos) e o outro representando um estado de co-existência, no qual os antígenos persistem diante da resposta imunológica. Usando análise de estabilidade linear, mostraram que o equilíbrio trivial é instável diante de uma pequena perturbação de antígenos, enquanto que o de co-existência, é estável.

Em 1994, Fishman e Perelson [10] apresentaram um trabalho em que analisaram a regulação cruzada (*cross regulation*), entre os subconjuntos de células auxiliaadoras Th1 e Th2, durante a resposta imunológica. Fundamentaram suas hipóteses no fato de que, quando um patógeno é encontrado, o organismo geralmente responde ou com a resposta humoral (através da produção de anticorpos), ou com a resposta célula-mediada (resposta DTH), porém, nunca com ambas. E a possível razão pela qual as respostas humoral e celular são mutuamente exclusivas é que, durante a resposta imunológica, as citocinas regulatórias do tipo interferon- $\gamma$  (IFN- $\gamma$ ), que são produzidas pelas células Th1, podem inibir a proliferação das células Th2, enquanto que a produção de interleucina-10 (IL-10) pelas células Th2, inibem a produção de citocinas pelas células Th1. Com isso, é de se esperar que uma delas domine a outra. Baseado nisso, propuseram um modelo que considera as células apresentadoras de antígeno APC (células dendríticas, células B ou macrófagos), as células Th1 e Th2 em repouso e ativadas, Th2 não proliferativas e Th1 não secretoras, e a concentração de interferon- $\gamma$ , interleucina-10 e de antígenos. O modelo foi adimensionalizado nos moldes do anterior e, após todas as simplificações, culminou em um sistema de EDO cujas variáveis adimensionais representam a concentração de locais (receptores) sobre as APC (tanto locais ocupados quanto desocupados), a população de células Th1 e Th2 em repouso e a concentração de antígenos. A população de células Th1 e Th2 ativadas, a população de células Th2 inibidas de proliferar e de células Th1 não secretoras são dadas em termos das variáveis do modelo. O sistema apresentou quatro pontos de equilíbrio, representando a ausência de antígeno, chamado trivial ( $E_{00}$ ), a ausência de células Th2 ( $E_{01}$ ), a ausência de células Th1 ( $E_{10}$ ) e a co-existência de todas as variáveis ( $E_{11}$ ). O estudo da estabilidade mostrou que  $E_{00}$  é instável independente dos parâmetros e que  $E_{01}$  é estável na região de parâmetros onde  $E_{10}$  é instável e vice-versa. A região de estabilidade foi determinada pelo parâmetro que mede a eficiência relativa entre a interação das células Th1 e Th2 com as apresentadoras de antígenos, APC. Assim, se as células Th2 interagem mais eficientemente com as APC, então a resposta é dominada por elas, enquanto que a resposta é dominada por Th1 se sua eficiência de interação for maior. Estudos numéricos mostraram que  $E_{11}$  é instável e bifurca de  $E_{01}$  para  $E_{10}$ , sendo que o parâmetro de bifurcação foi determinado explicitamente através da análise da estabilidade de  $E_{01}$  e  $E_{10}$ . Através da resolução numérica do sistema, pôde-se simular sua ação em doenças auto-imunes e parasitárias, leishmaniose e AIDS.

Em 1996, Oprea e Perelson [15] estudaram um modelo que explorou os mecanismos da resposta primária aos antígenos dependentes das células T, considerando a produção de anticorpos. Foram analisados dois casos: um onde as células B, uma vez ativadas, proliferam-se sob o controle de citocinas, sem nova interação direta com células T ativadas e outro, assumindo que as células B, após se dividirem, retornam ao estado de repouso e necessitam de um novo contato com as células T ativadas para se dividirem novamente. Os resultados do primeiro caso foram bem mais próximos aos obtidos em experimentos biológicos do que os resultados do segundo caso. O modelo considerou o crescimento e diferenciação das células B em um único foco de desenvolvimento da resposta primária (foi escolhido o baço) e, assim, a concentração de células representava o número de células por foco. Não foi considerada a ativação das células T através da apresentação de antígenos pelas células B, uma vez que é pouco provável que isso ocorra, já que as células dendríticas, sendo antígeno-não-específicas, estão presentes em um número muito maior que as células B antígeno-específicas e, assim, podem encontrar as células T antígeno-específicas muito mais freqüentemente. As células B podem contribuir significativamente para a apresentação de antígenos às células T “experientes”, ou seja, aquelas que já tiveram contato com o antígeno através das células dendríticas. Quando as células T ativadas interagem com as células B apresentadoras de antígeno, estas podem ser ativadas. A concentração inicial de células dendríticas em um foco foi tomada como função da dose de antígeno injetada, do número total de células dendríticas expostas ao antígeno, e do número médio de focos no baço. Foi assumido que tanto as células dendríticas apresentadoras de antígenos, quanto os antígenos em um foco, decaem exponencialmente. Para a dinâmica das células, a exemplo dos modelos citados anteriormente, foi considerado que cada célula dendrítica possui  $N$  locais onde as células T podem se ligar, uma vez que uma única célula dendrítica pode ativar de 10 a 20 células T em um dia. O modelo foi representado por um conjunto de EDO, acopladas de acordo com a interação entre as variáveis e considerando a dinâmica das variáveis dos compartimentos acima, mais a dinâmica das células que migram para outros focos, a concentração de anticorpos e também a ação das citocinas desencadeadoras de todo o processo. Na dinâmica das células B foram consideradas duas possibilidades. Dada a sua complexidade, o modelo foi estudado numericamente, através da solução numérica das equações. Os resultados mostraram que no caso em que as células B, uma vez ativadas, se proliferam sob o controle de citocinas, sem nova interação direta com células T ativadas, os níveis de anticorpos e o tamanho dos focos são similares aos observados em experimentos, enquanto que o segundo caso, ou seja, assumindo que as células B, após se dividirem, retornam ao estado de repouso e necessitam de um novo contato com as células T ativadas para se dividirem novamente, o tamanho dos focos foi quatro vezes menor e os níveis de anticorpos, uma ordem de grandeza menor que os encontrados nos experimentos. Isso sugere que as células B podem estar aptas a se dividirem sem contatos posteriores com as células T, uma vez que a interação cognata tenha ocorrido.

Em 1999, Fishman e Perelson [11] propuseram um modelo baseado nas mesmas premissas do modelo estudado anteriormente por eles, que trata da regulação cruzada das células T, porém a diferença básica é que este novo modelo considerou

também a diferenciação das células T em repouso, ou seja, supôs que as células T em repouso adquirem um fenótipo Th1 ou Th2 após a interação com as células apresentadoras de antígeno. As interações consideradas foram: células dendríticas periféricas (pDC) detectam e fagocitam patógenos invasores e os convertem em estímulos antigênicos na forma de complexos MHC-II. Estas células apresentadoras de antígenos migram para os órgãos linfóides secundários, maturando. Lá encontram uma célula T antígeno-específica e a interação de ambas provoca a ativação da células T, e a conseqüente produção de fenótipos tipo Th1 ou Th2. As células Th1 ativadas produzem interleucina-2 (IL-2) e interferon- $\gamma$  (IFN- $\gamma$ ), enquanto que as Th2 ativadas produzem interleucina-4 (IL-4) e interleucina-10 (IL-10). IL-4 e IL-10 inibem a produção de citocinas pelas células Th1 ativadas, resultando em células que não produzem citocinas (não secretoras) e, conseqüentemente, não podem se proliferar na ausência de fatores externos. Por outro lado, IFN- $\gamma$  inibe a utilização de citocinas pelas células Th2 ativadas, induzindo-as a produzirem citocinas, mas não podem utilizá-las. Estas células são ditas não-proliferativas. Um sistema de EDO foi utilizado para representar o modelo, e após adimensionalização como feita nos trabalhos anteriores, culminou em um sistema de cinco equações, cujas variáveis adimensionais representam os receptores sobre as APC, as células T, Th1, Th2 e os antígenos. Análise numérica indicou que, dependendo dos parâmetros e das condições iniciais, o sistema converge para uma das seguintes soluções de equilíbrio: trivial, co-existência dominada por Th1 ou co-existência dominada por Th2. Isso mostra que a consideração sobre a diferenciação das células T, que foi acrescentada ao modelo, não o alterou qualitativamente.

Seguindo esta mesma linha, o modelo da ação das células imunológicas diante de um ataque antigênico sem considerar os efeitos epidemiológicos desta ação, vários outros trabalhos foram desenvolvidos concomitantemente a estes citados. Adomian, por exemplo, em 1996 [1], apresentou um modelo para a resposta imunológica considerando a população de patógenos, P (que podem ser vírus, bactérias, antígenos ou células tumorais), e a efetividade ou capacidade do sistema imunológico de eliminá-los. Foi introduzida uma função para descrever o processo de reações do sistema imunológico diante da presença de células externas (diferenciação) e uma função para medir a taxa de crescimento das células efectoras durante o processo (proliferação). O objetivo deste trabalho foi encontrar uma solução analítica não perturbativa do sistema de EDO que representa o modelo, o que foi feito através de um método de decomposição de sistemas diferenciais desenvolvido pelo mesmo autor anteriormente [2].

Em 1998, Segel e Bar-Or [22] desenvolveram um modelo considerando o patógeno (P), as células efectoras do sistema imunológico que combatem este patógeno (E), e a medida dos danos que a ação imunológica causa ao indivíduo (N). Posteriormente foram introduzidos também os danos causados pelos patógenos. O estudo numérico do modelo mostrou que a partir de um certo tempo, as variáveis apresentam um comportamento oscilatório, porém com quantidades muito pequenas de patógenos, o que foi considerado sem significado epidemiológico. Também foi levado em conta a organização espacial das células do sistema imunológico e estudados critérios de otimização na escolha das células efectoras.

Em 2000, Pilyugin e Antia [21] propuseram dois modelos para a imunidade célula-mediada, sendo um para a resposta imunológica não-específica, e neste caso foi considerada a ação dos macrófagos e/ou outras células fagocitárias, e outro para a resposta imunológica específica, mediada por linfócitos T citotóxicos (CTL). Em ambos foi considerado o que se chamou de “handling time” no termo da equação que representa a destruição do patógeno pelas células imunes. O “handling time” foi definido como sendo o tempo de interação entre uma célula imunológica e uma sinalizada por conter um antígeno, durante o qual a célula imunológica não pode interagir e/ou destruir outras células sinalizadas. Inicialmente desenvolveu-se uma aproximação em estados quase-estacionários, o que permitiu estudar como o “handling time” pode ser incorporado em modelos já existentes. Então examinou-se as consequências desta variação para a dinâmica de infecções e, em particular, a vulnerabilidade da resposta imunológica dos hospedeiros. Para a resposta imunológica não-específica foram consideradas células fagocitárias em repouso ou inativas, ativadas livres e ativadas engajadas, com a suposição de estar em estados quase-estacionários. A vulnerabilidade desta resposta pôde ser medida através do estudo da estabilidade das soluções de equilíbrio do sistema. Assim, dependendo dos parâmetros, o sistema pode ter: (a) apenas a solução de equilíbrio trivial  $E_0$  (ausência de patógenos), (b) a solução trivial e uma não-trivial  $E_1$ , e (c) a solução trivial e duas não-triviais  $E_1$  e  $E_2$ . No caso (a) a solução é instável, ou seja, nestas condições de parâmetros os patógenos nunca serão eliminados pela resposta imunológica. No caso (b),  $E_0$  é estável e  $E_1$  é instável, havendo uma região à esquerda de  $E_1$  onde as soluções são limitadas e convergem para  $E_0$  e uma região de escape à direita, onde as soluções são ilimitadas. No caso (c),  $E_1$  é uma solução intermediária entre  $E_0$  e  $E_2$ , sendo que  $E_1$  é estável e as outras duas, instáveis, sendo que pequenas perturbações em  $E_2$  podem levar a solução para  $E_1$  ou para uma solução ilimitada (região de escape). Para a resposta específica, os patógenos foram representados por células infectadas e, também, foram consideradas as concentrações de células T livres e de complexos formados pelo encontro de uma célula T com as células infectadas. Este estudo verificou a existência de dois estados estacionários, sendo o trivial  $E_0$  (ausência de patógeno) e um de co-existência  $E_1$ . Dependendo dos parâmetros, quatro situações se apresentaram:  $E_0$  globalmente estável e  $E_1$  inexistente;  $E_0$  instável e  $E_1$  atrator;  $E_0$  e  $E_1$  instáveis e um círculo limite que atrai todas as soluções positivas em torno de  $E_1$ ;  $E_0$  e  $E_1$  instáveis, e todas as soluções são ilimitadas. Assim, o modelo mostrou que as infecções podem ou não ser controladas pela resposta imunológica, dependendo de alguns parâmetros e das condições iniciais.

Em 2001, Utzny e Burroughs [24] apresentaram um modelo para investigar a correlação entre as propriedades das células T de memória e os mecanismos de regulação das células (apoptose). Basearam-se nas seguintes hipóteses: uma infecção resulta na ativação de células T em repouso e de memória, com suas subseqüentes expansões clonais e a exterminação dos patógenos. Após esta exterminação, as células T ativadas se diferenciam em células de memória ou sofrem apoptose. Um evento infeccioso com eliminação dos patógenos e retorno ao estado de repouso constitui um ciclo de infecção-resolução que altera o número de linfócitos específicos ao patógeno. A questão central que este trabalho discute é se sucessivas infecções com

o mesmo patógeno resulta em um número sempre crescente de células de memória ou se este crescimento é limitado, ou seja, se a quantidade de células T de memória para cada patógeno é estável. O modelo proposto considera três sub-populações diferentes de células T: ativadas ( $T^*$ ), com memória específica para um particular patógeno (M) e com memória específica para todos os demais patógenos (T). O ciclo de infecção–resolução foi modelado em duas fases: a fase I considerou a ativação das células T e a eliminação dos patógenos, enquanto a fase II consistiu das células T ativadas sendo extintas e do processo de homeostase das células de memória. A fase II foi determinada por um sistema de EDO modelando apoptose, memória, diferenciação e homeostase das três sub-populações T, M e  $T^*$ . Esta fase foi o foco do trabalho e a teoria de ponto fixo foi utilizada para seu estudo. A principal conclusão que se obteve foi que as propriedades da memória imunológica surgem se a dinâmica da população que governa o processo é dependente da população em si. O controle de diferenciação em células de memória ou apoptose dependente da população permite ao sistema imunológico adaptar-se melhor às variações na resposta imunológica e à severidade da infecção. O tamanho do compartimento da memória específica gerada por uma infecção severa depende apenas fracamente do número de células ativadas envolvidas na resposta e, ainda mais, quase toda a memória protetora é gerada por infecções menores (vacinação). A análise matemática mostrou que o balanço entre apoptose e diferenciação em memória é essencial para a regulação do número de novas células T que entrarão para o “pool” da memória. Mostrou ainda que a memória diversificada é preservada somente se: primeiro, a regulação de novas entradas é dependente da população de células T ativadas; segundo, se a população de células T comunica-se com outras partes do sistema imunológico.

Em 2001, Bergmann, Van Hemmen e Segel desenvolveram um trabalho que estudou a decisão das células T auxiliaadoras em se diferenciarem em Th1 ou Th2, já que a primeira aproximação com o antígeno determina se a resposta será humoral ou celular, a qual é crucial para evitar a doença. O modelo considerou a variação das populações Th1 e Th2 de modo simples, tomando a forma

$$\begin{aligned} & \text{Taxa de variação das populações de células Th1/Th2} \\ & = \text{diferenciação} + \text{proliferação} - \text{morte}, \end{aligned}$$

onde nos termos de diferenciação e proliferação foram consideradas capacidade suporte. Para a taxa de variação de patógenos foram considerados seu crescimento e sua morte em função das células Th1 e Th2. O sistema foi estudado em equilíbrio e sete soluções foram encontradas, dependendo da variação dos parâmetros: a solução trivial, que é sempre instável; a cura com a prevalência de Th2, a doença crônica com a prevalência de Th2, a cura com prevalência de Th1, a doença crônica com a prevalência de Th1, a cura com concomitância de Th1 e Th2 e a doença crônica com concomitância de Th1 e Th2. As soluções com prevalência de Th1 ou Th2 podem ser instáveis ou estáveis dependendo da variação dos parâmetros, enquanto que a cura com ambas as células mostrou-se instável em um estudo numérico e a doença crônica com ambas as células não pôde ser calculada analiticamente.

É importante ressaltar que a compreensão do sistema imunológico influencia diretamente na melhora de modelos matemáticos epidemiológicos. Dushoff [8] apre-

sentou um trabalho que discute a incorporação de idéias oriundas de resultados imunológicos em modelos epidemiológicos, obtendo um bom resultado. O modelo desenvolvido por Dushof não trata especificamente de um modelo para a descrição de uma dinâmica do SI, mas sim, como um dado modelo utilizado na descrição de um fenômeno pode apresentar aspectos interessantes quando acrescentado de algumas considerações de resultados imunológicos. Em muitas doenças existem evidências de importantes interações entre a epidemiologia e a imunologia. Ambos, modelos e dados, sugerem que variáveis de caráter epidemiológico, como freqüência e intensidade de exposição, podem afetar resultados de caráter imunológico e, reciprocamente, os resultados da batalha imunológica entre hospedeiro e parasita determinam a habilidade da doença de se espalhar em uma comunidade. Ao considerar-se um modelo para o espalhamento de uma doença devida a helmintos, na presença de dois níveis de infecciosidade, a saber, baixa infectividade e alta infectividade (asserção esta de caráter imunológico), obtém-se a possibilidade de uma bifurcação do tipo “backward”, o que não se verifica, por exemplo, quando no modelo do mesmo fenômeno é considerada uma única classe de infecciosidade. Uma bifurcação “backward” caracteriza-se pela existência de um equilíbrio instável e que serve como um “break point”, isto é, abaixo deste valor a doença persiste e se estabelece, mesmo que não seja capaz de invadir. Assim, ao incorporarmos uma situação imunológica ao modelo epidemiológico possibilitamos uma explicação para mecanismos de múltiplos estados estacionários e para a persistência de certas doenças. Isso é particularmente interessante do ponto de vista de políticas públicas que envolvem programas de quimioterapias intensivas. Por exemplo, se os parâmetros da doença são tais que ela persiste na comunidade mas não seja capaz de invadi-la uma vez eliminada, temos a forte indicação de que a utilização de uma forte campanha quimioterápica pode ser capaz de erradicar a doença da comunidade, ao se curar os indivíduos que possibilitariam a sua persistência.

Salientamos também um trabalho de Bellomo e Preziosi [4] que trata da evolução de tumores e sua interação com o sistema imunológico. Os cânceres são provocados por crescimentos progressivos da progênie de uma única célula transformada, e, portanto, para curá-los são necessárias a remoção e/ou a destruição de todas as células malignas. Embora experimentos em animais venham demonstrando que as células T são um mediador crítico da imunidade tumoral (observações indicam que tumores expressam peptídeos antigênicos, MHC-I, que podem se tornar alvos para uma resposta de células T específicas para as células tumorais). No entanto, é importante salientar que a maioria dos tumores mostram poucas evidências de controle imunológico, pois, em geral, eles provavelmente não possuem peptídeos antigênicos distintos ou não possuem moléculas de adesão ou co-estimuladoras necessárias para ativar uma resposta primária das células T. Avanços na compreensão da apresentação de antígenos e das moléculas envolvidas na ativação das células T podem propor novas estratégias imunoterápicas, uma vez que uma estratégia de eliminação ou contenção seria induzir uma resposta imunológica contra o tumor capaz de discriminar entre as células tumorais e as sadias. A descrição da dinâmica do sistema imunológico versus células tumorais é tratada por Bellomo através da teoria cinética celular, que possibilitam uma descrição estatística da evolução de uma população

grande de células regida por interações cinéticas.

## 4. Modelo matemático

Um modelo matemático é apresentado para descrever uma ação do sistema imunológico: a resposta imunológica adaptativa mediada por células. Considera-se uma infecção em que as células B e T interagem com o antígeno, ocorrendo a ligação antígeno-anticorpo e a destruição de células alvo infectadas.

Apresenta-se, antes, o comportamento do micro-organismo em condições favoráveis sem a ação do sistema imunológico do hospedeiro, e de que maneira o sistema imunológico encontra-se em “repouso” no indivíduo. Estes estudos, em que os parasitas e o sistema imunológico não estão interagindo, ajudam na compreensão do sistema dinâmico complexo resultante da interação.

### 4.1. Antígeno e sistema imunológico isolados

Quando um micro-organismo invade o organismo humano que apresenta o sistema imunológico em estado de “dormência”, ou seja, incapaz de responder ao estímulo antigênico, espera-se que cresça sem nenhuma resistência, inicialmente. Porém, à medida em que a sua concentração no organismo do hospedeiro aumenta, exaure-se a capacidade de sua manutenção. Pois, todas as células alvo (envolvidas na multiplicação do micro-organismo) podem estar infectadas e/ou o hospedeiro debilita-se pela infecção, o que limitaria o seu crescimento.

Seja  $A(t)$  a concentração de um micro-organismo no hospedeiro no instante de tempo  $t$ , representado pelo antígeno identificado pelo sistema imunológico. A sua multiplicação está limitada pela disponibilidade das células alvo onde ocorrem as sínteses protéicas para a sua replicação. Esta disponibilidade é designada por  $C$ , que representa a capacidade de suporte do meio (o conjunto de todas as células alvo). Estes micro-organismos morrem (incapacidade de invadir células alvo ou replicação defeituosa) naturalmente a uma taxa constante  $\mu_A$ . Uma outra característica da infecção microbiana é a concentração de patógenos com que é inoculada no organismo. Assim, se uma concentração baixa é inoculada no indivíduo, a infecção não se estabelece, diante de adversidades como a dificuldade de encontro com as células alvo pelas inúmeras barreiras fisiológicas. Sendo, assim, necessária uma inoculação a uma concentração não muito baixa para ocorrer de fato a invasão e estabelecimento do parasita no hospedeiro.

Para levar em conta as duas características acima, uma descrição da dinâmica da replicação do parasita no hospedeiro sem a estimulação do sistema imunológico pode ser dada por

$$\frac{dA(t)}{dt} = \Phi[A(t)] A(t) - \mu_A A(t),$$

onde  $\Phi[A(t)]$  é a taxa de replicação per-capita do micro-organismo. Nesta taxa pode-se embutir as duas propriedades referentes ao crescimento do parasita. Por

isso, usa-se para esta taxa per-capita uma função dada por

$$\Phi[A(t)] = k_A A(t) \left[ 1 - \frac{A(t)}{C} \right],$$

com  $k_A$  sendo a taxa de replicação intrínseca do parasita no organismo invadido. Os parâmetros  $C$  e  $\mu_A$  são, respectivamente, a capacidade de suporte do meio e a taxa de inviabilização do micro-organismo. Em primeiro lugar, a função  $\Phi$  mostra que os parasitas podem se replicar, no máximo, até o valor da capacidade de meio  $C$ . Em segundo, a função  $\Phi$  cresce monotonicamente com  $A$  até o valor  $C/2$  e, deste valor em diante até  $C$ ,  $\Phi$  decresce monotonicamente. Isso mostra que a taxa de replicação per-capita apresenta valores nulos nos extremos ( $A(t) = 0$  e  $A(t) = C$ ), revelando tanto a busca ativa (dificuldade de se manter para  $A(t) \sim 0$ ) quanto a competição (exaure-se a disponibilidade dos “nutrientes” para  $A(t) \sim C$ ) pelas células alvo.

Assim, analisa-se a dinâmica do parasita descrita por

$$\frac{dA(t)}{dt} = k_A \left[ 1 - \frac{A(t)}{C} \right] A^2(t) - \mu_A A(t). \quad (4.1)$$

Os pontos de equilíbrio assintótico, isto é,  $A = \lim_{t \rightarrow \infty} A(t)$ , em que o valor de  $A$  é resultante da imposição  $\frac{dA(t)}{dt} = 0$ , são obtidos da solução da equação

$$A \left( A^2 - CA + \frac{C}{\phi^*} \right) = 0, \quad (4.2)$$

onde

$$\phi^* = \frac{k_A}{\mu_A}$$

é a capacidade reprodutiva do micro-organismo. Biologicamente, como  $\mu_A^{-1}$  é o período de tempo médio de viabilidade de um parasita,  $\mu_A^{-1} k_A$  é o número médio de novos parasitas que um parasita produz durante o período de tempo viável. Quanto maior for o período viável (logo, menor  $\mu_A$ ), o parasita terá mais chances de encontrar com sucesso a célula alvo e se replicar.

Uma vez obtido o ponto de equilíbrio, a sua estabilidade é determinada pelo auto-valor obtido da equação

$$f(A) = -\frac{3k_A}{C} A + 2k_A A - \mu_A,$$

que corresponde à primeira derivada do segundo membro da equação (4.1). Se o valor de  $f$  for negativo, então o ponto é assintoticamente estável (atrator de trajetórias); se for positivo, é instável (repulsor).

A equação (4.2) apresenta três possibilidades, dependendo do valor de  $\phi^*$  e  $C$ .

1) *Capacidade reprodutiva baixa*,  $\phi^* < 4/C$ . Neste caso, tem-se apenas o equilíbrio trivial,  $A = 0$ . O auto-valor correspondente a este ponto vale  $f(0) = -\mu_A$ . Portanto, quando o parasita apresenta baixa capacidade reprodutiva, ele não consegue manter-se no hospedeiro, mesmo que inocule concentrações elevadas do micro-organismo.

2) *Capacidade reprodutiva limiar*,  $\phi^* = 4/C$ . Neste caso, tem-se dois pontos de equilíbrio, o trivial  $A_1 = 0$  e o não-trivial,  $A_2 = C/2$ . Os auto-valores correspondentes a estes dois pontos valem  $f(0) = -\mu_A$  e  $f(C/2) = 0$ . O equilíbrio trivial é estável, enquanto nada se conclui a respeito do equilíbrio não-trivial. Para este ponto, estuda-se pequenos deslocamentos em torno do equilíbrio  $C/2$ , fazendo-se  $A(t) = a(t) + C/2$ , com  $a(t) \ll 1$ . Substituindo-a na equação (4.1), obtém-se

$$\frac{da(t)}{dt} = -\frac{3k_A}{2}a^2(t) + \frac{k_A C}{4}a(t).$$

Quando  $a(t)$  é muito pequeno, o segundo termo do segundo membro predomina, tendo-se, assim,  $\frac{da(t)}{dt} \sim \frac{k_A C}{4}a(t)$ . Logo, este ponto é instável. Portanto, se a inoculação inicial for baixa (menor que  $C/2$ ), a população de micro-organismo vai para a sua extinção; porém, se o valor da inoculação inicial aumentar (maior que  $C/2$ ), neste caso, o parasita estabelece-se no hospedeiro, tendendo para a sua capacidade de meio  $C$ . O ponto de equilíbrio  $A_2$  é denominado de ponto de ruptura (“break point”), pois delimita a concentração de antígenos em duas regiões com atratores distintos, o trivial e o não-trivial.

3) *Capacidade reprodutiva elevada*,  $\phi^* > 4/C$ . Neste caso, tem-se três pontos de equilíbrio, o trivial,  $A_1 = 0$ , e dois não-triviais,  $A_2 = \left[ C - \sqrt{C(C - 4/\phi^*)} \right] / 2$  e  $A_3 = \left[ C + \sqrt{C(C - 4/\phi^*)} \right] / 2$ , com  $A_2 < A_3$ . Os auto-valores correspondentes a estes três pontos valem  $f(0) = -\mu_A$ ,  $f(A_2) = \phi \left[ \sqrt{C(C - 4/\phi^*)} - (C - 4/\phi^*) \right] / 2$  e  $f(A_3) = -\phi \left[ \sqrt{C(C - 4/\phi^*)} + (C - 4/\phi^*) \right] / 2$ . Como  $C > 4/\phi^*$ , tem-se que  $f(A_2) > 0$  e  $f(A_3) < 0$ . Logo, os pontos de equilíbrio trivial  $A_1$  e não-trivial maior  $A_3$  são estáveis, enquanto o não-trivial menor  $A_2$  é instável. Portanto, a dinâmica neste caso apresenta dois atratores, dependendo da inoculação inicial de parasitas. Se o valor introdutório for menor que  $A_2$ , o micro-organismo vai para a extinção ( $A_1$ ), não se mantendo no hospedeiro; em caso contrário, para inoculação superior a  $A_2$ , o micro-organismo mantém-se no hospedeiro a uma concentração  $A_3$ .

O que ocorre se a capacidade de suporte do meio for ilimitado, ou seja, quando  $C \rightarrow \infty$ ? Biologicamente, quando ocorre infecção generalizada, o hospedeiro morre antes que a capacidade de suporte seja alcançada. Neste caso limite, as três possibilidades acima resumem-se em apenas duas possibilidades, pois  $\lim_{C \rightarrow \infty} 4/C = 0$ .

1) *Capacidade reprodutiva nula*,  $\phi^* = 0$ . Neste caso, tem-se somente o equilíbrio trivial  $A = 0$ , cuja situação não apresenta nenhum interesse biológico.

2) *Capacidade reprodutiva positiva*,  $\phi^* > 0$ . Neste caso, tem-se dois pontos de equilíbrio, o trivial  $A_1 = 0$  e o não-trivial  $A_2 = 1/\phi^*$ . Os auto-valores correspondentes a estes dois pontos valem  $f(0) = -\mu_A$  e  $f(1/\phi^*) = \mu_A$ . O equilíbrio trivial é estável, enquanto o equilíbrio não-trivial é instável. Portanto, a dinâmica neste caso apresenta um único atrator finito, dependendo da inoculação inicial de parasitas. Se o valor introdutório for menor que  $A_2$ , o micro-organismo vai para a extinção ( $A_1$ ), não se mantendo no hospedeiro; em caso contrário, para inoculação superior a  $A_2$ , o micro-organismo mantém-se no hospedeiro, cuja concentração cresce indefinidamente (este é outro “atrator”, porém no infinito). Para entender melhor, observe

que  $A_2$  e  $A_3$  da possibilidade 3) do caso  $C$  finito resultam em, quando  $C \rightarrow \infty$ ,  $A_2 = 1/\phi^*$  e  $A_3 \rightarrow \infty$ .

Como as possibilidades 3) do caso  $C$  finito e 2) do caso  $C \rightarrow \infty$  apresentam dinâmicas semelhantes, escolhe-se o caso  $C \rightarrow \infty$  por apresentar um estudo analítico mais simples. Porém, deve-se ter em mente que o crescimento do parasita é ilimitado quando a inoculação for relativamente grande (maior que  $A_2$ ). Novamente,  $A_2$  de 2) do caso  $C \rightarrow \infty$  é o ponto de ruptura.

Para o sistema imunológico do hospedeiro sem interação com o parasita, considera-se que as células B e T são produzidas a taxas constantes, respectivamente,  $k_B$  na medula óssea e  $k_T$  no timo. Estas células circulam no sangue durante períodos de tempo médios dados por  $\mu_B^{-1}$  e  $\mu_T^{-1}$ , onde  $\mu_B$  e  $\mu_T$  são as taxas de mortalidade, respectivamente, das células B e T. Assim, a dinâmica do sistema imunológico em “repouso” é descrita por

$$\begin{cases} \frac{dB(t)}{dt} = k_B - \mu_B B(t) \\ \frac{dT(t)}{dt} = k_T - \mu_T T(t), \end{cases} \quad (4.3)$$

onde  $B(t)$  e  $T(t)$  designam as células B e T no instante de tempo  $t$ .

O sistema dinâmico linear (4.3) apresenta um único ponto de equilíbrio não-trivial, dado por

$$\begin{cases} B = \frac{k_B}{\mu_B} \\ T = \frac{k_T}{\mu_T}, \end{cases}$$

que é um atrator global, pois deslocamentos em torno de equilíbrio, dados por  $B(t) = b(t) + k_B/\mu_B$  e  $T(t) = \tau(t) + k_T/\mu_T$ , quando substituídos no sistema dinâmico linear (4.3), têm as trajetórias dadas por

$$\begin{cases} \frac{db(t)}{dt} = -\mu_B b(t) \\ \frac{d\tau(t)}{dt} = -\mu_T \tau(t), \end{cases}$$

cujas soluções são funções exponenciais decrescentes. Espera-se que o sistema imunológico seja globalmente estável, independentemente dos parâmetros do modelo. Biologicamente, o sistema imunológico deve ser bastante robusto, de tal modo que, após debelar uma infecção, retorne rapidamente para os valores de equilíbrio.

O parasita invade e replica-se rapidamente para se estabelecer no organismo do hospedeiro. Concomitantemente, o hospedeiro reage ao corpo estranho produzindo reações do sistema imunológico. A batalha entre hospedeiro e parasita terá um desfecho: o invasor rechaçado ou uma infecção crônica.

## 4.2. Interação entre antígeno e sistema imunológico

Quando um parasita invade o organismo humano, o seu sistema imunológico é estimulado. Uma vez que seus antígenos são reconhecidos pelo sistema imunológico, há produção de anticorpos por parte das células B e destruição das células alvo infectadas pelas células T.

A biologia da estimulação do sistema imunológico é descrita sucintamente. Quando um patógeno invade o organismo, o sistema imunológico reconhece e identifica o antígeno. Este antígeno, após ser apresentado a uma célula T antígeno-específica, provoca a ativação desta (para fins de simplificação, o modelo não considera a dinâmica das células dendríticas apresentadoras de antígenos). Esta célula T ativada, ao encontrar uma célula B que também detectou este antígeno (denominada B apresentadora de antígeno) causa a ativação desta célula B, que se diferencia em células plasmáticas, as quais se proliferam e produzem os anticorpos que propiciarão o combate ao patógeno. As células T ativadas atuam na destruição das células alvo infectadas.

O modelo matemático, por simplificação, supõe que a ação dos anticorpos sobre o antígeno (patógeno) e a destruição das células infectadas pelas células T são descritas pelo encontro das células plasma com o antígeno. As células B apresentadoras de antígenos e plasma são agrupadas em uma única classe de células B ativadas. A desativação da resposta imunológica dá-se pela diminuição da concentração do antígeno detectado no organismo. Assim, os linfócitos efetores, após cumprirem seu papel, morrem por apoptose. O modelo não leva em conta as células de memória, tratando, assim, a defesa do organismo através de produção de anticorpos em uma primeira infecção. Toda a dinâmica das citocinas envolvidas nos fenômenos de ativação e desativação está englobada nos parâmetros.

Assim, o modelo considera que as células B em repouso (classe designada por  $B$ ), quando ativadas pelas células T ativadas pelo antígeno (classe designada por  $T^*$ ), produz anticorpos. De acordo com a simplificação, as células B apresentadoras de anticorpos e os que se diferenciam em células plasma são agrupadas em uma única classe, designada por  $B_p$ . A dinâmica da população de parasitas é representada pelos antígenos  $A$ , que considera a capacidade de suporte (oferecida pelo hospedeiro) ilimitada. Assim, a dinâmica da interação entre antígeno e anticorpo pode ser representada pelo sistema de equações diferenciais ordinárias

$$\begin{cases} \frac{dA(t)}{dt} = k_A A^2(t) - \mu_A A(t) - \alpha A(t) B_p(t) \\ \frac{dB(t)}{dt} = k_B - \mu_B B(t) - \beta B(t) T^*(t) \\ \frac{dB_p(t)}{dt} = \beta A(t) B(t) + \gamma A(t) B_p(t) - \mu'_B B_p(t) \\ \frac{dT(t)}{dt} = k_T - \mu_T T(t) - \delta A(t) T(t) \\ \frac{dT^*(t)}{dt} = \delta A(t) T(t) - \mu'_T T^*(t), \end{cases} \quad (4.4)$$

onde  $\mu'_B$  e  $\mu'_T$  são as taxas de mortalidade, respectivamente, das células B e T ativadas;  $\beta$  e  $\delta$  são as taxas de ativação, respectivamente, das células B por T ativadas e das células T por antígenos;  $\gamma$  é a taxa de proliferação (clonagem) de células B plasma; e  $\alpha$  é a taxa total de remoção do antígeno pela ação da resposta imunológica (anticorpos secretados pelas células B ativadas e eliminação de células infectadas pelas células T ativadas). Todos estes parâmetros são considerados constantes.

O sistema dinâmico (4.4) descreve a ação de anticorpos produzidos pelas células B plasma sobre antígenos, descrito pelo termo  $\alpha A(t) B_p(t)$ : a desativação de antígenos é feita pela ligação com os anticorpos, cuja concentração no sangue é proporcional à concentração de células B ativadas, e pela destruição de células alvo infec-

tadas pelas células T ativadas, cujo efeito é indireto, uma vez que são estas que estimulam e aumentam a concentração de células B plasma. As células B ativadas multiplicam-se proporcionalmente à concentração de antígenos, descrito pelo termo  $\gamma A(t) B_p(t)$ . Da mesma forma, a ativação das células T é proporcional à concentração de antígenos, descrito pelo termo  $\delta A(t) T(t)$ .

Os pontos de equilíbrio do sistema dinâmico (4.4) são os seguintes.

1) Equilíbrio trivial, dado pelos valores

$$\begin{cases} A = 0 \\ B = \frac{k_B}{\mu_B} \\ B_p = 0 \\ T = \frac{k_T}{\mu_T} \\ T^* = 0, \end{cases}$$

que corresponde ao retorno do sistema imunológico ao estado basal, após debelar a infecção. A estabilidade do equilíbrio trivial é dada pelos auto-valores associados à matriz Jacobiana (linearização do sistema dinâmico em torno do valor do equilíbrio) dada por

$$J = \begin{bmatrix} -\mu_A & 0 & 0 & 0 & 0 \\ 0 & -\mu_B & 0 & 0 & -\beta B \\ 0 & 0 & -\mu'_B & 0 & \beta B \\ -\delta T & 0 & 0 & -\mu_T & 0 \\ \delta T & 0 & 0 & 0 & -\mu'_T \end{bmatrix}.$$

Os auto-valores associados são  $\lambda_1 = -\mu_A$ ,  $\lambda_2 = -\mu_B$ ,  $\lambda_3 = -\mu'_B$ ,  $\lambda_4 = -\mu_T$  e  $\lambda_5 = -\mu'_T$ . Como todos os auto-valores são negativos, o ponto de equilíbrio trivial é localmente e assintoticamente estável. Portanto, um dos atratores é o sistema imunológico reassumir o estado de repouso, e em prontidão contra novas e outras infecções.

2) Equilíbrio não-trivial, dado pelos valores

$$\begin{cases} A = \frac{\mu_A + \alpha B_p}{k_A} \\ B = \frac{k_B \mu'_T [k_A \mu_T + \delta(\mu_A + \alpha B_p)]}{\mu_B \mu'_T k_A \mu_T + (\mu_A \mu'_T + \beta k_T) \delta(\mu_A + \alpha B_p)} \\ T = \frac{k_A k_T}{k_A \mu_T + \delta(\mu_A + \alpha B_p)} \\ T^* = \frac{k_T \delta(\mu_A + \alpha B_p)}{\mu'_T [k_A \mu_T + \delta(\mu_A + \alpha B_p)]}, \end{cases} \quad (4.5)$$

mais a equação para  $B_p$  dada por

$$c_0 + c_1 B_p + c_2 B_p^2 + c_3 B_p^3 = 0,$$

onde os coeficientes do polinômio de terceiro grau são

$$\begin{cases} c_0 = k_A k_B k_T \beta \delta \mu_A \\ c_1 = k_A k_B k_T \beta \delta \alpha + \left( \gamma \mu_A - k_A \mu'_B \right) \left[ \mu_B \mu'_T k_A \mu_T + \left( \mu_B \mu'_T + \beta k_T \right) \delta \mu_A \right] \\ c_2 = \left( \gamma \mu_A - k_A \mu'_B \right) \left( \mu_B \mu'_T + \beta k_T \right) \delta \alpha + \gamma \alpha \left[ \mu_B \mu'_T k_A \mu_T + \left( \mu_B \mu'_T + \beta k_T \right) \delta \mu_A \right] \\ c_3 = \left( \mu_B \mu'_T + \beta k_T \right) \gamma \delta \alpha^2, \end{cases} \quad (4.6)$$

com  $c_0 > 0$  e  $c_3 > 0$ . Dependendo do sinal dos coeficientes  $c_1$  e  $c_2$ , pela regra de Descartes, tem-se nenhuma raiz real positiva, ou duas raízes reais positivas (para  $B_p$ ).

Se  $\gamma \mu_A > k_A \mu'_B$ , os coeficientes  $c_1$  e  $c_2$  são também positivos. Logo, pela regra de Descartes, não se tem nenhuma raiz real positiva. Nesta situação, a infecção é sempre debelada pelo sistema imunológico. Para entender melhor, escreve-se a desigualdade na forma

$$\frac{1}{\mu'_B} \gamma > \frac{1}{\mu_A} k_A \equiv \phi^*.$$

O primeiro termo é o número médio de células B ativadas clonadas durante todo o período de circulação de uma célula B ativada. Assim, se a capacidade de clonagem das células B ativadas ( $\phi_p^* = \gamma / \mu'_B$ ) for suficientemente grande (maior do que a capacidade de replicação do parasita), então o sistema imunológico atua eficientemente na eliminação do micro-organismo invasor ( $B_p = 0$  implica em  $A = 0$ ), e o único ponto de equilíbrio é o trivial, ou seja, o sistema imunológico retorna ao seu estado basal.

Se  $\gamma \mu_A < k_A \mu'_B$ , os coeficientes  $c_1$  e  $c_2$  são ainda positivos, se os dois termos tiverem valores próximos. Suponha que o valor da capacidade de clonagem das células B ativadas  $\phi_p^*$  seja menor do que  $\phi^*$ , de modo que  $c_1$  e/ou  $c_2$  sejam negativos (de acordo com a equação (4.6), segunda e terceira definições). Neste caso, pela regra de Descartes, ou não se tem nenhuma raiz real positiva, ou pode ter duas raízes reais positivas. No caso de duas raízes reais positivas para  $B_p$ , baseado nos resultados anteriores (parasita isolado), conjectura-se que o ponto de equilíbrio com valor menor para  $B_p$  corresponde ao ponto de ruptura, enquanto o de valor maior é o atrator. Assim, para baixas concentrações de parasitas inoculadas (menores que o valor correspondente ao do ponto de ruptura), o sistema imunológico consegue debelar a infecção; porém, para valores elevados, o parasita co-existe no hospedeiro em elevada concentração, o que pode levar à falência do sistema imunológico, resultando na infecção crônica e permitindo infecções oportunistas.

Esta conjectura pode ser melhor entendida no caso especial, em que a clonagem de células B ativadas não depende da concentração de antígenos. Nesta situação, a terceira equação do sistema dinâmico (4.4) é escrita como

$$\frac{dB_p(t)}{dt} = \beta A(t) B(t) + \left( \gamma - \mu'_B \right) B_p(t),$$

e o ponto de equilíbrio não-trivial tem os valores dados pela equação (4.5) mais a

equação para  $B_p$  dada por

$$c_0 + c_1 B_p + c_2 B_p^2 = 0,$$

onde os coeficientes do polinômio de segundo grau são

$$\begin{cases} c_0 = k_B k_T \beta \delta \mu_A \\ c_1 = k_B k_T \beta \delta \alpha + (\gamma - \mu'_B) \left[ \mu_B \mu'_T k_A \mu_T + (\mu_B \mu'_T + \beta k_T) \delta \mu_A \right] \\ c_2 = (\gamma - \mu'_B) (\mu_B \mu'_T + \beta k_T) \delta \alpha, \end{cases}$$

com  $c_0 > 0$ .

Se  $\gamma > \mu'_B$ , os coeficientes  $c_1$  e  $c_2$  são também positivos. Logo, pela regra de Descartes, não se tem nenhuma raiz real positiva. Nesta situação, a infecção é sempre debelada pelo sistema imunológico. Porém, se  $\gamma < \mu'_B$ , um dos coeficientes  $c_1$  e  $c_2$  pode ser negativo (se  $\gamma \lesssim \mu'_B$ , então apenas  $c_1$  é negativo); mas existe um valor de  $\gamma$  baixo no qual ambos os coeficientes  $c_1$  e  $c_2$  são negativos. Assim, para  $\gamma < \mu'_B$  tem-se apenas uma única raiz real positiva para  $B_p$  dada por

$$B_p = \frac{c_1 + \sqrt{(c_1)^2 + 4c_0 |c_2|}}{2 |c_2|},$$

com

$$|c_2| = (\mu'_B - \gamma) (\mu_B \mu'_T + \beta k_T) \delta \alpha.$$

Este equilíbrio não-trivial é o ponto de ruptura. Assim, dependendo da concentração inicial, ou o sistema imunológico debela a infecção, ou ele vai a falência, e o parasita cresce indefinidamente, pois a clonagem de células B ativadas ocorre a uma taxa constante, independentemente da concentração de antígenos (estes crescem ilimitadamente).

No entanto, observe que, quando a taxa de clonagem depender da concentração de antígenos, caso descrito pelo sistema dinâmico (4.4), mesmo que o sistema imunológico não consiga debelar a infecção, a concentração de antígenos não cresce indefinidamente, mas até certo valor, correspondendo ao valor de equilíbrio estável  $B_p$  maior. Isto se deve ao fato de mais células B ativadas são clonadas quando se aumenta a concentração de antígenos, limitando o seu crescimento.

Quando um parasita invade hospedeiro que não reage imunologicamente, e encontrando capacidade de suporte ilimitada, verificamos a existência de uma concentração limiar  $A^0$ , ponto de equilíbrio não-trivial dado por

$$A^0 = \frac{1}{\phi^*} = \frac{\mu_A}{k_A},$$

que divide em dois o comportamento dinâmico para a concentração de antígenos: a diminuição até a sua total eliminação ou o crescimento ilimitado. A eliminação é a situação observada quando a quantidade de antígenos inoculados (no momento da infecção) é menor do que o valor limiar  $A^0$ , enquanto a explosão de antígenos

ocorre toda vez que a inoculação inicial estiver acima do valor limiar  $A^0$  (vide 2) do caso  $C \rightarrow \infty$ ).

Quando o sistema imunológico é estimulado e passa a eliminar os antígenos do organismo, ocorrem dois comportamentos dinâmicos para esta interação: eliminação do parasita ou a sua co-existência com o sistema imunológico estimulado (pode ocorrer a explosão, se a clonagem de células B ativadas não depender dos antígenos). Quando o sistema imunológico responde muito bem e/ou a capacidade reprodutiva do parasita for muito baixa, a eliminação do antígeno ocorre independentemente da carga parasitária introduzida no hospedeiro. Porém, em casos intermediários, ocorrem ou a eliminação ou a explosão do parasita, dependendo do valor da inoculação inicial do patógeno no hospedeiro. O valor de concentração limiar  $A^i$  corresponde ao ponto de equilíbrio não-trivial instável, que é o ponto de ruptura. Este ponto de ruptura correspondente ao sistema imunológico em interação com antígeno é maior do que o valor do ponto de ruptura do parasita isolado, ou seja,  $A^i > A^0$ .

Qual a função do sistema imunológico no combate às infecções? Como age o sistema imunológico? A função principal e primordial é a defesa do organismo contra os patógenos, resultando na eliminação total. Esta defesa do organismo ocorre em dois níveis. Primeiro, pela eliminação do antígeno pelos anticorpos secretados pelos linfócitos B ativados e pela destruição das células alvo infectadas pelos linfócitos T ativados. Segundo, o sistema imunológico normal é capaz de enfrentar e vencer concentrações elevadas de patógenos inoculados, e somente é levado à falência em casos de infecções de extrema severidade. Isso é possível pelo deslocamento do ponto de ruptura (do parasita isolado) para valores mais elevados (quando o sistema imunológico está atuando), permitindo que a eliminação de antígenos seja possível para inoculações elevadas, desde que ocorram abaixo do ponto de ruptura. Este valor de ruptura depende das características do sistema imunológico e do invasor, resumidas pelos parâmetros do modelo.

## 5. Conclusão

A imunologia tem avançado muito na descrição da interação do sistema imunológico com elementos estranhos que invadem o corpo. A forma de ação do sistema imunológico varia de acordo com o parasita invasor. Há mecanismos distintos da resposta imunitária quando se trata de macro-parasitas e micro-parasitas. Mesmo dentro da classe de micro-parasitas, a resposta do sistema imunológico difere em se tratando de fungos, bactérias e vírus. Mais ainda, dentro da classe de vírus, diferentes espécies de vírus estimulam diferentemente o sistema imunológico. Por exemplo o vírus HIV tem como células alvo os linfócitos T-CD4 para se replicar, enquanto vírus da varicela-zoster alojam-se em células dos gânglios dos nervos espinais ou cranianos, resultando em diferentes formas de estimulação antigênica.

Nos últimos anos, muitos trabalhos de caráter quantitativo têm surgido para explorar o sistema imunológico. Procurou-se, então, fazer uma síntese de trabalhos existentes na literatura que tratam o sistema imunológico através de uma abordagem matemática, em especial utilizando-se as equações diferenciais ordinárias

como instrumento para suas descrições. A modelagem matemática para descrever a fisiologia do sistema imunológico tem grande importância, tendo em vista a imensa quantidade de trabalhos quantitativos relativos ao sistema imunológico existentes na literatura, e a contribuição destes para melhor entender a resposta imunitária e, também, para uma aplicação mais eficiente da terapêutica disponível.

Um modelo matemático simplificado foi apresentado para descrever a interação do sistema imunológico com antígeno, escolhendo uma resposta imunológica mediada por células secretoras de anticorpos. Este modelo, mesmo sendo muito simples, descreveu como corpos estranhos são eliminados pelo sistema imunológico: pela inativação dos antígenos e pela capacidade aumentada do sistema imunológico em debelar os invasores mesmo em elevadas concentrações (deslocamento do ponto de ruptura para valores elevados).

**Abstract.** In the last two decades, advances in our understanding of the immunological system were driven by AIDS research. In particular, this knowledge could have a wider impact on the study of epidemiology by improving the treatment and prevention strategies of infectious diseases. One question that can be naturally posed is: can mathematical research, applied to immunology, generate further advances in the comprehension of the immunological system? Research in immunology might take advantage of this interaction with mathematics, even though mathematics cannot directly solve problems in immunology. Models and simulations of several aspects and behaviors of the immunological system might reduce the amount of experiments which are necessary for the development of new therapies and support the progress of theoretical immunology. Thus, the aims of this paper are to provide an analysis of the physiological and mathematical aspects of several immuno-mathematical models and their contributions to immunology and to present our own model in the same spirit.

## Referências

- [1] G. Adomian, Y. Cherruault e K. Abbaoui, A nonperturbative analytical solution of immune response with time-delays and possible generalization, *Math. Comput. Modeling*, **24** (10) (1996), 89-96.
- [2] G. Adomian, "Solving Frontier Problems of Physics: The Decomposition Method", Kluwer Publication, 1994.
- [3] R.M. Anderson, Complex dynamic behaviours in the interaction between parasite populations and the host's immune system, *Intern. Jour. Parasit.* **28** (1998), 551-566.
- [4] N. Bellomo e L. Preziosi, Modeling and mathematical problems related to tumor evolution and its interaction with the immune system, *Mathematical and Computer Modeling*, **32** (2000), 413-452.
- [5] C. Bergmann e J.L. van Hemmen, Th1 or Th2: how an appropriate T helper response can be made, *Bull. Math. Biol.*, **63** (2001), 405-430.

- [6] R.J. De Boer, e A.S. Perelson, Towards a general function describing T cell proliferation, *J. Theor. Biol.*, **175** (1995), 567-576.
- [7] R.J. De Boer, A.A. Freitas, e A.S. Perelson, Resource competition determines selection of B cell repertoires, *J. Theor. Biol.*, **212** (2001), 333-343.
- [8] J. Dushoff, Incorporating immunological ideas in epidemiological models, *J. Theor. Biol.*, **180** (1996), 181-187.
- [9] M.A. Fishman e A.S. Perelson, Modeling T cell-antigen presenting cell interactions, *J. Theor. Biol.*, **160** (1993), 311-342.
- [10] M.A. Fishman e A.S. Perelson, Th1/Th2 cross regulation, *J. Theor. Biol.*, **170** (1994), 25-56.
- [11] M.A. Fishman e A.S. Perelson, Th1/Th2 differentiations and cross regulation, *Bull. Math. Biol.*, **61** (1999), 403-436.
- [12] T.B. Kepler e A.S. Perelson, Somatic hypermutation in B cells: an optimal control treatment, *J. Theor. Biol.*, **164** (1993), 37-64.
- [13] D. Kirschner, Using mathematics to understand HIV immune dynamics, *Notices of the AMS*, **43** (2) (1996), 191-202 .
- [14] P.W. Nelson, "Mathematical Models of HIV Pathogenesis and Immunology", Tese de Doutorado, University of Washington, USA, 1998.
- [15] M. Oprea e A.S. Perelson, Exploring the mechanisms of primary antibody responses to T-cell dependent antigens, *J. Theor. Biol.*, **181** (1996), 215-236.
- [16] M. Oprea, E.W. Nimwegen e A.S. Perelson, Dynamics of one-pass germinal centers models: implications for affinity maturation, *Bull. Math. Biol.*, **62** (2000), 121-153.
- [17] A.S. Perelson, M. Mirmirani e G.F. Oster, Optimal strategies in immunology. I. B-cell differentiation and proliferation, *J. Math. Biol.*, **3** (1976), 325-267.
- [18] A.S. Perelson, M. Mirmirani e G.F. Oster, Optimal strategies in immunology. II. B memory cell production, *J. Math. Biol.*, **5** (1978), 213-256.
- [19] A.S. Perelson e P.W. Nelson, Mathematical analysis of HIV-I dynamics in vivo, *Society for Industrial and Applied Mathematics*, **41** (1999), 3-44.
- [20] D.M. Pierre, D. Goldman, Y. Bar-Yam e A.S. Perelson, Somatic evolution in the immune system: the need for germinal centers for efficient affinity maturation, *J. Theor. Biol.*, **186** (1997), 159-171.
- [21] S.S. Pilyugin e R. Antia, Modeling immune responses with handling time, *Bull. Math. Biol.*, **62** (2000), 869-890.

- [22] L.A. Segel e R.L. Bar-Or, “Immunology Viewed as the Study of an Autonomous Decentralized System. Artificial Immune System and their Application” (D. Dasgupta, ed.), Springer-Verlag, Berlin, 1998.
- [23] N.I. Stilianakis e J. Seydel, Modeling the T-cell dynamics and pathogenesis of HTLV-I infection, *Bull. Math. Biol.*, **61** (1999), 935-947.
- [24] C. Uetzny e N.J. Burroughs, Stability of a diverse immunological memory is determined by T cell population dynamics, *Bull. Math. Biol.*, **63** (2001), 685-713.

## Sobre A Erradicação de Doenças Infecciosas – Esforço de Vacinação

H.M. YANG<sup>1</sup>, Departamento de Matemática Aplicada, Instituto de Matemática, Estatística e Computação Científica, UNICAMP, CP: 6065, CEP: 13081-970, Campinas, SP, Brasil

L.K. HOTTA<sup>2</sup>, Departamento de Estatística, Instituto de Matemática, Estatística e Computação Científica, UNICAMP, CP: 6065, CEP: 13081-970, Campinas, SP, Brasil

**Resumo.** O acúmulo de conhecimentos relativos à transmissão de infecções resultou no desenvolvimento de modelos matemáticos para melhor estudar a sua dinâmica. Estes modelos descrevem a disseminação das doenças na comunidade e, também, podem ser preditivos quanto aos resultados de diferentes mecanismos de controle que são introduzidos nesta comunidade. Muitas infecções que ainda assolam a humanidade podem ser (e estão sendo) combatidas por meio de vacinações em massa. No entanto, mostra-se que o esforço mínimo de vacinação (para se obter a erradicação da doença) predito pelo modelo deve ser entendido com os devidos cuidados para evitar compreensões errôneas e aplicações equivocadas.

### 1. Introdução

A epidemia de doenças infecciosas depende fundamentalmente da interação entre o agente infeccioso e o sistema imunológico do hospedeiro. Desta interação resultam o padrão de distribuição da doença nas diversas faixas etárias e condições sociais e, ao mesmo tempo, a elaboração de mecanismos eficazes de controle.

Os agentes etiológicos de muitas doenças infecciosas são os seres microbianos, dos quais fazem parte as várias espécies de vírus. Estes seres, em geral, são menores do que as bactérias e são parasitas que têm toda a sua replicação ocorrendo dentro das células parasitadas. A partícula virótica completa, o vírion, é um bloco de material genético recoberto por uma capa que o protege do meio ambiente e que serve ainda de veículo de transmissão de uma célula hospedeira para outra. Os vírus contêm somente um tipo de ácido nucléico: ribonucléico (RNA) ou desoxirribonucléico (DNA). Devido ao fato de o vírus não possuir ribossomos, RNA de transferência, sistemas enzimáticos e geradores de trifosfato de adenosina (ATP), para a sua replicação

---

<sup>1</sup>hyunyang@ime.unicamp.br; apoio financeiro FAPESP e CNPq.

<sup>2</sup>hotta@ime.unicamp.br

necessita dos recursos do hospedeiro. Por isso, o vírus multiplica-se dentro do hospedeiro (no interior das células) por sínteses separadas com posterior reunião de seus componentes. Por outro lado, os vírus, fora das células, são metabolicamente inertes, sendo nada mais do que fragmentos de DNA (ou RNA) recobertos por uma capa. Estes seres inertes, quando infectam hospedeiros apropriados, causam moléstias infecciosas pelas suas atividades intracelulares.

Quando um agente microbiano, no caso um vírus, infecta (isto é, invade o organismo de) uma pessoa suscetível, esta torna-se infectada e, por um certo período, o vírus replica-se rapidamente no organismo humano, pois não encontra nenhuma resistência. Este curto período de tempo inicial é caracterizado por um crescimento exponencial do vírus (lei de Malthus). Simultaneamente à invasão do vírus, o indivíduo tem seu sistema imunológico estimulado e inicia-se a produção de anticorpos específicos contra o invasor. Após alguns dias, a concentração do vírus circulante no indivíduo começa a declinar à medida que ocorre o aumento da concentração de anticorpos produzidos. Decorridos mais alguns dias, praticamente não se encontra vírus circulante no indivíduo, quando então os níveis de anticorpos atingem valores máximos. Algumas infecções podem induzir uma imunidade perene, enquanto que outras, apenas temporária. Neste caso, a manutenção da concentração de anticorpos nos níveis apropriados pode ser explicada por infecções reincidentes, sendo que, na ausência destas, pode haver o declínio total.

O vírus apresenta a capacidade de se adaptar às modificações do meio ambiente (intensa atividade no interior das células ou permanecer inerte fora delas) para a sua perpetuação. Contudo, a capacidade de adaptação é verificada, também, nos seus hospedeiros, em especial o humano. Uma modificação adaptativa envolvendo a síntese protéica específica dos vertebrados, de enorme importância para a sua sobrevivência, é a resposta imunológica. Este é o principal mecanismo de defesa natural destes animais contra os micro-organismos patogênicos. Esta defesa imunológica pode ser induzida de outra forma. Jenner (1749-1823) inoculou um menino com raspado de pus da lesão de uma ordenhadora que adquirira varíola da vaca (uma moléstia benigna) e, posteriormente, inoculou, no mesmo menino, o pus de paciente no estágio ativo de varíola. Jenner observou que o menino não adoecera, demonstrando, assim, que a *vacinação* havia induzido a imunidade contra a varíola. Pasteur (1827-1875), fortuitamente, inoculou alguns animais com cultura envelhecida de agente etiológico da cólera nas galinhas. Estes animais não adoeceram quando foram inoculados com culturas comprovadamente virulentas, demonstrando, dessa forma, que as culturas envelhecidas perdem a sua virulência (patogenicidade) sem, contudo, perder a sua capacidade de induzir a imunidade.

A indução de imunidade por vírus atenuado, seja por meio de envelhecimento de vírus, seja pela passagem dos micro-organismos por *hospedeiros artificiais*, tem a finalidade de proteger as pessoas contra a forma virulenta. A vacinação, ou a proteção induzida por inoculação de vírus atenuados, reside no fato de estes vírus não produzirem a forma grave da moléstia; entretanto retêm a capacidade de induzir a imunidade no homem contra a forma virulenta. Isto é possível por ser o sistema imunológico altamente específico, ou seja, quando os antígenos (isto é, quaisquer substâncias capazes de, em condições normais ou apropriadas, induzir a

formação de anticorpos) são introduzidos no organismo, estes reagem de maneira altamente seletiva com os anticorpos correspondentes produzidos pelo sistema imunitário. Estes anticorpos são as proteínas séricas, as imunoglobulinas, formadas em resposta à administração de um antígeno. A conjugação da melhor compreensão do sistema imunológico humano com o avanço da engenharia genética tem levado, atualmente, à produção de vacinas sintetizadas nos laboratórios.

Em geral, a imunidade induzida por vírus circulantes na natureza é de longa duração e confere proteção durante quase a vida toda. Entretanto, as vacinas podem falhar em dois níveis. A falha primária da vacina consiste em não induzir a imunidade no indivíduo vacinado. Isto pode ocorrer pela falha na produção da vacina ou pela deficiência do sistema imunitário do indivíduo. A falha secundária ocorre quando o nível de imunidade vacinal é baixo, resultando, portanto, em uma perda dessa imunidade em um curto período. Neste caso, as reinfecções podem manter os níveis de anticorpos elevados.

A imunoepidemiologia leva em consideração a capacidade de estimulação do sistema imunológica resultante da interação entre o hospedeiro e o parasita para determinar o padrão de epidemia. Assim, a epidemiologia das doenças infecciosas pode ser feita de acordo com a concentração do vírus no organismo dos indivíduos com a evolução temporal, de acordo com a história natural da infecção. Os indivíduos suscetíveis, ao se infectarem com o vírus, passam por um período latente, isto é, um período de tempo durante o qual o vírus replica-se no interior das células parasitadas dos indivíduos. Estes são chamados de indivíduos expostos ou latentes (infectados, mas não infectantes). No período subsequente, desde quando as células parasitadas eliminam o vírus para o meio extracelular até o declínio quase completo da concentração de vírus neste meio, devido aos mecanismos de defesa, estes indivíduos são classificados como infectantes. Este intervalo em que o vírus encontra-se no meio extracelular é denominado período de recuperação (ou infeccioso), e é neste período que eles são eliminados para o meio ambiente por meio da saliva e excreções diversas (trato respiratório, fezes, urina). Durante este intervalo, os indivíduos infectantes podem transmitir o vírus aos indivíduos suscetíveis, quando se encontrarem próximos. Posteriormente, por causa da imunidade adquirida, os indivíduos deixam o estado infeccioso, sendo, portanto, chamados de recuperados ou imunes. Porém, em algumas infecções pode ocorrer a perda da imunidade adquirida, o que faz os indivíduos recuperados retornarem ao estado suscetível.

A importância da imunização pela vacina, do ponto de vista preventivo, é a possibilidade de transferir indivíduos suscetíveis diretamente para a classe dos imunes, escasseando, assim, a fonte para novas infecções e, ao mesmo tempo, os indivíduos desta classe conferem uma barreira protetora aos indivíduos suscetíveis. Desta forma, a vacinação tem sido uma grande aliada para prevenir a população contra — e até erradicar — algumas doenças infecciosas. Do ponto de vista profilático, os indivíduos vacinados não são acometidos pelas graves sintomas causadas pela infecção, reduzindo a morbidade.

Pode-se, assim, perceber que, no aspecto da Biologia da interação hospedeiro-parasita, a erradicação é um alvo plausível se houver vacina disponível contra a infecção. Surge, então, de forma natural, a questão do esforço da vacinação: qual é

a proporção de indivíduos que devem ser vacinados para resultar em um declínio da incidência, até a completa erradicação da doença? Portanto, se se tem por objetivo a erradicação de doenças infecciosas, deseja-se quantificar o esforço mínimo necessário para garantir o sucesso do mecanismo de controle a ser implantado. Este conhecimento prévio ajuda as autoridades sanitárias no planejamento e implementação de medidas preventivas. Entende-se por esforço de vacinação o número total de vacinas que devem ser aplicadas em uma comunidade.

Tendo por objetivo a determinação do número mínimo de indivíduos a serem vacinados em uma comunidade para a erradicação da doença, desenvolve-se um modelo matemático para este fim. Para tanto, o modelo matemático deve levar em consideração as características da transmissão da doença e o mecanismo de imunidade. Da análise do modelo, especialmente em equilíbrio assintótico, pode-se quantificar o esforço mínimo de vacinação para erradicar a infecção. Uma vez obtida a proporção crítica de indivíduos a serem vacinados, mostra-se como os resultados oriundos da epidemiologia matemática devem ser entendidos e aplicados na prática. Assim, discute-se extensamente a questão da extrema dificuldade em erradicar algumas doenças infecciosas, a despeito de modelos matemáticos vaticinarem a plausibilidade de obter a sua erradicação com uma certa facilidade.

## 2. Modelo matemático

Com o objetivo de estudar quando e como se obtém a erradicação de doenças infecciosas, desenvolve-se um modelo matemático baseado nos seguintes aspectos. O encadeamento do processo infeccioso inicia-se quando um indivíduo suscetível entra em contato com o agente infeccioso. Este indivíduo, denominado exposto, permanece neste estado desde o início do contato com o vírus até o momento em que se torna um agente transmissor da doença (infectante). Este intervalo de tempo é denominado período de incubação, denotado por  $\sigma^{-1}$ , onde o parâmetro  $\sigma$  é a taxa de incubação. Por sua vez, o organismo deste indivíduo, denominado de infectante, passa a combater o agente invasor através de produção de anticorpos específicos, que já se processava desde as primeiras horas após o contato com o vírus. Portanto, após um certo período, devido à ação do sistema imunitário, a concentração de vírus no indivíduo passa a ser zero ou praticamente nulo, situação em que não ocorre mais a eliminação do vírus para o meio ambiente. Este tempo é denominado de período de recuperação (ou infecção), designado por  $\gamma^{-1}$ , onde o parâmetro  $\gamma$  é a taxa de recuperação (ou infecção). Após o período de recuperação, o indivíduo passa a ser imune. Nesta abordagem não se considera a perda de imunidade.

Baseado na história natural da infecção, o modelo considera uma comunidade dividida em quatro compartimentos não interceptantes, representados por  $X(t, a)$ ,  $H(t, a)$ ,  $Y(t, a)$  e  $Z(t, a)$ , que são, respectivamente, as distribuições etárias  $a$  dos indivíduos suscetíveis, expostos, infectantes e recuperados no instante de tempo  $t$ . Estas distribuições etárias podem ser transformadas em quantidade de indivíduos. Por exemplo,  $X(t, a)da$  é o número de indivíduos suscetíveis compreendido entre as idades  $a$  e  $a + da$ , onde  $da$  é uma variação muito pequena (infinitesimal) de idade.

Assim, somando-se a densidade de indivíduos em todas as idades tem-se o número total de indivíduos em um compartimento, neste exemplo,  $\bar{X}(t) = \int_0^L X(t, a) da$  representa o número total de indivíduos suscetíveis, com o parâmetro  $L$  designando a idade máxima (longevidade) dos indivíduos.

O modelo é descrito pelas seguintes equações íntegro-diferenciais parciais

$$\begin{cases} \left( \frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) X(t, a) &= - [\lambda(t, a) + \nu(t, a) + \mu] X(t, a) \\ \left( \frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) H(t, a) &= \lambda(t, a) X(t, a) - (\sigma + \mu) H(t, a) \\ \left( \frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) Y(t, a) &= \sigma H(t, a) - (\gamma + \mu) Y(t, a) \\ \left( \frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) Z(t, a) &= \nu(t, a) X(t, a) + \gamma Y(t, a) - \mu Z(t, a), \end{cases} \quad (2.1)$$

onde  $\mu$  é a taxa de mortalidade natural (não se considera mortalidade diferenciada, ou seja, mortalidade adicional pela doença) dos indivíduos na comunidade,  $\nu(t, a)$  é a taxa de vacinação e  $\lambda(t, a)$  é a força de infecção, definida por

$$\lambda(t, a) = \int_0^L \beta(a, a') Y(t, a') da', \quad (2.2)$$

com  $\beta(a, a')$  sendo a taxa de contato entre indivíduos suscetíveis de idade  $a$  com infectantes de idade  $a'$ . Para a sua resolução, basta fornecer as condições iniciais ( $X(0, a) = X_0(a)$ ,  $H(0, a) = H_0(a)$ ,  $Y(0, a) = Y_0(a)$  e  $Z(0, a) = Z_0(a)$ ), obtidas, preferencialmente, em equilíbrio antes da introdução da vacinação, fazendo-se  $\nu(t, a) = 0$  e de contorno ( $X(t, 0) = \mu \bar{N}$ , onde  $\mu \bar{N}$  é a taxa de recém-nascidos quando se tem uma população constante dada por  $\bar{N}$ , e  $H(t, a) = Y(t, a) = Z(t, a) = 0$ , pois não são considerados nem anticorpos maternos, nem transmissão vertical). O índice zero nas variáveis indica uma situação epidêmica na ausência de vacinação.

Estuda-se o caso em que as taxas de vacinação e de contato são independentes de idade, assumindo valores iguais para todas as idades. Mais ainda, a vacinação é ininterrupta e com o mesmo valor em todo instante de tempo. Assim, tem-se  $\beta(a, a') = \beta$  e  $\nu(t, a) = \nu$ . Faz-se o estudo do equilíbrio assintótico, determinando os pontos de equilíbrio e a estabilidade. Posteriormente, estuda-se o esforço da vacinação.

O sistema de equações (2.1), que descreve uma comunidade sob um esquema de vacinação, pode também descrever uma outra situação distinta. A epidemiologia matemática da infecção sem a vacinação é descrita impondo-se  $\nu(t, a) = 0$  na primeira e na quarta equações do sistema (2.1), e adiciona-se o subscrito 0 nas variáveis dinâmicas. Suponha que a comunidade encontra-se em equilíbrio com a infecção, no momento imediatamente anterior à introdução da vacinação. A introdução da vacinação age como uma perturbação neste equilíbrio, que leva a comunidade a um novo valor de equilíbrio após decorrido um período de tempo. Matematicamente, este período é longo (infinito), por isso a nova distribuição da doença na comunidade é denominada de equilíbrio assintótico.

## 2.1. Equilíbrio assintótico – Pontos de equilíbrio

Quando as taxas de vacinação e de contato forem independentes de idade, o sistema de equações (2.1) e a equação (2.2) podem ser re-escritos em um único sistema, em termos de frações de indivíduos  $x(t)$ ,  $h(t)$ ,  $y(t)$  e  $z(t)$ , como

$$\begin{cases} \frac{d}{dt}x(t) &= \mu - [\beta y(t) + \nu + \mu] x(t) \\ \frac{d}{dt}h(t) &= \beta y(t)x(t) - (\sigma + \mu) h(t) \\ \frac{d}{dt}y(t) &= \sigma h(t) - (\gamma + \mu) y(t) \\ \frac{d}{dt}z(t) &= \nu x(t) + \gamma y(t) - \mu z(t), \end{cases} \quad (2.3)$$

onde, por exemplo,  $x(t) = \bar{X}(t)/\bar{N}$  é a fração de indivíduos suscetíveis, a força de infecção é  $\lambda(t) = \beta y(t)$  e tem-se a identidade  $x(t) + h(t) + y(t) + z(t) = 1$ .

Tem-se dois pontos de equilíbrio para o sistema de equações (2.3). O equilíbrio trivial, ou população livre da doença, tem as coordenadas

$$\begin{cases} x &= \frac{\mu}{\mu + \nu} \\ h &= 0 \\ y &= 0 \\ z &= \frac{\nu}{\mu + \nu}, \end{cases} \quad (2.4)$$

que corresponde à erradicação da doença pela vacinação. A vacinação transfere os indivíduos suscetíveis diretamente para a classe dos imunes. Cessando a vacinação, tem-se então a comunidade toda constituída apenas por indivíduos suscetíveis, com  $x = 1$  e  $h = y = z = 0$ .

O equilíbrio não-trivial, situação em que a doença mantém-se prevalente em uma comunidade mesmo sob vacinação, tem as coordenadas

$$\begin{cases} x &= \frac{(\mu + \sigma)(\mu + \gamma)}{\beta \sigma} \equiv \frac{1}{R_0} \\ h &= \frac{(\mu + \nu)(\mu + \gamma)}{\beta \sigma} (R_\nu - 1) \\ y &= \frac{\mu + \nu}{\beta} (R_\nu - 1) \\ z &= \frac{\nu}{\mu R_0} + \frac{\gamma(\mu + \nu)}{\mu \beta} (R_\nu - 1), \end{cases} \quad (2.5)$$

onde  $R_0$  e  $R_\nu$ , dadas por

$$R_0 = \frac{\beta \sigma}{(\mu + \sigma)(\mu + \gamma)} \quad (2.6)$$

e

$$R_\nu = R_0 \frac{\mu}{\mu + \nu}, \quad (2.7)$$

são, respectivamente, a razão de reprodutibilidade basal e a razão de reprodutibilidade (não mais basal, pois o valor é diminuído pela vacinação). O equilíbrio sem a vacinação pode ser obtido impondo-se  $\nu = 0$  em todas as equações.

Nas infecções por micro-parasitas, a razão de reprodutibilidade basal (sem vacinação) corresponde ao número médio de infecções secundárias que um caso primário produz, durante todo o seu período infeccioso, em uma comunidade totalmente suscetível na ausência de qualquer tipo de heterogeneidade [6]. Portanto, este número

fornece uma idéia de quão infectivo (contagioso) é uma infecção. Por heterogeneidade entende-se quaisquer fatores que possam alterar a distribuição homogênea da infecção, por exemplo, idade, condições sócio-econômicas e sazonalidade.

A dinâmica do sistema de equações (2.3) fica determinada se obter as trajetórias das variáveis dinâmicas ao longo do tempo, que se dirigem a um dos dois possíveis equilíbrios assintóticos. Pode-se mostrar que há apenas duas possíveis trajetórias: nóduo ou espiral [5].

## 2.2. Estabilidade dos pontos de equilíbrio assintótico

A estabilidade dos dois pontos de equilíbrio do sistema de equações (2.3) pode ser determinada pelos autovalores da equação  $\Phi(\phi) = \det(\mathbf{J}^* - \phi\mathbf{I}) = 0$ , onde  $\mathbf{I}$  é uma matriz identidade  $4 \times 4$  e  $\mathbf{J}^*$ , a matriz jacobiana do sistema de equações (2.3), é

$$\mathbf{J} = \begin{bmatrix} -(\mu + \nu) - \beta y(t) & 0 & -\beta x(t) & 0 \\ \beta y(t) & -(\mu + \sigma) & \beta x(t) & 0 \\ 0 & \sigma & -(\mu + \gamma) & 0 \\ \nu & 0 & \gamma & -\mu \end{bmatrix}$$

calculada nos pontos de equilíbrio.

Os autovalores correspondentes ao ponto de equilíbrio trivial, dado pela equação (2.4), são  $\phi_1 = -\mu$  e  $\phi_2 = -(\mu + \nu)$  mais as raízes do polinômio de segundo grau

$$\phi^2 + (2\mu + \sigma + \gamma)\phi + (\mu + \sigma)(\mu + \gamma)(1 - R_\nu) = 0.$$

Este polinômio tem duas raízes com parte real negativa se, e somente se,  $R_\nu < 1$  [3]. Portanto, o ponto de equilíbrio trivial é localmente e assintoticamente estável se  $R_\nu < 1$ ; e instável se  $R_\nu > 1$ .

No equilíbrio trivial sem a vacinação, tem-se  $x_0 = 1$  e outras frações nulas. Esta comunidade sem a infecção é possível se a razão de reprodutibilidade basal for menor que um, ou  $R_0 < 1$ . Se este valor ultrapassar unidade (isto é,  $R_0 > 1$ ), então este ponto é instável, isto é, a doença persiste na comunidade. Assim, a doença só se torna endêmica se um caso primário produz, em média, um número de casos secundários maior que unidade. Neste sentido, a vacinação permite que o valor da razão de reprodutibilidade  $R_\nu$  possa ser diminuído para valores menores que unidade pela retirada de indivíduos da classe dos suscetíveis.

Para determinar a estabilidade do ponto de equilíbrio não-trivial, substituem-se as coordenadas dadas pela equação (2.5) na matriz jacobiana. Os autovalores são  $\phi_1 = -\mu$  mais as raízes do polinômio de terceiro grau

$$\phi^3 + a\phi^2 + b\phi + c = 0,$$

onde os coeficientes  $a$ ,  $b$  e  $c$  são dados por

$$\begin{cases} a &= (3\mu + \nu + \sigma + \gamma) + (\mu + \nu)(R_\nu - 1) \\ b &= (2\mu + \sigma + \gamma)(\mu + \nu)R_\nu \\ c &= (\mu + \nu)(\mu + \sigma)(\mu + \gamma)(R_\nu - 1). \end{cases}$$

Este polinômio de terceiro grau obedece os critérios de Routh-Hurwitz se  $R_\nu > 1$ . Portanto, o ponto de equilíbrio não-trivial é localmente e assintoticamente estável se  $R_\nu > 1$  [3]; e se  $R_\nu < 1$ , então este ponto é instável. Analogamente, mostra-se a situação em que não há vacinação, com  $R_0$ .

Mostrou-se que o ponto de equilíbrio trivial é estável se  $R_0 < 1$ , e instável se  $R_0 > 1$ . Por outro lado, o ponto de equilíbrio não-trivial é estável se  $R_0 > 1$ , e instável se  $R_0 < 1$ . Em termo do valor de  $R_0$ , se este for maior que um, tem-se o equilíbrio não-trivial; em caso contrário, tem-se o equilíbrio trivial. Note que o valor da razão de reprodutibilidade basal  $R_0$  pode ser diminuído pela vacinação até atingir o valor de bifurcação, ou seja, existe um valor  $\nu_c$  tal que a razão de reprodutibilidade  $R_\nu$  atinge o valor unitário. Existem apenas dois pontos de equilíbrio, e o sistema dinâmico em estudo apresenta apenas um valor de bifurcação em relação ao parâmetro  $\nu$ , cujo valor é obtido impondo-se  $R_\nu = 1$ . Conseqüentemente, todos os valores de  $\nu$  maiores que  $\nu_c$  resultam em  $R_\nu < 1$ , sendo, assim, a vacinação uma possível estratégia de erradicação. Em outras palavras, à medida que o valor de  $\nu$  aumenta, a epidemia vai se tornando cada vez menos prevalente até se erradicar quando a taxa de vacinação atinge o seu valor crítico  $\nu_c$ .

Todos os resultados foram expressos em função da taxa de contato  $\beta$ . Portanto, determinando-se seu valor, toda a dinâmica fica conhecida. Porém, pode-se expressar a dinâmica em termo da força de infecção natural  $\lambda_0$ , definida pela equação (2.2), em vez de  $\beta$ . O adjetivo *natural* para a força de infecção designa uma comunidade sem intervenção de mecanismo de controle. Para tanto, basta multiplicar a terceira equação do sistema de soluções (2.5), impondo-se  $\nu = 0$ , por  $\beta$  nos dois membros, para obter

$$\lambda_0 = \mu (R_0 - 1). \quad (2.8)$$

Esta equação relaciona a força de infecção natural com a razão de reprodutibilidade basal. Uma outra relação importante é

$$R_0 = \frac{1}{x_0}, \quad (2.9)$$

proveniente da primeira equação do sistema de soluções (2.5). Pela simples observação, nota-se que a fração de indivíduos suscetíveis em equilíbrio antes da introdução da vacinação  $x_0$  é igual ao seu valor após a vacinação  $x$ . Esta igualdade só é válida se a força de infecção não for nula. Se uma vacinação erradicar a doença, isto é,  $\lambda = 0$ , então  $x_0 \neq x$ , como pode ser verificada comparando com a primeira equação do sistema de soluções (2.4).

### 3. Esforço de vacinação

A vacinação transfere indivíduos suscetíveis para o estado imune sem passar pelo estágio infectivo. Se a vacinação não erradicar a doença, então a fração de indivíduos suscetíveis em equilíbrio antes (indivíduos retirados deste compartimento unicamente pela infecção,  $\lambda_0$ ) é igual a pós-vacinação (indivíduos retirados deste compartimento pela infecção mais a vacinação,  $\lambda + \nu$ ). Para determinar o esforço

da vacinação, estuda-se a distribuição etária dos indivíduos suscetíveis antes da vacinação  $X_0(a)$ , governada pela equação

$$\frac{d}{da}X_0(a) = -(\lambda_0 + \mu)X_0(a), \quad (3.10)$$

e a distribuição após a vacinação  $X(a)$ , dada por

$$\frac{d}{da}X(a) = -(\lambda + \nu + \mu)X(a). \quad (3.11)$$

Ambas as equações têm a mesma condição inicial  $X_0(0) = X(0) = \mu\bar{N}$ . Note que a distribuição etária de todos os indivíduos é dada por

$$\frac{d}{da}N(a) = -\mu N(a),$$

cujas soluções são  $N(a) = \mu\bar{N}e^{-\mu a}$ , onde usou-se a condição inicial  $N(0) = \mu\bar{N}$ .

As distribuições etárias de indivíduos suscetíveis antes e após a vacinação, resolvendo, respectivamente, as equações (3.10) e (3.11), são

$$\begin{cases} X_0(a) = \mu\bar{N}e^{-(\lambda_0+\mu)a} \\ X(a) = \mu\bar{N}e^{-(\lambda+\nu+\mu)a}. \end{cases}$$

Assim, os números totais de indivíduos suscetíveis são

$$\begin{cases} \bar{X}_0 = \frac{\mu}{\lambda_0+\mu}\bar{N} \\ \bar{X} = \frac{\mu}{\lambda+\nu+\mu}\bar{N}, \end{cases}$$

calculados da definição  $\bar{X} = \int_0^L X(a)da$ , com  $L \rightarrow \infty$ . Para  $\lambda > 0$ , sabe-se que  $\bar{X}_0 = \bar{X}$ , portanto tem-se que

$$\lambda_0 = \lambda + \nu. \quad (3.12)$$

Esta identidade não se verifica quando a doença for erradicada pela vacinação, ou seja, quando  $\lambda = 0$ . Assim, a validade desta igualdade é garantida se  $\nu < \lambda_0$ , enquanto para  $\nu > \lambda_0$  não se verifica (ou melhor, não há nenhuma relação matemática que ligue situações antes e depois de vacinação). No limiar, isto é,  $\nu_c = \lambda_0$  (entendendo como o limite  $\lambda \rightarrow 0$ ), a fração de indivíduos suscetíveis pode ser dada pela primeira equação do sistema de soluções (2.4) com  $\nu = \nu_c \equiv \lambda_0$ . Portanto, restrito apenas aos equilíbrios natural (antes da vacinação) e assintótico (depois da introdução da vacinação), não é possível a determinação da proporção de indivíduos suscetíveis vacinados  $p$ , definida por

$$p = \frac{\bar{X}_0 - \bar{X}}{\bar{X}_0} = 1 - \frac{\bar{X}}{\bar{X}_0}, \quad (3.13)$$

pois a diferença  $\bar{X}_0 - \bar{X}$ , por força da equação (3.12), é sempre zero para  $\nu \leq \nu_c$ . Logo, não há como distinguir indivíduos suscetíveis transferidos para outros compartimentos imunes por vacina dos por infecção usando o sistema dinâmico representativo do modelo, já que a fração de indivíduos suscetíveis não sofre alteração

para  $\nu \leq \nu_c$ . Matematicamente, há dois parâmetros desconhecidos,  $\lambda$  e  $\nu$ , mas apenas uma única equação que os relaciona, dada pela equação (3.12), o que implica em infinitas possibilidades de soluções (vide [9] para uma discussão mais extensa). Para  $\nu > \nu_c$  não há nenhuma relação matemática entre  $\lambda_0$  e  $\nu$ .

A equação (2.7) relaciona os valores  $R_\nu$  e  $R_0$  por meio da taxa de vacinação  $\nu$ . Mostrou-se que para os valores  $\nu \leq \nu_c$ , não há como relacionar univocamente a taxa de vacinação  $\nu$  com a proporção vacinada  $p$ . Assim, não há como obter também o seu valor limiar  $p_c$ , que é o esforço mínimo de vacinação para a erradicação da doença, pois para valores de  $p \geq p_c$  ocorre a erradicação da doença. Note que para se obter  $p_c$ , deve-se usar  $\nu_c$ . A equação (3.12), para  $\nu \leq \nu_c$ , mostra que  $0 \leq \lambda \leq \lambda_0$ . Assim, relaciona-se  $p$  com  $\nu$  de tal modo que aquele esteja entre os valores mínimo e máximo. O valor mínimo  $p_m$  corresponderia a usar  $\lambda = \lambda_0$  na equação (3.11) e o valor máximo  $p_M$ , a usar  $\lambda = 0$  na equação (3.11). Matematicamente, resolve-se a questão da variável livre impondo a uma delas (no caso,  $\lambda$ ) um valor fixo.

Fazendo-se  $\lambda = 0$  na equação (3.11), e usando a equação (3.13), obtém-se

$$p_M = 1 - \frac{\mu}{\mu + \nu} = \frac{\nu}{\mu + \nu}. \quad (3.14)$$

Desta equação obtém-se a relação entre  $R_\nu$  e  $R_0$ , que é dada por

$$R_p = (1 - p) R_0, \quad (3.15)$$

onde  $R_\nu$  foi re-escrito como  $R_p$ . Fazendo-se  $\lambda = \lambda_0$  na equação (3.11), e usando a equação (3.13), obtém-se

$$p_m = 1 - \frac{\mu}{\mu + \lambda_0 + \nu} = \frac{\nu}{\mu + \lambda_0 + \nu}. \quad (3.16)$$

O valor da proporção exata de indivíduos suscetíveis do modelo está entre estes dois valores, ou seja,  $p_m \leq p \leq p_M$ .

O valor limiar da proporção de indivíduos suscetíveis vacinados para erradicar uma infecção  $p_c$  é obtido quando se impõe  $\nu = \nu_c = \lambda_0$  nas equações (3.14) e (3.16), resultando, respectivamente, em  $p_M^c = \lambda_0 / (\mu + \lambda_0)$  e  $p_m^c = \lambda_0 / (\mu + 2\lambda_0)$ . Usando as equações (2.8) e (2.9), pode-se relacionar os valores críticos extremos de  $p_c$  com a fração de indivíduos suscetíveis  $x_0$ . Tem-se, assim,

$$\begin{cases} p_m^c = \frac{1-x_0}{2-x_0} \\ p_M^c = 1 - x_0, \end{cases} \quad (3.17)$$

com  $p_m^c \leq p_c \leq p_M^c$ . Ambas as expressões são monotonicamente decrescentes para  $0 \leq x_0 \leq 1$ , conforme mostra Figura 1. Note que  $x_0 = 0$  equivale a ter  $R_0 \rightarrow \infty$ , correspondendo a uma doença extremamente contagiosa; enquanto  $x_0 = 1$  equivale a ter  $R_0 = 1$ , correspondendo a uma doença praticamente não-contagiosa (independentemente de número de casos introduzido na comunidade, esta doença vai naturalmente para a sua erradicação).

No entanto, vacinar uma proporção  $p_c$  de indivíduos suscetíveis equivale a vacinar uma fração  $f_c$  de todos os indivíduos da comunidade. Os valores extremos

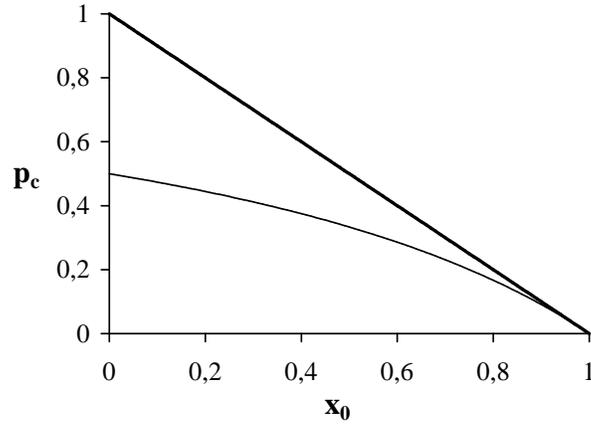


Figura 1: Os valores críticos extremos de  $p_c$  em função da fração de indivíduos suscetíveis  $x_0$ : as curvas referem-se ao extremo inferior  $p_m^c$  (fina) e ao extremo superior  $p_M^c$  (grossa). O valor exato de  $p_c$  situa-se entre as curvas de  $p_m^c$  e  $p_M^c$ .

correspondentes a esta fração são dadas por

$$\begin{cases} f_m^c = \frac{(1-x_0)x_0}{2-x_0} \\ f_M^c = (1-x_0)x_0, \end{cases} \quad (3.18)$$

com  $f_m^c \leq f_c \leq f_M^c$ . Ambas as expressões crescem e decrescem monotonicamente para  $0 \leq x_0 \leq 1$ , sendo que os valores máximos ocorrem para, respectivamente,  $x_0^* = 2 - \sqrt{2} \approx 0,586$  ( $f_m^c(x_0^*) \approx 0,17$ ) e  $x_0^* = 0,5$  ( $f_M^c(x_0^*) = 0,25$ ), conforme mostra Figura 2. Portanto, para obter a quantidade de vacinas a serem aplicadas na comunidade, designada por  $Q_c$ , basta multiplicar pelo número total de indivíduos na comunidade, cujos valores extremos são

$$\begin{cases} Q_m^c = f_m^c \bar{N} = \frac{(1-x_0)x_0}{2-x_0} \bar{N} \\ Q_M^c = f_M^c \bar{N} = (1-x_0)x_0 \bar{N}, \end{cases} \quad (3.19)$$

com  $Q_m^c \leq Q_c \leq Q_M^c$ .

Assim, o montante mínimo de vacinas a serem aplicadas para se erradicar uma infecção situa-se, no pior dos casos, entre 17% e 25% da quantidade de pessoas na comunidade, sendo que, para  $x_0$  diminuindo ou aumentando a partir deste valor, o montante de vacinas a serem aplicadas diminui. Observe que a proporção vacinada não depende da comunidade considerada, mas somente da infecção, tendo o mesmo valor para comunidades pequenas e grandes, enquanto o esforço de vacinação depende linearmente com o número de indivíduos de uma comunidade. Por exemplo, para uma doença com  $R_0 = 2$ , tem-se  $x_0 = 0,5$ , o que resulta em  $f_M^c = 0,25$ , ou seja, vacina-se 25% de indivíduos da população. Quanto ao esforço de vacinação,

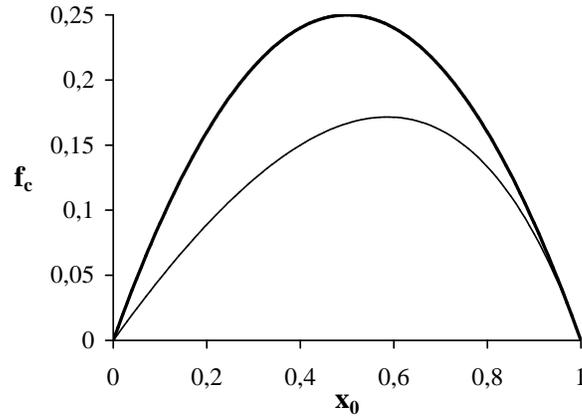


Figura 2: Os valores críticos extremos de  $f_c$  em função da fração de indivíduos suscetíveis  $x_0$ : as curvas referem-se ao extremo inferior  $f_m^c$  (fina) e ao extremo superior  $f_M^c$  (grossa). O valor exato de  $f_c$  situa-se entre as curvas de  $f_m^c$  e  $f_M^c$ .

ele, porém, aumenta com o tamanho da população considerada: se  $\bar{N} = 100$ , então o número mínimo de vacinas aplicadas é  $Q_M^c = 25$ ; enquanto se  $\bar{N} = 100.000$ , então o número mínimo de vacinas aplicadas é  $Q_M^c = 25.000$ .

Dependendo como todos estes resultados da epidemiologia matemática são compreendidos, a aplicação segue em duas formas distintas.

### 3.1. Aplicação direta dos resultados – Abordagem clássica

Na história da humanidade, infecções estiveram sempre presentes, e muitas delas perpetuaram-se até os dias atuais. Em relação às infecções que não deixam seqüelas (no sentido de deixar uma “marca” que identifique como tendo sido exposto à infecção), as pessoas não se preocuparam com o isolamento de indivíduos doentes. Porém, em relação às infecções que deixavam marcas, as pessoas aprenderam a se defender através de isolamento de indivíduos doentes, ou de vilas e cidades em que a doença era endêmica. Uma outra alternativa foi a emigração de indivíduos saudáveis (susceptíveis e recuperados). Assim, na ausência de vacina (tempos passados), o isolamento de indivíduos infectantes (sintomáticos, pois os assintomáticos não manifestam quadros clínicos da doença) tem quase os mesmos princípios de ação da vacina: evita-se que o agente infeccioso encontre indivíduos suscetíveis, sendo que o isolamento age como uma barreira protetora destes indivíduos.

Os resultados da epidemiologia matemática mostram que o esforço de vacinação é máximo para  $x_0$  em torno de 0,5, reduzindo-se à medida que decresce ou cresce a partir deste valor. Em termos de razão de reprodutibilidade basal, o esforço maior ocorre para  $R_0 \sim 2$ . Faz-se algumas asseverações epidemiológicas.

Quando as doenças infecciosas que deixavam seqüelas apresentavam alta ( $R_0$  muito elevado) ou baixa ( $R_0$  próximo de um) infectividade, de certa forma o isolamento foi um mecanismo de proteção eficiente. Como tem-se por objetivo a proteção de indivíduos suscetíveis, isola-se indivíduos infectantes e, assim, quanto maior o número destes indivíduos isolados, mais indivíduos suscetíveis estarão protegidos. Assim, proteger uma fração elevada de indivíduos da comunidade deve requerer um rígido isolamento de indivíduos infectantes, cujo esforço deve diminuir se o objetivo é proteger um número menor de indivíduos suscetíveis. Portanto, em casos extremos de infectividade, a fração de indivíduos suscetíveis a serem protegidos (isolados dos infectantes) era relativamente baixa — da mesma maneira que a ação de uma vacinação —, o que facilitaria a erradicação destas doenças. Dessa forma, doenças que deixam seqüelas, para se evadirem de isolamentos e se perpetuarem, precisavam apresentar um valor próximo de dois para a razão de reprodutibilidade basal, a fim de aumentar o esforço para isolar indivíduos sintomáticos. Portanto, doenças, que deixam seqüelas e que têm um valor para a razão de reprodutibilidade basal próximo de dois, podem ter perpetuado até os dias atuais e, juntamente com aquelas que não deixam nenhuma seqüela, são as moléstias-alvos das vacinas para a erradicação.

Um exemplo destas infecções pode ser a varíola, que foi considerada erradicada pela vacinação. As características epidemiológicas desta doença podem ser traduzidas como uma infecção com seqüelas que apresenta  $R_0$  próximo de 2, o que permitiu a sua erradicação através de vacinação em massa. Contudo, em vez de vacina, esta infecção pôde ser eliminada pelo isolamento em muitos países com boa infra-estrutura de saúde coletiva.

Mas, como se explica a persistência de algumas doenças infecciosas, como rubéola e sarampo, a despeito de intensas campanhas de vacinação, cujo esforço em muitos países, como Brasil, está muito acima do valor crítico preconizado pelos modelos matemáticos para se conseguir a erradicação? Antes de ser levado pelo impulso inicial de abandonar modelos matemáticos que fundamentam-se na lei da ação das massas, faz-se uma interpretação parcimoniosa dos resultados epidemiológicos matemáticos.

### 3.2. Aplicação parcimoniosa dos resultados – Abordagem alternativa

A abordagem anterior determinou o esforço de vacinação levando em consideração os indivíduos suscetíveis, sendo que nos resultados epidemiológicos estavam implícitos o fato de que os indivíduos suscetíveis podiam ser facilmente identificados. Em outras palavras, os indivíduos imunes trariam alguma marca que os distinguiam dos demais. Mas, se isto não for verdade, como elaborar e determinar o esforço de vacinação para erradicar uma doença? Deve-se mudar o modelo? Não, pois o modelo matemático está embasada na biologia (simplificada, é claro) da interação hospedeiro-parasita.

Em vez de supor que o modelo esteja errado, deve-se entender melhor os seus resultados. O modelo preconiza uma fração  $p_c$  de indivíduos suscetíveis (corres-

pondendo a uma fração  $f_c$  de todos os indivíduos da comunidade) que devem ser vacinados, se os indivíduos suscetíveis forem identificáveis. Mas, quando estes não puderem ser diferenciados dos demais (dos imunes, principalmente, pois estes formam geralmente a maior proporção na comunidade), em vez de entender o esforço mínimo de vacinação como  $f_c \bar{N} = Q_c$ , deve ser feita outra pergunta: *Quantas pessoas de uma comunidade devem ser vacinadas para que uma fração  $p_c$  de indivíduos suscetíveis seja efetivamente vacinada?* Para responder a esta pergunta, usa-se os métodos estatísticos.

Do ponto de vista estatístico, esta é uma pergunta sem resposta. Para que torne um problema estatístico bem posto, é preciso reformular a pergunta: *Quantas pessoas de uma comunidade com  $\bar{N}$  indivíduos devem ser vacinadas para que um número mínimo  $\bar{X}_c = p_c \bar{X}$  de indivíduos suscetíveis seja efetivamente vacinado com probabilidade de confiança  $1 - \alpha$ ?*

Para esta pergunta, a resposta para o caso de se querer ter 100% ( $\alpha = 0$ ) de chance de vacinar  $\bar{X}_c$  indivíduos suscetíveis é óbvia: deve-se vacinar um número de indivíduos dado por  $\bar{N} - \bar{X} + \bar{X}_c$ . Portanto, estuda-se o esforço de vacinação para valores de  $\alpha$  não-nulos considerando-se vários tamanhos de população.

### 3.2.1. Sobre a amostragem

Para se responder a pergunta do ponto de vista estatístico, entende-se a determinação do número de indivíduos da comunidade a serem vacinados como um problema de amostragem. Considere uma população onde os indivíduos podem ser classificados em 2 grupos: os suscetíveis e os imunes (os poucos indivíduos expostos e infectantes estão incluídos neste grupo). O interesse volta-se apenas para um dos grupos, o dos indivíduos suscetíveis, que serão vacinados. Assume-se que os indivíduos dos dois grupos não podem ser identificados apenas visualmente. Assim, diante da dificuldade de se descobrir quem pertence ao grupo dos suscetíveis que serão vacinados, escolhe-se aleatoriamente uma proporção de indivíduos da população toda para ser vacinada.

Determina-se o número de indivíduos vacinados de uma população  $\bar{N}$ , em que se tem  $\bar{X}_0$  indivíduos suscetíveis, para que se tenha uma proporção mínima  $\delta_0$  da população de suscetíveis vacinada. Para tanto, usa-se a fração de indivíduos suscetíveis na comunidade  $x_0 = \bar{X}_0/\bar{N}$ , a proporção da população vacinada  $\delta$ , o tamanho da amostra  $n = \bar{N}\delta$  e a variável aleatória representando o número de indivíduos suscetíveis na amostra  $K = \Gamma \bar{X}_0$ .

Na amostragem, a variável aleatória  $K$  tem distribuição exata hipergeométrica ( $Hi$ ), dada por

$$\Pr[K = k] = \frac{\binom{\bar{X}_0}{k} \binom{\bar{N} - \bar{X}_0}{n - k}}{\binom{\bar{N}}{n}}, \quad (3.20)$$

sendo denotada por  $Hi(\bar{N}, \bar{X}_0, n)$ , onde  $\Pr$  designa a probabilidade. Dependendo

de certas condições, ela pode ser aproximada pela distribuição normal

$$\mathfrak{N}\left(\frac{n\bar{X}_0}{\bar{N}}, n\frac{\bar{X}_0}{\bar{N}} \times \frac{\bar{N} - \bar{X}_0}{\bar{N}} \times \frac{\bar{N} - n}{\bar{N} - 1}\right).$$

Tomando-se  $\bar{N} - 1 \approx \bar{N}$  tem-se que  $\Gamma = K/\bar{X}_0$  tem distribuição aproximadamente normal dada por

$$\mathfrak{N}\left(\delta, \frac{(x_0^{-1} - 1)}{\bar{N}}\delta(1 - \delta)\right).$$

Portanto, a primeira questão relativa à amostragem, *qual deve ser o número ou a proporção da população total que deve ser vacinada (dada por  $\delta$ ), se se deseja que, com probabilidade no mínimo igual a  $(1 - \alpha)$ , a probabilidade de cobertura (a proporção vacinada da população de indivíduos suscetíveis) seja maior ou igual a  $\delta_0$* , é respondida.

O objetivo é encontrar  $\delta$  tal que

$$\Pr[\Gamma \geq \delta_0] \geq 1 - \alpha$$

ou

$$\Pr[\Gamma\bar{X}_0 \geq \delta_0\bar{X}_0] \geq 1 - \alpha.$$

A solução exata consiste em calcular o menor valor de  $n$  tal que

$$\Pr[Hi(\bar{N}, \bar{X}_0, n) \geq \delta_0\bar{X}_0] \geq 1 - \alpha. \quad (3.21)$$

Em muitas situações pode-se usar solução aproximada utilizando a aproximação normal. Pela aproximação normal, tem-se que a proporção da população total a ser vacinada é solução da equação do segundo grau (escolhendo-se o valor da solução entre 0 e 1)

$$\delta = \delta_0 - \frac{0,5}{\bar{X}_0} + z_\alpha \left[ \frac{(x_0^{-1} - 1)}{\bar{N}}\delta(1 - \delta) \right]^{1/2},$$

onde  $z_\alpha$  é o  $(1 - \alpha)$  quantil da distribuição normal padrão e o termo  $\bar{X}_0^{-1}/2$  é devida à correção de continuidade. Uma segunda aproximação é dada por

$$\delta = \delta_0 - \frac{0,5}{\bar{X}_0} + z_\alpha \left[ \frac{(x_0^{-1} - 1)}{\bar{N}}\delta_0(1 - \delta_0) \right]^{1/2}.$$

Esta expressão mostra de forma mais clara a dependência da quantidade adicional amostral em relação à nominal  $(\delta - \delta_0)$  em função de  $\bar{N}$ ,  $\bar{X}_0$ ,  $x_0$ ,  $\alpha$  e  $\delta_0$ . A proporção adicional é diretamente proporcional a  $|\delta - 1/2|$  e  $(1 - \alpha)$ , e inversamente proporcional a  $\bar{N}$  e  $x_0$ . A qualidade das aproximações é diretamente proporcional a  $\bar{N}$  e  $\alpha$ , e inversamente proporcional a  $|x_0 - 1/2|$ .

Uma outra questão relativa à amostragem, *sendo vacinada exatamente a proporção  $\delta$  da população total, qual a probabilidade de que a proporção vacinada da população de indivíduos suscetíveis seja menor ou igual a  $\delta_0$* , é repondida.

Este procedimento não será abordada neste trabalho, pois o objetivo é determinar o número de indivíduos a serem vacinados de uma comunidade para que se vacine um número desejado de indivíduos suscetíveis. Contudo, esta probabilidade é dada por

$$\beta = \Pr[K(\bar{X}_0, \bar{N}, \bar{N}\delta) \leq \delta_0 \bar{X}_0] = \Pr[\Gamma \leq \delta_0].$$

A aproximação normal, sem correção de continuidade e considerando  $\bar{N} \simeq \bar{N} - 1$ , permite ver de forma clara como os diversos parâmetros influenciam esta probabilidade. Ela é dada por

$$\beta \cong \Pr \left[ Z \leq \frac{\sqrt{\bar{N}} \left( \frac{\delta_0}{\delta} - 1 \right)}{\sqrt{(x_0^{-1} - 1)(\delta^{-1} - 1)}} \right].$$

Para o cálculo de probabilidade de sub-amostragem tem-se que  $(\delta_0/\delta) - 1 < 0$  e, portanto, esta probabilidade é inversamente proporcional ao tamanho da população  $\bar{N}$ , à proporção da população de indivíduos suscetíveis  $x_0$ , e à proporção da população amostrada  $\delta$ . Quando o tamanho da população ou a proporção da população de indivíduos suscetíveis é baixa a probabilidade de se ter uma sub-representação razoável não é desprezível.

### 3.2.2. Resultados

A questão da amostragem será estudada para dois valores de  $\alpha$  considerando-se 4 tamanhos de população. Tabela 1 apresenta, para alguns valores de  $x_0$  usados na amostragem, os valores extremos para  $p_c$  e  $f_c$  calculados pelas equações (3.17) e (3.18).

$x_0$	$p_m^c$ (%)	$p_M^c$ (%)	$f_m^c$ (%)	$f_M^c$ (%)
0,01	49,75	99	0,50	1
0,05	48,72	95	2,45	4,75
0,1	47,37	90	4,73	9
0,2	44,44	80	8,89	16
0,5	33,33	50	16,67	25
0,7	23,08	30	16,15	21
0,9	9,10	10	8,18	9

Tabela 1. Os valores críticos extremos para suscetíveis  $p_c$  e para população toda  $f_c$  em função de fração de indivíduos suscetíveis  $x_0$ .

Na Tabela 2 apresenta-se o valor do esforço de vacinação mínimo, relativo ao extremo inferior  $Q_m^c$ , correspondente aos valores de  $x_0$  dados na Tabela 1, para 4 diferentes tamanhos de população. O valor de  $Q_m^c$  é calculado usando-se a equação (3.19). Observe que para  $x_0 = 0,7$  tem-se valores de  $\bar{Q}_m^c$  próximos aos de  $x_0 = 0,5$ . Por isso este valor é suprimido no cálculo do tamanho da amostragem. Se o grupo de indivíduos suscetíveis puderem ser discriminados, a “amostragem” de indivíduos a serem vacinados é dada pela Tabela 2.

$x_0$	$N = 100$	$N = 1.000$	$N = 10.000$	$N = 100.000$
0,01	1	5	50	498
0,05	3	25	244	2.436
0,1	5	48	474	4.737
0,2	9	89	889	8.889
0,5	17	167	1.667	16.667
0,7	17	162	1.616	16.154
0,9	9	82	819	8.182

Tabela 2. O valor de  $\bar{Q}_m^c$ , a parte inteira de  $f_m^c \bar{N}$  mais 1, em função de fração de indivíduos suscetíveis  $x_0$ , para comunidades de diferentes tamanhos  $\bar{N}$ .

Tabela 2 fornece o número mínimo de vacinas (extremo inferior) que devem ser aplicados em uma comunidade para a erradicação da doença. Sendo os indivíduos suscetíveis passíveis de serem identificados, esta tabela fornece o número de indivíduos suscetíveis, escolhidos dentre todos os indivíduos da comunidade, que devem ser vacinados. Em caso contrário, resolve-se o seguinte problema de amostragem: determinar o número de indivíduos de uma comunidade que devem ser vacinados para se ter no mínimo  $\bar{X}$  indivíduos suscetíveis com uma probabilidade igual a  $1 - \alpha$ , em uma população com tamanho  $\bar{N}$ .

Inicialmente, trata-se de uma comunidade pequena, representativo de um aglomerado de famílias, considerando-se  $\bar{N} = 100$ . Tabela 3 resume o tamanho da amostra para diversos valores de  $x_0$ ,  $p$  e  $\alpha$ . O valor de  $p_m^c$  é dado na Tabela 1. Para um valor fixo de  $\alpha$  verifica-se, para  $p_m^c$  com diversos valores de  $x_0$ , que o esforço de vacinação decresce à medida que a fração de indivíduos suscetíveis aumenta (para  $\alpha = 0,10$  compare a segunda coluna da Tabela 2 com a terceira coluna da Tabela 3). Por exemplo, para  $x_0 = 0,01$  (doença altamente contagiosa,  $R_0 = 100$ ), deve-se vacinar apenas 1% (Tabela 2) da população, segundo a abordagem clássica; porém, na abordagem amostral, deve-se vacinar no mínimo 90% (Tabela 2). Contudo esta diferença diminui com o aumento de  $x_0$ . Por outro lado, o esforço de vacinação aumenta se a fração de indivíduos suscetíveis a serem vacinados aumentar, conforme mostram as segunda, quarta e sexta colunas para  $\alpha = 0,05$  da Tabela 2. Em geral, o esforço de vacinação situa-se acima da proporção de indivíduos suscetíveis que se deseja vacinar, sendo que a diferença é menor quanto maior for a proporção de suscetíveis (compare a sexta linha, correspondente a  $x_0 = 0,9$ ). Observe que o tamanho da amostra para uma proporção de indivíduos suscetíveis a serem vacinados a uma dada confiança sempre decresce com a proporção de indivíduos suscetíveis; exceto para proporções elevadas de vacinação e fração de indivíduos suscetíveis baixa. Por exemplo, na sexta e na sétima colunas da Tabela 3 o tamanho da amostra aumenta quando  $x_0$  aumenta de 0,01 para 0,05. Isso se deve ao fato de, para  $x_0 = 0,05$  deve-se vacinar todos os 5 indivíduos suscetíveis, que é mais difícil do que vacinar um único indivíduo suscetível para  $x_0 = 0,01$ .

$x_0$	$p_m^c(\alpha_1)$	$p_m^c(\alpha_2)$	$0,70(\alpha_1)$	$0,70(\alpha_2)$	$0,90(\alpha_1)$	$0,90(\alpha_2)$
0,01	95	90	95	90	95	90
0,05	81	70	92	89	99	98
0,1	69	64	85	81	96	95
0,2	59	56	81	79	96	94
0,5	42	40	77	75	94	93
0,9	12	11	72	72	91	91

Tabela 3. O tamanho da amostra necessário para se ter uma cobertura de vacinação  $p$  dos indivíduos suscetíveis, com probabilidade igual a  $1 - \alpha$ , em uma população com  $\bar{N} = 100$ . Usou-se três valores de  $p$  ( $p_m^c$ , 0,70 e 0,90) e dois, para  $\alpha$  ( $\alpha_1 = 0,05$  e  $\alpha_2 = 0,10$ ).

Trata-se agora de uma comunidade pequena, como um vilarejo, considerando-se  $\bar{N} = 1.000$ . Tabela 4 resume o tamanho da amostra para diversos valores de  $x_0$ ,  $p$  e  $\alpha$ . O valor de  $p_m^c$  é dado na Tabela 1.

$x_0$	$p_m^c(\alpha_1)$	$p_m^c(\alpha_2)$	$0,70(\alpha_1)$	$0,70(\alpha_2)$	$0,90(\alpha_1)$	$0,90(\alpha_2)$
0,01	696	645	850	812	963	946
0,05	602	578	786	766	946	935
0,1	553	536	763	749	936	928
0,2	495	483	744	734	926	920
0,5	358	353	723	718	914	911
0,9	96	95	708	706	905	904

Tabela 4. O tamanho da amostra necessário para se ter uma cobertura de vacinação  $p$  dos indivíduos suscetíveis, com probabilidade igual a  $1 - \alpha$ , em uma população com  $\bar{N} = 1.000$ . Usou-se três valores de  $p$  e dois, para  $\alpha$  (Tabela 3).

Trata-se de uma comunidade um pouco maior que um vilarejo, considerando-se  $\bar{N} = 10.000$ . Tabela 5 resume o tamanho da amostra para diversos valores de  $x_0$ ,  $p$  e  $\alpha$ . O valor de  $p_m^c$  é dado na Tabela 1.

$x_0$	$p_m^c(\alpha_1)$	$p_m^c(\alpha_2)$	$0,70(\alpha_1)$	$0,70(\alpha_2)$	$0,90(\alpha_1)$	$0,90(\alpha_2)$
0,01	5.762	5.585	7.657	7.507	9.369	9.286
0,05	5.229	5.150	7.311	7.241	9.190	9.148
0,1	4.982	4.928	7.218	7.169	9.136	9.106
0,2	4.607	4.571	7.147	7.114	9.093	9.073
0,5	3.411	3.394	7.074	7.058	9.048	9.037
0,9	926	922	7.025	7.019	9.016	9.013

Tabela 5. O tamanho da amostra necessário para se ter uma cobertura de vacinação  $p$  dos indivíduos suscetíveis, com probabilidade igual a  $1 - \alpha$ , em uma população com  $\bar{N} = 10.000$ . Usou-se três valores de  $p$  e dois, para  $\alpha$  (Tabela 3).

Finalmente, trata-se de uma comunidade grande, como uma pequena cidade ou um bairro de uma cidade grande, considerando-se  $\bar{N} = 100.000$ . Tabela 6 resume

o tamanho da amostra para diversos valores de  $x_0$ ,  $p$  e  $\alpha$ . O valor de  $p_m^c$  é dado na Tabela 1. O valor de  $\bar{Q}_m^c$ , para  $x_0 = 0,001$  que não consta na Tabela 2, é 498.

$x_0$	$p_m^c(\alpha_1)$	$p_m^c(\alpha_2)$	$0,70(\alpha_1)$	$0,70(\alpha_2)$	$0,90(\alpha_1)$	$0,90(\alpha_2)$
0,001	57.655	55.870	76.596	75.088	93.706	92.869
0,01	52.336	51.766	72.279	71.770	91.422	91.106
0,05	49.844	49.594	71.021	70.795	90.655	90.510
0,1	48.145	47.973	70.707	70.550	90.456	90.355
0,2	44.960	44.846	70.473	70.368	90.306	90.239
0,5	33.579	33.525	70.237	70.185	90.154	90.120
0,9	9.141	9.130	70.079	70.062	90.052	90.040

Tabela 6. O tamanho da amostra necessário para se ter uma cobertura de vacinação  $p$  dos indivíduos suscetíveis, com probabilidade igual a  $1 - \alpha$ , em uma população com  $\bar{N} = 100.000$ . Usou-se três valores de  $p$  e dois, para  $\alpha$  (Tabela 3).

As Tabelas 3 a 6 mostraram o número de indivíduos da população que devem ser vacinados para algumas combinações de  $\bar{N}$ ,  $x_0$ ,  $\delta$  e  $\alpha$ . Os valores encontrados mostram que os comentários sobre as aproximações são válidos (a primeira aproximação é boa mesmo para os casos extremos, embora neste caso ela tenha uma pequena sub-estimação, enquanto a segunda aproximação tende a super-estimar o tamanho da amostra necessário, mas pode ser considerada boa) e que a quantidade adicional amostral em relação á nominal em alguns casos pode ser bastante elevada.

## 4. Discussão

As infecções transmissíveis são as doenças causadas por agentes infecciosos transmitidas de uma fonte (indivíduos infectantes, animais, reservatórios) para um indivíduo suscetível direta (vias aéreas) ou indiretamente (hospedeiro ou vetor). Estas doenças, em geral as doenças tropicais, estão fortemente relacionadas com adversidades tanto ambientais quanto econômicas.

Em relação ao agente infeccioso, existem algumas propriedades epidemiológicas principais: a infectividade (ou transmissibilidade, ou contagiosidade), a patogenicidade, a virulência e a dose infectante. A infectividade está relacionada à capacidade do agente penetrar e multiplicar no novo hospedeiro, disseminando, assim, na comunidade. Por exemplo, o vírus da gripe é muito mais infectivo do que o vírus da rubéola. A patogenicidade é a capacidade do micro-organismo, uma vez instalado, resultar em sintomas em maior ou menor proporção. O sarampo, por exemplo, é muito patogênico, uma vez que praticamente todos os infectados apresentam sintomas e sinais específicos da doença. A virulência é a capacidade do vírus resultar em uma forma grave ou fatal da doença, por exemplo a raiva, em que todos os casos são fatais. Finalmente, a dose infectante é a quantidade do vírus necessária para iniciar uma infecção.

Em relação ao hospedeiro, este pode ser infectado por uma doença por causa da suscetibilidade, isto é, são os indivíduos que não possuem resistência ao agente

infeccioso, por isso podem contrair a doença quando em contato com o agente infeccioso. Outras peculiaridades são a resistência natural e a imunidade, sendo a primeira não-específica, e a outra, específica que resulta quando exposto ao agente. Dependendo das características fisiológicas e genéticas do hospedeiro pode resultar na presença de portadores (os infectados assintomáticos), que são os grandes responsáveis por manter a doença ao evadir-se dos mecanismos de controle. Por causa desta adaptação, pode-se induzir a uma resistência utilizando-se as vacinas.

As formas de interação do sistema hospedeiro-parasita influenciam na transmissão da infecção. Esta transmissão ocorre no meio-ambiente em que pessoas estão distribuídas geograficamente. Portanto, as flutuações do meio-ambiente, como as variações sazonais e presença de reservatórios fora o homem, e as condições sócio-econômicas, como a forma de agregação geográfica e o convívio social, alteram bruscamente as epidemias.

Todos estes fatores devem ser considerados em um modelo matemático para que se proponham medidas mais efetivas para a erradicação de infecções de transmissão direta. Mas modelos com um número elevado de parâmetros não são fáceis de analisar. Diante deste dilema, procura-se uma situação intermediária em que parâmetros considerados mais relevantes sejam escolhidos em detrimento de outros, para se obter um modelo matemático não tão complexo que permita um tratamento mais formal. Tendo em mente esta parcimônia matemática, estudou-se um modelo relativamente simples. Os seus resultados foram analisados para determinar as condições para a erradicação de uma epidemia.

Com o advento de vacinas como formas eficazes de controle de muitas doenças infecciosas, imaginou-se que elas poderiam ser erradicadas através de campanhas de imunização. Por exemplo, imaginava-se que a erradicação de sarampo poderia ser obtida no fim do século XX. Entretanto, a varíola foi a única doença considerada erradicada em todo o globo terrestre. Discute-se as razões que permitiram a sua erradicação, e compara-se com os resultados da epidemiologia matemática.

A varíola é causada por vírus (*Poxvirus variolae*) que é muito menos infectivo que outras infecções, como sarampo e rubéola. Acredita-se que um caso primário não resulte em mais que 3 a 5 casos, e em torno de 20% da forma severa resulta em morte. As infecções nos indivíduos apresentam duas viremias, e é na segunda viremia (nos órgãos ricos em tecido retículo-endotelial) que resulta em infecções leves e brandas ou em casos graves (o vírus aloja-se na pele). Somente casos clínicos de certa gravidade são realmente contagiosos, que são facilmente identificáveis pelos sintomas (assim como os casos benignos), sendo que indivíduos portadores (casos subclínicos) e os que apresentam forma leve da doença, assim como os convalescentes, não transmitem a infecção.

Estudos epidemiológicos mostram que a transmissão da varíola ocorre quando há contatos íntimos com pessoas doentes com certa gravidade, que são facilmente identificáveis pela intensidade do período pré-eruptivo, inclusive nas erupções abortivas. Por isso, as infecções secundárias ocorrem principalmente entre familiares, escolares e aqueles nas enfermarias dos hospitais, pois a única fonte de transmissão é o doente, não existindo nem hospedeiro intermediário, nem reservatório natural fora o homem.

Diante desse quadro, existem duas formas de controlar esta infecção. Em países

que apresentam um bom sistema de saúde pública, a doença pôde ser controlada através de isolamento de casos e a vigilância sanitária de contatos, como fizeram a Inglaterra e a Holanda. Em outros países, em geral, em desenvolvimento, o controle é feito através de campanhas de vacinação em massa. Porém, a vacinação pode resultar em complicações, podendo mesmo até matar. Por exemplo, em um surto em Nova York, em 1947, de um caso importado de México, ocorreram 14 casos graves com 2 mortes, tendo terminado por medidas de isolamento e vigilância sanitária. Entretanto, as vacinações em massa produziram no mínimo 48 mortes, dezenas de casos de vacínia generalizada e grave encefalite pós-vacínica, além de grande número de complicações dermatológicas. Em regiões onde a vacinação em massa foi introduzida, os novos casos surgem em decorrência de importação de doentes (ou enxertos). Nesta situação, as estratégias de controle podem ser as vacinações em anéis expansíveis, isolamento de casos ou vigilância sanitária de contatos.

Resumindo, a baixa transmissibilidade, doença facilmente identificável e que deixa marcas nos recuperados, a ausência de portadores e de reservatório natural e o período de incubação uniforme (10 a 14 dias) fizeram da varíola, talvez, uma infecção a mais fácil de ser combatida e erradicada do que outras doenças contagiosas. Assim, de certo modo, os resultados clássicos da epidemiologia matemática são verificados, isto é, um esforço de vacinação relativamente baixo.

Dessa forma, para que uma doença infecciosa possa se manter na natureza, o agente causador precisa ter a capacidade de evasão contra medidas de controle. Por exemplo, as infecções que resultam na morte adicional são as menos favorecidas para se perpetuarem, pois a letalidade da doença age contra a sua manutenção [8]. Por sua vez, as que não induzem à uma mortalidade adicional, porém deixam seqüelas estão na faixa intermediária quanto à capacidade de se perpetuarem na natureza. Estas infecções, para a sua erradicação, exigem um isolamento dos indivíduos infectantes, sendo este mais rígido quanto mais o valor de  $R_0$  estiver próximo de 2. Finalmente, as mais capacitadas para se perpetuarem na natureza estão as infecções que não deixam nenhuma marca e, ao mesmo tempo, devem ser brandas, isto é, apresentam sintomas suaves para confundir com outras infecções virais (como as exantemáticas). A brandura de uma infecção está relacionada com a forma de interação entre o parasita e o hospedeiro, sendo que a indução de imunidade é tanto um mecanismo de defesa do último quanto uma forma de perpetuação do primeiro. Mesmo que os resultados clássicos apontem a possibilidade de erradicação destas doenças, as infecções com estas características estão resistindo a intensos esforços de erradicação. Por isso, para estas doenças devem ser adotadas nova perspectiva de aplicação dos resultados epidemiológicos. Em outras palavras, deve-se entender o esforço de vacinação não ao número de indivíduos suscetíveis que devem ser vacinados, mas em determinar uma amostra que contenha exatamente o número de indivíduos suscetíveis que devem ser vacinados.

Assim, levando-se em consideração apenas o valor da razão de reprodutibilidade basal ( $R_0$ ) e o esforço necessário para a erradicação, dada pela porcentagem de indivíduos em uma comunidade ( $f_c \times 100\%$ ) que devem ser vacinados (ou isolados), pôde-se conjecturar que as infecções que perpetuaram e estão presentes nos dias atuais são aquelas brandas com seqüelas com  $R_0$  próximo de 2, ou não deixam seqüela

nenhuma. Pois, as doenças infecciosas sem estas características foram erradicadas na história da humanidade através de intervenções como isolamento ou emigração dos indivíduos saudáveis de locais de contaminação. A erradicação da varíola e a grande dificuldade de controlar sarampo coadunam-se com esta hipótese.

Sabe-se que a indução de imunidade por parte dos parasitas pode ser usada como forma de controle, através de vacinação, diminuindo sensivelmente a fonte de indivíduos suscetíveis. Em contrapartida, as doenças infecciosas brandas, para se perpetuarem na comunidade em que a vacinação tenha sido introduzida, deve apresentar valores elevados para a razão de reprodutibilidade basal. Pois quanto maior for o valor de  $R_0$ , maior será a dificuldade de efetivamente encontrar e vacinar, mesmo que seja em proporções baixas, os raros indivíduos suscetíveis em um ‘mar’ de indivíduos imunes. Atualmente, a rubéola é considerada erradicada em muitas regiões do mundo, enquanto o sarampo ainda resiste a intensas campanhas de vacinação. Pode-se explicar esta situação vigente comparando-se os valores da razão de reprodutibilidade basal. Usando-se os valores  $\mu = 0,017$  e para as forças de infecção para rubéola e sarampo, respectivamente,  $\lambda_0 = 0,0766$  e  $\lambda_0 = 0,25$  (todos os parâmetros, em  $anos^{-1}$ ) [1]; e a equação  $R_0 = 1 + \lambda_0/\mu$  [6], modificação da equação (2.6), tem-se, para a razão de reprodutibilidade basal, os seguintes valores: para sarampo,  $R_0 = 15,7$ , e para rubéola,  $R_0 = 5,5$ . Portanto, a fração de indivíduos suscetíveis para sarampo é  $x_0 = 0,064$ , e para rubéola,  $x_0 = 0,18$ . Estas são as doenças exantemáticas, em geral brandas e assintomáticas, que se confundem com outras exantemas. Por isso, deve-se aplicar a abordagem da amostragem. Assim, as Tabelas 3 a 6 mostram que o esforço de erradicação é muito grande, sendo muito maior para o sarampo para baixas coberturas ( $p$  pequeno, em torno de  $p_m^c$ ), porém diminuindo à medida em que se aumenta o esforço de vacinação ( $p = 0,90$ ).

As infecções virais, aquelas brandas, podem ser controladas por vacinas, se estas existirem. Em geral, quando se deseja introduzir campanhas de vacinação, as únicas informações disponíveis são relativas às forças de infecção. Deste valor pode-se calcular a fração de indivíduos suscetíveis, utilizando-se os resultados obtidos de modelo matemático. No início, sem informação, deve-se levar em consideração o esforço mínimo preconizado pela abordagem de amostragem. À medida que a vacinação ocorre na comunidade, os indivíduos suscetíveis tornam-se discriminados pelo fato de não terem registro na carteira de vacinação. Mesmo assim, ainda persiste o problema, pois nem todos que apresentam registro de vacinação na carteira de fato são imunes, devido às falhas primária e secundária das vacinas. Por isso, o esforço mínimo deve ser ainda embasado na metodologia de amostragem, porém dando mais peso para os valores obtidos usando-se o extremo inferior da proporção crítica  $p_m^c$ , do que os do extremo superior da proporção crítica  $p_M^c$ . Assim, apresentar o valor crítico  $p_c$  dentro de intervalo,  $p_m^c \leq p_c \leq p_M^c$ , não é uma fraqueza do modelo, pelo contrário, permite uma maior flexibilidade na aplicação dos resultados matemáticos para elaborar estratégias de vacinação.

Toda esta discussão foi feita considerando-se apenas o equilíbrio assintótico. Mostrou-se que, quando se aplica uma vacinação constante (a taxa de vacinação  $\nu$  é a mesma em todo tempo), duas possíveis configurações epidêmicas são possíveis: a erradicação da doença ( $\nu \geq \nu_c$ ) ou a persistência da doença em nível de prevalência

menor que a encontrada antes da vacinação ( $\nu < \nu_c$ ). Matematicamente, o equilíbrio assintótico é obtido após um tempo muito longo,  $t \rightarrow \infty$ , pois as variáveis podem estar sofrendo variações tão pequenas (como bilionésimos) sem alcançar a condição de equilíbrio. Biologicamente, a unidade de infecção é um indivíduo, e não fração de indivíduos. Por isso, o equilíbrio assintótico pode ser obtido em tempos finitos. O lapso de tempo entre o início da vacinação e a situação de equilíbrio diminui com o aumento do valor da taxa de vacinação  $\nu$  [7]; por isso, para  $\nu_c$  o tempo decorrido para resultar na erradicação da doença é a maior possível, diminuindo-se progressivamente com o aumento de  $\nu$ .

Finalmente, as equações (3.14) e (3.16) foram obtidas considerando-se que a força de infecção pós-vacinação em equilíbrio estava limitada entre dois valores extremos, ou  $0 \leq \lambda \leq \lambda_0$ . Por isso, impôs-se  $\lambda = 0$  e  $\lambda = \lambda_0$  para se obter estas equações. Como está-se estudando uma situação ideal (vacinação constante ao longo de tempo), a força de infecção (mesmo que possa oscilar, a sua média anual decresce) diminui gradativamente ao longo do tempo, por isso o valor de proporção crítica (média anual) ao longo do tempo situa-se entre os dois extremos dados pelas equações (3.14) e (3.16).

## 5. Conclusão

O modelo matemático não considerou fatores importantes na transmissão de infecções. Não se considerou, por exemplo, a mutação e a variabilidade antigênica de vírus, tão essenciais para evadir-se diante da resposta imunitária. Nem a perda de imunidade, e nem a presença de portadores, pois a presença dos transmissores “silenciosos”, assim como os reservatórios naturais, mantém a endemia em baixos níveis, enquanto a perda de imunidade diminui a pressão sobre o vírus para a sua perpetuação [6]. A influenza, um exemplo de vírus (influenza A, B e C) com variabilidade antigênica, e a gripe comum (rhinovírus), um vírus com centenas de sorotipos sem nenhuma reação cruzada, são doenças cuja erradicação é praticamente impossível.

Mesmo considerando uma única cepa de vírus, e todos os fatores relacionados à transmissão sendo homogêneos, foi possível mostrar a grande dificuldade para se controlar algumas doenças infecciosas. Portanto, os resultados matemáticos apresentados aqui representam a situação mais favorável para a erradicação das infecções, pois as características acima não foram consideradas. Por exemplo, a presença dos portadores e a perda de imunidade estão frustrando todas as previsões otimistas para a erradicação de muitas infecções. Questões sócio-demográficas, devido à intensa mobilidade resultante da globalização das economias dos países, são novos fatores complicadores no combate às doenças infecciosas.

Todos os fatores heterogêneos podem e devem ser levados em consideração para refinamento do modelo, com o propósito de obter resultados epidemiológicos mais acurados. Porém, os seus resultados desviarão um pouco do resultado basal obtido de modelo simplificado. Assim, resultados epidemiológicos obtidos de modelos simplificados, com hipótese de homegeneidade em todos os fatores, podem ser aplicados para elaborar esquemas de vacinação para erradicação de doenças. Conforme dis-

cutido neste trabalho, pode-se usar ou a abordagem clássica ou a metodologia de amostragem para determinar o esforço mínimo da vacinação, dependendo das características da doença entre os indivíduos na comunidade.

**Abstract.** Mathematical models, based on the accumulated knowledge related to the disease's transmission mechanisms, can be a useful tool, not only to describe the distribution of the infection in a community, but also to provide scenarios when controlling mechanisms, as vaccinations, are introduced in this community. One of the fundamental results provided by the model is related to the minimum vaccination effort, that is, the minimum number of individuals vaccinated to obtain the eradication of the infection. But, this result must be dealt with care in order to avoid misunderstandings.

## Referências

- [1] R.M. Anderson and R.M. May, "Infectious Diseases of Humans – Dynamics and Control", Oxford University Press, Oxford, New York, 1992.
- [2] J.J. Angulo, Varíola, em "Doenças infecciosas e parasitárias" (R. Veronezi, ed.), pp. 55-63, Ed. Guanabara Koogan S.A., Rio de Janeiro, 1991.
- [3] M.B.F. Leite, R.C. Bassanezi e H.M. Yang, The basic reproduction ratio for a model of directly transmitted infections considering the virus charge and the immunological response, *IMA J. Math. Appl. Med. Biol.*, **17 (1)** (2000), 15-31.
- [4] M.Z. Rouquayrol e N. Almeida, "Epidemiologia e Saúde", Ed. Médica e Cient. Ltda., Rio de Janeiro, 1999.
- [5] H.M. Yang, Modelling vaccination strategy against directly transmitted diseases using a series of pulses, *Jour. Biol. Systems*, **6 (2)** (1998), 187-212.
- [6] H.M. Yang, "Epidemiologia Matemática – Estudo dos Efeitos da Vacinação em Doenças de Transmissão Direta", EDUNICAMP e FAPESP, Campinas e São Paulo, 2001.
- [7] H.M. Yang, Modeling directly transmitted infections in a routinely vaccinated population – The force of infection described by Volterra integral equation, *Appl. Math. Comput.*, **122 (1)** (2001), 27-58.
- [8] H.M. Yang, Dinâmica de propagação de vírus: transmissibilidade, virulência e mecanismos de controle, em "Seleta do XXIV CNMAC" (A. Sri Ranga, ed.), TEMA, Vol. 3.1, pp. 223-232, SBMAC e FAPESP, São Carlos e São Paulo, 2002.
- [9] H.M. Yang, The effects of re-infection in directly transmitted infections modelled with vaccination, *IMA Jour. Math. Appl. Med. Biol.*, **19** (2002), 113-135.