

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE FILOSOFIA
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

Álvaro Hadad Filho

**Considerações sobre a estatística médica:
uma análise crítica do movimento “Medicina baseada em
evidências”**

São Paulo
2018

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE FILOSOFIA
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA DA CIÊNCIA

**Considerações sobre a estatística médica:
uma análise crítica do movimento “Medicina baseada em
evidências”**

Álvaro Hadad Filho

Dissertação apresentada ao Programa de Pós-Graduação em Filosofia do Departamento de Filosofia da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, para a obtenção do título de Mestre em Filosofia sob orientação do Prof. Dr. Pablo Rubén Mariconda

São Paulo
2018

Folha de Aprovação

HADAD FILHO, A. Considerações sobre a estatística médica: uma análise crítica do movimento “Medicina baseada em evidências”. Dissertação (Mestrado) – Faculdade de Filosofia, Letras e Ciências Humanas. Departamento de Filosofia, Universidade de São Paulo, São Paulo, 2018.

Este trabalho foi realizado com auxílio de bolsa da CNPq.

RESUMO

HADAD FILHO, A. Considerações sobre a estatística médica: uma análise crítica do movimento “Medicina baseada em evidências”. Dissertação (Mestrado) – Faculdade de Filosofia, Letras e Ciências Humanas. Departamento de Filosofia, Universidade de São Paulo, São Paulo, 2018.

O movimento “Medicina baseada em evidências” (EBM), surgido na década de 1990, encontrou rápida aceitação por parte da comunidade médica e dos sistemas de saúde. Entre suas principais características, encontram-se a exigência de que a prática clínica seja baseada na “melhor evidência disponível”, a hierarquização da evidência, a valorização dos ensaios clínicos e, sobretudo, o recurso extensivo a procedimentos de análise estatística. Neste trabalho, apresentamos a EBM, descrevemos seus conceitos e procedimentos centrais e indicamos alguns de seus antecedentes históricos. Damos especial atenção aos conceitos de randomização, significância estatística, evidência científica e eficácia terapêutica. Finalmente, desenvolvemos uma crítica às concepções de cientificidade e progresso defendidas pela EBM e a utilizamos como ponto de partida para tecermos considerações gerais acerca do estatuto epistemológico da medicina, do progresso médico e das funções que a estatística desempenha na medicina contemporânea.

Palavras-chave: Filosofia da ciência, Medicina baseada em evidências, estatística médica, eficácia terapêutica.

ABSTRACT

HADAD FILHO, A. Considerações sobre a estatística médica: uma análise crítica do movimento “Medicina baseada em evidências”. Dissertação (Mestrado) – Faculdade de Filosofia, Letras e Ciências Humanas. Departamento de Filosofia, Universidade de São Paulo, São Paulo, 2018.

“Evidence-based medicine” (EBM) is a medical movement whose first appearance dates back to the 1990’s. Since then, it has received wide acceptance from the medical community and international health systems. Among its most important characteristics, it is possible to indicate the demand to base the clinical practice on the “best current evidence”, the hierarchies of evidence, the valorisation of the randomized-controlled trials, and, especially, the extensive recourse to procedures of statistical analysis. This Master’s dissertation is intended to present the EBM movement, describe its main concepts and procedures, and identify some of its historical backgrounds. Special consideration is given to the concepts of randomization, statistical significance, scientific evidence, and therapeutic efficacy. Finally, we present some criticisms on the conceptions of medical science and medical progress defended by EBM proponents. We then use them as a starting point for the development of our own considerations about the epistemological status of medicine, the medical progress and the advancement of knowledge in the contemporary medical sciences.

Keywords: Philosophy of science, Evidence-based medicine, medical statistics, therapeutic efficacy.

SUMÁRIO

Introdução

Capítulo 1 - O que é “medicina baseada em evidência”?

1.1. Introdução

1.2 Motivações da EBM

1.3 Hierarquia da evidência científica

1.4 Ensaio clínico randomizado

1.5 Qualidade da evidência no início do século XXI

1.6 EBM, novo paradigma?

Capítulo 2 - Evidência científica e evidência na medicina

2.1 Introdução

2.2. Evidência científica

2.3 Evidência na medicina

2.4 Evidência e testes de hipóteses

2.5 Conclusão

Capítulo 3 - Ciência e progresso na medicina contemporânea

3.1 O estatuto epistemológico da medicina

3.2 Duas noções de progresso médico

3.3 As relações entre pesquisa laboratorial biomédica, pesquisa clínica e prática clínica

3.4 Conclusão

Conclusão

Referências bibliográficas

INTRODUÇÃO

Há um quadro do pintor belga James Ensor (1860-1949) chamado *Les mauvais médecins*. Foi exibido pela primeira vez em 1893, no Salão dos XX, em Bruxelas. Mostra um doente deitado em seu leito, vestido com trajes carnavalescos. Cinco médicos, portando por sua vez fraques e cartolas e munidos de instrumentos bizarros, como facas, serrotes e um saca-rolhas, retiram do abdome do doente um verme enorme, que se enrola nas pernas dos cirurgiões. São professores eminentes da Faculdade de Medicina da Universidade Livre de Bruxelas (Bluard, 1998). Um deles está guardando o pagamento pelos serviços no bolso do casaco de um colega. A morte observa pacientemente, parada à porta do quarto, esperando que os médicos façam por ela o seu trabalho. Em um caderno de notas largado no chão, lê-se: “Esqueci uma esponja no abdome; peritonite se manifestará” (*J’ai laissé éponge dans le ventre péritonite se déclarera*).

Toda uma série de registros da desconfiança em relação aos médicos e à medicina poderia ser levantada, mas por ora nos limitaremos a indicar dois exemplos ilustres. Na literatura, são muito conhecidas as sátiras de Molière (1622-1673) aos médicos da corte francesa, com seus doentes imaginários e explicações tautológicas da ação sonífera do ópio por uma suposta propriedade dormitiva. Já na filosofia, encontram-se nos ensaios de Montaigne (1533-1592) graves acusações de que os médicos deliberadamente prescrevem substâncias nocivas para seus clientes, a fim de torná-los de fato doentes e dependentes de seus cuidados.

Mas as críticas à medicina e as indicações da fragilidade do conhecimento médico não se limitam à opinião de “leigos”, ou às obras de filósofos e artistas. De fato, são recorrentes nos escritos médicos as afirmações de que certos tratamentos levariam à morte dos pacientes, além de exortações para uma retificação radical da prática. Assim, por exemplo, em 1731 o médico inglês Francis Clifton (†1736) observa os perigos do emprego de tratamentos cujos efeitos não foram cuidadosamente observados:

Se um Homem promete a cura das doenças, sem conhecer sua natureza, suas aparências e suas consequências, além dos melhores remédios em uso; ou se ele sabe dessas coisas, mas ainda negligencia a observar o caso como deveria, será uma sorte muito grande se ele não prejudicar a todos, exceto ele. (Clifton, 1731, p. 7)

Para Clifton, a medicina moderna encontrava-se em um estado deplorável cuja razão era o excesso de teorias e a busca por causas primeiras, que eram

inadequadamente tomadas como fundamento da prática. Situação que poderia ser resolvida pela retomada do método de investigação dos antigos, isto é, a observação e descrição sistemática dos casos dos pacientes e dos efeitos dos remédios. Esse apelo por um retorno à observação é ecoado pelo cirurgião escocês James Lind (1716-1794), outro importante representante da tradição empirista britânica do século XVIII:

Médicos mais velhos, referindo doenças muito variadas e incomuns a tais causas imaginárias, privam o mundo do verdadeiro avanço da sua arte: o qual só pode ser esperado de histórias acuradas de diferentes casos, relatadas fidedigna e honestamente, e distinguidas umas das outras com a mesma acurácia observada por escritores de botânica na descrição de diferentes plantas (Lind, 1757 [1753], p. 41).

Menos de um século depois dos escritos desses autores ingleses, a simples observação e descrição do curso das doenças e das ações dos medicamentos já não eram mais consideradas suficientes para garantir o avanço da medicina, ainda que continuassem a ser vistas como necessárias. Era preciso, por um lado, relacionar as diferentes manifestações mórbidas apresentadas em vida com as lesões anatômicas encontradas à dissecação do cadáver. Por outro, era preciso contar o número de lesões, de sintomas, de casos bem sucedidos e de óbitos de acordo com cada tratamento empregado. Somente após a coleta de uma série de casos, do registro cuidadoso de todas as variáveis relevantes e da comparação dos diversos casos de uma mesma doença com casos de outras doenças, seria possível fazer afirmações acerca do diagnóstico, do prognóstico e do tratamento das inúmeras entidades mórbidas. Isto é, a partir do século XIX não é mais suficiente para a tradição empirista a descrição qualitativa da doença. Uma ferramenta mais poderosa e precisa deve ser utilizada para auxiliar o médico na construção do conhecimento e na sua tomada de decisões: o cálculo de probabilidades.

Essas novas exigências são muito bem ilustradas pela obra do patologista francês P. C. A. Louis (1787-1872), que as aplicou de forma rigorosa em seus trabalhos. Para Louis (1837), a medicina poderia aspirar ao mesmo grau de exatidão da física e da química, mas, para isso, seria forçada a lançar mão, como essas ciências, da análise numérica. Para realizar um diagnóstico seguro, os médicos deveriam determinar a proporção em que casos de uma determinada doença apresentam um sintoma específico. Por exemplo, comparando a proporção de indivíduos tuberculosos que apresentam hemoptise com a proporção de indivíduos com outras doenças pulmonares que apresentam o mesmo sintoma, pode-se

concluir que a presença de sangue no escarro torna o diagnóstico de tuberculose extremamente provável. Similarmente, para decidir a respeito da eficácia de um método terapêutico, o número de óbitos entre doentes submetidos a esse tratamento deveria ser confrontado com o número de óbitos entre indivíduos em que a mesma doença seguiu a sua marcha natural. Procedimentos análogos valeriam também para a resolução de problemas de prognóstico e etiologia, demonstrando que o recurso à estatística seria adequado à resolução de qualquer questão de importância médica, o que a tornaria a única via para garantir o estatuto científico da medicina e assegurar o avanço do seu conhecimento.

Com essas breves referências, não queremos de forma alguma sugerir que as mesmas questões acerca do conhecimento médico e dos fundamentos da prática tenham sido colocadas por autores tão diversos, em períodos também muito distintos. De fato, podemos encontrar na formulação de problemas acerca da ciência médica e nas propostas de solução características típicas da medicina de uma época. O movimento “Medicina baseada em evidências”, surgido na década de 1990, que pretendemos estudar neste trabalho, pode ser visto como uma tentativa, entre tantas outras, de assentar a prática clínica sobre bases seguras, racionais e científicas.

Também os autores desse movimento apontam as insuficiências metodológicas da medicina de sua época, afirmam que o emprego de tratamentos cuja eficácia e segurança não foi demonstrada colocaria a vida dos pacientes em risco e apelam para uma reforma profunda da prática. David Sackett (1934-2015), um de seus principais teóricos, descreve da seguinte forma as suas motivações pessoais para ingressar na pesquisa clínica:

Durante meu treinamento de pós-graduação em medicina interna, quanto melhor eu me tornava em diagnosticar as enfermidades de meus pacientes, mais frustrado eu me tornava com a ignorância coletiva da minha profissão em relação a como eu deveria tratá-los, ou se eu deveria absolutamente tratá-los (Sackett, 2015, p 325).

Similarmente, Brian Haynes teria se aproximado intelectualmente do movimento durante uma aula sobre a teoria freudiana em que, perguntando ao palestrante pelas evidências a favor da teoria, obteve como resposta que não havia nenhuma. O jovem estudante sentiu um “formigamento” pelo corpo e se “(...) perguntou quanto da sua educação médica era baseada em teorias sem prova” (Smith; Rennie, 2014). Obviamente, esses autores não consideram que a medicina seja um caso perdido, mas encontram na sofisticação dos métodos de análise estatística, na

hierarquização da evidência e na execução de ensaios clínicos a melhor forma para assegurar a cientificidade da prática clínica.

De nossa parte, também tivemos a chance de nos espantar com a fragilidade do conhecimento médico e com a dificuldade de indicar um tratamento sem a certeza de seu sucesso, ou, no mínimo, um garantia razoável de sua eficácia. Nesse sentido, consideramos bem vinda a iniciativa da medicina baseada em evidências. Mas não podemos deixar de nos espantar também com a maneira ingênua e acrítica com que as propostas do movimento foram acatadas pela comunidade médica, talvez seduzida pela complexidade de suas fórmulas e linguajar matemático. Com este trabalho, pretendemos contribuir para a retificação de mais um engano, que consiste em atribuir à estatística uma importância para a medicina maior do que lhe cabe de fato e de direito. Para isso, inicialmente nos colocaremos dentro do discurso científico, para então nos afastarmos progressivamente dele.

Como sugerem as indicações acima, a história da medicina poderia ser contada pelo viés do erro, da ignorância e da luta contra o charlatanismo, e acreditamos que tal abordagem teria um poderoso alcance heurístico. No entanto, adotamos neste trabalho o procedimento oposto e tentamos retirar toda referência à possibilidade de desonestidade, má formação e má fé por parte dos médicos, além de não nos dedicarmos a questões de ordem material, a determinantes sociais e econômicos da medicina contemporânea. Reconhecemos de antemão a pertinência de qualquer crítica que por acaso considere as nossas análises excessivamente idealizadas. Mas interpretamos este trabalho como o primeiro momento de uma pesquisa mais ampla acerca da medicina contemporânea e de seus antecedentes históricos, em que tentamos separar, ainda que muito arbitrariamente, as questões mais diretamente relacionadas com o conhecimento de questões de natureza extra-epistêmica. Com essa escolha, pretendemos evitar que uma pertinente crítica ao afastamento da medicina em relação aos interesses diretos dos indivíduos doentes e sua possível subordinação a interesses comerciais nos fizesse cair em um relativismo fácil ou no niilismo terapêutico. De qualquer forma, acreditamos não ter deformado excessivamente o quadro da medicina atual. De fato, parece-nos que a pintura de Ensor poderia chamar-se, com justiça, simplesmente *Médecins*.

CAPÍTULO 1: O que é “medicina baseada em evidência”?

1.1 Introdução

A expressão “medicina baseada em evidência” (“*evidence-based medicine*”, doravante EBM) aparece pela primeira vez na literatura médica em 1991, em um editorial de Gordon Guyatt (1953-) para o *American College of Physicians Journal Club*. O autor descreve duas formas distintas de resolver uma questão clínica comum em hospitais, a determinação da causa da anemia de um paciente idoso. O “caminho do passado” consistiria em solicitar exames tradicionalmente utilizados para esse fim, cuja indicação o médico responsável pelo paciente havia aprendido desde a sua formação acadêmica. Caso houvesse alguma dúvida em relação aos resultados, o profissional poderia ainda recorrer à opinião de algum colega mais experiente, ou solicitar o parecer de uma autoridade no assunto, um especialista em doenças do sangue.

Por outro lado, no que descreve como o “caminho do futuro”, o médico procuraria ativamente em uma base de dados informações acerca das propriedades diagnósticas dos testes solicitados, como sensibilidade e especificidade, para decidir qual deles seria o mais adequado para o paciente em questão e para interpretar corretamente os possíveis resultados. A diferença entre a EBM, o “caminho do futuro”, e o procedimento tradicional poderia ser descrita nos seguintes termos:

O caminho do futuro descrito acima representa um importante avanço na inclusão de novas evidências na prática clínica. Anteriormente, os médicos eram ensinados a procurar por autoridades (seja um livro texto, um especialista, ou um médico local mais experiente) para resolver questões acerca do manejo de pacientes. A EBM utiliza estratégias adicionais, incluindo o rastreamento rápido de publicações de estudos que são diretamente relevantes para o problema clínico, a avaliação crítica desses estudos e a aplicação dos resultados dos melhores estudos para o problema clínico em questão (Guyatt, 1991, p. A16).¹

No ano seguinte, Guyatt publica, junto com alguns colaboradores, um artigo em que desenvolve mais detalhadamente as características da nova abordagem da prática e do ensino da medicina. Esse texto também se inicia com um problema clínico comum: um residente, diante de um paciente adulto com o primeiro episódio de convulsão, precisa informar-lhe os riscos de recorrência, isto é, de apresentar novos episódios de convulsão, e justificar a necessidade de utilizar antiepilépticos a longo prazo. Um médico formado pelo antigo procedimento recorreria a seus

¹ Todas as traduções apresentadas neste trabalho, com exceção de casos explicitamente indicados, são de nossa autoria.

preceptores para obter essa informação, confiando em sua resposta acerca do risco de recorrência, descrito simplesmente como “alto”. Para o médico formado pela nova abordagem, no entanto, essa resposta não seria satisfatória, o que motivaria uma pesquisa pela informação relevante. Dessa vez não se trata de procurar dados acerca da acurácia de testes diagnósticos, mas de encontrar estudos adequados à determinação do prognóstico de uma doença específica. Novamente, uma rápida e eficiente utilização de um banco de dados fornece uma resposta adequada: um estudo acompanhou pacientes até 4 anos após o primeiro episódio de convulsão e descobriu que o risco de recorrência em um ano está entre 43 e 51%, alto o suficiente para justificar a prescrição do anticonvulsivante.

Nesse artigo de 1992, os autores descrevem a EMB, o “caminho do futuro”, como um “novo paradigma”, fazendo referência direta às ideias de Thomas Kuhn (1922-1996):

A prática médica está mudando, e a mudança, que envolve utilizar a literatura médica de forma mais eficiente em guiar a prática médica, é suficientemente profunda para que possa ser adequadamente chamada de uma *mudança de paradigma* [paradigm shift] (Evidence-Based Medicine Working Group, 1992, p. 2420, grifo nosso).

No “antigo paradigma”, segundo a sua descrição, considerava-se suficiente como base do conhecimento médico as observações e a experiência clínica não sistematizadas, a opinião de especialistas e o raciocínio fisiopatológico oriundo das disciplinas biológicas básicas, como fisiologia, patologia e farmacologia. Essa base informaria tanto as decisões terapêuticas, quanto as considerações acerca da validade de testes diagnósticos e do prognóstico dos pacientes. O “novo paradigma”, por sua vez, apesar de reconhecer a importância da formação tradicional, considera a experiência e o raciocínio fisiopatológico insuficientes para a tomada de decisão na prática clínica e sublinha a necessidade de avaliar criticamente a literatura científica disponível, à procura de um conhecimento sistematizado, reproduzível e livre de vieses.

Neste primeiro capítulo, pretendemos apresentar a EBM e discutir seus princípios metodológicos e conceitos científicos, recorrendo a textos de reconhecida importância histórica em que esses conceitos e princípios receberam a sua formulação inicial, ou mais influente. Esperamos demonstrar que a sua caracterização como um “novo paradigma” não é adequada por, basicamente, três motivos: (1) a EBM não cria novos problemas para a pesquisa clínica, pois

permanecem os mesmos para cada especialidade particular, nem oferece novos meios de solucioná-los, como se esperaria de um novo paradigma no sentido kuhniano, mas apenas sistematiza uma metodologia apropriada a uma ampla área de pesquisa; (2) a aplicação à medicina de métodos estatísticos e epidemiológicos, uma das características fundamentais da EBM, vinha ocorrendo, no mínimo, desde a década de 1820, o que nos impede de chamá-los de “novos”; e (3) mesmo nas primeiras duas décadas do século XXI, a maior parte das condutas tomadas na prática clínica não são referendadas por estudos com o rigor metodológico exigido pela EBM, mas continuam sendo amplamente sustentados pela experiência clínica e pela opinião de especialistas; ou seja, não houve a substituição do paradigma vigente, mas há convivência das práticas tradicionais com a abordagem baseada em evidência. Ainda que uma análise profunda do conceito kuhniano de paradigma científico, discutido exaustivamente na literatura filosófica, não seja o foco do nosso trabalho, pretendemos utilizá-lo como um contraponto para, partindo da discussão do que a EBM não é, chegar a uma caracterização mais precisa do movimento.

Certamente não somos os primeiros a oferecer uma caracterização da EBM. Em 1996, por exemplo, David Sackett (1934-2015), epidemiologista clínico de reconhecida importância por desenvolver princípios de avaliação crítica da literatura que seriam centrais para EBM, como teremos a oportunidade de discutir mais à frente, publica um artigo intitulado justamente *Medicina baseada em evidência: o que ela é, o que ela não é*. Assumindo um tom mais moderado do que o de Guyatt, que havia sido seu aluno, inicia o artigo reconhecendo “as origens filosóficas da EBM na Paris de meados dos século XIX” (Sackett *et al.*, 1996, p. 71) e indicando o grande impacto que o movimento havia causado no meio médico desde a publicação dos seus primeiros documentos. Em apenas 5 anos, cursos de graduação e pós graduação haviam incorporado seus princípios, palestras sobre como praticá-la e ensiná-la eram frequentes, e jornais científicos haviam se adaptado para publicar a informação nos moldes propostos. Mas diversas críticas foram feitas, e era necessário responder a elas. Basicamente, pode-se dizer que a maior parte referia-se a uma depreciação da experiência clínica, o que o autor se apressa para esclarecer, reconhecendo a sua importância, mas reiterando que a simples experiência e a observação não eram suficientes para o esclarecimento da prática, sendo necessário recorrer a procedimentos experimentais.

Porém, o que mais nos interessa é a caracterização do movimento proposta por Sackett e que transcrevemos a seguir:

EBM é o uso consciencioso, explícito e judicioso da melhor evidência vigente [*current*] na tomada de decisões acerca do cuidado de pacientes individuais. (...) Por melhor evidência clínica externa disponível, nós queremos dizer pesquisa clinicamente relevante, frequentemente oriunda das ciências básicas da medicina, mas especialmente da pesquisa clínica centrada no paciente sobre a acurácia e precisão de testes diagnósticos, o poder de marcadores prognósticos e a eficácia e segurança de regimes terapêuticos, preventivos e de reabilitação. (Sackett et al., 1996, p.71-2)

Como já deve ter ficado claro, a exigência de utilização da melhor evidência científica disponível abrange as três funções fundamentais da medicina: diagnóstico, prognóstico e terapêutica. A procura pela evidência visa à aplicação ao cuidado do paciente individual, e inicialmente não às políticas públicas ou à formação médica por si mesma. Mas a noção de “melhor evidência disponível” não é trivial, e é necessário analisar o que os autores entendem de fato com essa expressão para compreender finalmente em que evidência a medicina deveria ser baseada. Mas antes de passarmos à discussão das hierarquias de evidência, gostaríamos de dizer algumas palavras sobre as motivações relatadas para o desenvolvimento da EBM.

1.2 Motivações da EBM

O reconhecimento de que uma porção expressiva da pesquisa médica continha graves erros no que concerne ao desenho dos estudos e à aplicação de métodos estatísticos se tornou particularmente alarmante na década de 1980. Por exemplo, em estudos que visavam determinar o risco de recorrência de convulsões febris em crianças, a maior parte dos autores não se atentava para as diferenças nos resultados decorrentes da seleção da população estudada. Assim, os dados eram extremamente conflitantes, porque alguns dos artigos baseavam-se em crianças referenciadas para centros de alta complexidade, cuja gravidade inicial da doença já colocava em maior risco de recorrência, enquanto outros baseavam-se na população geral, com um risco de recorrência mais baixo (Altman, 1982, p. 61). Ou seja, a própria elaboração dos estudos dava margem ao surgimento de vieses, isto é, tendências a desvios sistemáticos nos resultados que não deviam ser atribuídos à variável em questão; nesse caso específico, um viés de seleção da amostra (*sample selection bias*). Apesar disso, alguns autores indicavam o uso profilático de anticonvulsivantes para todas as crianças com convulsões febris, uma indicação que se tornava incorreta à análise específica de cada população. Similarmente, estudos

que pretendiam determinar o risco de efeitos colaterais de medicações poderiam ter conclusões falsamente negativas por incluir uma amostra muito pequena de pacientes, insuficiente para indicar efeitos colaterais graves, mas relativamente raros.

Um exemplo da aplicação inadequada de ferramentas estatísticas levantado na época é a utilização do teste de hipóteses para avaliação de múltiplos desfechos clínicos em um mesmo ensaio. No caso do ensaio clínico de um medicamento, espera-se demonstrar que existe uma diferença estatisticamente significativa entre o grupo submetido ao tratamento pesquisado e um grupo controle quanto a um desfecho clínico importante, como invalidez, exacerbações da doença ou morte. Uma diferença estatisticamente significativa é arbitrariamente definida, na maior parte dos casos, como uma diferença que não se deve ao acaso com mais de 95% de chance. Dito de outro modo, em apenas 5% dos casos os resultados do ensaio seriam produzidos se os dois grupos fossem retirados da mesma população, ou seja, fossem iguais.² As hipóteses concorrentes são chamadas “hipótese principal”, que afirma não haver diferença entre os grupos, e “hipótese alternativa”, que afirma o contrário, que há diferença entre os grupos. Nesse caso, a chance de se cometer o erro denominado “tipo I” e rejeitar a hipótese principal quando não há diferença real entre os dois grupos é de 5% para cada desfecho analisado. No entanto, quando se aumenta o número de testes realizados, aumenta paralelamente o risco de cometer o erro tipo I e erroneamente atribuir eficácia ao medicamento em pelo menos um dos desfechos (Pocock; Hughes; Lee, 1987). Assim, a testagem de dois desfechos aumentaria esse risco para cerca de 10% e de três desfechos para cerca de 14%.

Esse tipo de problema colocava em dúvida a validade e a credibilidade de muitos estudos cujos resultados poderiam ter consequências indesejadas, se levados para a prática clínica. Além disso, os autores da EBM preocupavam-se com a grande variabilidade das condutas médicas, mesmo no manejo de doenças muito comuns, como a hipertensão. Encaminhando pacientes com diagnóstico recente de hipertensão para tratamento, observou-se que um terço deles não recebia anti-

² “(...) significância *estatística* nos informa meramente se uma diferença é provavelmente *real*, e não se ela é importante ou grande. Mais precisamente, a significância estatística de uma diferença não é nada além de uma afirmação da probabilidade dessa diferença se dever ao simples acaso.” (Department of clinical epidemiology and biostatistics, McMaster University, 1981b, p. 1159, grifo dos autores).

hipertensivos. A pesquisa pelas razões dessa discrepância demonstrou que, além de questões técnicas, como os níveis iniciais de pressão arterial, um determinante da prescrição ou não de medicamentos era o ano de formação do médico responsável. Médicos com formação mais antiga tendiam a prescrever menos anti-hipertensivos, mesmo quando indicados, o que demonstrava uma rápida deterioração nos conhecimentos com os anos de formação. Somando isso ao crescimento exorbitante da literatura científica, tornava-se necessário desenvolver meios de facilitar o acesso à nova informação veiculada nos periódicos. A partir de 1981, Sackett publica, através do departamento de epidemiologia clínica e bioestatística de Universidade de McMaster, uma série de artigos para capacitar os profissionais da saúde a realizar uma triagem eficaz da literatura, à procura da pequena parcela de estudos com metodologia adequada e com resultados potencialmente aplicáveis à prática clínica. Esses artigos, intitulados *How to read clinical journals*, apresentam de forma sistematizada os critérios de qualidade dos tipos específicos de estudos clínicos, com algoritmos que permitiam aos leitores excluir estudos com falhas metodológicas sem precisar ler todo o seu conteúdo. Esse procedimento, denominado “avaliação crítica” (*critical appraisal*), de análise da qualidade da literatura científica se tornaria uma estratégia central da EBM.

A criação em 1991 do *American College of Physicians Journal Club*, a que nos referimos no início deste capítulo, era uma resposta a essa dificuldade de se manter atualizado em relação às publicações da literatura médica. Os editores pretendiam servir-se da estratégia de avaliação crítica de estudos para selecionar os artigos de melhor qualidade técnica e maior aplicabilidade na área de clínica médica e, assim, abreviar o trabalho dos profissionais que normalmente dispunham de pouco tempo para a leitura de periódicos especializados (Haynes, 1991). Além disso, a divulgação dos artigos era pensada de forma a tornar mais rápida e eficiente a leitura, introduzindo no título informações importantes, como o tipo de desenho do estudo, e publicando um sumário com os principais resultados e aspectos metodológicos.

Particularmente reiterada era a necessidade de se abandonar a experiência e o pensamento fisiopatológico como bases suficientes para prática clínica. Um dos exemplos que os autores utilizavam frequentemente para ilustrar esse ponto era a administração rotineira de anti-arrítmicos em casos de infarto do miocárdio com

arritmias ventriculares. Partindo da evidência de que pacientes que manifestavam tais arritmias apresentavam uma mortalidade maior nos dias subsequentes ao infarto, parecia lógico deduzir que o uso de substâncias capazes de reverter as alterações do ritmo diminuiria essa taxa. E, de fato, essa foi uma prática disseminada, até que se demonstrou, através de um ensaio clínico randomizado que, ao contrário do que se esperava, a mortalidade era ainda maior nos pacientes submetidos a esse tratamento (Sackett, 1995). Similarmente, considerando que os altos níveis de colesterol eram um dos principais fatores de risco para infarto do miocárdio, esperava-se diminuí-lo através de medicações hipolipemiantes, como o clofibrato. Mas também nesse caso foi demonstrado que, apesar da eficácia em diminuir os níveis sanguíneos de lipídios, o uso de clofibrato era associado com um aumento da mortalidade total (Sackett; Haynes; Tugwell, 1985, p. 180-1). Exemplos como esses sublinhavam a diferença entre o raciocínio fisiopatológico, típico das ciências biológicas básicas, e o raciocínio clínico, e justificavam a exigência de ensaios clínicos randomizados para a utilização rotineira de novas terapêuticas.

Portanto, podemos encontrar na literatura o relato de 4 motivações básicas para o desenvolvimento da EBM: (1) o reconhecimento da baixa qualidade metodológica da maior parte dos estudos clínicos publicados na literatura médica; (2) a necessidade de habilitar os profissionais com estratégias de análise crítica da literatura para responder a questões levantadas pela prática clínica e (3) de mantê-los atualizados em suas áreas de atuação; e, finalmente, (4) a adaptação dos periódicos aos critérios de qualidade defendidos pelo movimento. Indicados os problemas metodológicos a que o movimento visava responder, podemos passar ao estudo da hierarquização da evidência.

1.3 Hierarquia da evidência científica

A melhor evidência depende, em primeiro lugar, do tipo de pergunta clínica a ser respondida. Caso se trate de uma questão de diagnóstico, como a de saber se um novo exame laboratorial é adequado para a detecção precoce de determinada doença, deve-se recorrer ao desenho (*design*) em que há comparação do exame em análise com um método diagnóstico considerado “padrão ouro” (*gold standard*) para

aquela doença.³ Idealmente, os examinadores devem estar “cegos” (*blind*) quanto ao resultado do exame padrão ouro quando avaliam os resultados do novo teste, para que não sejam de alguma forma influenciados por essa informação. Caso se trate de uma questão de prognóstico, como a de saber, por exemplo, qual a expectativa de vida de um paciente com câncer, deve-se procurar por um coorte, um estudo em que diversos indivíduos com o mesmo diagnóstico são acompanhados ao longo do tempo desde um estágio inicial da doença, e em que todos os desfechos relevantes são registrados. Finalmente, em se tratando de uma questão terapêutica, como a de decidir se um medicamento é realmente eficaz para a cura de uma doença ou para a prevenção de suas complicações, considera-se como o estudo mais forte o ensaio clínico randomizado (*randomized control trial*, doravante RCT).

Como existem diversas nuances da análise e da estruturação de cada um desses desenhos específicos, limitaremos a discussão aos testes terapêuticos, sobretudo o ensaio clínico. Essa escolha é motivada pelo reconhecimento de que a terapêutica eficaz é o que normalmente se espera com mais fervor da medicina contemporânea, além ser nessa área que surgem as questões mais polêmicas acerca da evidência.

Uma das primeiras tentativas de hierarquização da evidência clínica foi publicada em 1979 em um documento da *Canadian Task Force on the Periodic Health Examination*. Os membros do grupo, de que fazia parte David Sackett, realizaram uma revisão da literatura sobre o exame periódico de saúde, prática através da qual são solicitados exames para diagnóstico precoce de pacientes inicialmente assintomáticos. Naquela época, como até hoje em muitos casos, a maior parte dos médicos solicitava uma bateria de exames indiscriminadamente para uma ampla faixa etária de pacientes, com uma frequência bastante elevada, ou anual. O grupo demonstra não haver embasamento para tal prática e defende um uso racional dos exames preventivos, baseado nas informações de estudos clínicos, e individualizado para cada faixa etária e gênero de pacientes. São analisados mais de 70 procedimentos e exames de triagem, e as recomendações são classificadas

³ Os exames considerados padrão ouro em geral já tem a acurácia diagnóstica bem estabelecida e normalmente são o exame anatomopatológico, a autópsia, ou algum exame de imagem invasivo, como a angiografia. Espera-se de um novo procedimento diagnóstico que seja menos invasivo, possua menos efeitos colaterais ou que consiga detectar a doença precocemente.

paralelamente de acordo com a efetividade (*effectiveness*) e o suporte da intervenção.⁴

“A efetividade da intervenção foi graduada de acordo com a qualidade da evidência obtida (...)” (Canadian Task Force on the Periodic Health Examination, 1979, p. 1195). Eram classificadas como “I” as recomendações que provinham de pelo menos um RCT; como “II-1”, de coortes ou estudos caso-controle; como “II-2”, de ensaios com controle histórico; e como “III” aquelas que eram sustentadas somente pela opinião de especialistas e pela experiência clínica.

Dedicaremos uma seção ao RCT, mas adiantaremos alguns pontos básicos da sua estrutura para melhor compreensão da classificação apresentada. O ensaio clínico pretende demonstrar a relação causal entre uma intervenção, como a administração de um anti-hipertensivo, e um desfecho clínico pré-determinado, como a diminuição dos níveis pressóricos. No mínimo dois grupos são comparados: o primeiro, chamado “grupo teste”, recebe a intervenção em análise, e o segundo, chamado “grupo controle”, não a recebe. Espera-se que os dois grupos sejam idênticos em todos os aspectos (idade, gênero, gravidade da doença etc.) e difiram apenas quanto à intervenção estudada. Para que isso ocorra, utiliza-se o recurso da randomização (*randomization*), que consiste em alocar aleatoriamente os participantes nos dois grupos, ou, dito de outro modo, em garantir que cada um dos participantes tenha a mesma chance de 50% de receber ou não a intervenção.

Os estudos do tipo coorte acompanham pacientes durante um longo período a partir do diagnóstico ou do início do tratamento, sendo, por isso, descritos como estudos *prospectivos*. Normalmente utilizados para esclarecer aspectos de prognóstico, podem fornecer informações acerca da eficácia terapêutica quando conseguem comparar grupos que, apesar de não terem sido deliberadamente alocados para isso, receberam tratamentos distintos. Isso ocorre porque, não havendo um consenso quanto ao melhor tratamento para algumas doenças e diferindo os recursos técnicos de instituições e países, pacientes com o mesmo diagnóstico são manejados de formas distintas. Já o estudo caso-controle, contrariamente ao coorte, é um estudo *retrospectivo*, ou seja, parte de determinado desfecho, como a cura ou a morte, e analisa a quais intervenções os pacientes

⁴ Em um glossário dos termos utilizados ao longo do texto, os autores expõem o que entendem por *efetividade*: “(...) o atributo de uma intervenção ou manobra que resulta em mais benefício do que malefício [*more good than harm*] para aqueles a quem elas são oferecidas.” (Canadian Task Force on the Periodic Health Examination, 1979, p. 1203).

havia sido submetidos, procurando diferenças estatisticamente significantes entre grupos. Como a alocação dos indivíduos para receber um ou outro tipo de tratamento não é randomizada, os coortes e casos-controles são sujeitos a inúmeros vieses, que diminuem a força da evidência que produzem.

Quanto ao segundo critério de classificação, as recomendações eram agrupadas de “A” a “E”: (A) “boa” (*good*) evidência disponível para indicar uma intervenção; (B) evidência “razoável” (*fair*) para a intervenção; (D) evidência razoável para excluir a intervenção do exame periódico de saúde; e (E) boa evidência para a exclusão. Em relação às recomendações classificadas como “C”, o grupo compreendia que a evidência disponível não era suficiente nem para indicar uma intervenção, nem para contra-indicá-la. Os autores procuraram conferir uma recomendação “A” somente a intervenções que tinham sido sustentadas por RCT, ou cuja efetividade era de classe “I”, o que poderia tornar as duas categorias redundantes. Mas, diante de doenças com alta mortalidade para as quais havia um tratamento eficaz, ainda que nunca avaliado por um RCT, julgaram apropriado conferir-lhes uma recomendação “A”, mesmo que recebessem classificação “II”. De qualquer modo, o que se percebe nessa classificação, excluindo-se os aspectos mais técnicos, é a tentativa de sistematização da qualidade da evidência para a prática clínica e o reconhecimento do RCT como a fonte de informação mais forte e confiável para o seu embasamento.

Diversas tentativas de aprimorar a hierarquização da evidência se seguiram na literatura médica. Em 1989, Sackett publica um documento referente ao uso de anticoagulantes, que mantém a dupla classificação das recomendações quanto ao seu embasamento e indicação. Agora, a primeira classificação, que diz respeito à validade e à credibilidade dos estudos que sustentam uma recomendação, é chamada de “nível de evidência” (*level of evidence*), e a segunda, que explicita a conclusão do grupo em relação à força de uma recomendação, ganha o título de “grau de recomendação” (*grade of recommendation*). Inúmeras variações dessa categorização foram publicadas por sociedades de especialistas, visando atender às especificidades de sua área de atuação, mas a estrutura geral da classificação dupla e paralela em níveis de evidência e graus de recomendação se mantém.

Esse aprimoramento da categorização da evidência ocorreu simultaneamente a uma tomada de consciência por parte dos proponentes da EBM de que realizar

uma avaliação crítica da literatura para todas as decisões importantes na prática clínica seria impossível para cada profissional isoladamente. Isso reforçou a necessidade de se publicar textos em que um grupo de autores revisa a literatura relevante para um assunto específico, como o tratamento da hipertensão, e publica recomendações sobre o manejo dessa doença. Tais textos, chamados “diretrizes clínicas”, já existiam antes da EBM, mas normalmente expressavam as opiniões dos especialistas da área e eram considerados enviesados, por não expressarem claramente os métodos de pesquisa da informação e não oferecerem uma sistematização da evidência. A partir dos textos a que nos referimos acima, paulatinamente as sociedades de especialistas vão se adaptando a esse modelo e passam a publicar os seus protocolos discriminando a qualidade da evidência e a força das recomendações.

Atualmente, o sistema de classificação da evidência mais influente é conhecido pelo acrônimo GRADE (*Grades of recommendation, assessment, development, and evaluation*). Publicado em sua primeira versão em 2004, ele responde a algumas deficiências dos modelos pregressos, sobretudo à crítica de que simplesmente considerar o RCT como melhor fonte de evidência poderia levar a erros grosseiros, já que muitos desses estudos apresentam graves falhas metodológicas. O GRADE continua considerando os estudos experimentais, isto é, os RCTs, como melhor fonte de evidência, comparativamente aos estudos observacionais – coortes e casos-controles –, mas passa a operar por um sistema de pontuações em que cada estudo perde ou ganha pontos de acordo com critérios específicos.⁵ A qualidade final da evidência é o resultado dessa soma e subtração de pontos. Além disso, o GRADE apresenta uma elaboração mais refinada dos critérios a serem levados em consideração para definir a força de uma recomendação, como a relação entre benefícios e efeitos colaterais, os custos da intervenção e as preferências dos pacientes. Assim, uma intervenção pode ter a eficácia sustentada por um bom estudo clínico, mas receber uma recomendação mais fraca por se considerar seus efeitos colaterais excessivamente altos.

É interessante notar que a maior parte dos sistemas de classificação da evidência atualmente disponíveis considera os resultados de estudos nas ciências

⁵ Entre esses critérios positivos, poderíamos citar a forte associação estatística entre a intervenção e o desfecho clínico e gradiente dose resposta; e como critérios de subtração de pontos, inconsistência e alta probabilidade de viés.

biológicas básicas como a evidência mais fraca para guiar a prática clínica. De fato, algumas representações da hierarquia da evidência em forma de pirâmides colocam esses estudos em sua base, ao passo que os RCT ocupam o seu topo. Discutiremos no capítulo 3 como essa avaliação da importância dos estudos básicos difere de uma concepção muito influente no século XIX, que considerava a pesquisa fisiológica laboratorial como o fundamento para o avanço do conhecimento médico.

Apresentados os aspectos essenciais do sistema de classificação e hierarquização da informação científica desenvolvido pela EBM, podemos passar ao estudo mais detalhado do RCT, considerado a fonte de evidência mais forte para a prática clínica. Iniciaremos o estudo por um breve histórico dos métodos de comparação entre grupos e interpretação estatística dos resultados, para então prosseguirmos com a análise de suas características.

1.4 Ensaio clínico randomizado

Em 1747, James Lind (1716-1794), cirurgião escocês, realiza a bordo do navio *Salisbury* da marinha britânica uma das primeiras comparações de tratamentos de que se tem registro na literatura médica. Em meados do século XVIII, médicos e marinheiros deparavam-se com uma doença relativamente nova, o escorbuto, que começou a ser extensamente descrita a partir do final do século XV, com o início das grandes navegações. Atualmente compreendido como uma deficiência de vitamina C, resultado da dieta de longos meses em alto mar, carente em frutas e verduras frescas, o escorbuto dizimou a tripulação de importantes navios, mas as principais autoridades médicas discordavam acerca de suas possíveis causas e tratamentos. Lind pretende indicar as suas observações coletadas em alto mar, e não “(...) propor nada ditado meramente pela teoria; mas confirmar tudo pela experiência e pelos fatos, os mais seguros e infalíveis guias” (Lind, 1757, p. 147). Tomando 12 pacientes cujos “(...) casos eram tão similares como poderia ter” (Lind, 1757, p. 149), Lind divide-os em seis grupos de dois e oferece a cada um desses grupos um tratamento distinto: sidra (*cider*), elixir vitriol, laranja e limão, vinagre, água do mar, e um electuário preparado com alho, semente de mostarda e bálsamo do Peru. Além de assegurar-se que a gravidade dos casos era comparável, Lind toma precauções para que outros fatores não influenciassem nos tratamentos propostos. Assim, aloja os pacientes no mesmo ambiente e administra exatamente a mesma dieta, com

exceção dos tratamentos específicos. Esses cuidados podem ser compreendidos quando se atenta para as teorias mais fortes da época, compartilhadas inclusive por Lind, que encontravam nas condições climáticas, como frio e umidade excessiva, as principais causas para essa doença.

Com exceção do grupo a que se administrou frutas cítricas, o tratamento durou 14 dias (*a fortnight*). Apesar das frutas terem acabado em apenas seis dias, a melhora dos pacientes submetidos a esse regime era clara: “A consequência foi que os efeitos positivos mais imediatos e visíveis foram percebidos no uso de laranjas e limões; um desses que os recebeu, estando pronto para o trabalho no final de seis dias.” (Lind, 1757, p. 150). O segundo paciente desse grupo também apresentou uma melhora expressiva e passou a ajudar no cuidado dos demais doentes.

Apesar desse importante resultado, Lind não desenvolve uma teoria que associa o escorbuto a uma simples deficiência alimentar e continua compreendendo a doença dentro do quadro de uma teoria humorista, em que as condições climáticas eram as principais responsáveis pela supressão da perspiração (*perspiration*), que levariam a um acúmulo de humores e sua consequente putrefação no organismo. As frutas e verduras teriam sua influência explicada pela propriedade de dissolver esses humores e facilitar a sua excreção. Além disso – e o que mais nos interessa no momento –, Lind tampouco transforma os resultados da sua pesquisa em preceitos metodológicos e não afirma que os tratamentos devessem ser todos submetidos a um teste parecido, continuando a indicar diversos medicamentos que não havia testado, como bem observado por Ulrich Tröhler (2005).

Também falta à obra de Lind o recurso à análise matemática que caracterizaria os ensaios terapêuticos dos séculos seguintes. A aplicação de tal expediente pode ser encontrada de forma sistematizada já em 1825, na obra do patologista francês Pierre-Charles-Alexandre Louis (1787-1872), intitulada *Recherches anatomico-pathologiques sur la phthisie*. Nessa obra, Louis procura esclarecer as lesões extra pulmonares que acometem os pacientes tuberculosos, considerando que o comprometimento dos pulmões já estava bem estabelecido. Recorrendo a um procedimento típico da clínica, que consiste em relacionar lesões encontradas à autópsia com a sintomatologia apresentada em vida pelo paciente, Louis demonstra que as lesões características da tísica, os tubérculos, podem ser encontrados nos mais diversos órgãos, como intestino, cérebro e baço. No entanto,

o que esse texto apresenta de mais importante é o recurso constante a uma avaliação numérica dos casos, na tentativa de determinar a frequência relativa em que cada lesão ou sintoma era encontrado nos pacientes tuberculosos: “Nós relatamos a *proporção* na qual cada uma das lesões tinha sido observada; de modo que nossa obra é, em todas as suas partes, uma espécie de *estatística da física*” (Louis, 1825, p. xxii, grifo nosso). Além disso, Louis também lança mão da comparação entre grupos, já que investiga as lesões e sintomas de pacientes mortos por outras causas, a fim de saber quais alterações poderiam ser consideradas específicas de cada doença.

Assim como Lind, Louis afirma ater-se aos fatos, depreciando de maneira ainda mais forte os sistemas médicos que tiram conclusões incompletamente sustentadas pela experiência.⁶ Mas, ao contrário daquele, pretende conferir às suas pesquisas o rigor de um método. Confiando que a análise de um grande número de casos poderia compensar as incertezas oriundas das variações do quadro clínico dos doentes, afirma que o seu método poderia produzir resultados exatos e oferecer uma certeza comparável à da matemática, já que

(...) o edifício da medicina repousa inteiramente sobre os fatos e que a verdade só pode emergir daqueles que foram bem e completamente observados. Então, e somente então, pode-se ver, numa série de observações, os dados de um problema desconhecido para muitos, do qual é necessário encontrar o valor: e como na matemática esse valor não muda com as pessoas que se ocupam da solução do problema, deve-se necessariamente, também na medicina, obter resultados idênticos da análise de muitas observações (Louis, 1825, p. xvii).

Se as investigações de 1825 acerca da tuberculose interessam sobretudo ao diagnóstico da doença, vemos em 1828 Louis expandir o seu método para questões de tratamento. No início do século XIX, as sangrias haviam se tornado uma das principais ferramentas terapêuticas dos médicos parisienses, intensamente influenciados pelo sistema de François Broussais (1772-1838). A retirada do sangue dos órgãos inflamados, através da secção de veias ou da aplicação de sanguessugas, era uma consequência de suas teorias, que encontravam nas inflamações a causa de diversas doenças. Louis pesquisa a eficácia desse recurso

⁶ Segundo Woillez (1873), a crítica de Louis aos sistemas médicos visava especialmente a fisiologia de Broussais, extremamente influente na França do começo do século XIX.

terapêutico no tratamento da pneumonia, comparando a mortalidade de pacientes submetidos à sangria nos 4 primeiros dias de doença e após esse período.⁷

Os cuidados de Louis na preparação dos grupos é notável. Após esclarecer os critérios para o diagnóstico de pneumonia (febre, dor torácica, catarro purulento) e da convalescência (retomada da alimentação, defervescência por 3 dias), informa que a média de idade dos pacientes nos dois grupos era comparável (33 e 36 anos), assim como a gravidade da doença. Como resultado, Louis fornece uma mortalidade de 3/7 para os paciente sangrados até 4 dias, contra uma mortalidade de 1/4, para aqueles sangrados após essa data. Apesar desse resultado parecer hoje extremamente plausível, contrariava uma prática amplamente disseminada e sustentada por uma das teorias patológicas mais importantes da época.

O “método numérico”, como Louis o chamava, também deveria ser aplicado a todos os campos do conhecimento clínico: diagnóstico, prognóstico e terapêutica. Isso fica claro em um texto de caráter metodológico publicado pelo autor em 1837, intitulado *De l'examen des malades et de la recherche des faits généraux*. Sendo uma ciência de observação, a medicina só poderia progredir através do registro de toda informação relevante, para que, a partir de um grande número de casos, fosse possível ascender a leis gerais.⁸ No caso do diagnóstico, o registro e a comparação da frequência relativa dos sintomas em cada doença poderia tornar sua detecção mais segura. Por exemplo, observando que a quase totalidade dos pacientes que apresentaram hemoptise em vida tinha tubérculos pulmonares à autópsia, ao passo que essas lesões não eram encontradas em pacientes com outras queixas respiratórias, mas sem hemoptise, conclui-se que a presença de sangue no escarro torna a existência de tuberculose “infinitamente provável”.

No caso da avaliação de um tratamento específico, por sua vez, seria necessário

(...) sobretudo indicar o resultado de sua ação em relação à mortalidade, à marcha mais ou menos rápida das doenças em direção a um fim favorável ou desfavorável [heureuse ou malheureuse]: e, para atingir esse objetivo, é necessário comparar entre eles um número suficientemente grande de casos de uma afecção, de mesmo grau; uns relativos a indivíduos em que a doença terá sido abandonada a ela mesmo, outros, a indivíduos a quem tais ou tais medicamentos terão sido administrados. (Louis, 1837, p. 36).

⁷ Além da mortalidade, Louis analisa a ação das sangrias sobre o que chamaríamos hoje de outro “desfecho clínico”, o alívio dos sintomas.

⁸ “(...) a enumeração completa, ou a análise rigorosa de todos os fatos exatos de mesmo gênero que foi possível encontrar, é o único meio de chegar ao conhecimento das leis de nossa economia; preceito que a lógica impõe e o uso rejeita como muito embaraçoso.: (Louis, 1837, p. 23).

E Louis ainda complementa que essa comparação, uma vez sendo feita, deveria ser ampliada para determinar a influência do tratamento em diferentes graus da doença, em menores ou maiores doses e na associação com outros tratamentos. É difícil não notar a semelhança com os princípios da EBM.

Certamente Louis não foi o primeiro nem seria o único a aplicar a análise matemática às informações clínicas e a comparar grupos. Tröhler (2000) relata diversos exemplos de análises matemáticas visando a avaliação de intervenções terapêuticas que ocorreram ainda no século XVIII, sobretudo no Reino Unido. Tal procedimento tornar-se-ia cada vez mais utilizado ao longo do século XIX, tanto no que concerne a questões epidemiológicas, como a relação apontada por John Snow (1849) entre a epidemia de cólera em Londres e a utilização de água contaminada por esgoto, quanto a questões preventivas e terapêuticas, como a demonstração por Ignaz Semmelweis (1850) de que a higiene das mãos dos médicos diminuía a mortalidade por infecção puerperal.

Comparação entre grupos e análise estatística dos resultados se consolidariam como as duas principais características do ensaio clínico. No início do século XX, as contribuições mais importantes para o aprimoramento de suas técnicas se relacionariam com a formação dos grupos e a análise dos resultados. Limitaremos a discussão a duas delas: teste de significância e randomização.⁹

Quando, em meados do século XIX, a eficácia de uma intervenção era avaliada, comparando-se um grupo tratado com um grupo não tratado, a análise dos resultados não procedia por métodos mais sofisticados do que uma simples razão de casos. Por exemplo, quando Louis compara a mortalidade de pacientes submetidos precoce ou tardiamente à sangria, apenas informa uma proporção dos óbitos para cada um dos grupos e, intuitivamente, afirma que a mortalidade do grupo submetido à intervenção era maior. No entanto, restava a possibilidade dessa diferença não dever-se de fato à intervenção estudada, mas sim ao acaso. O teste de significância visa oferecer um recurso matemático que assegure, com

⁹ Certamente, limitando a discussão à randomização e ao teste de hipótese, excluímos importantes avanços na estruturação do ensaio clínico, como a necessidade de se comparar o grupo experimental com um grupo que recebeu uma intervenção ineficaz, ou placebo. A necessidade de se ofertar algum tipo de tratamento foi progressivamente reconhecida com a publicação de estudos demonstrando que a simples assistência médica era capaz de gerar melhoras estatisticamente significantes em desfechos clínicos importantes.

considerável grau de confiabilidade, que a diferença entre os grupos não é um mero artefato do experimento.

Dada a amostra de um experimento, é possível calcular a chance dos resultados obtidos serem produtos do acaso. Essa probabilidade foi chamada, por Ronald Fisher (1890-1962), de “valor-p”. No modelo de Fisher, desenvolvido nas décadas de 1920 e 1930, o experimento visa produzir evidência para a refutar uma hipótese nula, que afirma não haver diferença entre os grupos comparados.¹⁰ Quanto menor o valor-p, menor a probabilidade dos resultados serem gerados quando a hipótese nula é verdadeira. O valor de 5% foi proposto como nível de significância adequado para a maior parte dos experimentos que Fisher tinha em mente quando desenvolveu suas ideias, mas poderia ser ajustado de acordo com as particularidades de cada experimento.¹¹ Se o valor-p do experimento for menor do que o nível de significância estabelecido, considera-se que há boa evidência para rejeitar a hipótese nula e concluir pela diferença entre os grupos. No caso do ensaio clínico, normalmente se conclui que a intervenção estudada – um medicamento, uma cirurgia, uma terapia reabilitadora etc. – produz diferenças quanto ao desfecho visado, diferenças essas que podem ser tanto positivas, quanto negativas. Assim, pode-se concluir que determinada substância, como a sinvastatina, diminui o risco cardiovascular, ao passo que outra substância, como a sibutramina, aumenta esse risco. Como veremos no próximo capítulo, essa inferência não é isenta de polêmicas.

A randomização é um procedimento que visa alocar aleatoriamente os elementos da amostra nos grupos comparados, gerando a mesma probabilidade de pertencerem a cada um deles. Segundo Fisher, ela garantiria a validade do teste de significância, ao assegurar que a distribuição das diferenças potencialmente importantes para o desfecho foram distribuídas homoganeamente: “Esse é, de fato, o único ponto do procedimento experimental em que as leis do acaso, que se espera estejam em controle exclusivo da distribuição de frequências, foram explicitamente introduzidas” (Fisher, 1935, p. 22). Com isso, institui-se um procedimento

¹⁰ Segundo Fisher (1933), a hipótese nula nunca é refutada, mas é possível recolher evidência suficiente para dizer que é extremamente improvável.

¹¹ “É usual e conveniente para experimentadores tomarem 5 por cento como um nível de significância padrão, no mesmo sentido em que estão preparados para ignorar todos os resultados que não conseguem atingir esse padrão e, desse modo, eliminar de discussões adicionais a maior parte das flutuações que causas ocasionais [chance causes] introduziram nos resultados experimentais” (Fisher, 1935, p. 15-6).

padronizado para o controle de vieses, já que é possível que existam variáveis potencialmente importantes que são desconhecidas ao se planejar o experimento. Além disso, a randomização permite que nem os pesquisadores, nem os pacientes saibam se pertencem ao grupo experimental ou ao grupo controle, o que poderia influenciar os resultados do teste. Essa característica dos estudos é normalmente chamada de duplo cego (*double-blind*).

Apesar de Fisher não ter pensado especificamente na pesquisa clínica ao desenvolver o teste de significância, estando preocupado mais diretamente com experimentos para acessar a produtividade de variedades agrícolas, sua teoria foi aplicada com êxito ao ensaio clínico. Os primeiros estudos que utilizam tanto a randomização, quanto o teste de significância datam da década de 1940, sendo o teste da estreptomicina no tratamento da tuberculose o mais conhecido.¹² Esse ensaio, organizado pelo *Medical Research Council* (MRC) britânico e publicado em 1948, tornou-se célebre também pela clareza da descrição de seus procedimentos experimentais, como os critérios de seleção dos pacientes e os meios de avaliação da resposta ao tratamento, que era realizada independentemente por observadores que desconheciam se o caso havia recebido estreptomicina, ou não. Os autores explicitam o método de alocação dos pacientes nos grupos desde o início: “A determinação de se um paciente seria tratado com estreptomicina e repouso (caso S) ou apenas com repouso (caso C) foi feita por referência a séries estatísticas baseadas em números de amostragem aleatória (...)” (Medical Research Council, 1948, p. 770). E a significância estatística dos resultados também é sublinhada para cada um dos desfechos avaliados. Em relação à melhora radiológica, por exemplo, os autores observam: “28 dos pacientes S (51%) e somente 4 dos pacientes C (8%) melhoraram consideravelmente (*a probabilidade de tal diferença ocorrer por acaso é menos do que uma em um milhão*)” (Medical Research Council, 1948, p. 772, grifo nosso). Apesar desses efeitos animadores, os pesquisadores reforçam a necessidade de utilizar controles em testes de medicamentos, já que uma parcela pequena, mas digna de nota, melhorou apenas com o repouso.

¹² Antes da randomização ser sistematicamente adotada, o método de alocação dos pacientes nos grupos comparados era a *alternância*: o primeiro paciente ia para o grupo A, o segundo para o grupo B, o terceiro para o A, o quarto para o B, e assim sucessivamente. Assim, em um ensaio clínico também do MRC para avaliação da eficácia de soro anti-estreptocócico para tratamento de pneumonia, realizado cerca de dez anos antes, descreve-se da seguinte forma a alocação dos participantes: “O método conseqüentemente acordado (...) foi que casos alternados de pneumonia, tomados simplesmente na ordem de sua admissão no hospital, deveriam ser usados respectivamente para tratamento com soro e como controles.” (Medical Research Council, 1934, p. 242).

Em 1962, com a Emenda “Kefauver-Harris” à *Lei federal de alimentos, medicamentos e cosméticos (Food, drugs, and cosmetic act)* americana de 1938, viria o reconhecimento institucional do ensaio clínico como requisito obrigatório para demonstrar-se a eficácia de uma droga. A partir de então, a *Food and Drug Administration* (FDA) passa a exigir, para liberar a comercialização de uma nova substância com fins terapêuticos, que o laboratório comprove a sua eficácia e segurança através de testes clínicos controlados. A secretaria responsável pelo processo de liberação de uma nova droga poderia negar a sua licença caso compreendesse que não havia “evidência substancial” demonstrando que teria os efeitos esperados nas situações indicadas. E o documento esclarece o que entende por esse termo:

(...) ‘*evidência substancial*’ significa evidência consistindo de *investigações adequadas e bem controladas*, incluindo *investigações clínicas*, [realizadas] por especialistas qualificados por treinamento científico e experiência para avaliar a *eficiência [effectiveness]* da droga envolvida, com base nas quais poderia ser concluído, justa e responsabilmente, por esses especialistas que a droga terá o efeito que alega ter nas condições de uso prescritas, recomendadas ou sugeridas na sua rotulação ou rotulação proposta (Estados Unidos da América, 1962, Sec. 102 (d), grifo nosso).

Para a EBM, o RCT é a fonte de melhor evidência científica para a prática médica porque é o desenho que está menos sujeito a vieses que acometem frequentemente os estudos não experimentais, ou epidemiológicos. Quando uma mesma questão clínica é avaliada independentemente por estudos observacionais – como coortes ou caso controle – e por RCT, normalmente se encontram efeitos muito maiores nos primeiros, exatamente pela presença de fatores de confusão que influenciam sistematicamente os resultados.

Particularmente importantes são os vieses evitados pelo duplo cego, que a randomização possibilita. Durante o teste de um novo tratamento, impedir que os pesquisadores saibam em qual dos grupos os pacientes estão sendo alocados é desejável para que, deliberada ou inconscientemente, não selecionem para o grupo experimental aqueles que obteriam os maiores benefícios da nova intervenção. Sackett (2015) oferece o exemplo de um teste da vacina BCG contra a tuberculose, realizado na década de 1930. Um estudo não randomizado obteve resultados bastante expressivos na redução da mortalidade de crianças que receberam a vacina, resultados que não se mantiveram quando avaliados por um ensaio randomizado. A explicação da divergência dos dois estudos está na seleção dos pacientes: no primeiro caso, foram selecionadas sistematicamente crianças com

melhor status social para receber a vacina, e a diferença da mortalidade devia-se a essa característica, e não à BCG. Esse viés é chamado “viés de susceptibilidade” (*susceptibility bias*) e ocorre quando os grupos comparados diferem quanto a uma característica basal capaz de influenciar o desfecho em análise.

Do ponto de vista do paciente, também é importante que não saibam a qual intervenção estão sendo submetidos, já que o simples cuidado médico é capaz de gerar efeitos expressivos sobre os sintomas. Novamente, Sackett (2015) ilustra esse problema, chamado viés de expectativa positiva (*positive expectancy bias*), com um exemplo real: na década de 1950, a ligadura da artéria mamária interna era frequentemente realizada para o alívio da angina do peito. Um cirurgião alocou aleatoriamente pacientes para terem essa artéria ligada, ou para receberem apenas uma incisão no peito, sem saberem que o vaso permanecia patente. O acompanhamento clínico era realizado por um médico que também não sabia se o paciente havia sido submetido à ligadura, ou não. Os resultados são surpreendentes, porque uma parcela muito expressiva dos doentes que haviam sofrido apenas uma incisão no peito, mas pensavam que suas artérias haviam sido ligadas, apresentou uma melhora importante dos sintomas de angina.¹³

Os estudos observacionais deveriam limitar-se, por isso, ao levantamento inicial de possíveis associações entre intervenções e desfechos, que esperariam ser adequadamente avaliadas por RCT posteriores; ao estudo de condições tão raras que a condução de um RCT seria impraticável; ao estudo de tratamentos cujo eficácia, geralmente aceita pelo meio médico, tornaria a realização de um estudo experimental antiética.¹⁴

Recapitulando o que apresentamos com exemplos históricos ao longo desta seção, podemos dizer que o ensaio clínico é um procedimento experimental típico das ciências da saúde que visa demonstrar a associação causal entre uma intervenção e um desfecho específico. Clínico porque utiliza como participantes

¹³ “Pacientes que gostam do tratamento de fato reportam desfechos melhores, independentes da verdadeira eficácia do seu tratamento. É por isso que nossos RCT empregaram (sempre que possível e possivelmente mais do que os detratores podem pensar) ocultação [*blinding*] dos pacientes em teste sobre o seu tratamento, desfechos ‘duros’, como mortalidade total, e o julgamento ‘cego’ de desfechos mais leves [*softer outcomes*]” (Sackett, 2015, p. 328).

¹⁴ Já que o RCT visa demonstrar uma relação causal, poderia ser aplicado para o esclarecimento de questões de etiologia. No entanto, como a realização de tal experimento exigiria a administração de substâncias ou outros agentes que se suspeita serem a causa de alguma doença, sua execução é considerada na maior parte das vezes antiética.

indivíduos humanos acometidos por doenças reconhecidas e manejadas na prática médica e porque pretendem produzir informação para esclarecer essa prática. Nesse sentido, deve ser diferenciado da pesquisa laboratorial básica das diversas áreas da biologia, como microbiologia, fisiologia e farmacologia, que utilizam modelos animais, culturas celulares e análises moleculares para esclarecer processos envolvidos nos estados de saúde e de doença.¹⁵ O sucesso da pesquisa no nível laboratorial não garante que sua aplicação na prática seja bem sucedida, o que determina a necessidade de realização do ensaio clínico. Trata-se de um procedimento experimental porque controla as condições de manifestação do fenômeno estudado, geralmente a marcha natural de uma doença, e intervém sistematicamente sobre esse fenômeno na tentativa de testar uma hipótese. Na maior parte das vezes, a hipótese afirma que um agente A (como um medicamento) é capaz de produzir efeitos consideráveis sobre um desfecho clínico predeterminado B (por exemplo, a mortalidade de uma doença).

Se a EBM recomenda que se utilize a melhor evidência disponível para a tomada de decisão no manejo de um paciente, podemos nos perguntar com que frequência o RCT, considerado a fonte de melhor evidência, está disponível para um clínico em sua prática. Na próxima seção, utilizaremos os recursos de avaliação crítica da literatura desenvolvidos pela EBM para analisar a qualidade do conhecimento científico que serve de base à prática médica atual.

1.5 Diretrizes clínicas e qualidade da evidência

Dissemos na seção 1.1 que, inicialmente, a EBM limitava-se a decisões no manejo do paciente individual e não se interessava por políticas de saúde. No entanto, com a percepção cada vez mais clara de que a avaliação crítica da literatura defendida pelo movimento era impossível para o clínico na sua prática diária, os autores da EBM começam a aplicar seus procedimentos à confecção de diretrizes práticas (*practice policies*), documentos que trazem recomendações sobre o que se considera os melhores procedimentos em saúde. Esses documentos podem ser confeccionados por sociedades de especialistas, agências reguladoras, organizações internacionais de saúde, ou qualquer outro grupo de profissionais direta ou indiretamente implicados no cuidado de pacientes. De fato, uma das

¹⁵ Retornaremos a essa questão na seção 2.3, onde discutiremos melhor a distinção entre pesquisa laboratorial biomédica e pesquisa clínica.

características das diretrizes práticas é uma crescente interdisciplinaridade, já que tendem a ser produzidas por comitês de formação bastante diversificada, como médicos generalistas e especialistas, estatísticos, metodologistas, representantes do Estado e, recentemente, usuários dos serviços de saúde.

As diretrizes práticas, assim como os RCTs, não são uma criação da EBM. Diversas motivações internas e externas à medicina concorreram para o seu crescimento, que se iniciou na década de 1960. George Weisz e colaboradores (2007) discutem várias delas, mas gostaríamos de destacar as seguintes: o controle dos custos dos serviços de saúde subsidiados tanto pelos Estados, quanto por seguradoras privadas; a regulação do manuseio de instrumentos diagnósticos e terapêuticos potencialmente perigosos, como a radioterapia; a padronização de categorias diagnósticas e de procedimentos técnicos de mensuração, que permitiria a comparação dos dados de pesquisa produzidos em centros distantes; a uniformização de medidas terapêuticas e preventivas de doenças complexas, como o câncer; e a sistematização do conhecimento científico, que crescia exponencialmente nas ciências da saúde. O próprio desenvolvimento do RCT e sua aceitação crescente como a fonte de evidência mais forte para a prática clínica contribuiu para a disseminação das diretrizes:

A conclusão lógica para muitos foi que as práticas diferindo daquelas validadas por RCTs são desvios errôneos do procedimento clínico correto, e que a solução percebida é difundir o conhecimento das práticas corretas de várias formas, incluindo diretrizes práticas (Weisz *et al.*, 2007, p. 704).

Esses textos podem tratar das mais diversas etapas e atividades dos serviços de saúde, como diagnóstico, prevenção e promoção da saúde, triagem (*screening*) e medidas de saúde pública. Atualmente, existem diretrizes metodológicas até mesmo para a confecção das próprias diretrizes, sendo talvez a mais importante delas conhecida pelo acrônimo AGREE (*Appraisal of guidelines for research and evaluation*). Como vimos na seção 1.3, a hierarquização da evidência clínica aparece pela primeira vez na literatura médica em um texto desse tipo, uma diretriz acerca do exame periódico de saúde. Mas especialmente importantes são aquelas que trazem recomendações sobre o manejo de pacientes com doenças específicas, chamadas de “diretrizes da prática clínica” (*clinical practice guidelines*). As recomendações veiculadas nas diretrizes clínicas orientam as condutas a serem tomadas no manejo de pacientes, desde o diagnóstico da doença até seu tratamento. Por exemplo, lê-se numa recomendação retirada de uma diretriz para o

manejo da hipertensão: “Em adultos com diabetes mellitus e hipertensão, o tratamento com drogas anti-hipertensivas deve ser iniciado com a pressão arterial de 130/80 mmHg, ou mais alta, com um objetivo de tratamento de menos de 130/80 mmHg” (Whelton *et al.*, 2017, p. 116).

Como dissemos, apesar das diretrizes clínicas antecederem a EBM, esses documentos foram progressivamente sendo adaptados ao modelo proposto pelo movimento, e hoje em dia virtualmente todas as diretrizes publicadas discriminam a qualidade da evidência das suas recomendações, seguindo o mesmo princípio de hierarquização da evidência apresentado na seção 1.3. Ao lado da recomendação citada acima como exemplo, é possível encontrar a sua classificação: “classe de recomendação I, nível de evidência B”. Isso quer dizer, como explicitado no começo da diretriz, que o controle medicamentoso da pressão arterial de pacientes diabéticos, visando uma pressão arterial de 130/80 mmHg, é fortemente recomendada, e que essa recomendação é sustentada por, no mínimo, um RCT de moderada qualidade.

Mas esse exemplo representa adequadamente a qualidade do conjunto das recomendações publicadas em diretrizes clínicas? Alguns estudos foram feitos para responder a essa pergunta, e gostaríamos de relatar os seus resultados, já que, de certo modo, oferecem um diagnóstico do estado do conhecimento médico atual. O primeiro deles (Tricocci *et al.*, 2009) avaliou todas as recomendações (7196, no total) publicadas de 1984 a 2008 pela *American Heart Association* – certamente a sociedade de cardiologia mais influente no mundo. Apenas 11% das recomendações classificadas eram sustentadas por mais de dois RCTs (nível de evidência A), enquanto 48% de todas as recomendações era baseada apenas na opinião de especialistas (nível de evidência C). Similarmente, Poonacha e Go (2011) mostram que, das recomendações publicadas em diretrizes da *National Comprehensive Cancer Network* sobre o tratamento de diversos tipos de câncer, apenas 6% são consideradas sustentadas por evidência de alta qualidade e fortemente recomendadas. A maior parte das recomendações (83%) era indicada apesar de sustentada por evidência considerada de baixa qualidade, e portanto referendada sobretudo pela opinião e experiência dos especialistas que contribuíram para a produção das diretrizes. Lembremo-nos que uma das intenções da EBM era justamente diminuir, na tomada de decisões sobre o manejo de pacientes, a

influência do julgamento baseado na experiência individual e na opinião de especialistas.

Em mais um estudo, Lee e Vielemeyer (2011) mostram que mais da metade (55%) das recomendações sobre o manejo de doenças infecciosas são consideradas de baixa qualidade científica. Outra informação importante é que, quando se compara diretrizes que foram submetidas a atualizações com suas versões anteriores, percebe-se que não houve melhora na qualidade da evidência, contrariando a expectativa de que a continuidade das pesquisas resolveria as incertezas clínicas. Normalmente, os novos documentos tendem a fazer recomendações adicionais, mas também sustentadas por evidência de baixa qualidade. Esses resultados formalizam uma percepção comum entre os médicos no exercício da profissão e uma suspeita dos pacientes tratados por eles: a maior parte das práticas clínicas correntemente em uso não tem a eficácia demonstrada. E, já antecipando a discussão da próxima seção, deixam claro que não houve a substituição de um “antigo paradigma” caracterizado pela importância dada à opinião de especialistas na tomada de decisões, já que essa é a base para a maior parte das recomendações atualmente vigentes.

De qualquer modo, apesar da aderência profissional às diretrizes também ser uma questão a ser discutida, já que costuma ser bem mais baixa do que o esperado, elas de fato representam uma mudança de foco e escala na autoridade médica. Por um lado, houve uma diminuição da importância da experiência e do julgamento individual na tomada de decisão clínica e sua transferência para grupos de especialistas. Por outro, os padrões de prática deixam de ser tão regionalizados e passam a ser mais homogeneizados tanto nacional, quanto internacionalmente, já que os estudos utilizados para a sua formulação são mais ou menos limitados e amplamente disponíveis. Além disso, o prestígio de que gozam algumas instituições (como a OMS, por exemplo) faz com que suas recomendações sejam seguidas mundialmente.

As diretrizes clínicas são o último elemento central da EBM que pretendíamos apresentar neste capítulo. Ainda que breve, acreditamos que essa introdução ao nosso objeto de estudo é suficiente para tirarmos algumas conclusões iniciais sobre ele.

1.6 EBM, novo paradigma?

Tendo apresentado as principais características da EBM e indicado brevemente seus antecedentes históricos, podemos retomar a questão colocada no início deste capítulo. O que é a EBM? Devemos aceitar a caracterização proposta por Guyatt e considerá-la uma revolução que introduziu um novo paradigma no pensamento e na prática médica?

Ainda que o termo “paradigma” tenha sofrido uma grande vulgarização, sendo utilizado de formas muito distintas daquelas propostas pelo próprio Kuhn, resolvemos levar a sério as afirmações dos defensores da EBM e responder à sua apropriação do termo. Isso se torna especialmente relevante quando observamos que diversos autores acataram e reproduziram de forma pouco crítica as afirmações do movimento. Por exemplo, Lewis e Orland (2004, p. S4) afirmam que “A EBM é, de fato, o novo paradigma para a prática médica”, retomam quase literalmente os desenvolvimentos do artigo do Evidence-Based Medicine Working Group (1992) e defendem que o ensino da EBM se torne obrigatório nas escolas de medicina. Em um artigo mais recente, que faz um balanço do movimento 20 anos após o seu surgimento, lemos que a EBM é “(...) um paradigma que agora está encravado [*embedded*] no ensino, na prática, na pesquisa, na escrita e na gestão do cuidado em saúde (...)” (Seshia; Young, 2013, p. 465). Poderíamos reproduzir outros exemplos como esses, mas consideramos desnecessário fazê-lo. É interessante apenas notar que mesmo alguns trabalhos que fazem críticas à EBM, como o artigo de Keane e Berg (2016), não questionam a sua auto caracterização como um “novo paradigma”.¹⁶ Ainda que muitas dessas utilizações do termo sejam claramente pouco rigorosas, a sua reprodução na literatura médica é tão disseminada que merece uma retificação. Como dissemos, a argumentação do que a EBM não é nos levará ainda a uma caracterização mais adequada do movimento. Além disso, o recurso às ideias de Kuhn será útil novamente na seção 3.1, quando recorreremos a elas para compreender alguns tipos específicos de atividades médicas.

Começemos por uma revisão da concepção de paradigma científico, conforme apresentada por Thomas Kuhn na obra *The structure of scientific revolutions*, de 1962, e posteriormente desenvolvida em outros trabalhos. Segundo

¹⁶ Obviamente existem exceções a essa afirmação. Por exemplo, Crawley (1993), em uma carta aos editores do *Journal of the American Medical Association*, afirma que a EBM não pode ser considerada um novo paradigma da prática clínica porque não modifica os conceitos que organizam a visão de mundo médica. Nas páginas que se seguem, daremos argumentos que apontam na mesma direção.

Kuhn (1970a [1962]), a maior parte das atividades do que comumente se entende por ciência consiste na tentativa de resolver problemas específicos, ou quebra-cabeças (*puzzles*), em sua terminologia. Esses problemas e as regras para a sua resolução são definidos através de êxitos científicos pregressos, que constituem um modelo para a prática subsequente de todo um grupo de cientistas. Os trabalhos de Isaac Newton (1643-1727) acerca da luz, por exemplo, lançaram as bases para inúmeras pesquisas posteriores, que se caracterizavam por compartilhar uma perspectiva que interpretava os fenômenos ópticos como resultado do comportamento de inúmeros e ínfimos corpúsculos. Para Kuhn, os trabalhos de Newton constituem o paradigma da pesquisa em óptica durante o século XVIII e parte do século XIX, até a sua substituição por um outro paradigma, que atribui à luz e seus fenômenos uma natureza ondulatória.

Antes da obtenção de um paradigma, a pesquisa em uma área específica do conhecimento se caracteriza por um debate constante entre diferentes escolas, que divergem em relação aos seus princípios básicos, aos fenômenos cujo estudo apresenta maior relevância e aos procedimentos legítimos de investigação. A aquisição de um paradigma torna desnecessário o debate acerca de assunções fundamentais, que passam a ser aceitas por toda a comunidade relevante, e possibilita a concentração dos esforços intelectuais na solução de problemas muito específicos e bem delimitados. Além disso, a emergência de um paradigma altera as fronteiras disciplinares e seus respectivos aparatos institucionais:

Nas ciências (ainda que não em áreas como medicina, tecnologia e direito, cuja principal razão de ser [*raison d'être*] é uma necessidade social externa), a formação de jornais especializados, a fundação de sociedades de especialistas e a exigência por um lugar especial no currículo são normalmente associadas com a primeira recepção de um paradigma único por um grupo (Kuhn, 1970 [1962], p. 19).

A esse tipo de atividade científica altamente especializada e eficiente, que a aceitação de um paradigma inaugura, Kuhn se refere como *ciência normal*.

No seção 3.1, discutiremos melhor a citação acima, que já esboça o ponto de vista de Kuhn em relação à medicina, mas por ora continuamos com a exposição de suas ideias. Como dissemos, a ciência normal, guiada por um paradigma, caracteriza-se por um escopo de pesquisa bastante restrito. Segundo Kuhn (1970 [1962], cap. 3), os cientistas concentram a maior parte dos seus esforços no aumento da precisão com que apreendem fatos já conhecidos e de especial importância dentro do paradigma, na determinação de valores numéricos, como

constantes e leis universais, e na tentativa de aplicação dos conhecimentos disponíveis a fenômenos similares. Isto é, a parte mais substancial da atividade científica não visa à descoberta de fatos inesperados ou ao desenvolvimento de teorias e conceitos novos, mas a articulações teóricas e discretas expansões do paradigma vigente.

O outro lado da atividade científica, ou seu aspecto revolucionário, muito conhecido e especialmente enfatizado pelo senso comum, inicia-se quando o paradigma se torna reiteradamente incapaz de resolver problemas, ou explicar fenômenos, cuja elucidação se espera dele. Diante do que Kuhn denomina de *anomalias*, os cientistas tentam inicialmente realizar pequenas adaptações do paradigma, que frequentemente resolvem as inconsistências observadas. No entanto, é possível que essas anomalias resistam mesmo a grandes modificações e flexibilizações do paradigma e que instaurem um período de *crise* na disciplina em questão. Nessas circunstâncias, um outro tipo de pesquisa, denominada por Kuhn de pesquisa extraordinária, instaura-se, podendo levar finalmente à emergência de um novo paradigma, que substitui o paradigma anterior e dá início a um novo período de pesquisa normal, com novos problemas, métodos e aparatos técnicos e institucionais. Mas mais do que isso, a substituição de um paradigma por outro, típica de uma revolução científica, modifica substancialmente a maneira com que os cientistas enxergam o mundo, pois novas entidades passam a habitá-lo e novas relações se instauram entre as entidades existentes. Para Kuhn (1970 [1962], cap. 10), essa mudança na visão científica de mundo é análoga a uma mudança gestáltica (*Gestalt switch*), através da qual uma figura distinta é percebida diante dos mesmos estímulos sensoriais.

Deve ter ficado claro na apresentação acima que a EBM não modifica as entidades nosológicas que habitam o pensamento e a prática médica, não propõe uma nova classificação das doenças, ou um conceito diferente de saúde. Pelo contrário, para que a pesquisa epidemiológica e clínica seja realizada, é necessário que essas entidades sejam previamente reconhecidas e amplamente aceitas. Analisaremos a natureza dinâmica da doença quando discutirmos as noções de progresso médico no capítulo 3, mas podemos dizer, por enquanto, que os procedimentos para a classificação de doenças permanecem essencialmente os mesmos desde o final do século XVIII e o início do século XIX e baseiam-se na

correlação dos sinais e sintomas apresentados em vida com as lesões encontradas no cadáver. O próprio Kuhn, em uma revisão da sua noção de revoluções científicas, sublinha a importância das mudanças taxonômicas: “O que caracteriza as revoluções é (...) a mudança em diversas das categorias taxonômicas, pré requisitos para as generalizações e descrições científicas” (Kuhn, 1987, p. 20). Uma indicação de como as categorias diagnósticas permanecem inalteradas é o fato da EBM considerar o exame anatomopatológico, a biópsia e a necrópsia, como o critério de comparação “padrão-ouro” para avaliar a acurácia de novos exames. O movimento se limita a fazer considerações acerca da melhor forma de se obter conhecimento sobre entidades pré-estabelecidas. Portanto, não houve nada que se pareça com uma mudança gestáltica na visão de mundo médica.

Do ponto de vista sociológico, a EBM tampouco cria novas instituições, novos periódicos, novas sociedades de especialistas. As diversas referências que fizemos a textos da EBM mostram que a maior parte das publicações dos seus autores foi, e continua sendo, veiculada por periódicos muito mais antigos, como o *New England Journal of Medicine*, o *Canadian Medical Association Journal* e o *Journal of the American Medical Association*.¹⁷ De fato, houve uma adaptação dos artigos publicados nesses periódicos ao modelo proposto pelo movimento, como já indicamos, mas todo o aparato institucional permanece o mesmo. Similarmente, não encontramos nenhuma mudança na organização e na divisão das especialidades médicas, que continuam sendo definidas pelas mesmas características disciplinares e mantêm o mesmo programa de pesquisa. O objeto de estudo das sociedades de cardiologia, por exemplo, como a hipertensão, o risco cardiovascular, as arritmias cardíacas, permanecem os principais focos de sua atividade científica. Os recursos metodológicos desenvolvidos pelo movimento podem ser e são aplicados a uma ampla gama de pesquisa clínica e abrangem diversas especialidades. Uma característica importante de uma revolução científica, pelo contrário, é alterar as divisões disciplinares e seus respectivos objetos de pesquisa.

Mas resta a possibilidade da EBM ter introduzido uma maneira suficientemente nova de resolver os problemas colocados pela clínica para que seja considerada revolucionária. Certamente reconhecemos uma ruptura nas ciências da saúde que se relaciona com a aplicação de métodos numéricos. Inicialmente, o

¹⁷ O *American College of Physicians Journal Club*, a que nos referimos no início do capítulo, criado como um apêndice de um periódico já existente, é uma exceção a essa constatação.

levantamento das estatísticas de mortalidade possibilita delinear um prognóstico muito mais seguro do que a experiência pessoal de cada clínico permitiria, além de iluminar questões importantes de etiologia. Mas, principalmente, a aplicação sistemática da comparação numérica de grupos permite a detecção de efeitos terapêuticos difíceis de verificar de outra forma, fazendo com que qualquer afirmação acerca de eficácia de uma intervenção deva ser referendada por dados numéricos, e não simplesmente baseada na teoria médica vigente. No entanto, essa descontinuidade – caso seja preferível por um motivo ou outro evitar o termo “revolução” – ocorreu em algum momento da primeira metade do século XIX, ou talvez até no final do século XVIII. A determinação desse momento, no entanto, exigiria um aprofundado trabalho histórico, que foge às pretensões desta dissertação. De qualquer modo, aceitando-se isso, fica claro que o que ocorreu no final do século XX com o aparecimento da EBM não pode ser considerado como um evento revolucionário, mas representa uma etapa evolutiva de um processo de quantificação da medicina que vinha, há dois séculos, desenvolvendo-se cada vez mais.

As motivações para o surgimento da EBM relatadas na seção 1.2 poderiam ainda parecer com a crise que precede uma revolução científica e que se caracteriza pelo crescente reconhecimento de que os meios oferecidos pelo modelo vigente não são mais capazes de responder às questões que inicialmente resolviam. No entanto, o que os autores envolvidos apontam é a aplicação insatisfatória dos recursos metodológicos e estatísticos existentes, insistindo por uma retificação da sua utilização. Quanto à substituição do pensamento fisiopatológico e da experiência clínica como base para a tomada de decisões práticas, as considerações da seção 1.5 devem ter sugerido que está longe de ocorrer plenamente, já que continuam guiando o manejo dos pacientes nos inúmeros casos em que falta evidência científica de boa qualidade para embasar as condutas.

Após todas essas considerações predominantemente negativas, teríamos alguma conclusão positiva para propor em relação à melhor forma de caracterizar a EBM? Diríamos que a EBM é um movimento de normatização da prática clínica, que exige que as condutas médicas sejam sustentadas por estudos que atendam aos critérios de qualidade metodológicos definidos por seus proponentes. Sua principal contribuição para o conhecimento médico relaciona-se com o desenvolvimento do

arsenal conceitual de avaliação crítica da pesquisa epidemiológica e clínica, mesmo que a maior parte desses conceitos não tenha sido criada pelo movimento, mas apropriada de desenvolvimentos progressos oriundos principalmente da estatística e da epidemiologia. Nesse sentido, pode ser interpretada como a radicalização de uma concepção específica de progresso, que encontra na matematização estatística a via mais segura para garantir o estatuto científico da medicina e resolver as inúmeras incertezas relacionadas com a prática clínica.

Mas essa concepção de progresso não é a única existente na medicina e talvez não seja a mais adequada, como discutiremos no capítulo 3. No entanto, antes de abordarmos esse problema, precisamos nos debruçar sobre a noção de evidência e compreender melhor como ela se aplica às diversas atividades de pesquisa e prática médica. Dedicaremos o próximo capítulo a essas questões.

CAPÍTULO 2: Evidência científica e evidência na medicina

2.1 Introdução

Ao longo do primeiro capítulo deste trabalho, em que apresentamos o nosso objeto de estudo, a EBM, e lhe propusemos uma caracterização inicial, utilizamos expressões como “melhor evidência disponível”, “evidência científica” e “evidência mais forte”. Ainda que tenhamos esclarecido o que o movimento entende por “melhor evidência disponível”, através da apresentação das hierarquias de evidência para a prática clínica, uma tarefa importante do nosso trabalho resta por fazer: tentar esclarecer a própria noção de evidência. Este capítulo será dedicado a essa tarefa. Inicialmente, apresentaremos e discutiremos algumas definições e conceitos de evidência científica comumente defendidos por filósofos da ciência. Em seguida, tentaremos aplicar essas definições e conceitos à medicina. Finalmente, discutiremos como os testes de hipóteses estatísticas, utilizados para a produção de boa parte da evidência na pesquisa clínica, devem ser interpretados à luz dos conceitos e definições trabalhados. É preciso notar que não pretendemos resolver as questões levantadas pelas evidências científica e médica, mas apenas delimitar alguns pontos que julgamos mais prementes e indicar caminhos possíveis para soluções e desenvolvimentos futuros.

2.2 Evidência científica

Nesta seção, discutiremos algumas concepções filosóficas acerca da evidência científica. Seguiremos a exposição apresentada por Peter Achinstein (1935-) na obra *The book of evidence*, de 2001, em que realiza uma revisão das diversas teorias filosóficas da evidência e desenvolve uma teoria própria. Além de reconhecermos a importância do seu extenso estudo na área de evidência e confirmação, o recurso à obra de Achinstein se justifica por esse autor defender uma posição pluralista em relação à evidência, que julgamos especialmente apropriada ao caso da medicina. Como discutiremos na próxima seção, parece-nos que dois tipos de conceitos distintos devem ser utilizados para caracterizar o que correntemente se entende por evidência médica.

Inicialmente, Achinstein (2001, cap.3) discute duas definições probabilísticas de evidência bastante influentes: (i) evidência como informação que aumenta a

probabilidade de uma hipótese; e (ii) evidência como informação com base na qual uma hipótese apresenta alta probabilidade.

De acordo com a primeira definição, que chamaremos de *relevância positiva*, qualquer dado ou informação e capaz de realizar um incremento na probabilidade $p(h)$ de uma hipótese h , deve ser considerada evidência a favor dessa hipótese. Em notação probabilística, podemos definir evidência, nessa perspectiva, da seguinte forma:

(1) e é evidência a favor da hipótese h se e somente se $p(h|e) > p(h)$,

onde se compreende por $p(h|e)$, a probabilidade de h dado e , ou sob condição da ocorrência de e . $p(h|e)$ é um tipo de probabilidade condicional, pois a probabilidade do evento ou da hipótese de interesse depende da ocorrência ou é influenciada pela ocorrência prévia de um outro evento.¹⁸ Probabilidades condicionais $p(a|b)$ podem ser diferenciadas, nesse sentido, de probabilidades categoriais (Hacking, 2001, cap. 4), notadas simplesmente como $p(a)$, onde a e b são quaisquer eventos, e onde $p(a)$ independe da ocorrência, ou não leva em consideração a ocorrência de qualquer evento b .

A fim de tornar a noção de condicionalidade mais clara, imaginemos o evento a “retirar uma carta de número par dentre cartas de 1 a 10 do naipe de ouros”. A probabilidade $p(a)$ é de $1/2$, já que, entre as 10 cartas, 5 são pares. Imaginemos um outro evento b , que afirma que “uma carta de número ímpar foi retirada na primeira tentativa e que essa carta não foi repostada no monte”. Se utilizarmos o mesmo grupo de cartas e retirarmos mais uma, a probabilidade $p(a|b)$ passa a ser $5/9$, já que restam apenas 9 cartas, 5 das quais são pares. Isto é, a probabilidade de retirar uma carta de número par é influenciada pela condição de uma carta de número ímpar já ter sido retirada. É interessante notar que a condicionalidade não exige que a condição seja verdadeira, mas apenas supô-la verdadeira. Esse detalhe será importante quando discutirmos os diferentes conceitos de evidência.

De acordo com a definição de evidência como incremento da probabilidade da hipótese, se reformularmos o evento a como a hipótese h de que “uma carta de número par será retirada na próxima tentativa”, o evento b deve ser considerado evidência a favor de h , já que $5/9 > 1/2$, ou $p(h|b) > p(h)$.

¹⁸ Obviamente, a verdade de uma hipótese pode ser compreendida como um evento, e a $p(h)$ interpretada como a probabilidade da hipótese h ser verdadeira.

Podemos ainda oferecer alguns exemplos para ilustrar como essa definição se aplica ao caso das ciências. Achinstein (2001) refere-se, ao longo da obra, aos experimentos de J. J. Thomson (1856-1940) acerca dos raios catódicos. No final do século XIX, havia uma controvérsia em torno da natureza desses fenômenos, se seriam constituídos por partículas ou ondas e se teriam cargas elétricas, ou não. Alguns experimentos prévios, notadamente os de Heinrich Hertz (1857-1894), foram incapazes de demonstrar qualquer comportamento dos raios catódicos que sugerisse a presença de cargas elétricas. No entanto, Thomson consegue demonstrar em 1897 que os raios catódicos são desviados por campos magnéticos e elétricos, o que pode ser interpretado como evidência a favor da hipótese de serem constituídos por partículas eletrizadas. Um defensor da definição de evidência como relevância positiva diria que os resultados dos experimentos de Thomson são evidência a favor dessa hipótese porque a tornam mais provável.

Para fazermos uma primeira aproximação desse conceito à pesquisa em ciências biológicas, podemos utilizar o caso da descoberta da produção de glicose pelo fígado, apresentada por Claude Bernard (1813-1878) em 1848 à *Société de Biologie* de Paris. A demonstração experimental de uma maior concentração sanguínea de glicose no sangue que sai do fígado, comparativamente ao sangue que entra no órgão, era evidência a favor da hipótese da produção hepática dessa substância, mesmo que essa não tenha sido a hipótese inicial de Bernard. Um defensor da definição de evidência como relevância positiva, diria, mais uma vez, que esse achado experimental era evidência a favor da hipótese da glicogênese hepática porque a tornava mais provável. Isso valeria também para os resultados de pesquisas subsequentes realizadas por Bernard sobre o mesmo tema. A demonstração de que no tecido hepático havia uma substância capaz de gerar glicose aumentaria ainda mais a probabilidade da hipótese da produção hepática de compostos glicídicos.

Essa primeira definição de evidência, porém, está sujeita a contra-exemplos que demonstram que, em muitas situações, não estamos dispostos a considerar um evento como evidência a favor de uma hipótese, apesar desse evento tornar a hipótese mais provável. Na maior parte dos casos, isso ocorre quando o incremento na probabilidade da hipótese trazido pela nova informação é pequeno.

Achinstein (2001, cap.4) oferece diversos exemplos contrários à definição de evidência como relevância positiva, cujo critério não considera nem necessário, nem suficiente. A apresentação de apenas alguns deles basta para indicar as suas possíveis fraquezas.¹⁹ Imaginemos uma loteria justa em que 1 milhão de bilhetes foram vendidos, dos quais apenas 1 bilhete será sorteado, e a hipótese h que afirma “João será ganhador”. A informação e de que “João comprou 1 bilhete dessa loteria” aumenta a probabilidade de h ser verdadeira, isto é, de João ser vencedor, já que a probabilidade de ganhar possuindo um bilhete (1/1 milhão) é maior do que a probabilidade de ganhar sem possuir bilhete algum (zero, se todos os possíveis fatores intervenientes forem excluídos). No entanto, em nenhum uso corrente do termo “evidência”, o fato de alguém possuir um bilhete é tomado como evidência de que será campeão em tal loteria, pois a probabilidade de h ocorrer é extremamente pequena. Similarmente, se se considera que a frequência de acidentes em elevadores é cerca de 1:1 bilhão, o fato de alguém realizar uma viagem em elevador deveria ser considerado evidência de que essa pessoa sofrerá um acidente de elevador, já que a probabilidade de sofrer um acidente se torna maior ao utilizar esse meio de transporte. Uma indicação de que, normalmente, não se toma esse evento como evidência da hipótese do acidente é o fato de não alterar em absolutamente nada a conduta de alguém, já que a maior parte das pessoas anda de elevador sem temer que um acidente lhe ocorra.

Inversamente, Achinstein (2001, p. 70) tenta produzir uma situação em que, apesar de uma nova informação ou evento não aumentar a probabilidade da hipótese em relação ao que se sabia anteriormente, atribuímos o título de evidência a esse evento. Trata-se novamente de uma loteria justa, mas desta vez apenas 1000 bilhetes serão vendidos, e apenas um será sorteado. e_1 afirma que “Bill Clinton possui 999 bilhetes de uma loteria qualquer”, enquanto e_2 afirma que “Bill Clinton possui 999 bilhetes daquela loteria em questão”. A probabilidade da hipótese h “Bill Clinton será vencedor na loteria” seria, segundo Achinstein, igual diante de e_1 e diante de $e_1 \& e_2$.²⁰ Ainda assim, consideraríamos e_2 evidência a favor de h .

¹⁹ Realizamos pequenas alterações nos exemplos fornecidos por Achinstein, a fim de tornar a exposição mais simples. Essas alterações, no entanto, não afetam de forma alguma o núcleo do argumento e nem geram implicações que o autor não aceitaria.

²⁰ Isto é, para Achinstein, $p(h|e_1) = p(h|e_1 \& e_2) = 999/1000$.

Não concordamos com esse exemplo, pois julgamos que a probabilidade de h dado e_1 é menor do que a probabilidade de h dado $e_1 \& e_2$, já que a primeira deveria ser multiplicada pela chance dos bilhetes de Bill Clinton pertencerem à loteria em questão. Se, por exemplo, houvesse três loterias similares, uma das quais seria responsável por sortear os bilhetes de Bill Clinton, e não tivéssemos nenhuma outra informação, deveríamos multiplicar $p(h|e_1)$ por $1/3$, isto é, a chance dos bilhetes de Bill Clinton pertencerem à loteria em questão. No entanto, concedemos que incrementos muito pequenos na probabilidade de uma hipótese, como ilustrado pelo primeiro exemplo, podem não ser intuitivamente suficientes para constituir evidência a seu favor.

Antes de apresentarmos a segunda definição de evidência, queremos aproveitar os exemplos acima para introduzir mais um conceito que será importante nas discussões que se seguem. Quando descrevemos os casos dos bilhetes das loterias, precisamos afirmar, em um dos casos, que há “uma loteria justa, cujo total de bilhetes é de 1 milhão, dos quais apenas um será sorteado”. A informação e “João possui um bilhete” só pode ser interpretada como evidência, ou não, a favor da hipótese h “João será vencedor” diante dessa *informação de fundo* (*background information*). Chamaremos a essa informação de fundo de f . Em notação probabilística, a probabilidade condicional de uma hipótese h diante do evento e e da informação f passa a ser escrita da seguinte forma: $p(h|e \& f)$. Isso pode ser interpretado como a probabilidade de uma hipótese sob condição de ocorrência da conjunção de e e f . Se, por outro lado, quisermos tratar as informações de fundo como verdadeiras e relativizar a probabilidade da hipótese apenas à ocorrência do evento e , notaremos a probabilidade condicional da seguinte forma: $p_f(h|e)$. Por ora, contentamo-nos em indicar que a maior parte das avaliações de uma informação como evidência para uma hipótese científica depende fundamentalmente de informações de fundo tomadas como verdadeiras.

Uma maneira de evitar os contra-exemplos à definição de evidência como relevância positiva apresentados acima é afirmar que, para que e seja considerado evidência a favor de h , é necessário que um limiar de probabilidade seja ultrapassado. Isso nos fornece a segunda definição padrão de evidência, a que chamaremos de *alta probabilidade*:

(2) e é evidência a favor da hipótese h se e somente se $p(h|e) > k$, onde k é um número racional entre 0 e 1.

Retomando o exemplo de João e da loteria, em que a mesma informação de fundo f apresentada acima é válida, podemos dizer, de acordo com a nova definição, que a informação e que “João possui 950 mil bilhetes da loteria” constitui evidência a favor da hipótese h de “João ser o vencedor na loteria”, já que a probabilidade da hipótese ocorrer, dada a informação, é bastante alta (19/20). No entanto, resta a questão de determinar o valor de k . Achinstein (2001, cap.6), que aceita o critério de alta probabilidade, apesar de não considerá-lo suficiente, defende que k seja 1/2.

Vejamos por quê. Achinstein (2001, cap. 5) considera a sua teoria da evidência como uma teoria de “probabilidade epistêmica objetiva” (*objective epistemic probability*). Discutiremos melhor a noção de objetividade mais à frente, mas por ora adiantamos que se refere à independência da evidência em relação ao conhecimento ou à crença de qualquer indivíduo ou grupo de indivíduos. A teoria seria epistêmica por interessar-se em esclarecer o que Achinstein (2001, p. 104) chama de “graus de razoabilidade de crença” (*degrees of reasonableness of belief*). Esses graus são probabilísticos, compreendidos entre 0 e 1, onde 0 indica o menor grau possível de razoabilidade de uma proposição, ao passo que 1 indica o seu mais alto grau. Para Achinstein uma evidência objetiva oferece uma boa razão para acreditar em uma hipótese, ou uma razão mais forte para acreditar em uma hipótese h do que em uma hipótese alternativa $\neg h$. Se existe uma boa razão para acreditar em uma hipótese, a probabilidade dessa hipótese ser verdadeira deve ser maior do que a probabilidade de ser falsa, ou seja, deve ser maior do que 1/2.

Como discutiremos na próxima seção, talvez o valor de 1/2 produza uma definição de evidência forte demais quando aplicada à medicina, já que muitas das informações que correntemente são consideradas evidências não tornam a hipótese provável em tão alto grau. Além disso, a definição de evidência como alta probabilidade também está sujeita a contra-exemplos, já que, diferentemente da definição de relevância positiva, não exige que uma nova informação ou evento aumente a probabilidade da hipótese para que sejam considerados evidência a seu favor. É possível, portanto, produzir situações absurdas através da simples conjunção de um evento irrelevante com informações de fundo diante das quais a hipótese já apresenta alta probabilidade. Achinstein (2001) utiliza em várias

situações o exemplo de Michael Jordan para ilustrar esse ponto. Dada a informação de fundo f que “Michael Jordan é um astro do basquete masculino”, a hipótese h que afirma que “Michael Jordan não engravidará” é bastante alta (Achinstein, 2001, p. 71). Quando se acrescenta a f a informação e que “Michael Jordan come os cereais *Wheaties*”, a probabilidade de h permanece extremamente elevada. Como a probabilidade condicional $p(h|f&e)$ é maior do que qualquer valor k muito alto, e deve ser considerado evidência a favor de h , já que satisfaz à definição acima. Em nenhum emprego corrente do termo “evidência”, porém, alguém defenderia que e constitui evidência a favor de h , pois a informação de que Michael Jordan come os cereais *Wheaties* é irrelevante para a hipótese em questão.

Para evitar esses contra-exemplos, Achinstein (2001, cap. 7) acrescenta mais um critério à sua definição de evidência: além de exigir alta probabilidade (isto é, maior do que $1/2$), é necessário que haja um vínculo explicativo entre a hipótese e o evento. Para Achinstein, esse vínculo poderia ser tanto a explicação da hipótese pela evidência, da evidência pela hipótese, ou de ambas por um terceiro fator. Assim, retomando o exemplo da loteria da qual João possui 950 mil bilhetes, essa informação constitui evidência a favor da hipótese de que João será ganhador porque, além de tornar a probabilidade da hipótese muito alta, explica por que a hipótese apresenta alta probabilidade de ocorrer.

Mas para que a definição de evidência defendida por Achinstein (2001, cap. 8) seja completa, é necessário acrescentar ainda um terceiro critério: a não-implicação. Para que e seja evidência a favor de h , e não pode implicar h . Esse critério marca a distinção entre a lógica dedutiva, em que se fala de implicação lógica, e a lógica indutiva, em que se fala de grau de confirmação de uma hipótese. Não faz sentido dizer que um evento e é evidência a favor de uma hipótese h implicada dedutivamente por e , já que, nesse caso, a ocorrência de e torna a ocorrência de h uma certeza. A lógica indutiva, pelo contrário, interessa-se por casos em que a ocorrência de um evento b é indicada apenas de forma mais ou menos provável pela ocorrência de um outro evento a , o que Rudolf Carnap (1891-1970) chamava de “implicação parcial” (1950, p. 206).

Em resumo, para Achinstein (2001, p. 170) a evidência é definida pelos seguintes critérios: e é evidência a favor de h se e somente se (i) $p(h|e) > 1/2$; (ii)

existe uma conexão explicativa entre e e h ; (iii) não há uma relação de implicação entre e e h .

Há ainda uma terceira definição de evidência que, apesar de ser apenas brevemente discutida no livro de Achinstein, gostaríamos de indicar aqui. Essa definição decorre da teoria de Ian Hacking (1936-) acerca da fundamentação lógica da estatística, desenvolvida na obra *Logic of statistical inference*, de 1965. Além disso, teria sido defendida por autores como A. W. F. Edwards e Richard Royall (Achinstein, 2001, p. 125). Hacking (1965, cap.5-6) recorre a uma lógica de suporte para explicar que um evento tenha maior *verossimilhança* (*likelihood*) dada uma hipótese h_1 do que dada uma hipótese concorrente h_2 . Verossimilhança, também um tipo de probabilidade condicional, pode ser definida como a chance de um evento a ocorrer sob condição de uma hipótese h .²¹ Quanto maior a verossimilhança de um evento a dada uma hipótese h , maior a chance desse evento ocorrer caso a hipótese h seja verdadeira. Hacking não desenvolve sua teoria em termos de evidência, mas, como dissemos, é possível extrair uma definição de sua obra. Essa definição, do que poderíamos chamar de *evidência de verossimilhança*, afirmaria que

(3) e é evidência a favor de h_1 se e somente se $p(e|h_1) > p(e|h_2)$.

Ou seja, o evento e será evidência a favor da hipótese h_1 que torna a ocorrência de e mais provável.

Essa definição de evidência, diferentemente das duas definições apresentadas anteriormente, exige que duas hipóteses sejam comparadas para que se afirme que e é evidência a favor de h , caracterizando, portanto, uma evidência comparativa. A ela pode-se claramente objetar que é possível afirmar que um evento e é evidência a favor de uma hipótese h_1 , sem que uma hipótese h_2 seja levada em consideração, ou mesmo que uma hipótese alternativa seja conhecida. Mas, ainda mais grave, utilizando-se dessa definição é possível inferir que e é evidência a favor de qualquer hipótese h_1 , desde que se compare h_1 com uma hipótese h_2 absurda. Essas características tornam a definição de verossimilhança inadequada para os nossos propósitos; por isso, não retornaremos a ela nas discussões que se seguem.

As duas definições de evidência, relevância positiva e alta probabilidade, sendo probabilísticas, devem obedecer às leis fundamentais da probabilidade. Como

²¹ Em notação probabilística $p(a|h)$.

se sabe, existe uma grande controvérsia na interpretação do sentido de uma proposição probabilística, já que, muito esquematicamente, algumas escolas defendem que a probabilidade indica o grau de confiabilidade de uma proposição, enquanto outras afirmam que uma proposição pode ser somente verdadeira ou falsa e que a probabilidade se refere somente à frequência relativa de eventos. Segundo Carnap (1950, §9), o próprio termo “probabilidade” é ambíguo, já que pode ser tomado em seu sentido historicamente mais antigo, e que indica o grau de confirmação de uma proposição, ou em um sentido mais recente, que indica a frequência relativa de eventos em longas sequências. Para distinguir as duas utilizações, que seriam consistentes, Carnap refere-se ao grau de confirmação como “probabilidade₁” e à frequência relativa como “probabilidade₂”.

Apesar dessas controvérsias, alguns postulados são aceitos pela maior maior parte das teorias probabilísticas; esses postulados são os axiomas de Kolmogorov (Hacking, 1965, p. 18-9). O primeiro deles afirma que a probabilidade de um evento a deve estar compreendida entre zero e um:

$$(I) 0 \leq p(a) \leq 1.$$

O segundo axioma afirma que a probabilidade de um evento certo Ω é igual a 1:

$$(II) p(\Omega) = 1.$$

E o terceiro afirma que, para dois eventos a e b mutuamente exclusivos, a probabilidade da disjunção dos eventos é igual à soma das probabilidades de cada um dos eventos:

$$(III) p(a \cup b) = p(a) + p(b).$$

As diferentes interpretações dos postulados básicos da probabilidade e de suas possíveis derivações geram conceitos de evidência distintos. Achinstein (2001, cap. 2) distingue quatro deles: (a) evidência de situação epistêmica (*epistemic situation evidence*, doravante evidência ES); (b) evidência subjetiva; (c) evidência potencial; e (d) evidência verídica. Esses conceitos diferem uns dos outros de acordo com algumas características, como relativização do conhecimento da evidência a uma pessoa, grupo ou situação epistêmica; objetividade e subjetividade; caráter *a priori* ou empírico; e exigência da verdade da evidência e da hipótese. A seguir dedicaremos algumas linhas à exposição de cada um deles.

Começemos pela evidência subjetiva (Achinstein, 2001, p.22-4). Uma de suas principais características é a relativização do conhecimento da informação ou evento

a uma pessoa ou grupo de pessoas em um tempo determinado. Essa pessoa ou grupo deve acreditar que e é verdadeiro e que h é verdadeiro ou provavelmente verdadeiro por causa de e . Ou seja, para que uma informação ou evento seja evidência subjetiva, é necessário que uma pessoa ou um grupo de pessoas saibam dessa informação ou evento e acreditem haver uma relação entre eles e a provável verdade de uma hipótese, mesmo que essa crença não seja racionalmente justificada.

A evidência ES (Achinstein, 2001, p.19-22), por sua vez, relativiza o conhecimento a uma situação epistêmica, e não a uma pessoa ou um grupo. Uma situação epistêmica pode ser compreendida como uma situação em que qualquer indivíduo estaria racionalmente justificado em crer na verdade de uma hipótese e leva em consideração o que esse indivíduo estaria em condição ou não de saber. Porém, diferentemente da evidência subjetiva, a evidência ES não exige que exista, de fato, alguém nessa situação.²² Também contrariamente à evidência subjetiva, que não exige a verdade de e , mas simplesmente que alguém acredite em e , a verdade de e é necessária para constituir evidência ES. Apesar disso, a verdade da hipótese não é exigida, o que faz da evidência ES um conceito falibilista, já que a atribuição de um valor de verdade a uma hipótese pode ser modificada com a mudança da situação epistêmica.

Os dois últimos conceitos de evidência, potencial e verídica, não incluem nenhuma relativização, ao contrário dos dois primeiros conceitos apresentados. Assim como a evidência ES, são conceitos ditos objetivos, pois não dependem da crença ou do conhecimento de nenhum indivíduo ou grupo de indivíduos para que um evento constitua evidência a favor de uma hipótese. Se no caso da evidência ES dissemos que um indivíduo em determinada situação epistêmica está justificado em acreditar na verdade de uma hipótese, no caso das evidências potencial e verídica diremos que a evidência constitui uma boa razão para acreditar na verdade de uma hipótese. Uma boa razão para acreditar em uma hipótese pode ser interpretada como uma relação entre a evidência e a hipótese. É possível que alguém esteja justificado em acreditar em uma hipótese, mas que não tenha boas razões para

²² Na explicação de Achinstein, uma situação epistêmica seria "(...) um tipo abstrato de situação em que, entre outras coisas, alguém sabe ou acredita que algumas proposições são verdadeiras, alguém não está em posição de saber ou de acreditar que outras [proposições] são [verdadeiras], e alguém sabe (ou não sabe) como raciocinar a partir dessas [proposições] para a hipótese de interesse, mesmo se essa situação não existe de fato para ninguém" (2001, p. 20).

fazê-lo, já que lhe faltam informações suficientes para realizar um julgamento apropriado acerca dela. Como dissemos, um dos motivos pelos quais Achinstein julga necessário que o valor k da definição (2) de evidência seja maior do que $1/2$ é que a partir desse limiar existe uma boa razão para acreditar em uma hipótese.

Basicamente, o que distingue as evidências potencial e verdadeira é a exigência da verdade da hipótese. Para que e constitua evidência verdadeira a favor de h , é necessário que e e h sejam verdadeiros. Por outro lado, a evidência potencial exige, além das outras condições apresentadas acima, a verdade de e , mas não a verdade de h . Segundo Achinstein (2000), quando os cientistas referem-se a evidências, utilizam normalmente um desses dois últimos conceitos.

Apesar de termos indicado a distinção conceitual defendida por esse autor, não a utilizaremos dessa forma nas discussões que se seguem. Nosso intuito com essa exposição era, além de fazer jus à teoria que vínhamos apresentando até então, indicar quais são os critérios relevantes para a caracterização conceitual da evidência. Dela, manteremos apenas a distinção mais básica entre evidência subjetiva e objetiva. Ainda que reconheçamos a fragilidade de tal distinção, que pode remeter a comprometimentos metafísicos de difícil defesa, consideramos útil adotá-la, pois facilitará a caracterização de níveis diferentes de atividades médicas. Além disso, como tentaremos indicar, essa é uma distinção amplamente utilizada na literatura especializada, o que oferece, no mínimo, a vantagem de uma uniformidade terminológica. Deixaremos de lado as demais subcategorias porque a sua adoção implicaria aceitar também o valor de $1/2$ como limiar de probabilidade para a definição de evidência como alta probabilidade. Como já antecipamos e desenvolveremos melhor mais à frente, consideramos esse valor alto demais para a utilização que normalmente se faz do termo “evidência” em medicina.

Tendo discutido brevemente as duas definições de evidência, isto é, relevância positiva e alta probabilidade, e alguns conceitos de evidência, julgamo-nos em condições de abordar a questão da evidência na medicina. A próxima seção será dedicada a essa tarefa.

2.3 Evidência na medicina

As relações entre a noção de evidência e o pensamento médico remontam ao período de surgimento das primeiras teorias da probabilidade na literatura filosófica

e científica ocidental. De fato, Hacking (1984, cap. 3-4) afirma que o conceito de evidência foi uma condição para a emergência da probabilidade como a compreendemos atualmente e que esse conceito origina-se nas ciências “menores”, como a química e a medicina. Durante a Idade Média e o Renascimento, distinguia-se ciência e opinião através da ideia de que, na ciência, o conhecimento é demonstrado dedutivamente, ao passo que as opiniões são apenas mais ou menos prováveis, de acordo com as autoridades que se posicionam a favor delas. “Provável”, nesse sentido, não indica o que tem mais chance de acontecer ou de ser verdadeiro, mas o que é mais louvável ou meritório. Ao longo do século XVI e no início do século XVII, ocorre uma mudança de sentido do termo “probabilidade”, pois uma afirmação provável deixa progressivamente de ser aquela a favor da qual conta o peso da autoridade, para se tornar aquela mais suportada pelos sinais da natureza, ou pela evidência. Compreensivelmente, os sinais mais fortes são aqueles que ocorrem com maior frequência, o que já nos aproxima das concepções modernas de probabilidade. Como a medicina não poderia ser considerada uma ciência, no sentido de uma ciência dedutiva conforme ao pensamento medieval, devia ater-se aos sinais fornecidos pela natureza e, portanto, à evidência.

Com essa brevíssima referência, quisemos apenas indicar que a utilização do termo evidência na medicina é historicamente adequada e reforçar o argumento defendido no capítulo 1 de que as discussões acerca da evidência na medicina não se originam com a EBM. Resta saber se essa utilização é adequada também de uma perspectiva analítica.

Como primeiro passo para uma análise da evidência médica, gostaríamos de propor uma distinção entre três níveis de atividades do que normalmente se entende por “medicina”: (i) a pesquisa laboratorial em ciências biológicas básicas, orientada para o esclarecimento de questões de interesse médico, como a fisiopatologia das doenças, e para a investigação de substâncias com potencial terapêutico, a que chamaremos de *pesquisa laboratorial biomédica*; (ii) a investigação da eficácia de intervenções terapêuticas, que utiliza comparação entre grupos de pacientes com entidades clínicas determinadas e recorre à análise estatística dos resultados, a que chamaremos de *pesquisa clínica*; e (iii) e a atenção e o cuidado com pacientes individuais, nos consultórios, clínicas e hospitais, a que chamaremos de *prática clínica*.

Consideramos essa distinção adequada por indicar uma divisão de atividades que ocorre de fato na medicina contemporânea, pois os especialistas tendem a dedicar a maior parte de seus esforços produtivos a um desses níveis em particular. Não afirmamos, no entanto, que sejam estanques e reconhecemos haver certa mobilidade de profissionais nos diversos níveis, já que é possível, por exemplo, que um profissional da prática clínica se dedique concomitantemente à pesquisa clínica ou à pesquisa biomédica. Apesar disso, as atribuições de cada um desses níveis de atividade são dificilmente confundidas. Não se espera que um clínico contribua para o avanço do conhecimento médico, mas que ofereça um cuidado humano e tecnicamente adequado a seus pacientes. Inversamente, não é função dos pesquisadores prestar assistência a indivíduos doentes, mas sim contribuir de maneira significativa na solução de problemas científicos. Inclusive, em muitos casos o cuidado de pacientes por pesquisadores seria legalmente proscrito, pois muitos profissionais de pesquisa clínica e biomédica não receberam uma formação médica; além disso, é possível que haja conflitos de interesses entre as duas atividades. As próprias questões éticas suscitadas pelos dois níveis de pesquisa e pela prática são diferentes, já que no caso da pesquisa biomédica relacionam-se com o manejo de animais de laboratório (Festing, Wilkinson, 2007), ao passo que na pesquisa clínica envolvem os riscos potencialmente muito elevados de submeter seres humanos a novos tratamentos experimentais. Na prática clínica, por sua vez, as questões éticas relacionam-se com a conduta individual do médico, com a honestidade de sua prática, com o domínio técnico dos tratamentos e intervenções que oferece e com a relação médico-paciente.

O desenvolvimento de uma nova intervenção terapêutica segue fases análogas à divisão que propomos. Como se sabe, antes que uma droga seja testada em seres humanos, sua segurança deve ter sido demonstrada em estudos pré-clínicos em modelos animais. Apenas após extensa coleta de informações sobre segurança, farmacocinética e possíveis mecanismos de ação, novas substâncias são liberadas para pesquisa clínica. Similarmente, somente após longos anos de pesquisas clínicas que demonstram a segurança e a eficácia dessa substância em seres humanos, a nova droga pode ser aprovada para ampla utilização na prática clínica (Umscheid, Margolis, Grossman, 2011). A separação entre os diferentes tipos de atividades pode ser ainda indicada pela dificuldade da estruturação de um estudo

clínico a partir dos dados da pesquisa laboratorial e pela dificuldade de aplicação dos resultados dos ensaios clínicos à prática clínica. De fato, as diferenças de abordagem dos dois níveis de pesquisa motivou a criação do conceito de *pesquisa translacional*, que visa justamente desenvolver desenhos de ensaios clínicos a partir das informações obtidas pela pesquisa laboratorial, levando em consideração que a eficácia terapêutica e a toxicidade de uma nova substância podem ser extremamente diferentes em animais e humanos. Retornaremos a esse conceito quando discutirmos, no capítulo 3, as noções de progresso médico.

Além de adequada, essa distinção nos parece produtiva, já que possibilita compreender as características particulares de cada um desses níveis de atividade e as maneiras com que se articulam. Como desenvolveremos a seguir, um conceito objetivo de evidência é mais adequado para a descrição da evidência gerada pelos primeiros dois níveis de pesquisa, ao passo que um conceito subjetivo é necessário para esclarecer a utilização da evidência pelo profissional da prática clínica no manejo de seu paciente individual. Além disso, parece-nos que a evidência produzida pela pesquisa clínica dependa em grande medida de informações de fundo produzidas pela pesquisa biomédica.

Finalmente, acreditamos que, historicamente, os diversos níveis apresentaram padrões de desenvolvimento distintos uns dos outros. A pesquisa biomédica desenvolveu boa parte de sua metodologia e de seu instrumental material e conceitual já em meados do século XIX. A pesquisa clínica, por sua vez, apesar de remontar ao início do século XIX, como indicamos no primeiro capítulo, só atingiria um patamar de maturidade científica, do nosso ponto de vista, nas décadas de 1920 e 1930, pois dependia da elaboração de teorias consistentes de testagem de hipóteses estatísticas. Já a prática clínica pode ser considerada como a atividade médica mais fundamental, estando presente sempre que um médico oferece algum tratamento ao seu paciente, ainda que essa descrição seja bastante grosseira. No entanto, neste trabalho não apresentaremos argumentos históricos suficientes para defender essa afirmação e reservaremos essas questões para desenvolvimentos futuros.

Passemos ao estudo um pouco mais detalhado de cada um desses níveis de atividades. Por *pesquisa laboratorial biomédica*, compreendemos qualquer atividade pré-clínica de pesquisa desenvolvida em laboratórios e que utilize o instrumental

conceitual e material característico das ciências biológicas básicas, como patologia, microbiologia, bioquímica e fisiologia, visando o esclarecimento dos mecanismos subjacentes às doenças e o desenvolvimento de novos recursos terapêuticos. Consideramos que também pertencem a essa categoria as pesquisas que, ainda que não tenham sido claramente motivadas por um problema clínico, como uma doença específica, possam ser utilizadas proveitosamente nos dois outros níveis de atividades médicas. Destacamos os modelos animais de doenças humanas e as técnicas de engenharia molecular, que permitem caracterizar as etapas críticas da fisiopatologia da doença, identificar as moléculas especialmente ativas nessas etapas e visá-las como alvos privilegiados da ação terapêutica, através do desenho de outras moléculas capazes de se ligar especificamente às primeiras (McGonigle, Ruggeri, 2014). Essa é uma via particularmente importante – ainda que não seja a única – de desenvolvimento de novas classes farmacológicas e de progresso na terapêutica, como demonstram os novos anticorpos alvo-moleculares utilizados no tratamento do câncer e de doenças auto-imunes e alguns anti-hipertensivos, como o captopril, cuja história esboçaremos abaixo.

Por *pesquisa clínica*, compreendemos qualquer investigação que tome como objeto indivíduos humanos acometidos por entidades mórbidas bem definidas, no intuito de caracterizá-las ou de testar seus possíveis tratamentos. Essa categoria pode ser ainda subdividida em (i) *pesquisa clínica observacional* (ou *epidemiológica*), que indica a associação entre fatores etiológicos, demográficos e terapêuticos, e variações no curso das doenças, sobretudo através de análises retrospectivas; e (ii) *pesquisa clínica experimental*, em que a intervenção é planejada e aplicada ativamente sob condições controladas, e em que os dados são coletados prospectivamente. A evidência produzida por estudos observacionais tende a ser considerada mais fraca do que a evidência produzida por estudos experimentais, cujo exemplo padrão é o RCT, que discutimos detalhadamente na seção 1.4. É uma característica proeminente dos dois tipos de pesquisa clínica o recurso a procedimentos estatísticos, no sentido de demonstrar a associação entre duas ou mais variáveis consideradas importantes. Ainda que a pesquisa laboratorial biomédica também recorra à estatística, sobretudo para lidar com questões de medida, o pensamento dos dois níveis de pesquisa é essencialmente diferente, já que a pesquisa laboratorial procura explicar a doença e a ação de medicamentos

através de seus mecanismos subjacentes, enquanto a pesquisa clínica aponta relações que, a rigor, não é capaz de explicar. Retornaremos também a essa questão quando discutirmos, no capítulo 3, duas noções distintas de progresso na medicina.

Finalmente, por *prática clínica* nos referimos ao cuidado médico em nível individual, à tentativa de formulação de um diagnóstico correto com base na história clínica de um paciente que procura ajuda especializada e à proposta de uma terapêutica adequada. Esse nível de atividade deve levar em consideração as particularidades do indivíduo doente, sua situação psicossocial, a presença de comorbidades, suas preferências e valores pessoais. Além disso, depende também da experiência pessoal do médico, da sua segurança com determinados tratamentos, da sua formação acadêmica e de seus valores. Apesar de informar-se dos resultados das pesquisas biomédica e clínica, a prática clínica muitas vezes se distancia delas, porque as situações reais de exercício da medicina diferem amplamente das situações ideais de pesquisa e porque em muitas situações o médico percebe que o tratamento padrão não será satisfatório para seu paciente em particular, percepção que pode não ser explicitamente formulada.

Como descrever, através das definições e conceitos trabalhados acima, a evidência nos três níveis de atividades médicas? Começemos por uma breve história do anti-hipertensivo captopril, que ilustra muito bem as etapas ideais de desenvolvimento de um novo fármaco.

Na década de 1950, alguns pesquisadores debruçaram-se sobre a importância do sistema renina-angiotensina na patologia da hipertensão arterial. Sabia-se que a renina, uma substância produzida pelos rins, era responsável por formar, a partir de um precursor presente no sangue, uma substância capaz de elevar os níveis pressóricos através da contração do músculo liso dos vasos sanguíneos. Essa substância formada pela renina era conhecida na época pelo nome de hipertensina, mas foi posteriormente rebatizada com o nome de angiotensina, pelo qual a conhecemos hoje. A renina é liberada sempre que o fluxo sanguíneo dos rins diminuída, seja por causa de diminuições do volume sanguíneo circulante, seja por obstruções mecânicas nas artérias renais. Apesar disso ser um mecanismo fisiológico, que permite a manutenção do equilíbrio dos níveis pressóricos apesar de fatores adversos, como desidratação e perda sanguínea,

observou-se que os níveis de hipertensina estavam anormalmente elevados em indivíduos com hipertensão arterial. Skeggs e colaboradores (1954) demonstram que havia duas formas distintas de hipertensina, chamadas hipertensina I e II (angiotensina I e II, na nomenclatura atual), e que os efeitos vasopressores da hipertensina II eram mais poderosos do que os da hipertensina I. A transformação de hipertensina I em hipertensina II era dependente da ação de uma enzima que foi caracterizada (Skeggs, Kahn, Shumway, 1956) e que posteriormente recebeu o nome de enzima conversora de angiotensina (ECA).

Em 1965, o médico e pesquisador brasileiro Sérgio Henrique Ferreira (1934-2016) demonstra que o extrato de veneno da serpente *Bothrops jararaca* era capaz de inibir a destruição de uma outra substância responsável pela regulação da pressão arterial, a bradicinina. Outros pesquisadores (Ng, Vane, 1970) observam que esses extratos de veneno de jararaca inibiam também a enzima conversora de angiotensina e que os efeitos hipertensivos da angiotensina I podiam ser bloqueados pela utilização desse extrato. O isolamento dos seus componentes ativos ocorreu nos anos seguintes, resultando na identificação do teprotide, a primeira substância com potencial de utilização clínica como anti-hipertensivo cujo mecanismo de ação era a inibição da ECA. No entanto, o teprotide era ativo apenas através da administração endovenosa, o que limitava a sua aplicação na prática clínica. Em 1977, Ondetti e colaboradores conseguem produzir, com técnicas de engenharia molecular, uma substância análoga ao teprotide, mas que poderia ser utilizada pela via oral. Essa substância, que chamaram inicialmente de SQ 14,225, ficaria conhecida como captopril.

Em 1977, Murthy e colaboradores testam o captopril em ratos normotensos e demonstram que a sua administração bloqueia o aumento da pressão arterial que normalmente se segue à injeção de angiotensina I. No ano seguinte, Laffan e colaboradores (1978) reproduzem esses resultados em diversos modelos experimentais de hipertensão cujo mecanismo presumido era uma ativação exacerbada do sistema renina-angiotensina, também concluindo que captopril é eficaz nessas situações. Todos esses estudos que relatamos até aqui são exemplos do que chamamos de pesquisa laboratorial biomédica.

Nos anos seguintes, são realizados os primeiros estudos clínicos acerca da eficácia do captopril, entrando no segundo nível de atividades que distinguimos. Em

1979, Brunner e colaboradores testam a eficácia do captopril em 22 pacientes com hipertensão arterial e concluem que a medicação reduz, de forma estatisticamente significativa, os seus níveis pressóricos. Esses resultados são reproduzidos em novos ensaios clínicos, utilizando amostras maiores de pacientes. Podemos citar o estudo do *Veterans Administration Cooperative Study Group on Antihypertensive Agents* (1984), que testou o captopril em 495 indivíduos hipertensos e que se destaca por ter utilizado os recursos de randomização e duplo cego que discutimos na seção 1.4, sendo considerado, por isso, de melhor qualidade metodológica.

Partindo da hipótese h que afirma que “captopril reduz os níveis pressóricos de pacientes humanos com hipertensão arterial”, normalmente considera-se que tanto as pesquisas laboratoriais, quanto as pesquisas clínicas apresentadas acima constituam evidência a seu favor. No entanto, de acordo com as hierarquias de evidência defendidas pela EBM e apresentadas na seção 1.3, a evidência e_2 produzida por ensaio clínicos deve ser considerada mais forte do que a evidência e_1 , produzida pela pesquisa com modelos animais. Utilizando a definição de evidência como relevância positiva, podemos dizer que e_2 é evidência mais forte do que e_1 se o incremento da probabilidade da hipótese h produzido por e_2 for maior do que o incremento na probabilidade da hipótese h produzido por e_1 . Além disso, é necessário que tanto e_1 , quanto e_2 produzam algum incremento na probabilidade de h . Em notação probabilística, de acordo com a noção de relevância positiva, poderíamos definir evidência mais forte da seguinte forma:

(IV) e_2 é evidência mais forte do que e_1 a favor da mesma hipótese h se e somente se $p(h|e_2) - p(h) > p(h|e_1) - p(h)$, onde tanto $p(h|e_1) > p(h)$ quanto $p(h|e_2) > p(h)$.

A primeira parte da definição obviamente equivale a $p(h|e_2) > p(h|e_1)$.

Uma definição da força da evidência também é possível quando se adota uma noção de alta probabilidade. e_2 será evidência mais forte do que e_1 a favor da hipótese h , se a probabilidade de h dado e_2 for maior do que a probabilidade de h dado e_1 , sendo que tanto a probabilidade de h dado e_1 , quanto a probabilidade de h dado e_2 devem ser maiores do que um limiar k . Em notação probabilística:

(V) e_2 é evidência mais forte do que e_1 a favor da mesma hipótese h se e somente se $p(h|e_2) > p(h|e_1)$, onde tanto $p(h|e_1) > k$ quanto $p(h|e_2) > k$.

Nessa definição, a força de uma evidência é compreendida em termos da diferença da probabilidade da hipótese em relação a um limiar k . Quanto maior for a diferença em relação a k da probabilidade da hipótese dada a evidência, mais forte será a evidência.

Como havíamos antecipado, o valor de $1/2$ para o limiar k proposto por Achinstein (2001) nos parece inadequado para a caracterização do conceito de evidência conforme utilizado na medicina. Se concedermos que os resultados positivos dos estudos utilizando modelos animais de hipertensão arterial constituem evidência a favor da eficácia do captopril na redução da pressão arterial, teremos que afirmar também, seguindo esse ponto de vista, que esses resultados tornam a probabilidade da eficácia do captopril maior do que $1/2$. No entanto, acreditamos que nenhum proponente da EBM defenderia tal afirmação, porque poderia motivar a utilização na prática clínica de substâncias que não foram testadas em ensaios clínicos, o que é exatamente o contrário do que pretendem. Isto é, 50% de chance de determinada substância ser eficaz pode ser considerada uma chance suficientemente alta de sucesso para motivar a sua utilização na prática clínica, principalmente quando não são conhecidas alternativas.

Uma possibilidade de lidar com esse problema encontra-se no recurso ao conceito de informação de fundo. Pode-se dizer, por exemplo, que na passagem do nível pré-clínico para o nível clínico uma nova informação de fundo f modifica a interpretação dos resultados da pesquisa. Essa informação f afirma que “as respostas a substâncias experimentais são diferentes em animais e humanos”. Assim, a probabilidade $p(h|e_1)$ de que captopril reduza os níveis pressóricos em humanos, dada os resultados e_1 dos experimentos com animais, é maior do que $1/2$. No entanto, a probabilidade $p_f(h|e_1)$, que leva em consideração a informação de fundo f , é menor do que $1/2$, o que faz com que e_1 deixe de ser considerado evidência a favor de h , se utilizarmos a definição de alta probabilidade e se aceitarmos o valor de $1/2$ para k . Mas formulamos a hipótese h afirmando justamente que captopril é eficaz em reduzir os níveis pressóricos *de humanos* e continuamos considerando que as informações obtidas de estudos com animais constituam evidência a seu favor. Fica claro que essa também é a posição da EBM quando se observa que os resultados de pesquisas laboratoriais estão inseridos nas

suas hierarquias de evidência, ainda que como evidência mais fraca do que a produzida por ensaios clínicos.

Quanto aos diferentes conceitos de evidência que discutimos acima, parece-nos que existe uma importante diferença entre os conceitos adequados para a descrição da evidência na pesquisa laboratorial biomédica e na pesquisa clínica, por um lado, e na prática clínica, por outro. Argumentaremos que nos primeiros dois níveis de atividades médicas um conceito objetivo de evidência é utilizado, ao passo que um conceito subjetivo de evidência é empregado na prática clínica. Com o conceito de evidência objetiva, queremos indicar, basicamente, uma relação entre duas proposições ou entre dois eventos que não depende do conhecimento de um indivíduo ou grupo de indivíduos, ou que não é relativizada a esse conhecimento.²³ Por exemplo, o fato dos resultados dos experimentos com animais de laboratório constituir evidência a favor da eficácia anti-hipertensiva do captopril independe da crença ou do conhecimento de qualquer indivíduo, mas pode ser compreendido como uma relação lógica entre as proposições que expressam a evidência e a hipótese, ou como uma relação explicativa entre elas.

A primeira posição é defendida por autores como Carnap (1950), que desenvolveu um sistema de lógica indutiva para explicar o que chamava de “probabilidade₁”, ou grau de confirmação. Para ele, interessava compreender como um valor r poderia ser atribuído à relação entre evidência e hipótese em proposições probabilísticas do tipo $p(h|e) = r$. Esse valor dependia, basicamente, do sistema semântico em que as duas proposições e e h haviam sido formuladas e da função de confirmação escolhida. Apesar dessas proposições poderem ser formuladas com base na experiência, a relação entre elas é *a priori*, do mesmo modo que em lógica dedutiva as relações entre as premissas e a conclusão dependem apenas de sua forma lógica. A verdade da hipótese e da evidência, portanto, não são necessárias para que a relação expressa por uma proposição probabilística seja válida. Assim, se a evidência e afirma, por exemplo, que “um dado é simétrico”, e a hipótese h

²³ Segundo Achinstein, uma interpretação objetiva da evidência afirma que uma proposição probabilística ser verdadeira “(...) é um fato objetivo que não tem nada a ver com o que alguma pessoa sabe, ou acredita. Se uma sentença dessa forma $[p(h|e) = r]$ é verdadeira ou falsa, isso é assim quer alguém acredite ou não, ou mesmo saiba sobre e ou h ou sua relação (...)” (2001, p. 48-9). A distinção de Carnap é bastante parecida: “As relações são objetivas, não subjetivas, neste sentido: se uma dessas relações é verdadeira ou não em um caso concreto não depende de se ou do que alguém pode calhar de imaginar, pensar, acreditar, ou saber sobre essas sentenças” (1950, §11). Obviamente, essa posição não é aceita pelas teorias ditas subjetivistas.

afirma que “o número três cairá na próxima jogada com esse dado”, o grau de confirmação $p(h|e)$ é igual a $1/6$, devido a uma relação *a priori* e objetiva entre essas sentenças. No entanto, nas ciências de um modo geral, e também na medicina, não se considera que a relação entre a evidência e a hipótese seja *a priori*, mas sim empírica, pois a validade de uma proposição expressando o grau de confirmação de uma hipótese decorre de características factuais do mundo físico, apreendidas observacional ou experimentalmente.

A posição que defende que as relações entre hipótese e evidência são de tipo explicativo pode ser classificada como objetiva e empírica. Um autor que trabalhou as questões da probabilidade através de uma perspectiva empírica é Hacking (1965), que se propôs a esclarecer a noção de chance, compreendida como uma propriedade disposicional do mundo físico. Para Hacking, a chance do lançamento de uma moeda resultar em cara depende de propriedades físicas da moeda, como peso, simetria, distribuição da massa, e do seu dispositivo de lançamento, de modo que essa propriedade pode ser apreendida experimentalmente, através de longas sequências de lançamentos. Ainda que Hacking não tenha desenvolvido sua teoria em termos de evidência, como observamos na seção precedente, é possível extraí-la de sua obra. Se for aceito que uma proposição probabilística refere-se a uma propriedade disposicional do mundo físico, pode-se interpretar uma proposição do tipo $p(h|e)$, que relaciona uma evidência a uma hipótese, como também indicando uma relação física entre elas.

Para Achinstein (2001, cap. 5) a verdade de uma proposição desse tipo pode ser tanto *a priori*, quanto empírica, dependendo de como são formuladas as proposições e e h . Por exemplo, se e_1 afirma que “uma moeda é simétrica”, e h afirma que “a próxima jogada resultará em cara”, a probabilidade $p(h|e_1)$ pode ser determinada *a priori* como $1/2$. No entanto, se e_2 afirma que “em uma sequência de 100 lançamentos com essa moeda, obteve-se cara apenas 20 vezes”, e h afirma que “a próxima jogada resultará em cara”, a probabilidade $p(h|e_2)$ passa a ser interpretada como empírica e recebe um valor de $1/5$.

Voltando ao caso do captopril, afirmamos que, do nosso ponto de vista, a relação das evidências laboratorial e clínica a favor da hipótese de sua eficácia pode ser compreendida como uma relação empírica, em oposição a uma relação *a priori*. Isto é, o fato da administração de captopril reduzir a pressão arterial de ratos e de

humanos pode ser explicado por um mecanismo fisiológico subjacente, que esclarece como o captopril bloqueia a ação de uma enzima (ECA) responsável pela produção de uma substância (angiotensina II) com forte ação vasoconstritora e, portanto, hipertensiva. Essa relação é, como dissemos, objetiva, pois independe da crença ou do conhecimento de qualquer indivíduo ou grupo, e empírica, porque depende de propriedades do mundo físico, passíveis de serem acessadas observacional e experimentalmente. Isso é válido mesmo para os casos de pesquisas clínicas que apontam uma relação estatística entre duas variáveis, como, por exemplo, um agente ambiental e uma doença, sem que haja uma explicação do mecanismo de interação entre as duas variáveis. No entanto, presume-se que tal mecanismo exista e que seja passível de determinação ou esclarecimento através de pesquisas adicionais.

A situação é completamente distinta na prática clínica, em que sempre é necessário relativizar a evidência ao conhecimento ou à crença de um indivíduo: o médico. Ao informar-se dos resultados das pesquisas laboratorial e clínica e ao utilizá-los para formular um diagnóstico e propor uma terapia, o médico os toma como sua evidência pessoal, ou subjetiva. Isto é, esses resultados constituem a razão pela qual o médico *sabe* ou *acredita* estar realizando um diagnóstico e prescrevendo um tratamento corretamente. Dizemos, nesse caso, que para o médico A o fato de captopril ter reduzido a pressão arterial de pacientes hipertensos em um RCT constitui evidência para guiar a sua conduta diante do paciente B com níveis pressóricos elevados. Mas certamente essa não será a única informação considerada na sua tomada de decisão. O médico A avalia ainda a sua experiência pessoal com o captopril, a história pregressa do paciente, suas comorbidades, entre outros fatores. Pode ser que se lembre, por exemplo, já ter oferecido captopril ao paciente B, e que a medicação não produziu os efeitos esperados. Isso não muda o fato de captopril ser eficaz na redução dos níveis pressóricos de pacientes hipertensos, nem o fato dos estudos citados acima constituírem evidência objetiva a favor da hipótese da eficácia de captopril; muda apenas a evidência pessoal do médico A em relação à eficácia de captopril para o paciente B. Ou seja, a abstração do contexto, característica de uma interpretação objetiva da evidência, não pode ocorrer quando um conhecimento é aplicado. Isso explica, entre outras coisas, que

dois médicos cheguem a conclusões distintas diante de um mesmo caso, o que ocorre com bastante frequência na prática clínica.

É preciso notar que a interpretação subjetiva da evidência na prática clínica não decorre da necessidade de levar em consideração as possíveis consequências de uma decisão e de avaliá-las segundo os valores do médico e do paciente. Obviamente essa avaliação é extremamente importante, mas ela não afeta o suporte que uma evidência confere a uma hipótese. Em teoria da decisão, costuma-se definir o valor esperado de uma ação como o a soma dos produtos das probabilidades de cada uma de suas consequências pela suas respectivas utilidades (Hacking, 2001, p. 82), conforme a seguinte equação:

$$(V) \text{ Valor esperado da ação } A = \sum p(C_n) \times u(C_n),$$

onde C_n indica as n possíveis consequências da ação A , $p(C_n)$ a sua probabilidade dada a ocorrência de A , e $u(C_n)$ a utilidade de cada uma dessas consequências. A utilidade pode ser interpretada como o valor que se atribui a uma determinada consequência, sendo expressa na maior parte das vezes por um valor monetário, a fim de facilitar a sua quantificação.

Mais uma vez devemos recorrer a um exemplo retirado de jogos de azar para ilustrar esses conceitos. Imaginemos uma loteria cujo total de bilhetes vendidos é 100, cujo preço por bilhete é \$10,00 e cujo prêmio é \$1000,00. Para calcularmos o valor esperado da ação A “comprar um bilhete dessa loteria”, devemos multiplicar a chance de ganharmos com um bilhete (0,01) pela utilidade desse resultado, isto é, receber \$1000,00, e somá-lo com o produto da chance de não ganharmos (0,99) pela sua utilidade (\$0,00), além de subtrair o produto de uma utilidade de \$10,00 (o preço do bilhete) pela chance de perdermos essa quantia (1,00, pois é preciso comprar o bilhete para concorrer na loteria). O valor esperado dessa ação A seria, portanto:

$$\text{Valor esperado de } A = (0,01 \times \$1000,00) + (0,99 \times 0) + [1,00 \times (-\$10,00)] = 0,$$

o que faz dessa loteria um jogo justo, ou seja, um jogo em que não se está fadado a ter um valor esperado negativo, independente dos possíveis resultados.

De uma perspectiva subjetivista, não se pode dizer que a probabilidade que alguém atribui a um evento seja incorreta, pois trata-se de uma probabilidade pessoal. No entanto, pode-se exigir que as probabilidades atribuídas a eventos alternativos sejam coerentes, isto é, que obedeçam aos axiomas de Kolmogorov

apresentados acima. Para dois eventos a e $\neg a$ mutuamente exclusivos e exaustivos, por exemplo, se se atribui um valor de 0,6 à probabilidade $p(a)$, deve-se coerentemente atribuir um valor de 0,4 à probabilidade $p(\neg a)$. A validade dessa exigência pode ser vista exatamente quando se aplica as probabilidades subjetivas à tomada de decisões, já que, caso as probabilidades não sejam coerentes, é possível envolver-se em uma situação em que se está fadado a ter valor um esperado negativo.

Neste capítulo, analisamos apenas a primeira variável da equação (V), isto é, $p(C_n)$, que calculamos com base na evidência disponível, já que não faz sentido falar em evidência em relação a utilidades, ou valores. A determinação das consequências de uma decisão clínica, levando em consideração os valores do paciente, parece-nos de uma importância fundamental, mas certamente extrapola as pretensões deste trabalho. Ainda assim, julgamos ter dado alguns passos importantes no esclarecimento da noção de evidência, conforme empregada na medicina. Antes de concluirmos o capítulo, gostaríamos apenas de fazer algumas observações sobre o problema da atribuição de um valor r a uma probabilidade condicional $p(h|e)$, e sobre como esse problema relaciona-se com os testes de hipóteses estatísticas.

2.4 Evidência e testes de hipóteses

O recurso a exemplos simples, como o lançamento de um dado, a retirada de uma bola de uma urna, ou outros procedimentos típicos de jogos de azar, apesar de ser amplamente utilizado na literatura especializada, pode levar a alguns mal entendidos. O principal deles é julgar que em casos de extrema complexidade, como a avaliação de hipóteses ou teorias competidoras à luz da evidência disponível, seja sempre possível determinar os fatores envolvidos nessa avaliação. Isso é especialmente verdadeiro quando se analisa a atribuição de um valor a uma proposição probabilística de tipo $p(h|e) = r$, ou seja, o grau de confirmação de uma hipótese por uma evidência. Diante da evidência atualmente disponível, qual é o grau de confirmação da hipótese de que o hábito de fumar cause câncer de pulmão? Ou da hipótese que afirma que captopril é eficaz em reduzir os níveis pressóricos de pacientes hipertensos? Nos dois casos poderíamos responder que o grau de confirmação das hipóteses é bastante alto. Mas quão alto?

Na pesquisa clínica, que recorre incessantemente aos testes de hipóteses estatísticas para demonstrar a relação entre duas variáveis, o nível de significância dos testes poderia ser erroneamente tomado como o grau de confirmação de uma hipótese. Como discutimos na seção 1.2, um nível de significância de 0,05 indica que em apenas 5% dos casos os resultados do estudo seriam obtidos caso não houvesse diferença real entre os grupos comparados. Como essa chance é muito baixa, normalmente se infere que os grupos são de fato diferentes e que a diferença se deve a um fator específico, como um medicamento que havia sido administrado a apenas um desses grupos.

Por exemplo, no estudo de Brunner e colaboradores (1979) a que nos referimos acima, a média da pressão arterial de oito pacientes com hipertensão arterial de origem vascular foi comparada antes e após a administração de captopril. Os pesquisadores realizam um teste de hipóteses e concluem que a diferença das médias é estatisticamente significativa, com um nível de significância de 0,001. Isso indica que o grau de confirmação da hipótese de eficácia do captopril com base nos resultados desse estudo é de 0,999? Certamente não.

O outro ensaio clínico que mencionamos, realizado pelo *Veterans Administration Cooperative Study Group* (1984), compara a média de pressão arterial de 427 pacientes, que haviam sido alocados aleatoriamente para receber captopril ou placebo. Um teste de hipóteses estatísticas demonstra que a diferença entre os grupos apresenta um nível de significância de 0,003, menos rigoroso do que o do estudo mencionado acima. Deveríamos concluir, levando em consideração a diferença de significância dos dois estudos, que os resultados do estudo de 1979 conferem um suporte mais forte à hipótese de eficácia do captopril do que os resultados do estudo de 1984? Mais uma vez, certamente não. Os dois estudos apresentam importantes diferenças metodológicas, que são avaliadas quando se considera o grau de confirmação que os seus resultados emprestam à hipótese.

O primeiro deles é o tamanho da amostra. Normalmente, considera-se que quanto maior o número de indivíduos estudados, maior é a força evidencial de um estudo. Isso pode ser justificado recorrendo à ideia de que quanto maior uma amostra, maior a chance da frequência relativa observada de um evento aproximar-se da frequência relativa real do evento na população de onde a amostra foi retirada. Amostras maiores possibilitam também um maior grau de precisão de estimativas, já

que o erro padrão de uma medida é inversamente proporcional ao tamanho da amostra (Altman, Bland, 2005). Além disso, amostras maiores são mais apropriadas para a detecção de eventos relativamente raros, como alguns efeitos colaterais, cuja chance de ocorrência em amostras pequenas é bastante baixa.

Por outro lado, o tipo de amostragem também influencia a interpretação dos resultados de um estudo. Idealmente, espera-se que os indivíduos captados para a pesquisa tenham sido escolhidos aleatoriamente entre todos os indivíduos que constituem a população de interesse. No entanto, como na maior parte das situações práticas a população não está inteiramente disponível para que cada um de seus indivíduos tenha a mesma chance de ser escolhido, esse tipo de amostragem pode ser considerado apenas um ideal teórico. Ao contrário, a alocação aleatória, ou randômica, dos indivíduos da amostra nos diferentes grupos do estudo é considerada um procedimento exequível e que garante a validade dos testes estatísticos, como discutido na seção 1.4. A randomização, portanto, impactaria positivamente a força da evidência produzida pelos estudos que se utilizam dela. No caso dos dois estudos citados acima, o fato do ensaio de 1984 ter sido randomizado faz com que os seus resultados sejam considerados muito mais fortes do que os resultados do ensaio de 1979.

Considerações de ordem prática, como a avaliação das consequências da aceitação ou da rejeição de uma hipótese, apesar de, a rigor, não influenciarem o seu grau de confirmação por uma evidência, também são feitas quando se interpreta os testes de hipóteses. Jerzy Neyman (1894-1981), um dos autores responsáveis pelo desenvolvimento de uma teoria de testagem de hipóteses que seria extremamente influente na pesquisa clínica, tratando da inadequação de se assumir que um mesmo nível de significância, como 0,05, seja apropriado para todos os testes de hipóteses, afirma que:

“De fato, parece ser desejável determinar o nível de significância de acordo com algumas circunstâncias que variam de um problema particular para o próximo. Uma é a consequência de cometer um erro (chamado de [erro] do primeiro tipo), quando a hipótese [nula] testada é de fato verdadeira e é rejeitada. Dependendo do problema, as consequências podem ser graves e, nesse caso, um nível de significância mais rigoroso, talvez 0,001, será indicado.” (Neyman, 1957, p. 12)

Isso indica como questões extrínsecas ao teste de hipóteses podem influenciar a escolha do nível de significância para um estudo em particular e que fatores de natureza não epistêmica, como os valores, estão implicados nas decisões práticas, conforme observamos na seção precedente.

Poderíamos multiplicar os fatores considerados quando se pesa os resultados de um estudo. Mas parece-nos que já está suficientemente claro que os testes de hipóteses são apenas um deles. O sistema de avaliação da evidência proposta pelo GRADE, que apresentamos na seção 1.3, tenta responder a esses problemas, através da atribuição de pontos positivos ou negativos aos estudos, considerando algumas de suas características metodológicas. De fato, as contribuições da EMB nos parecem em certa medida limitadas a essa avaliação das questões metodológicas da pesquisa clínica, conforme concluímos no capítulo 1. No entanto, o problema de atribuir um valor específico r ao grau de confirmação de uma hipótese não é solucionado pelas hierarquias de evidência propostos pelo movimento, já que o resultado da computação de pontos de um estudo não gera a probabilidade que uma hipótese adquira com base nos seus resultados. A consideração do grau de confirmação de hipóteses médicas permanece, portanto, como um julgamento intuitivo e não quantificado, apesar desses esforços em tornar a avaliação dos fatores mais sistematizada e reproduzível.

Poderíamos nos perguntar ainda se a determinação quantitativa do grau de confirmação de uma hipótese é possível na medicina. Do nosso ponto de vista, a complexidade dos fatores considerados quando se avalia a força evidencial a favor de uma hipótese médica torna uma determinação exata extremamente difícil, ainda que, em princípio, possível. De qualquer forma, essa dificuldade não impede que tentativas de solução do problema, ainda que não sejam completamente adequadas, tenham implicações positivas na teoria e na prática médica.

2.5 Conclusão

Neste capítulo, propusemos um esclarecimento inicial da noção de evidência correntemente utilizada na medicina, de uma maneira geral, e pela EBM, em particular. Para isso, precisamos discutir inicialmente os conceitos e definições de evidência já trabalhados na literatura filosófica, a fim de poder aplicá-los posteriormente ao caso da medicina. Adotamos uma posição pluralista em relação ao conceito filosófico de evidência científica e defendemos a hipótese de que dois tipos diferentes de conceitos de evidência são adequados para a compreensão de grupos distintos de atividades médicas. Deve ter ficado claro que a nossa intenção não era desenvolver uma teoria normativa de evidência, mas de descrever como

esses conceitos são de fato utilizados na pesquisa e na prática médica. No entanto, a utilização de conceitos de evidência científica para a descrição da evidência médica pressupõe que a medicina possa ser bem caracterizada como uma ciência, mas esse ponto está longe de ser resolvido na filosofia da ciência e no próprio pensamento médico. No próximo capítulo, discutiremos, entre outras questões, o estatuto epistemológico da medicina.

Capítulo 3 - Ciência e progresso na medicina contemporânea

3.1 O estatuto epistemológico da medicina

No capítulo 1, discutimos a adequação da descrição da EBM como um “novo paradigma” médico e defendemos uma resposta negativa a essa questão.²⁴ Naquela ocasião, evitamos deliberadamente tocar em um ponto que obviamente se impunha. Deveríamos ter nos perguntado se, independentemente da EBM, a noção kuhniana de paradigma científico é adequada à compreensão da medicina. No entanto, como essa questão de certo modo relaciona-se com o estatuto epistemológico da medicina, consideramos mais pertinente reservá-la para o momento em que desenvolvêssemos a nossa posição acerca da cientificidade médica. Apesar de Kuhn não ter desenvolvido a sua teoria visando um critério de demarcação, o compartilhamento de um paradigma por uma comunidade pode ser considerado uma característica tão proeminente das disciplinas científicas que, se negarmos a existência de um ou mais paradigmas na medicina, apontaremos de certa forma para uma negação de seu estatuto científico. Inversamente, se conseguirmos demonstrar a vigência de paradigmas na pesquisa e na prática médica contemporânea, daremos um primeiro passo no sentido da defesa de sua cientificidade. Esta seção será dedicada ao desenvolvimento dessas questões.

É preciso observar, antes de iniciarmos as discussões, que não apresentaremos neste trabalho um critério de demarcação formal para esclarecer o estatuto epistemológico da medicina. Antes, partiremos da noção de paradigma para realizar uma caracterização sociológica de algumas atividades médicas que apresentam características suficientemente fortes para que sejam consideradas científicas. A tentativa de fazer qualquer julgamento acerca do estatuto epistemológico da medicina através de um critério formal de demarcação exigiria a formalização prévia da medicina contemporânea, ou ao menos de alguma de suas áreas de conhecimento, mas esse trabalho, se possível, parece-nos ainda longe de ser realizado. Feitas essas considerações iniciais, podemos passar à abordagem do nosso problema.

O modelo de desenvolvimento científico proposto por Kuhn baseia-se sobretudo nas ciências da matéria bruta, como a física e a química, ainda que em algumas passagens o autor dê indicações de que considera a sua aplicação

²⁴ Essas discussões encontram-se na seção 1.6.

adequada às ciências biológicas e, eventualmente, a alguns ramos das ciências sociais. No que concerne à medicina, Kuhn nunca aborda diretamente a questão de seu estatuto epistemológico, mas se refere a ela apenas de passagem. Na citação apresentada na p. 37, por exemplo, podemos apreender que a razão da organização dos médicos em comunidades de especialistas foi “uma necessidade social externa”, e não a posse de um paradigma em comum. Algumas parágrafos antes da passagem citada, Kuhn (1970a [1962], p. 15) havia mencionado a medicina entre outros “ofícios” (*crafts*), como a metalurgia e a fabricação de calendários. Em outro texto (Kuhn, 1970b), a identificação da medicina com um ofício reaparece, justamente no contexto de uma discussão sobre o uso da noção de paradigma como um critério de demarcação. Após defender a solução de problemas (*puzzle-solving*) específicos como a característica demarcadora de uma disciplina científica, Kuhn afirma que a astrologia não possuía tais problemas e que, por isso,

(...) a astrologia nunca foi uma ciência. Em vez disso, era um ofício, uma das artes práticas, com estreita semelhança à engenharia, à meteorologia e à medicina, como essas artes foram praticadas até pouco mais do que um século atrás. Os paralelos com uma medicina mais antiga e com a psicanálise contemporânea são (...) particularmente próximos (Kuhn, 1970a, p. 8).

Essa última passagem, no entanto, contém qualificações temporais que nos parecem especialmente importantes. A caracterização da medicina como um ofício seria válida apenas para a medicina como praticada até meados do século XIX? O que teria ocorrido nesse período que justificaria uma revisão do seu estatuto epistemológico? A medicina do final do século XIX e início do século XX deixou de ser um ofício para tornar-se o que? Uma ciência? Ainda que o próprio Kuhn não tenha respondido a essas perguntas, vejamos como outros autores abordaram a questão, seja fazendo referência direta à noção de paradigma, seja baseando-se em outros pressupostos teóricos.

Karl Rothschild (1977), avaliando a aplicabilidade do modelo kuhniano à medicina clínica, observa que uma importante mudança na compreensão da doença ocorre desde meados do século XVII. Muito esquematicamente, pode-se dizer que, nesse período, uma concepção estrutural da doença começa a substituir uma concepção humoral, que havia dominado o pensamento médico ocidental desde a Antiguidade clássica. Da perspectiva da patologia humoral, as doenças eram o resultado de um desequilíbrio dos quatro humores constituintes do organismo, isto é, sangue, fleuma, bile e atrabile, e de suas respectivas qualidades. Enquanto o estado

de saúde caracterizava-se pela harmonia desses humores, a doença advinha da preponderância de um deles sobre os outros três. A terapêutica, consistentemente, visava à restituição do equilíbrio humoral, seja através de medicações que estimulassem a liberação do humor excessivo através dos órgãos excretores, seja através da aplicação de substâncias e procedimentos com qualidades contrárias ao humor dominante. Assim, como uma das qualidades da atrabile era o frio, uma terapia possível para as doenças melancólicas seria a mudança para regiões mais quentes.

Por outro lado, a patologia mecânica, que Rothschuh (1977, p. 78) considera um segundo paradigma, explica tanto o funcionamento do organismo saudável, quanto os fenômenos mórbidos, recorrendo a partículas microscópicas presentes nos fluidos corporais, sobretudo no sangue. Essas partículas, de formas, movimentos e tamanhos diversos, atravessam os poros dos vasos sanguíneos e são depositadas para formar os órgãos normais e produzir fenômenos como a contração muscular. No entanto, distúrbios no seu movimento através desses mesmos vasos poderiam causar inúmeras alterações mórbidas, como edemas, hemorragias, câimbras e dor. Uma terapia possível dentro desse quadro teórico, ainda que não discutida por Rothschuh, é a sangria, indicada para os casos em que se considerava que a doença advinha da estagnação da circulação sanguínea local.

Entre os diversos fatores que concorreram para a emergência dessa nova patologia, Rothschuh (1977, p. 80-1) aponta a descrição da circulação sanguínea por William Harvey (1578-1657) em 1628, a demonstração microscópica da existência de corpúsculos sanguíneos e o sucesso de dispositivos mecânicos, como lentes, engrenagens e autômatos, na solução de problemas da vida diária. Era um princípio da nova medicina que os fenômenos naturais e técnicos eram submetidos aos mesmos mecanismos.

Ao invés da fundação de uma área específica de pesquisa, com problemas e métodos próprios, a transição para uma patologia mecânica representa uma mudança nos pressupostos metafísicos da medicina, todavia ainda incapaz de organizar a profissão ao redor de um paradigma comum.²⁵ Uma parcela expressiva da literatura médica até o final do século XVIII permanece eminentemente

²⁵ Levando em consideração a coexistência de concepções diversas de doença, inclusive em meados do século XX, e a incapacidade da comunidade médica em se organizar ao redor de uma patologia única, Rothschuh (1977, p. 85) conclui que não se pode falar em revoluções em medicina, mas apenas em *inovações*.

especulativa, com pouco ou nenhum recurso à observação de pacientes ou experimentos como meios de aumentar o conhecimento médico. Programas de pesquisa em subdisciplinas médicas surgiram apenas ao longo do século XIX, influenciados por novos desenvolvimentos da concepção estrutural de doença, sobretudo através da emergência da anatomia patológica. No final do século XVIII e início do século XIX, a doença passa a ser identificada com uma lesão orgânica específica, que fornece ao mesmo tempo um critério seguro para o seu diagnóstico e uma explicação para os fenômenos mórbidos.²⁶ O trabalho do anátomo-patologista consiste em buscar a lesão que corresponde a um conjunto de sinais e sintomas bem determinados clinicamente, tomando como referência os êxitos de autores como Xavier Bichat (1771-1802) e René Lennec (1781-1826) no esclarecimento da lesão típica de algumas doenças específicas, como a tuberculose pulmonar. A partir da segunda metade do século XIX, os procedimentos e problemas de pesquisa da patologia tornam-se ainda mais específicos, com a mudança de escala no foco de investigação, que deixa as estruturas macroscópicas para dedicar-se a estruturas microscópicas, como os tecidos e as células.

Similarmente, a fisiologia é outra área que encontra focos de pesquisa cada vez mais bem delimitados ao longo do século XIX. A determinação dos constituintes normais dos líquidos orgânicos, como sangue e urina, a investigação da função de órgãos, como o fígado, o pâncreas e a tireóide, o mapeamento das estruturas nervosas e suas respectivas atividades, entre outros temas, compreendem boa parte da literatura médica científica da época. Particularmente importante é o recurso por parte de fisiologistas a procedimentos experimentais cada vez mais sofisticados, no intuito de responder às questões levantadas. Mas retornaremos à fisiologia e à experimentação na próxima seção, quando discutirmos a concepção bernardiana de progresso médico. De qualquer forma, essa especialização crescente ao longo do século XIX pode ser também observada em outras disciplinas biomédicas, como a microbiologia, a farmacologia e a bioquímica.

Georges Canguilhem (1988, p. 18), também tratando do estatuto epistemológico da medicina, aponta o período compreendido entre 1799 e 1837, como o período em que a medicina se torna uma “ciência positiva”. Diferentemente

²⁶ Foucault (2011 [1962]) apresenta uma discussão aprofundada acerca da emergência da anatomia patológica e da transição de uma patologia humorista para uma patologia estrutural no final do século XVIII e início do século XIX.

das discussões acima e de boa parte de seus próprios trabalhos anteriores, que focam nas atividades médicas de pesquisa laboratorial, Canguilhem sublinha nesse texto as contribuições da matematização estatística para a cientificidade médica. Inclusive, perguntando-se pelos autores que haviam contribuído para esse processo de positivação da medicina, indica o nome de Louis – cuja obra apresentamos rapidamente na seção 1.5 – como um de seus principais representantes. A aplicação do cálculo de probabilidades à medicina é interpretada por Canguilhem (1988) como uma forma de identificar os riscos e expectativas associados com uma intervenção e de avaliar objetivamente procedimentos terapêuticos. Além disso, os procedimentos estatísticos seriam capazes de trazer maior segurança às previsões prognósticas e ao diagnóstico etiológico das doenças. Independentemente de divergências na explicação desse processo, fica claro que há um certo consenso entre os historiadores e filósofos da medicina de que um importante avanço ocorreu na medicina da primeira metade do século XIX e que justifica atribuir-lhe um estatuto científico.

De nossa parte, carecendo de um estudo histórico mais aprofundado das mudanças ocorridas na medicina do século XIX, preferimos abordar o problema através da medicina contemporânea. Para isso, retomaremos a distinção de atividades médicas proposta na seção 2.3, pois nos parece que as discussões apresentadas acima são prejudicadas por uma confusão entre os diferentes níveis de atividades médicas. Rothschild (1977, p. 74-3), por exemplo, afirma que a medicina não é propriamente uma disciplina científica, porque não busca o aumento do conhecimento, mas apenas a aplicação do conhecimento científico no intuito de realizar uma terapêutica eficaz. Obviamente, dentro da divisão que propomos, essa afirmação valeria apenas para a prática clínica, mas não para as pesquisas laboratorial e clínica. Do nosso ponto de vista, o fato das investigações nessas duas grandes áreas poder ser guiada por considerações de ordem prática, como o tratamento de doenças, não impede que tanto o processo de obtenção do conhecimento, quanto os seus resultados, sejam legitimamente qualificados como científicos.

Começamos pela pesquisa laboratorial biomédica. É possível afirmar que, atualmente, esse nível de atividade médica possui problemas, procedimentos, comunidades e aparatos institucionais científicos? Acreditamos que sim.

Recorramos mais uma vez às pesquisas acerca da hipertensão arterial para ilustrar esse ponto. A rápida leitura de um artigo (Lerman *et al.*, 2005) tratando dos modelos animais de hipertensão permite identificar alguns dos *pressupostos teóricos compartilhados* pela comunidade que estuda a fisiopatologia e o tratamento dessa doença: (i) a hipertensão é uma doença multifatorial, que envolve contribuições ambientais e genéticas; (ii) apesar disso, é possível isolar experimentalmente as principais vias implicadas na regulação da pressão arterial normal e envolvidas na hipertensão essencial; (iii) uma das consequências do aumento sustentado dos níveis pressóricos é o surgimento de lesões secundárias em órgãos alvos, como rins, coração e vasos sanguíneos; (iv) os modelos animais de hipertensão reproduzem, de maneira mais ou menos fidedigna, os mecanismos e os efeitos da hipertensão de humanos; (v) a intervenção específica sobre os mecanismos desregulados permite induzir a diminuição dos níveis pressóricos e evitar o surgimento de lesões secundárias; e (vi) o sucesso do tratamento da hipertensão em modelos animais permite esperar um sucesso no tratamento da hipertensão humana. O estabelecimento de cada um desses pontos certamente demandou tempo e intensa aplicação de recursos intelectuais e materiais, de modo que não ocorre a nenhum dos membros da comunidade a que o artigo se destina questionar ou demonstrar o seu fundamento. Ao contrário, a resolução das questões que essa comunidade se impõe exige a sua aceitação prévia.

Os êxitos progressivos de alguns dos membros dessa comunidade fornecem o modelo para o seu trabalho subsequente. Em 1934, por exemplo, Harry Goldblatt (1891-1977) e colaboradores, trabalhando com a *hipótese* de que a isquemia renal é responsável pelo aumento da pressão arterial em alguns tipos de doença renal, são bem sucedidos em produzir o primeiro modelo duradouro de hipertensão arterial em animais. Seu procedimento consistia, basicamente, em clampar as artérias renais de cães, de modo a produzir uma semi-obstrução ao fluxo de sangue que chega ao órgão, e posteriormente *medir* a pressão arterial através de um manômetro inserido na artéria carótida. O fluxo de sangue que saía do rim manipulado também era medido, no intuito de estabelecer uma *relação* entre a diminuição do fluxo sanguíneo renal e o aumento da pressão arterial sistêmica.

O estudo de Goldblatt pode ser interpretado como o paradigma da pesquisa laboratorial em hipertensão arterial. Boa parte do trabalho subsequente dessa

comunidade de pesquisadores consistiu em *reproduzir* os achados de Goldblatt em outras espécies animais, como porcos, coelhos, macacos e, principalmente, ratos. A tentativa de *explicação* dos mecanismos de aumento da pressão arterial após isquemia renal também atraiu a atenção de muitos pesquisadores, que precisaram identificar as substâncias vasopressoras liberadas pelo rim, além das alterações estruturais microscópicas do órgão. O sistema renina-angiotensina, de que tratamos brevemente na seção 2.3, recebeu grande esclarecimento desse empreendimento, que finalmente culminou no desenvolvimento de uma nova classe de anti-hipertensivos, os inibidores da enzima conversora de angiotensina (iECA).

Como uma confirmação dos mecanismos propostos, o aumento da pressão arterial de origem renal foi obtido através de inúmeras *variações experimentais*, como a retirada cirúrgica de parte dos rins, a embolização das artérias renais, a infusão intravenosa de componentes do sistema renina-angiotensina e, mais recentemente, a manipulação genética das cobaias, induzindo a super expressão dos genes responsáveis pela codificação das enzimas e precursores ativos nesse sistema (Lerman, 2005). Todo esse trabalho da comunidade relevante contribuiu para aperfeiçoar significativamente os modelos animais de hipertensão arterial, capazes de *mimetizar* de forma cada vez mais fidedigna a fisiopatologia da hipertensão humana. Obviamente, inúmeros problemas e inadequações existem, mas são justamente eles que mantêm a atividade de pesquisa nessa área.

Compartilhamento de pressupostos teóricos por uma comunidade de especialistas, teste de hipóteses experimentais, manipulação laboratorial do objeto de estudo, reprodução artificial de fenômenos naturais, recurso à medição de variáveis fisiológicas, determinação numérica da relação entre duas ou mais variáveis, replicação de resultados através de variações do procedimento experimental e tentativa de explicação dos fenômenos experimentais. Todas essas características nos parecem ser suficientemente fortes para classificarmos o tipo de atividade descrita acima como uma atividade científica. Seguramente, essas considerações acerca da pesquisa em modelos animais de hipertensão se aplicam a uma grande parcela do que viemos chamando de pesquisa laboratorial biomédica.

Pode-se dizer o mesmo em relação à pesquisa clínica? Em termos gerais, acreditamos que sim, ainda que existam diferenças fundamentais entre as atividades laboratoriais e clínicas, que fazem contribuições distintas à cientificidade médica.

Utilizemos mais uma vez um exemplo concreto para ilustrar a discussão. Os pesquisadores que se dedicam ao estudo clínico das doenças cardiovasculares também compartilham de pressupostos teóricos que podem ser facilmente apontados através da leitura de um de seus artigos (Kannel; McGee; Gordon, 1976): (i) as doenças cardiovasculares são entidades clínicas relacionadas com algumas características biológicas e comportamentais;²⁷ (ii) é possível identificar, através de estudos populacionais comparando grupos de indivíduos doentes e saudáveis, as variáveis mais significativamente associadas com essas doenças; (iii) as informações obtidas nos estudos populacionais permitem estimar o risco cardiovascular de um indivíduo, isto é, a sua chance de desenvolver um evento cardiovascular ao longo do tempo; (iv) a intervenção sobre os fatores de risco identificados, através de mudanças no estilo de vida e de medicamentos, é, ao menos em hipótese, uma forma de diminuir o risco cardiovascular de um indivíduo; (v) essa hipótese é passível de teste, novamente através da comparação de dois grupos de indivíduos expostos aos mesmos fatores de riscos, um dos quais será submetido às intervenções investigadas, enquanto o outro servirá como controle.

O estudo de Framingham pode ser considerado o paradigma dos estudos epidemiológicos acerca da história natural das doenças cardiovasculares e de seus fatores de risco.²⁸ Iniciado em 1948, esse estudo acompanhou em sua primeira fase 5209 moradores da cidade de Framingham, no estado de Massachussets, ao longo dos anos. Trata-se de um estudo de tipo coorte, caracterizado por admitir indivíduos que ainda não apresentavam o evento de interesse, isto é, indivíduos que ainda não haviam desenvolvido doença cardiovascular. As características desses indivíduos, como idade, sexo, peso, colesterol sérico, pressão arterial, tabagismo, eram avaliadas – inclusive recorrendo a instrumentos de *medição* – no início do estudo e subsequentemente a cada 2 anos, até que desenvolvessem um dos eventos de interesse, como infarto agudo do miocárdio, acidente vascular cerebral e morte. Utilizando algumas variáveis específicas, os pesquisadores de Framingham

²⁷ Doença cardiovascular foi definida de maneiras distintas por diferentes autores. Segundo Kannel, McGee e Gordon (1976, p. 46), trata-se de um grupo de doenças, que compreende doença arterial coronariana, insuficiência cardíaca congestiva, doença cerebrovascular e claudicação intermitente.

²⁸ Não pretendemos fazer uma análise aprofundada do estudo de Framingham, que já foi extensamente debatido, tanto na literatura técnica médica, quanto por autores com interesses metodológicos e filosóficos. Para uma breve história do estudo e de sua influência na pesquisa em cardiologia, cf. Mahmood et al. (2014). Para uma discussão filosófica do impacto do conceito de fatores de risco nas próprias noções de doença e normalidade, cf. Giroux (2013).

propuseram uma equação capaz de fornecer a *previsão* de ocorrência de um evento cardiovascular para um indivíduo em um prazo de 8 anos. Ou seja, a equação de Framingham oferece uma *estimativa* do risco cardiovascular individual, compreendido como a probabilidade condicional de ocorrência de um evento cardiovascular, dada a ocorrência prévia de outros eventos específicos, isto é, a presença e a intensidade de um ou mais fatores de risco.²⁹

O trabalho em risco cardiovascular subsequente ao estudo de Framingham consistiu em *replicar* os resultados do estudo e aplicá-los a outras populações, em avaliar a contribuição relativa de outras variáveis para o risco cardiovascular e em aprimorar a equação para o seu cálculo, no sentido de aumentar a *precisão* da estimativa. Além disso, outra parcela importante de estudos avaliou a eficácia do controle dos fatores de risco na prevenção de eventos cardiovasculares, seja através de exercícios físicos e mudanças alimentares, ou de medicações anti-hipertensivas e hipolipemiantes. Como a hipertensão arterial se mostrou um dos fatores de risco mais fortemente relacionados com as doenças cardiovasculares, inúmeros estudos buscaram *testar experimentalmente* a hipótese de que a diminuição dos níveis pressóricos diminuiria a mortalidade por doenças cardiovasculares. Especialmente importante é a evidência gerada por RCTs, que estabeleceram de forma consistente que a diminuição dos níveis pressóricos poderia levar a uma diminuição da mortalidade por eventos cardiovasculares em pacientes hipertensos. Permanecem abertas questões relativas ao nível de pressão arterial desejável em pacientes com alto risco, à adequabilidade dos resultados dos estudos clássicos para populações específicas, como idosos, negros e diabéticos, e ao melhor modelo para a estimativa do risco cardiovascular.

Esperamos ter deixado claro com esse exemplo que a comunidade de pesquisadores envolvidos com a pesquisa clínica em doenças cardiovasculares possui problemas muito específicos, cuja resolução é a razão de sua coesão como um grupo de cientistas. O mesmo pode ser dito de outros grupos que trabalham em pesquisa clínica e que se dedicam à resolução de questões relativas a doenças como a asma, o Alzheimer, e os diversos tipos de câncer. São características claramente científicas de suas atividades o recurso à medição, a determinação numérica das relações entre variáveis populacionais, o esforço de replicação de

²⁹ os fatores de risco

resultados, a realização de previsões cada vez mais precisas acerca de eventos futuros e o teste de hipóteses experimentais sob condições controladas. Mas é preciso notar que, diferentemente da pesquisa laboratorial biomédica, a pesquisa clínica não pretende, nem poderia por si mesma, propor uma explicação para os fenômenos que estuda. Mas retornaremos a essa questão, que julgamos extremamente importante, nas seções que se seguem.

Antes, devemos analisar as características da prática clínica. Utilizemos desta vez um relato de caso (Stoller *et al.*, 1988) para ilustrá-la. Uma paciente de 59 anos, com queixa de tosse de 10 meses de duração, é encaminhada a um serviço de pneumologia para avaliação. Ela já havia sido submetida a radiografia de tórax e espirometria, exames que normalmente se indicam como propedêutica de tosse, mas ambos apresentavam resultados normais. Ainda assim, os médicos assistentes propuseram tratamento de teste com broncodilatadores, supondo que a paciente tivesse asma. Após mais 10 meses de utilização dessas substâncias, ela ainda tossia, e os exames complementares permaneciam normais. Os médicos então atentaram para a possibilidade de que sua tosse fosse um efeito colateral de um novo medicamento, o captopril, que a paciente vinha utilizando para tratamento de hipertensão. Uma semana após a suspensão da medicação, a tosse havia desaparecido.

Ainda que exista embasamento teórico e alguma sustentação científica para a abordagem da paciente acima, a decisão dos médicos em realizar exames complementares e iniciar um tratamento para uma doença específica, como a asma, supondo seu diagnóstico adequado ao caso particular, é uma decisão individualizada, que depende basicamente da tradição de cuidado vigente na época e no lugar e da experiência pessoal do médico. Diante de uma queixa, como “tosse”, o médico vislumbra uma série de diagnósticos possíveis e talvez os organize de acordo com sua maior ou menor frequência. Se as informações do exame clínico não forem suficientemente fortes para firmar um diagnóstico, o médico pode solicitar os exames cujos resultados podem auxiliá-lo em sua decisão. Isto é, caso a paciente apresente à radiografia uma escavação pulmonar ou uma massa mediastinal, pode-se presumir que a causa de sua tosse é tuberculose, ou câncer de pulmão, e seguir com novos exames e tratamentos indicados. Caso os exames não auxiliem no diagnóstico, pode-se indicar um tratamento de teste, pois a queixa do paciente exige

uma resposta, mesmo que eventualmente não haja grande suporte científico para a opção terapêutica disponível. A retirada do captopril no caso relatado acima é um exemplo em que esse tipo de abordagem foi bem sucedida, mesmo que ainda não houvesse uma explicação razoável para a associação entre o medicamento e a tosse. De fato, a observação clínica se torna fonte de uma informação valiosa que havia passado despercebida para os pesquisadores envolvidos com o desenvolvimento e teste laboratorial dessa substância, criando um novo problema para a sua pesquisa, isto é, o esclarecimento do mecanismo através do qual o captopril gera tosse. Além disso, com a multiplicação de relatos similares, a utilização de iECAs passa a constar no diagnóstico diferencial de tosse crônica.

Pode parecer que a comunidade de médicos especialistas que se reúnem em torno de uma disciplina, como cardiologia e pneumologia, constituam uma comunidade científica como os pesquisadores que trabalham na pesquisa biomédica e clínica. No entanto, diferentemente desses últimos, os médicos que cuidam de pacientes não pretendem ampliar o conhecimento médico – ainda que eventualmente possam fazê-lo –, nem resolver problemas científicos específicos. Ao contrário, pretendem resolver problemas práticos pontuais, instâncias particulares de problemas gerais, cuja resolução definitiva esperam das pesquisas biomédica e clínica. Mas, obviamente, a prática clínica não se limita a ofertar soluções bem estabelecidas cientificamente e muitas vezes se antecipa ao esclarecimento laboratorial e estatístico dos recursos terapêuticos que emprega, visando atender à demanda de cuidado por parte dos pacientes. Como os médicos na prática não podem esperar que haja demonstração rigorosa da eficácia de todos os tratamentos de que dispõem antes de indicá-los a um paciente, é compreensível que uma parcela substancial de suas condutas seja baseada em evidência considerada de baixa qualidade, como discutido na seção 1.5.

O grau de confirmação de uma hipótese na prática clínica, como saber se é “asma” o diagnóstico correto do paciente, deve ser sempre relativizado à crença ou ao conhecimento de um indivíduo, o médico, que leva em consideração os resultados da pesquisa científica e sua experiência pessoal com casos similares, muitas vezes de forma apenas parcialmente consciente. Além da probabilidade de uma hipótese, a tomada de decisões clínicas precisa pesar como as consequências

possíveis das ações podem ou não satisfazer os valores visados, como a cura, o aumento da expectativa de vida, ou o alívio dos sintomas do paciente.

Isso faz com que o modo de avaliação dos desfechos na prática clínica também seja próprio. Ainda que o médico possa recorrer a instrumentos de medição, como um esfigmomanômetro, para determinar se a pressão arterial de um paciente encontra-se no intervalo desejado, a avaliação do sucesso de um tratamento na prática clínica leva em consideração outros critérios, muitos dos quais são eminentemente qualitativos. Por exemplo, ainda que o captopril tenha sido eficaz em diminuir a pressão de um paciente hipertenso, talvez o médico não considere o tratamento bem sucedido, por essa medicação ter ocasionado um aumento perigoso nos seus níveis de potássio. Essa avaliação envolve, ainda e fundamentalmente, a percepção do paciente em relação aos efeitos da medicação administrada, ao alívio dos sintomas de que se queixava, à gravidade dos efeitos colaterais esperados, ao impacto do tratamento em sua vida diária. Como se trata da utilização prática de uma intervenção, não é suficiente que a sua eficácia tenha sido rigorosamente demonstrada, mas é necessário que o paciente seja capaz de aderir ao tratamento de forma adequada. Caso contrário, diremos que o tratamento foi ineficiente.

Com essas considerações, pretendemos indicar que a prática clínica não é adequadamente caracterizada como uma atividade científica, mas exige uma outra categoria epistemológica. Do nosso ponto de vista, a melhor forma de descrevê-la é como uma técnica ou uma tecnologia, desde que se compreenda por este termo um conjunto de técnicas cientificamente informadas, mas não a simples aplicação de um conhecimento científico prévio.

Como outras atividades técnicas, a prática clínica – ao menos no que concerne à sua ação terapêutica – consiste em uma série de procedimentos práticos visando a um fim específico. Muitas vezes, esse fim é um produto concreto. No caso da metalurgia, por exemplo, trata-se de produzir um utensílio de metal, ao passo que na poesia pretende-se produzir um poema. Analogamente, na prática clínica visa-se à restituição da saúde de um indivíduo doente ou, para falarmos como Canguilhem (2013 [1943], p. 8), a “restauração do normal”. Como em outras técnicas, o aprendizado da clínica se dá majoritariamente em campo, sob a orientação de um profissional mais experiente, através da atenção aos seus conselhos e da reprodução de suas condutas. Ainda que seja possível que esses profissionais

tenham uma teoria sistemática da sua atividade, ela não é estritamente necessária para que sejam bem sucedidos na prática. Além disso, o simples compartilhamento de princípios práticos não é capaz de tornar um estudante apto a exercer a profissão médica, que exige tempo para a consolidação da *experiência pessoal*. Ao longo de sua carreira, o médico se torna capaz de manejar um arsenal terapêutico mais ou menos extenso, através do conhecimento da posologia de medicamentos, de suas indicações específicas, de seus efeitos colaterais esperados. Ao indicar um tratamento para um novo paciente, o médico leva em consideração, além de seus conhecimentos teóricos, os sucessos e fracassos pregressos que obteve com seu emprego nos mais diversos tipos de casos.

Mais uma característica compartilhada entre a prática clínica e outras atividades técnicas é a sua irredutibilidade aos conhecimentos científicos de que possam fazer uso. Esse ponto foi particularmente bem trabalhado por Canguilhem, que também descrevia a medicina como uma técnica. Antes mesmo de sua famosa tese de medicina, de 1943, intitulada *Essai sur quelques problèmes concernant le normal et le pathologique*, Canguilhem já se preocupava com as relações entre técnica e ciência e já enxergava na medicina uma área privilegiada para o estudo dessa relação. Em um artigo de 1937 sobre a concepção de técnica de René Descartes (1596-1650), Canguilhem se posiciona contra o ponto de vista positivista, segundo o qual as técnicas seriam uma simples aplicação de um conhecimento científico prévio. Na leitura que Canguilhem faz de Descartes, a técnica se torna uma atividade de criação, que decorre das necessidades do ser vivo. Assim, as lentes teriam sido desenvolvidas, na tentativa de apurar a visão humana, antes de uma teoria óptica que oferecesse os passos para a sua construção, ou explicasse os seus fenômenos. Mas mais do que isso, a novidade técnica e, principalmente, as resistências encontradas ao seu livre exercício tornam-se uma fonte de problemas para a pesquisa científica, que deixa de ter a precedência histórica e lógica em relação à técnica:

A ciência procede da técnica não como se o verdadeiro fosse uma codificação do útil, um registro do sucesso, mas ao contrário na medida em que o embaraço técnico, o insucesso e o fracasso convidam o espírito a se interrogar sobre a natureza das resistências encontradas pela arte humana, a conceber o obstáculo como objeto independente dos desejos humanos e a buscar um conhecimento verdadeiro (Canguilhem, 1982 [1937], p 121).

Essa perspectiva é retomada e desenvolvida na tese de 1943. Segundo Canguilhem (2013 [1943]), a clínica, em sua atividade terapêutica, não pode ser

vista como a simples aplicação de um conhecimento científico prévio, oriundo das ciências biológicas básicas, como a fisiologia. Ao contrário, são essas ciências que decorrem da clínica, campo onde os fenômenos mórbidos se manifestam pela primeira vez e são trazidos à atenção do médico pelas queixas dos indivíduos doentes. É apenas após a constatação de uma disfunção específica que a sua função normal correspondente pode ser esclarecida pela análise experimental, o que faz com que o estudo dos fenômenos mórbidos preceda historicamente o estudo dos fenômenos ditos normais. Mas a clínica tampouco espera a conclusão dessa análise antes de propor uma terapêutica ao doente. As doenças, compreendidas como obstáculos ao livre exercício das capacidades humanas, das variações de comportamento e da amplitude de ação, motivam o emprego de recursos terapêuticos que visam restituir ao indivíduo sua capacidade de resposta e adaptação diante das flutuações do meio, mesmo que não haja demonstração experimental de sua eficácia ou mesmo compreensão do seu mecanismo de ação. Diante dessas características, Canguilhem chega a uma conclusão que se aproxima da nossa, ao menos no que tange ao aspecto terapêutico da prática clínica:

(...) a clínica não é uma ciência e não será jamais uma ciência, mesmo que venha a usar meios de eficácia sempre mais garantida cientificamente. A clínica não se separa da terapêutica, e a terapêutica é uma técnica de instauração ou restauração do normal, cujo fim – a satisfação subjetiva de saber que uma norma é instaurada – escapa à jurisdição do saber objetivo (Canguilhem, 2013 [1943], p. 200).

Entretanto, a descrição da prática clínica como uma técnica pode passar a falsa impressão de que não há nenhuma preocupação por parte dos clínicos em fundamentar as suas condutas em princípios e conhecimentos científicos sólidos. Como vimos ao longo do capítulo 1, a própria EBM é uma tentativa de sistematizar princípios norteadores da prática clínica, através do recurso constante à literatura médico-científica e da exigência de que as condutas tenham a sua eficácia demonstrada por estudos clínicos de boa qualidade metodológica. Certamente, a tentativa de basear a clínica em princípios teóricos e resultados científicos não é uma iniciativa da EBM, já que é possível argumentar que, em maior ou menor grau de acordo com o período histórico, os médicos em campo sempre estiveram atentos aos desenvolvimentos científicos de sua época. A prática das sangrias no início do século XIX, por exemplo, era baseada em especulações que enxergavam nas inflamações a causa de uma ampla gama de doenças e que esperavam que a remoção do sangue excessivo pudesse trazer alívio aos órgãos inflamados. Mesmo

que a sua eficácia tenha sido posteriormente posta em dúvida, a prática das sangrias era completamente plausível dentro do quadro teórico dominante na época e condizente com os achados de disciplinas médicas de base, especialmente a anatomia patológica.

O fato de muitas condutas médicas serem empregadas antes de terem sido esclarecidas experimentalmente não impede que resultados obtidos nas pesquisas biomédica e clínica sejam aplicadas na prática. De fato, muitos desses resultados levam à introdução de novas terapêuticas no arsenal médico – como foi o caso do captopril –, ou mesmo à interrupção de determinadas práticas – como foi o caso da sangria, que se provou mais maléfica do que benéfica no tratamento da maioria das doenças. A pesquisa científica, tanto laboratorial, quanto clínica, retifica, informa, enriquece e apura a prática. É nesse sentido que podemos dizer que a prática clínica é uma *tecnologia*, entendendo por esse termo uma técnica cientificamente esclarecida. É também nesse sentido que compreendemos as ideias de outros autores acerca do estatuto epistemológico da medicina, como as de Rothschild (1977, p. 86), que a considerava uma "ciência operacional" (*operationale Wissenschaft*), ou mesmo as de Canguilhem (1988, p. 29), que a redefiniu em um texto do final de sua carreira como uma "soma evolutiva de ciências aplicadas".³⁰ As descrições desses autores são, portanto, adequadas, mas apenas parcialmente, pois se aplicam exclusivamente a um dos aspectos do complexo de atividades que constitui a medicina. Do nosso ponto de vista, nenhuma classificação pode abranger e expressar, a uma só vez, todas essas facetas.

De fato, há ainda uma outra atividade fundamental da prática clínica, que não é adequadamente apreendida pelas classificações que apresentamos acima. Independentemente de sua capacidade de intervenção sobre as doenças, a clínica pode ser interpretada como a observação e classificação sistemática dos fenômenos mórbidos e, portanto, como uma *taxonomia*. Qualquer tentativa de modificação da doença pressupõe a identificação da entidade mórbida em questão, mesmo que esta se resuma a apenas um sintoma. É por isso que a descrição do que chamaríamos hoje de *quadro clínico* apresenta desde a medicina hipocrática um lugar de

³⁰ Com a expressão "soma de ciências aplicadas", Canguilhem (1988, p. 25-9) pretende indicar a utilização e a unificação pela medicina de informações produzidas por ciências diversas, como a física e a química, que possuem um programa próprio de obtenção do conhecimento. Por sua vez, a qualificação "evolutiva" designa a sua capacidade de integrar conhecimentos novos, à medida em que vão sendo produzidos por essas outras ciências, ao seu projeto de cuidado de indivíduos doentes.

proeminência na prática médica, pois é através dela que as diversas doenças são identificadas, permitindo firmar o diagnóstico e especular o prognóstico nos casos individuais. Os desenvolvimentos posteriores da taxonomia clínica – ou da *nosologia*, se se quiser utilizar um termo mais frequente na literatura médica – consistiram em complexificar a organização sistemática das doenças, levando em consideração não apenas sintomas isolados, mas agrupamentos de sintomas e seu curso ao longo do tempo.³¹ Essa abordagem é particularmente apropriada quando se percebe que, na maior parte dos casos, não há um sinal ou sintoma individual que seja capaz de firmar o diagnóstico da doença, mas que, pelo contrário, um mesmo sintoma pode se manifestar em doenças muito distintas. A tosse, por exemplo, pode ser indicativa de uma doença alérgica, como a asma, de uma doença infecciosa, como a pneumonia, ou de uma neoplasia, como o câncer de pulmão. Para firmar o diagnóstico, o clínico levará em consideração os outros sinais e sintomas que se manifestam simultânea ou adjacientemente a ela.

Assim como outros sistemas de classificação, a taxonomia médica organiza as doenças em grupos progressivamente mais abrangentes de doenças. Isso se torna especialmente claro quando recorremos a um sistema de classificação de doenças muito utilizado na prática clínica atual, a *International Classification of Diseases* (ICD), da Organização Mundial de Saúde. Somos ali informados de que a hipertensão essencial pertence ao grupo das doenças hipertensivas, que também abrange a doença cardíaca hipertensiva e a hipertensão secundária. Observamos também que as doenças hipertensivas fazem parte do grupo mais amplo de doenças do sistema circulatório, ao lado de outros grupos de doenças, como a insuficiência cardíaca e as arritmias cardíacas. O ICD fornece ainda *critérios* para o diagnóstico da hipertensão arterial essencial e para a sua diferenciação de outras doenças. Lemos, por exemplo, que: “Hipertensão essencial (primária), que corresponde a 95% de todos os casos de hipertensão, é definida como pressão alta para a qual uma causa secundária não foi encontrada” e que esse diagnóstico exclui as doenças cerebrovasculares (World Health Organization, 2018, BA00). Como discutiremos melhor na seção 3.3, as pesquisas biomédica e clínica dependem da aceitação da validade dessas categorias diagnósticas.

³¹ Poderíamos indicar esses desenvolvimentos como a transição de uma nosologia *sintomática* para uma nosologia *sindrômica*.

Com isso, não queremos sugerir que seja possível realizar uma distinção rígida entre os componentes puramente observacionais da prática clínica e seus componentes terapêuticos. Obviamente, considerações de ordem terapêutica também podem ser apreendidas nos sistemas de classificação das doenças, o que não é de forma alguma surpreendente, já que o clínico não busca o conhecimento das doenças por si mesmo, mas sim visando restituir a saúde a um indivíduo doente.³² Essa característica dos sistemas de nosologia foi sublinhada por Alvan Feinstein (1925-2001), um autor que se dedicou de forma particularmente profunda à compreensão da taxonomia médica:

Um clínico faz mais do que identificar; ele intervém. Sua taxonomia médica precisa, portanto, ser adequada não somente para o diagnóstico da doença, mas para identificar as pessoas e os fenômenos de manejo clínico encontrados no tratamento profilático, no tratamento remedial e no cuidado dos pacientes (Feinstein, 1972, p. 126).

Um exemplo atual de como a resposta a um medicamento pode ser levada em consideração na definição de uma doença é a própria asma, já que um dos seus critérios diagnósticos é justamente o aumento do fluxo expiratório após utilização de medicações broncodilatadoras (Global Initiative for Asthma, 2018, p. 6). Assim, no relato de caso que apresentamos acima, se a paciente com tosse tivesse evidência de resposta a essas medicações, os médicos provavelmente firmariam o diagnóstico de asma.

De nosso ponto de vista, a análise dos sistemas atuais de nosologia é especialmente difícil, pois há neles um complexo cruzamento de classificações baseadas em variáveis morfológicas, funcionais e etiológicas. O esclarecimento da importância relativa de cada uma dessas variáveis na nosologia contemporânea, dos padrões de semelhança e diferença utilizados para delimitar entidades clínicas e das contribuições dos resultados das pesquisas laboratorial e clínica para a alteração das entidades existentes e para a criação de novas entidades são questões que

³² De fato, o diagnóstico de certo modo possibilita a *especificidade* do tratamento. Assim, as febres, como eram conhecidas no final do século XVIII o que hoje chamaríamos de doenças infecciosas, deviam ser tratadas com a casca de quina (*Cinchona spp.*); as inflamações, como a peripneumonia de Broussais, demandavam a sangria; e a pneumonia bacteriana, por sua vez, exigia um antimicrobiano.

consideramos abertas e dignas de investigações adicionais.³³ No entanto, uma discussão aprofundada desses problemas extrapola os limites desta dissertação.

Como conclusão a esta seção, gostaríamos de observar que nossas discussões acerca do estatuto científico das pesquisas biomédica e clínica, por um lado, e eminentemente técnico da prática clínica, por outro, parecem-nos refletir as conclusões a que havíamos chegado, na seção 2.3, acerca do tipo de evidência empregado em cada um desses níveis de atividade. Parece-nos esperado que uma atividade científica requeira uma noção objetiva de evidência, isto é, uma interpretação da evidência como uma relação factual entre um evento e uma hipótese, ou como uma relação lógica entre as proposições que os expressam. Essa relação seria independente das crenças e opiniões pessoais e, por isso, objetiva. Ao contrário, o emprego de um conjunto de técnicas visando um fim específico, como a restituição da saúde a um indivíduo doente, típico da prática clínica, exige uma interpretação subjetiva da evidência, já que esta deve ser sempre relativizada ao conhecimento ou à crença de um outro indivíduo, o médico. É nesse nível de atividades que os recursos da teoria da decisão encontram a sua aplicação mais frutífera, já que pretendem informar as decisões práticas com base na evidência disponível e nas expectativas envolvidas com cada uma das consequências possíveis das ações.

Mas se a prática clínica é uma técnica, e se a técnica é uma atividade criativa de resposta a necessidades vitais, devemos esperar que o seu avanço dependa da criatividade dos médicos em campo e das contingências do meio? Mas se também é verdade que a prática pode informar-se e enriquecer-se com os resultados das pesquisas laboratorial e clínica, não poderíamos esperar que avanços nessas áreas sejam posteriormente transmitidos à prática? Na próxima seção, retomaremos outra questão que deixamos aberta no capítulo 1 e discutiremos o progresso na medicina.

3.2 Duas noções de progresso médico

Na conclusão do capítulo 1, afirmamos que a EBM representa uma noção específica de progresso na medicina, que encontra na aplicação extensiva de métodos

³³ De fato, as contribuições das pesquisas em nível laboratorial e populacional tendem a ser cada vez maiores na nosologia contemporânea. Os próprios valores de pressão arterial considerados normais são constantemente sujeitos a revisões, levando em consideração estudos sobre o risco cardiovascular. Essa tendência permite-nos perguntar se a prática clínica deve manter a sua prioridade como campo de delimitação das entidades mórbidas, ou se transmitirá progressivamente a função de definição da saúde e da doença para os outros níveis de atividades médicas.

estatísticos a via mais segura para o avanço do conhecimento médico. Sugerimos também que essa não era a única concepção de progresso médico existente e que talvez não fosse a mais adequada. Nesta seção, apresentaremos as ideias de Claude Bernard acerca da medicina experimental e as utilizaremos como ponto de partida para tecermos as nossas críticas em relação à estatística médica, em geral, e à EBM, em particular.

A obra de Claude Bernard apresenta um duplo interesse para o estudo da medicina. Por um lado, suas contribuições para a fisiologia do século XIX são amplamente reconhecidas como acréscimos importantes ao conhecimento das ciências biológicas e como fonte de conceitos fundamentais para várias áreas de pesquisa subsequente.³⁴ Entre os seus êxitos científicos mais importantes, podemos citar a descoberta da função glicogênica do fígado, a demonstração da contração ativa da musculatura lisa vascular e de seu papel na regulação da circulação, e o esclarecimento do mecanismo de ação de diversos fármacos e venenos, como o monóxido de carbono e o curare. Por outro lado, Bernard desenvolveu ao longo de sua vida importantes considerações acerca da metodologia de pesquisa em ciências biológicas e da sua função para a cientificidade médica. É a esse segundo aspecto de sua obra que nos dedicaremos nas próximas páginas, reconhecendo, no entanto, que uma compreensão completa de sua metodologia exigiria também levar em consideração seu trabalho como experimentador.

Em 1865, tendo já realizado boa parte das descobertas a que nos referimos acima, Bernard publica uma obra intitulada *Introduction à l'étude de la médecine expérimentale*, onde expõe sua perspectiva em relação ao método experimental e à sua aplicação à medicina. Segundo Bernard (1952 [1865], p.111), o estudo dos fenômenos vitais deve seguir os mesmos princípios que o estudo da matéria bruta, levando em consideração, no entanto, as particularidades dos primeiros. Entre esses princípios, Bernard destaca o que chama de *princípio do determinismo*, que afirma que os mesmos fenômenos devem necessariamente se seguir às mesmas condições iniciais, ou, inversamente, que qualquer diferença na manifestação de um fenômeno se deve a uma diferença nas condições materiais de sua produção. Afastando-se de uma tradição vitalista, para a qual os fenômenos vitais seriam

³⁴ Por exemplo, o conceito bernardiano de *secreção interna*, que designa a produção de uma substância fisiologicamente ativa por uma glândula e sua liberação na corrente sanguínea, serviu de ponto de partida para uma parcela importante da pesquisa em endocrinologia da segunda metade do século XIX.

submetidos a uma força própria dotada de espontaneidade, Bernard afirma reiteradamente a possibilidade de esclarecimento experimental das condições de manifestação dos fenômenos orgânicos, completamente submetidos aos processos físico-químicos.³⁵ Essas condições podem ser vistas como a causa próxima dos fenômenos, e a análise experimental seria capaz de revelar as suas leis, isto é, as relações constantes e necessárias entre as causas e seus efeitos.

Além disso, para Bernard, o princípio do determinismo não vale apenas para os fenômenos ditos normais, ou fisiológicos, mas também e igualmente para os fenômenos mórbidos. Esse ponto de vista marca mais um afastamento em relação ao vitalismo de sua época, que considerava a doença como a ausência absoluta de leis, ou como o caos orgânico completo. Mas mais do que simplesmente postular a submissão dos fenômenos patológicos a leis, Bernard afirma que elas são as mesmas leis que regem os fenômenos normais e que esses dois tipos de fenômenos se diferenciariam apenas pelas condições de sua manifestação:

(...) não há, na natureza, nada de perturbado, nem de anormal; tudo se passa segundo leis que são absolutas, isto é, sempre normais e determinadas. Os efeitos variam em razão das condições que os manifestam, mas as leis não variam. O estado fisiológico e o estado patológico são regidos pelas mesmas forças, e eles diferem tão somente pela pelas condições particulares nas quais a lei vital se manifesta (Bernard, 1952 [1865], p. 45).

O exemplo a que Bernard recorre ao longo de sua obra para ilustrar a concepção de identidade dos estados normais e patológicos é o diabetes. A presença de glicose no sangue de animais e humanos não é um sinal de doença, mas, ao contrário, representa uma das principais fontes de energia para a atividade dos diversos tecidos e órgãos. Nos casos de diabetes, não se trata do aparecimento de uma nova substância no organismo, mas simplesmente de um aumento da concentração sanguínea de açúcar acima dos níveis usuais. Esse aumento da glicemia é responsável pelas diversas manifestações clínicas da doença, sobretudo a presença de açúcar na urina, tida como sinal típico da doença em sua época. A partir de um determinado limiar de glicemia, os rins não são mais capazes de reter a glicose no sangue, resultando em seu extravasamento através do filtro renal. A perda de açúcar na urina e de grandes quantidades de líquidos arrastados

³⁵ Ao criticar o vitalismo, Bernard (1952 [1865], p. 102) refere-se diretamente ao nome de Georges Cuvier (1769-1832), que teria afirmado a impossibilidade de aplicação dos métodos das ciências da matéria bruta ao estudo dos seres organizados. Segundo essa tradição, a tentativa de análise experimental dos seres vivos desprezaria seu caráter organizado, em que cada fenômeno mobiliza o organismo como um todo e depende da concorrência de outras funções,

osmoticamente por ele explicam as outras manifestações características da doença, como emagrecimento acentuado e sede intensa. O diabetes, portanto, não institui uma ordem orgânica nova, mas apenas representa uma série de fenômenos que passam a se manifestar a partir do momento em que um constituinte normal do sangue encontra-se em níveis excessivamente elevados.

Como fica claro no exemplo acima, a afirmação de continuidade dos fenômenos normais e patológicos exige diferenciá-los quantitativamente. De fato, para Bernard a doença é caracterizada pela diminuição ou pela exacerbação de uma função orgânica normal. Esse aspecto do pensamento bernardiano foi discutido em profundidade por Canguilhem em sua tese de medicina, onde também são apontadas as insuficiências dessa concepção. Além de levantar alguns exemplos de doenças que não podem ser consideradas como simples excessos ou carências orgânicas, como a alcaptonúria, Canguilhem observa que essa concepção tende a confundir dois estados que são claramente distintos.³⁶ Para o médico em campo e principalmente para o paciente enfermo, os estados de saúde e doença são qualitativamente opostos, já que a doença é avaliada negativamente como um encolhimento do campo de possibilidades do indivíduo e como a privação de sua atividade plena. É justamente essa oposição *valorativa* entre saúde e doença que motiva a procura de ajuda por parte do paciente e a tentativa de intervenção por parte do médico.

De nossa parte, reconhecemos haver uma tensão entre a perspectiva clínica, centrada na experiência pessoal do paciente, e a perspectiva fisiológica, que tenta reduzir os fenômenos mórbidos a processos físico-químicos; mas uma tensão não é uma contradição. A tentativa de análise e explicação do fenômeno mórbido é feita em benefício do paciente, já que com elas pretende-se chegar a um tratamento eficaz para a sua doença, e essa é justamente uma das expectativas que o paciente nutre em relação à medicina. De Bernard, reteremos não a identidade quantitativa

³⁶ A alcaptonúria é uma doença rara, de herança autossômica recessiva, causada pelo defeito de metabolismo do ácido homogentísico, um derivado da tirosina. A deficiência da enzima responsável pela sua degradação leva ao acúmulo da substância nos tecidos e ao seu aparecimento na urina, que se torna caracteristicamente negra (Mistry, Bukhari, Taylor, 2013). Segundo Canguilhem (2013 [1943]), esse é um exemplo de doença em que as manifestações clínicas não podem ser interpretadas como consequência de um excesso ou carência de uma função normal, já que o ácido homogentísico não é um componente normal da urina. Não podemos deixar de expressar o nosso espanto diante do argumento de Canguilhem. O fato do ácido homogentísico não ser um constituinte normal da urina não nos parece capaz de infirmar a concepção bernardiana, pois as manifestações mórbidas da alcaptonúria podem ainda ser interpretadas como consequência do excesso de um constituinte normal do organismo.

dos estados normais e patológicos, que realmente não se sustenta após análise mais fina, mas sim a sua exigência de fundamentação da terapêutica no conhecimento prévio desses estados.

De fato, para Bernard, a continuidade entre os fenômenos fisiológicos e patológicos se estende também para os fenômenos terapêuticos. Da mesma forma como os estados patológicos podem ser compreendidos como variações de uma função normal, a ação dos medicamentos deve ser interpretada de acordo com a sua influência sobre essas funções:

Nós não saberíamos encontrar (...) nenhuma diferença radical entre a natureza dos fenômenos fisiológicos, patológicos e terapêuticos. Todos esses fenômenos derivam de leis que, sendo próprias à matéria vivente, são idênticos em sua essência e variam tão somente pelas condições diversas nas quais os fenômenos se manifestam. (...) a verdadeira base científica da terapêutica deve ser dada pelo conhecimento da ação fisiológica dos casos mórbidos, dos medicamentos ou dos venenos, o que é exatamente a mesma coisa (Bernard, 1952 [1865], p. 260-1).

Daí a prioridade da fisiologia no pensamento bernardiano: o esclarecimento das funções normais do organismo permite compreender as manifestações patológicas e a ação curativa dos medicamentos. A *medicina experimental* bernardiana compreende, portanto, a fisiologia, a patologia e a terapêutica e considera como a sua base racional a primeira dessas disciplinas. O avanço do conhecimento médico será proporcional à extensão e à profundidade com que a análise experimental tiver sido levada nessas disciplinas.

Bernard propõe um esquema do desenvolvimento das ciências e da medicina, dividido em etapas análogas aos três estados de Auguste Comte (1798-1857).³⁷ Esse esquema já está presente na *Introduction* de 1865, mas é mais claramente exposto em uma obra intitulada *Principes de médecine expérimentale*. Trata-se de uma obra inacabada, publicada postumamente em 1947, e que vinha sendo escrita ao longo de sua vida. A célebre *Introduction* deveria ser um preâmbulo a essa obra maior, onde os princípios da medicina experimental seriam de fato lançados. Ali, Bernard (1947, cap. 3) nos informa que o primeiro estágio da medicina é um estado *ante-científico*, teológico, ou heróico, caracterizado por uma concepção ontológica da doença como um ente maligno que vem alojar-se no corpo e no espírito humano. Coerentemente, o cuidado dos doentes era tarefa dos sacerdotes, que recebiam

³⁷ No entanto, diferentemente de Comte, Bernard não acredita que, ao atingir um estágio mais avançado, uma ciência positiva supere completamente os estágios anteriores. Pelo contrário, haveria coexistência dos diversos estágios em uma mesma ciência, já que em algumas de suas questões o determinismo foi encontrado, enquanto outras ainda tentam atingi-lo.

orientações curativas de outros entes divinos. Na história ocidental, esse estado corresponderia à medicina pré-hipocrática. A esse estágio, segue-se o segundo período da medicina ocidental, o *empirismo*, caracterizado pela observação não sistemática dos fenômenos mórbidos e dos efeitos de diversos medicamentos sobre o curso das doenças. É nesse período que uma grande quantidade de fatos são coletados, mas ainda não há uma tentativa de explicá-los racionalmente. Lembremos que empirismo, na história da medicina, designa uma escola ou doutrina médica que afirma que o conhecimento das causas das doenças e da ação dos medicamentos é desnecessário para a prática e que esta depende apenas do conhecimento dos efeitos dessas medicações sobre os diversos tipos mórbidos.

Pelo contrário, o terceiro e último estágio da medicina, dito *científico* ou *experimental*, caracteriza-se pela tentativa de esclarecimento da causa próxima dos fenômenos de interesse médico. Nesse estágio, o médico não se contenta em constatar os fatos, mas tece intuitivamente hipóteses acerca deles e deduz dessas hipóteses novos fatos passíveis de verificação experimental. É através desse procedimento que o experimentador se torna capaz (i) de *explicar* os fenômenos que estuda, (ii) de *prever* o seu curso natural e (iii) de *modificá-lo* de acordo com seus propósitos.³⁸ Apesar de Bernard não oferecer uma definição explícita do que entende por explicação científica, fica claro ao longo de seus textos que tem em mente algo como o esclarecimento do *mecanismo* de produção dos fenômenos. A noção de mecanismo, por sua vez, pressupõe a identificação dos elementos orgânicos básicos, da sua ação mútua e das etapas de sua articulação temporal. O mecanismo de produção de açúcar pelo fígado, por exemplo, inclui a ação de um fermento orgânico – ou uma enzima, como diríamos hoje – sobre uma substância armazenada nas células hepáticas, o glicogênio. O fermento orgânico realiza a quebra do glicogênio em seus elementos constituintes, as moléculas de glicose, que são então liberadas na corrente sanguínea através da grande veia que recebe o sangue do fígado, a veia porta. É justamente comparando o teor de açúcar no sangue que chega e sai do fígado que o experimentador tem um primeiro indício de sua produção hepática. Outros procedimentos experimentais devem então se suceder para verificar essa hipótese e aprofundar o entendimento do seu processo de produção.

³⁸ De fato, para Bernard, assim como para Comte, a capacidade de previsão e controle dos fenômenos naturais são características distintivas das ciências maduras.

Nessa perspectiva, a estatística médica não pode ser considerada uma ciência, porque não explica os fenômenos, não consegue prevêê-los e não pode modificá-los. Antes, pertence à etapa menos desenvolvida do empirismo:

De fato, a estatística (...) oferece, assim como o empirismo, apenas o *estado das coisas, status rerum*, ao passo que a ciência ou o racionalismo oferece a *razão das coisas, ratiocinatio rerum*, quer dizer, a explicação dos fatos com a relação racional que os associa a sua causa próxima. A estatística é, definitivamente, apenas um recenseamento [*dénombrement*] que fornece a expressão empírica dos fatos e ela não poderá jamais dar origem a uma ciência definitiva e precisa, nem a *uma ciência de observação que consiste em prever corretamente* em um dado caso, nem *uma ciência experimental, que consiste em prever e agir corretamente em um caso dado* (Bernard, 1947, p. 59-60; grifo do autor).

Trata-se de uma afirmação extremamente importante, que gostaríamos de analisar com cuidado. Começemos pela incapacidade da estatística em explicar os fenômenos.

Para Bernard (1952 [1865], p. 188-96), a estatística não fornece o determinismo dos fenômenos, isto é, não identifica as suas condições de manifestação e as leis que relacionam causa e efeito. Por exemplo, a média de glicose sanguínea de um paciente é incapaz de expressar as variações dessa variável ao longo do dia e as influências que sofre de condições orgânicas como a alimentação e o jejum. Além disso, não desvela os órgãos responsáveis por sua absorção, distribuição e consumo, justamente as informações que mais interessam ao fisiologista e ao médico, pois indicam o caminho para o esclarecimento de disfunções mórbidas correspondentes, como o diabetes. Similarmente, a simples relação da quantidade de saliva produzida por uma glândula salivar com o seu peso é enganadora, pois não expressa as flutuações rítmicas da secreção glandular, influenciadas por excitantes como a visão do alimento, ou inibidas durante períodos de relativa inatividade orgânica, como o sono. Ou seja, a estatística torna estáticos fenômenos que são essencialmente dinâmicos.

Consideramos a crítica justa. Em termos gerais, os resultados produzidos pela análise estatística permitem-nos afirmar que duas ou mais variáveis estão relacionadas e quantificar a chance dessa associação dever-se ao acaso. O valor-p, que discutimos rapidamente na seção 1.5, é uma medida dessa chance. No entanto, a estatística não informa sobre qual tipo de relação se trata, isto é, se uma variável é causa de outra, ou se são ambas produzidas por uma causa comum. Além disso, não elucida o mecanismo de articulação de duas variáveis, ou seja, não esclarece *como* ou *por que* duas ou mais variáveis se relacionam.

Para Bernard (1947, p. 62-3), a incapacidade em desvelar o determinismo dos fenômenos duplica-se na incapacidade em prever a sua marcha necessária. Caso uma mulher grávida pergunte a um médico qual a chance de seu filho ser homem, o médico pode, utilizando as últimas estatísticas de nascimentos na França, afirmar-lhe que essa chance é de $3/5$. No entanto, o médico não pode dizer-lhe se em seu caso particular o filho será homem, ou mulher, pois desconhece a lei que determina o gênero de uma criança. Colocando a questão em termos contemporâneos e muito simplistas, poderíamos dizer que a estatística de nascimentos não dá nenhuma informação acerca dos cromossomos e genes relacionados com a diferenciação sexual, cuja demonstração em um embrião possibilitaria uma previsão segura acerca dessas características.

No entanto, do nosso ponto de vista, o prognóstico é uma das funções médicas em que o aporte das informações estatísticas é de maior valor. Como as características de base relacionadas com a marcha de uma doença são potencialmente inumeráveis e como muitas vezes essas características são independentes da doença principal, não é possível esperar que o conhecimento de seus mecanismos fundamente uma previsão absoluta nos casos individuais. A observação da marcha da doença em um grande número de indivíduos e a sua estratificação segundo variáveis de reconhecida importância prognóstica nos parece a melhor forma de estabelecer um prognóstico mais ou menos preciso, que inclusive sirva de base à tomada de decisões clínicas. Isso não impede, porém, que um aprofundamento do conhecimento fisiopatológico da doença guie a busca de informações em nível populacional. Esse processo de esclarecimento mútuo entre a pesquisa clínica e a pesquisa biomédica é especialmente claro na oncologia. Normalmente, o tempo de sobrevivência de pacientes com câncer é obtido através de estudos epidemiológicos que dividem os paciente em estágios de gravidade comparável. No entanto, a presença ou ausência de marcadores tumorais específicos, que muitas vezes são moléculas envolvidas nas vias de replicação celular, como alguns receptores hormonais, são cada vez mais reconhecidas como determinantes da sobrevivência. Naturalmente, essas mesmas moléculas se tornam um alvo preferencial da ação terapêutica.

Isso nos leva ao último ponto da crítica bernardiana à estatística médica. Quando um médico afirma, utilizando informações estatísticas, que um tratamento é

capaz de curar uma doença em tal proporção de casos, o motivo pelo qual o tratamento é bem sucedido em uma parcela dos casos e mal sucedido em outra permanece indeterminado. Ao contrário, se conhece a causa próxima da doença e o mecanismo de ação da droga, pode afirmar que o tratamento terá êxito. Bernard (1952 [1865], p. 114; 1947, p. 73-4) utiliza recorrentemente o exemplo da sarna para ilustrar essa concepção. Enquanto se desconhecia o parasita causador dessa doença, os mais diversos tratamentos eram indicados empiricamente para combatê-la, e eram levantadas as estatísticas de sucesso. A partir do momento em que se reconhece o ácaro causador dessa afecção e encontram-se os meios de combatê-lo, o sucesso torna-se seguro: "Outrora, podíamos dizer (...) que ora curávamos a sarna, ora não a curávamos; mas hoje que nos dirigimos à causa determinada dessa doença nós a curamos *sempre*" (Bernard, 1952 [1865], p. 114; grifo do autor). Essa garantia de sucesso do tratamento decorreria do conhecimento do mecanismo da doença e da ação do fármaco, que agiria especificamente sobre as funções desreguladas. Portanto, a efetividade de um tratamento seria consequência de sua racionalidade, compreendida como a análise dos fenômenos mórbidos e terapêuticos através de princípios científicos, ou experimentais.

A diferença entre a perspectiva bernardiana e a dos proponentes da EBM não poderia ser maior. Lembremo-nos de que os defensores do movimento consideram a aplicação de métodos estatísticos a fonte de evidência mais forte para a tomada de decisões na prática clínica, ao passo que os resultados da pesquisa laboratorial seriam a fonte de evidência mais fraca. Coerentemente, o progresso da medicina dependeria da aplicação extensiva de estudos clínicos, sobretudo do RCT, e do apuro dos métodos de análise estatística. Bernard, pelo contrário, afirma que o progresso na medicina e a garantia de seu estatuto científico dependem fundamentalmente da fisiologia, ou do que chamamos neste trabalho de pesquisa laboratorial biomédica.

Mas resta saber o que queremos dizer quando nos referimos ao progresso na medicina. Consideremos que o propósito da pesquisa científica médica é o desenvolvimento de tratamentos que levem efetivamente à cura ou ao alívio dos sintomas dos pacientes. Podemos expressar esse propósito através do conceito de *eficácia*, que compreendemos como a capacidade objetivamente demonstrável e sistematicamente reproduzível de uma intervenção artificial alterar em um mesmo

sentido a marcha de um fenômeno natural. A *eficácia terapêutica*, por sua vez, seria a capacidade de uma intervenção terapêutica – medicamentosa, cirúrgica ou fisioterápica – alterar a história natural da doença, isto é, as etapas de desenvolvimento que normalmente ocorrem caso a doença permaneça sem tratamento. A demonstração objetiva da ação terapêutica sobre uma variável orgânica, como pressão arterial e glicemia, ou sintoma subjetivo, como dor, ocorre através da medição dessas variáveis ou do relato do paciente. Nesse último caso, a medida objetiva da influência do tratamento sobre a variável de interesse torna-se especialmente difícil, o que justifica a aplicação de ferramentas para avaliação de sintomas subjetivos, como escalas visuais de dor e questionários de qualidade de vida. Para que um tratamento seja considerado eficaz, é necessário que seus efeitos sejam reprodutíveis, isto é, que a sua utilização em indivíduos e momentos distintos levem aos mesmos efeitos; além disso, a intervenção não pode ora causar o aumento de uma variável, ora a sua diminuição. Por exemplo, para que um anti-hipertensivo seja considerado eficaz é preciso que seja consistentemente capaz de diminuir a pressão arterial do paciente e nunca elevá-la. Finalmente, a comparação com indivíduos em que o tratamento não foi aplicado possibilita atribuir-lhe as alterações observadas na marcha da doença.

Ainda que os métodos estatísticos, em geral, e o RCT, em particular, sejam capazes de demonstrar a eficácia de um tratamento, não são capazes de indicar um caminho para o desenvolvimento de novos tratamentos eficazes. Em uma situação hipotética em que não houvesse informação biológica acerca da fisiopatologia da doença e do mecanismo de ação dos fármacos, a pesquisa clínica, reduzida à aplicação de procedimentos estatísticos, estaria obrigada a testar aleatoriamente a eficácia de diversas substâncias para o tratamento das inúmeras doenças. Ainda que, eventualmente, algum tratamento se mostrasse eficaz através desse método e, nesse sentido, representasse um acréscimo real ao arsenal terapêutico vigente, tal método seria extremamente ineficiente. Ao contrário, a maior parte dos tratamentos de teste que chegam à pesquisa clínica possuem um fundamento biológico mais ou menos forte. Se interpretarmos o progresso médico como o *aumento da proporção de tratamentos eficazes*, compreendemos que esse progresso depende muito mais da continuidade das pesquisas em nível laboratorial biomédico do que em nível populacional.

De fato, parece-nos que, ao contrário do que defendem os proponentes da EBM, a estatística possua uma importância *terminal* para o desenvolvimento da medicina. Por um lado, o raciocínio biológico serve como referência para a interpretação dos resultados da pesquisa clínica, já que um resultado estatístico que não é biologicamente plausível pode sempre ser colocado em dúvida. Nesse sentido, a pesquisa biomédica é, de fato, a base da hierarquia de evidência na medicina, mas não como quer a EBM, como a fonte de evidência mais fraca. Ao contrário, deve ser tomada como a informação de fundo diante da qual os resultados da pesquisa clínica podem ser considerados fonte de evidência forte. Por outro lado, é a pesquisa biomédica que produz novas intervenções experimentais cuja chance de serem realmente eficazes é suficientemente alta para justificar a realização de grandes ensaios clínicos.

Com isso, não queremos sugerir que os estudos populacionais sejam inúteis e que a demonstração de eficácia em nível laboratorial seja suficiente para imediata aplicação de um novo tratamento na prática clínica. Reconhecemos plenamente que o sucesso em nível laboratorial, sobretudo em modelos animais, não garante a eficácia em nível clínico, ou populacional. É obviamente necessário que novas intervenções sejam testadas em seres humanos, mas apenas a pesquisa em ciências biológicas básicas, orientada para o esclarecimento de questões clínicas, pode *guiar* as pesquisas clínicas, através do desenvolvimento de novos recursos terapêuticos. Talvez devêssemos mesmo dizer *criação* de novos recursos terapêuticos, pois é de fato uma atividade criadora que testemunhamos nas ciências biomédicas contemporâneas. Como quisemos ilustrar com o caso do captopril, o esclarecimento dos mecanismos patológicos disfuncionais e, principalmente, o reconhecimento de etapas cruciais da fisiopatologia, possibilita a produção ativa de novas substâncias com potencial terapêutico, através, por exemplo, de técnicas de engenharia molecular. Diante de resultados promissores na pesquisa de base, cabe à pesquisa clínica, munida de procedimentos estatísticos, julgar se o novo tratamento experimental cumpre em nível populacional as promessas que havia feito em nível laboratorial. Todas essas considerações nos obrigam a atribuir à estatística médica uma importância muito mais *reguladora* do que *diretiva* para o desenvolvimento da medicina.

Para concluir esta seção, gostaríamos apenas de observar que a medicina é o exemplo de uma disciplina científica em que o ideal de progresso, cuja decadência já foi proclamada no meio filosófico, permanece plenamente vigente.³⁹ A expectativa por tratamentos eficazes – curativos ou sintomáticos – parece-nos legitimada pelos interesses de um ser que quer continuar vivo, ativo e funcionante. Mas mais do que isso, essa é também uma expectativa racional, porque, por um lado, a teoria nos permite vislumbrar a possibilidade de um tal progresso e, por outro, a história nos dá exemplos de casos em que ele foi efetivamente atingido.

Após toda essas discussões, será conveniente recapitular alguns pontos que foram desenvolvidos neste trabalho e que consideramos de especial importância.

3.3 As relações entre pesquisa laboratorial biomédica, pesquisa clínica e prática clínica

Nesta seção final, apresentaremos esquematicamente as relações entre os três níveis de atividades médicas que distinguimos na seção 2.3. Não introduziremos nenhum conteúdo propriamente novo, mas apenas reorganizaremos parte do que já foi dito, a fim de deixar claras as contribuições relativas de cada uma dessas atividades para a medicina contemporânea.

Começemos pela prática clínica. Como observado na seção 3.2, uma de suas funções é a identificação dos padrões mórbidos e a classificação sistemática das doenças, seguindo critérios de semelhança e diferença. Diversos sintomas são agrupados de acordo com a frequência com que aparecem paralela ou sucessivamente no curso da doença. A identificação de categorias diagnósticas, como “hipertensão arterial”, “diabetes”, “asma”, “pneumonia” etc., é o primeiro passo para a instituição dos procedimentos técnicos da prática clínica. Em um caso ainda mais simples, a categoria diagnóstica pode reduzir-se a um único sintoma, como “dor”, “febre”, “fadiga”. De qualquer forma, é somente a partir da identificação dessas categorias diagnósticas e da subsunção do caso particular de um paciente a uma delas que um prognóstico e um tratamento específico podem ser propostos. Esse tratamento pode ser empírico, isto é, referendado apenas pela tradição ou pela

³⁹ O próprio Canguilhem (1987) foi um autor que afirmou que idéia de progresso, típica do pensamento humanista do século XVIII, sofreu uma profunda decadência durante o século XIX, até a sua derrocada definitiva em meados do século XX, após a Segunda Guerra Mundial.

experiência pessoal do médico, ou ter tido a sua eficácia demonstrada experimentalmente.

Obviamente, a identificação das categorias diagnósticas também precede a tentativa de demonstração de eficácia de qualquer tratamento, seja através de procedimentos laboratoriais, seja através de estudos populacionais. De fato, tanto a pesquisa laboratorial biomédica, quanto a pesquisa clínica, dependem da aceitação prévia da validade das categorias diagnósticas observadas na prática clínica. Por exemplo, tanto os pesquisadores que trabalham com modelos animais de hipertensão, quanto aqueles que estudam risco cardiovascular, precisam aceitar que exista “hipertensão arterial” e que ela seja suficientemente caracterizada pela simples medida da pressão arterial. Normalmente, a pesquisa clínica recebe o influxo da pesquisa laboratorial biomédica, onde são desenvolvidos e inicialmente avaliados os novos tratamentos experimentais antes de serem testados em humanos. Caso os resultados das pesquisas laboratorial e clínica forneçam evidências a favor da hipótese de eficácia de um fármaco ou outro tipo de intervenção, esse tratamento passa a ser utilizado na prática clínica.

Isso não quer dizer que apenas tratamentos de eficácia demonstrada sejam utilizados na prática clínica. Na verdade, como também já indicado na seção 1.5, poderíamos afirmar a eficácia de uma parcela relativamente pequena das intervenções terapêuticas atualmente empregadas. Mas como, se as agências reguladoras, como o FDA, exigem que a eficácia de um tratamento seja demonstrada para que liberem a sua utilização? Para que uma medicação seja posta no mercado, é necessário demonstrar a sua segurança e eficácia para o tratamento de *uma* doença, mas a partir do momento em que essa medicação passa a ser comercializada, fica a cargo do médico empregá-la ou não em outras circunstâncias. É isto que efetivamente acontece em muitas situações clínicas: diante de uma doença para a qual não se encontrou um tratamento adequado, os médicos normalmente utilizam medicações cujo uso foi aprovado para outras doenças. Acreditamos que, a despeito de todo o progresso possível e do enriquecimento constante da prática pela experimentação, sempre existirá uma importante defasagem entre a terapêutica instituída e a sua validação científica. Além da complexidade das questões médicas, o seu objeto de estudo não é um ente estático esperando para ser conhecido. Pelo contrário, as doenças são entidades

dinâmicas, que mudam com o tempo. Novas doenças emergem, como a síndrome da imunodeficiência adquirida, e as próprias intervenções de saúde levam à modificação das entidades mórbidas, como no caso da resistência bacteriana. Uma possibilidade de demonstração experimental completa da eficácia terapêutica exigiria a fixidez da doença e das categorias diagnósticas, o que, claramente, já está descartado.

Existe, portanto, uma circularidade nas relações entre os três níveis de atividades médicas, pois a prática clínica recebe informações de volta das pesquisas laboratorial e clínica após ter-lhes emprestado as categorias diagnósticas que servem de base para o seu trabalho. Não se trata, no entanto, de uma circularidade viciosa, pois as pesquisas biomédica e clínica não visam demonstrar a validade das categorias diagnósticas cuja aceitação a sua própria atividade pressupõe, mas alcançar novos conhecimentos a partir delas. Fica claro que, ao contrário do que acredita a EBM, a estatística não pode ser tomada como o fundamento da prática clínica. O mesmo deve ser dito em relação às investigações médicas nas ciências biológicas básicas. Se há algum fundamento para a prática clínica, ele deve ser buscado na própria prática clínica. Mas infelizmente está acima das pretensões deste trabalho tentar resolver essa questão.

3.4 Conclusão

O que comumente chamamos de medicina compreende atividades bastante distintas, unificadas pelo propósito de manutenção e restituição da saúde de indivíduos humanos. Essa diversidade de atividades justifica o *estatuto epistemológico híbrido* da medicina, em parte ciência experimental, em parte técnica, em parte observação e classificação sistemática de fenômenos naturais. Os diferentes níveis de atividade também concorrem de formas distintas para a obtenção desse propósito: enquanto a pesquisa científica visa à demonstração da eficácia de intervenções terapêuticas, é a prática clínica que institui essas intervenções, tendo sido a sua eficácia cientificamente validada, ou não. Além disso, é na prática clínica que surgem as categorias diagnósticas, que podem ser interpretadas como pré-requisitos para a instituição das três funções médicas básicas, isto é, diagnóstico, prognóstico e terapêutica. Ainda que a posição de vanguarda da clínica em relação à pesquisa científica e a própria natureza dinâmica

dos fenômenos mórbidos impeçam uma demonstração completa da eficácia das terapias empregadas na prática, o progresso na medicina é possível, se for compreendido como o aumento da proporção das terapias com eficácia demonstrada. Acreditamos que esse progresso deva ser esperado muito mais da pesquisa laboratorial em ciências biológicas básicas, orientada para o esclarecimento de questões de interesse clínico, do que da pesquisa em nível populacional, amplamente baseada em procedimentos estatísticos. Nesse sentido, consideramos a concepção bernardiana de progresso na medicina mais profunda e adequada do que aquela defendida pelos proponentes da EBM.

CONCLUSÃO

A EBM, interpretada como um movimento de normatização metodológica da medicina contemporânea, levanta importantes questões acerca do conhecimento médico, da dinâmica de produção e teste de hipóteses experimentais e da melhor informação para a tomada de decisões na prática clínica. Certamente reconhecemos a relevância das contribuições do movimento para o debate médico científico, sobretudo no que concerne à sofisticação dos procedimentos de análise estatística, à avaliação da evidência médica e à investigação sistemática da literatura disponível. No entanto, consideramos as suas propostas claramente insuficientes para compreender a complexidade da medicina contemporânea e para indicar uma via única para o seu progresso futuro.

Entre as diversas críticas que fizemos às concepções do movimento, destacamos a redução das atividades médicas ao nível da pesquisa clínica, onde procedimentos de análise estatística são majoritariamente mobilizados. Trata-se de uma visão empobrecida, pois desconsidera, por um lado, o aporte das ciências biológicas básicas, responsáveis pela produção de hipóteses causais acerca dos fenômenos mórbidos e pelo desenvolvimento de novas intervenções com potencial terapêutico. Por outro lado, o movimento também negligencia as características eminentemente técnicas da prática clínica e a sua atuação na observação e delimitação de entidades mórbidas, atividades para as quais a estatística é de pouca valia. Ou seja, a EBM erra, acima de tudo, por tomar a estatística como a base da medicina, quando deveria interpretá-la como apenas mais um entre os diversos instrumentos que a medicina mobiliza a favor do seu projeto de restituição e manutenção da saúde. Além disso, em relação à própria pesquisa clínica, – nível de atividades médicas onde as maiores contribuições do movimento se localizam – e à tomada de decisões prática, um aspecto fundamental é pouco trabalhado, já que seus proponentes não oferecem recursos para avaliar criticamente o valor esperado das condutas médicas. É necessário, no entanto, estender essa última crítica ao nosso próprio trabalho, pois, ao colocarmos a eficácia terapêutica como a direção do progresso médico, deixamos de avançar a discussão no sentido das consequências da intervenção terapêutica. Isto é, deveríamos ter nos perguntado se o aumento da proporção de terapias com eficácia demonstrada significa uma melhora da qualidade do cuidado prestado a indivíduos enfermos. Mas, como dissemos, essa investigação

exigiria a identificação prévia dos valores envolvidos com a prática clínica, o que ainda não foi rigorosamente realizado. No entanto, esperamos poder preencher essa lacuna em trabalhos posteriores.

Com referências pontuais, mas recorrentes, à história da medicina, esperamos ter deixado claro que a EBM pode ser inscrita na tradição do empirismo médico, compreendido como a doutrina que sustenta que o esclarecimento das causas e mecanismos dos fenômenos fisiológicos, mórbidos e terapêuticos apresenta uma importância secundária para a prática clínica. Nesse sentido, o nosso trabalho pode ser interpretado como o desenvolvimento de argumentos em defesa do racionalismo médico, compreendido, por oposição ao empirismo médico, como a doutrina que reconhece na formulação e teste de hipóteses e teorias biológicas causais a base do estatuto científico da medicina e a via fundamental para o avanço do seu conhecimento.

REFERÊNCIAS BIBLIOGRÁFICAS

ACHINSTEIN, P. *The book of evidence*. Oxford: Oxford University Press. 2001.

ALTMAN, D. Statistics in medical journals. *Statistics in Medicine*, v. 1, p. 59-71, 1982.

ALTMAN, D.; BLAND, J. Statistics notes: standard deviations and standard errors. *British Medical Journal*, v. 331, p. 903, 2005.

ATKINS, D. et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches. The GRADE working group. *BMC Health Services Research*, v. 4, n. 38, p., 2004.

BERNARD, C. De l'origine du sucre dans l'économie animale. *Archives générales de médecine*, 4^a série, t. XVIII, p. 303. Set, 1848.

_____. *Introduction à l'étude de la médecine expérimentale*. Paris: Flammarion. 1952 [1865].

_____. *Principes de médecine expérimentale*. Paris: Presses Universitaires de France, 1947.

BLUARD, C. James Ensor et les médecins. *Histoire des Sciences Médicales*, v. 23, n. 2, p. 129-31, 1998.

BRUNNER, H. et al. Oral angiotensin-converting enzyme inhibitor in long-term treatment of hypertensive patients. *Annals of Internal Medicine*, v. 90, p. 19-23, 1979.

BURNS, P.; ROHRICH, R.; CHUNG, K. The levels of evidence and their roles in evidence-based medicine. *Plastic Reconstructive Surgery*, v. 128, n. 1, p. 305-10, 2011.

CANADIAN TASK FORCE ON THE PERIODIC HEALTH EXAMINATION. The periodic health examination. *Canadian Medical Association Journal*, v. 121, p. 1193-254, 1979.

CANGUILHEM, G. Descartes e a técnica. Tradução de Lígia Fraga Silveira. *Trans/Form/Ação*, v. 5, p. 111-22, 1982 [1937].

_____. Essai sur quelques problèmes concernant le normal et le pathologique. In: *Le normal et le pathologique*. Paris: Presses Universitaires de France, 2013 [1943].

_____. Pathologie et the physiologie de la thyroïde au XIX^e siècle. In: *Études d'histoire et de philosophie des sciences*. Paris: J Vrin, 1975.

_____. La décadence de l'idée de Progrès. *Revue de Métaphysique et de Morale*, v. 92, n. 4, p. 437-54, 1987.

_____. Le statut épistémologique de la médecine. *History and Philosophy of the Life Sciences*, v. 10, supl., p. 15-29, 1988.

CARNAP, R. *Logical foundations of probability*. Chicago: The University of Chicago Press, 1950.

CLIFTON, F. *Tabular observations recommended as the plainest and surest way of practising and improving physick, in a letter to a friend*. Londres: J. Brindley, 1731.

CRAWLEY, L. Evidence-based medicine: a new paradigm for the patient. *Journal of the American Medical Association*, v. 269, n. 10, p. 1253, 1993.

DEPARTMENT OF CLINICAL EPIDEMIOLOGY AND BIOSTATISTICS, MCMASTER UNIVERSITY. How to read clinical journals: I. Why to read them and how to start reading them critically. *Canadian Medical Association Journal*, v. 124, p. 555-8, 1981a.

_____. How to read clinical journals: V. To distinguish useful from useless or even harmful therapy. *Canadian Medical Association Journal*, v. 124, p. 1156-62, 1981b.

ESTADOS UNIDOS DA AMÉRICA. Lei Federal de Alimentos, Medicamentos e Cosméticos dos Estados Unidos, Emenda "Kefauver Harris" de 10 de outubro de 1962 (Estatutos em Geral 76, Lei Pública 87-781, pág. 780-796).

EVIDENCE-BASED MEDICINE WORKING GROUP. Evidence-based medicine: a new approach to teaching the practice of medicine. *Journal of the American Medical Association*, v. 286, n. 17, p. 2420-5, 1992.

FEINSTEIN, A. Clinical biostatistics XIII. On homogeneity, taxonomy, and nosography. *Clinical Pharmacology and Therapeutics*, v. 13, n. 1, p. 114-29, 1972.

FERREIRA, S. A bradykinin-potentiating factor (BPF) present in the venom of *Bothrops jararaca*. *British Journal of Pharmacology*, v. 24, p. 163-9, 1965.

FESTING, S.; WILKINSON, R. The ethics of animal research. *EMBO reports*, v. 8, n. 6, p. 526-30, 2007.

FISHER, R. *The design of experiments*. Edimburgo: Oliver and Boyd, 1935.

FOUCAULT, M. *O nascimento da clínica*. Rio de Janeiro: Forense Universitária, 2011 [1963].

GIROUX, E. Les modèles de risque en médecine. Quelles conséquences pour la définition des normes et pour le jugement clinique? Exemple du calcul du risque cardiovasculaire global. In: FARGOT-LARGEAULT, A. (ed.). *L'émergence de la médecine scientifique*. Paris: Éditions Matériologiques, 2013.

GLOBAL INITIATIVE FOR ASTHMA. *A pocket guide for asthma management and prevention (for adults and children older than 5 years)*. 2018. Disponível em: <https://ginasthma.org/2018-pocket-guide-for-asthma-management-and-prevention/>

GOLDBLATT, H. et al. The production of persistent elevation of systolic blood pressure by means of renal ischemia. *Journal of Experimental Medicine*, v. 59, n. 3, p. 347-79, 1934.

GUYATT, G. Evidence-based medicine. *American College of Physicians Journal Club*, v. 114, n. 2, p. A16, 1991.

GUYATT, G. et al. Users' guides to the medical literature: IX. A method for grading health care recommendations. *Journal of the American Medical Association*, v. 274, n. 22, p. 1800-4, 1995.

_____. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal*, v. 336, p. 924-6, 2008.

HACKING, I. *Logic of statistical inference*. Cambridge: Cambridge University Press. 1965.

_____. *The emergence of probability: a philosophical study of early ideas about probability, induction and statistical inference*. Cambridge: Cambridge University Press. 1984.

_____. *An introduction to probability and inductive logic*. Cambridge: Cambridge University Press. 2001.

HAYNES, B. The origins and aspirations of ACP Journal Club. *American College of Physicians Journal Club*, v. 114, n.1, p. A18, 1991.

KANNEL, W.; MCGEE, D.; GORDON, T. A general cardiovascular risk profile: the Framingham study. *The American Journal of Cardiology*, v. 38, p. 46-51, 1976.

KEANE, M.; BERG, C. Evidence-based medicine: a predictably flawed paradigm. *Trends in Anaesthesia and Critical Care*, v. 9, p. 49-52, 2016.

KUHN, T. *The structure of scientific revolutions*. Chicago: The University of Chicago Press, 1970a [1962].

_____. Logic of discovery or psychology of research. In: LAKATOS, I.; MUSGRAVE, A. (ed.). *Criticism and the growth of knowledge*. Cambridge: Cambridge University Press. 1970b.

_____. What are scientific revolutions? In: KRÜGER, L.; DASTON, L.; HEIDELBERGER, M. (ed.). *The probabilistic revolution: volume I. Ideas in history*. Cambridge: The MIT Press. 1987.

LAFFAN, R. et al. Antihypertensive activity in rats of SQ 14,255, an orally active inhibitor of angiotensin I converting enzyme. *The Journal of Pharmacology and Experimental Therapeutics*, v. 204, n. 2, p. 281-8, 1978.

LEE, D.; VIELEMEYER, O. Analysis of overall level of evidence behind Infectious Diseases Society of America practice guidelines. *Archives of Internal Medicine*, v. 171, n. 1, p. 18-22, 2011.

LERMAN, L. et al. Animal models of hypertension: an overview. *Journal of Laboratory and Clinical Medicine*, v. 146, n. 3, p. 160-73, 2005.

LEWIS, S.; ORLAND B. The importance and impact of evidence-based medicine. *Journal of Managed Care Pharmacy*, v. 10, n. 5, s. a, S3-5, 2004.

LIND, J. *A treatise on the scurvy*. In three parts. Containing an inquiry into the nature, causes and cure, of that disease. Together with a critical and chronological view of what has been published on the subject. Londres, 1757 [1753].

LOUIS, P. *Recherches anatomico-pathologiques sur la phthisie*. Paris: Chez Gabon et Compagnie, 1825.

_____. *Recherches sur les effets de la saignée dans quelques maladies inflammatoires, et sur l'action de l'émétique et des vésicatoires dans la pneumonie*. Paris: J. B. Baillière, 1835.

_____. De l'examen des malades et de la recherche des faits généraux. *Mémoires de la société médicale d'observation*, t. I, p. 1-63, 1837.

MAHMOOD, S. et al. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. *The Lancet*, v. 383, p. 999-1008, 2014.

MCGONIGLE, P.; RUGGERI, B. Animal models of human disease: challenges in enabling translation. *Biochemical Pharmacology*, v. 87, n. 1, p. 162-71, 2014.

MEDICAL RESEARCH COUNCIL. The serum treatment of lobar pneumonia. *The British Medical Journal*, v.1, p. 241-5, 1934.

_____. Streptomycin treatment of pulmonary tuberculosis. *British Medical Journal*, v. 2, p. 769-82, 1948.

MISTRY, J.; BUKHARI, M.; TAYLOR, A. Alkaptonuria. *Rare Diseases*, v.1, n.1, p. 1-7, 2013.

NEYMAN, J. "Inductive behavior" as a basic concept of philosophy of science. *Revue de l'Institut International de Statistique*, v. 25, n. 1/3, p. 7-22. 1957.

NG, K.; VANE, J. Some properties of angiotensin converting enzyme in the lung in vivo. *Nature*, v. 225, p. 1142-4, 1970.

ONDETTI, M.; RUBIN, B.; CUSHMAN, D. Design of specific inhibitors of angiotensin-converting enzyme: new class of orally active antihypertensive agents. *Science*, v. 196, p. 441-3. 1977.

POCOCK, S.; HUGHES, M.; LEE, R. Statistical problems in the reporting of clinical trials. *The New England Journal of Medicine*, v. 317, n. 7, p. 426-32, 1987.

POONACHA, T.K., GO, R.S. Level of Scientific Evidence Underlying Recommendations Arising From the National Comprehensive Cancer Network Clinical Practice Guidelines. *Journal of Clinical Oncology*, American Society of Clinical Oncology, v. 29, n. 2, p. 186-91, 2011.

ROTHSCHUH, K. Ist das Kuhnsche Erklärungsmodell wissenschaftlicher Wandlungen mit Gewinn auf die Konzept der Klinische Medizin anwendbar? In: DIEMER, A. (ed.). *Die Struktur wissenschaftlicher Revolutionen und die Geschichte der Wissenschaften*. Meisenheim: Verlag Antonin Hain, p. 73-90, 1977.

SACKETT, D. Bias in analytic research. *Journal of Chronic Diseases*, v. 32, p. 51-63, 1979.

_____. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest*, v. 95, n. 2 (suplemento), 1989.

_____. Why did I become a clinician-trialist?. *Journal of the Royal Society of Medicine*, v. 108, n. 8, p. 325-30, 2015.

SACKETT, D.; ROSENBERG, W. The need for evidence-based medicine. *Journal of the Royal Society of Medicine*, v.88, p. 620-4, 1995.

SACKETT, D.; HAYNES, B.; TUGWELL, P. *Clinical epidemiology: a basic science for clinical medicine*. Boston: Little, Brown and Company, 1985.

SACKETT, D. et al. Evidence-based medicine: what it is and what it isn't. *British Medical Journal*, v. 312, p. 71-2, 1996.

SEISHA, S.; YOUNG, G.,. The evidence-based medicine paradigm: where are we 20 years later? Part 1. *Canadian Journal of Neurological Sciences*, v. 40, p. 465-74, 2013.

SEMMELEWEIS, I. (1850) Lecture on the genesis of puerperal fever (childbed fever). In: BROCK, (ed.) *Milestones in microbiology*. Londres: Prentice-Hall International, 1961.

SKEGGS, L. et al. The existence of two forms of hypertensin. *Journal of Experimental Medicine*, v. 99, n. 3, p. 275-82. 1954.

SKEGGS, L.; KAHN, J.; SHUMWAY, N. The preparation and function of the hypertensin-converting enzyme. *Journal of Experimental Medicina*, v. 103, n. 3, p. 295-9, 1956.

SMITH, R.; RENNIE, D. Evidence based medicine – an oral history. Provides the background to one of modern medicine's greatest achievements. *The British Medical Journal*, v. 348, 2014.

- SNOW, J. *On the mode of communication of cholera*. Londres: John Churchill, 1849.
- STOLLER, J. et al. Captopril-induced cough. *Chest*, v. 93, n.3, p. 659-61, 1988.
- TRICOCCI, P et al. Scientific evidence underlying the ACC/AHA clinical practice guidelines. *The Journal of the American Medical Association*, v. 301, n. 8, p. 831-841. Fev. 2009.
- TRÖHLER, U. "To improve the evidence of medicine": The 18th century British origins of a critical approach. Edimburgo: Royal College of Physicians of Edinburgh. 2000.
- _____. Lind and scurvy: 1747 to 1795. *Journal of the Royal Society of Medicine*, v. 98, p. 519-22, 2005.
- UMSHEID, C; MARGOLIS, D.; GROSSMAM, C. Key concepts of clinical trials: a narrative review. *Postgraduate Medical Journal*, v. 123, n. 5, p. 164-204. 2011.
- VETERANS ADMINISTRATION COOPERATIVE STUDY GROUP ON ANTI-HYPERTENSIVE AGENTS. Low-dose captopril for the treatment of mild to moderate hypertension. I: *Results of a 14-week trial*. *Archives of Internal Medicine*, v. 144, n. 10, p. 1947-53, 1984.
- WEISZ, G. et al. The emergence of clinical practice guidelines. *The Milbank Quarterly*, v. 85, n. 4, p. 691-727, 2007.
- WHELTON, P et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Hypertension*, 2017.
- WORLD HEALTH ORGANIZATION. *International Classification of Diseases 11th Edition*. 2018. Disponível em: <https://icd.who.int/browse11/l-m/en>