CUIDADO COM A MARÉ

FOTOS LÉO RAMOS

Relatório propõe limites para o uso de indicadores na avaliação científica no Reino Unido

Fabrício Marques

O debate sobre a confiabilidade de métodos quantitativos para medir o impacto da produção científica e acadêmica ganhou um novo capítulo com a divulgação, em julho, de um relatório encomendado pelo Higher Education Funding Council for England (Hefce), órgão responsável por financiar e avaliar o sistema universitário e de pesquisa da Inglaterra. Fruto de 15 meses de trabalho de uma equipe interdisciplinar independente, o documento *The metric tide* (A maré das métricas) aborda a utilidade e o uso abusivo de indicadores no julgamento do mérito de universidades e grupos de pesquisa. Diante da constatação de que se disseminaram parâmetros como indicadores de impacto e *rankings* universitários, o grupo sugere parcimônia ao empregá-los. "As métricas precisam ser escolhidas com cuidado e devem sempre suplementar e apoiar o julgamento de especialistas, em vez de substituí-lo", diz Richard Jones, pró-reitor de Pesquisa e Inovação da Universidade de Sheffield, membro do painel que produziu o documento.

O grupo apresentou o conceito de "métrica responsável", baseado em cinco pontos. Um deles é a humildade, entendida como o reconhecimento de que a avaliação por pares, embora imperfeita e sujeita a equívocos, é capaz de enxergar de forma ampla a qualidade da produção científica, algo que indicadores isolados ainda não conseguem fazer. O segundo ponto é a robustez, condição que exclui o uso de dados descontextualizados ou não suficientemente representativos num processo de avaliação. Segundo o relatório, a ênfase em parâmetros "estreitos e mal desenhados" produz consequências negativas. Um exemplo desaconselhado é a utilização do fator de impacto de uma revista científica para definir a qualidade de qualquer trabalho publicado nela ou o mérito de seus autores. Isso porque tais índices apenas espelham médias observadas em conjuntos de artigos publicados em períodos anteriores. O documento também menciona o uso de citações de artigos como um critério universal de qualidade, sem contemplar as realidades distintas das disciplinas.

O terceiro ponto é a transparência, garantindo que a coleta de dados e suas análises sejam abertas e compreensíveis para os pesquisadores e a sociedade. A utilização massiva de *rankings* universitários é criticada pelo relatório, com o argumento de que falta em muitos deles transparência sobre a escolha de indicadores. O quarto ponto é a diversidade, esforço para adotar um conjunto de indicadores capaz de abranger aspectos variados da contribuição dos pesquisadores. Por fim, o quinto elemento é a reflexividade, compreendida como a preocupação de identificar prontamente efeitos indesejados que o uso de indicadores possa gerar e a disposição para corrigi-los.

"A atração pelas métricas só tende a aumentar", escreveu na revista *Nature* James Wilsdon, professor da Universidade de Sussex e líder do painel que produziu o relatório. Segundo ele, há demandas crescentes para avaliar o investimento público em pesquisa e educação superior e, ao mesmo tempo, a quantidade de dados sobre o desempenho científico e a capacidade de analisá-los se multiplicaram. "As instituições precisam administrar suas estratégias de pesquisa e, simultaneamente, competir por prestígio, estudantes e recursos."

O assunto é especialmente sensível no Reino Unido porque, a cada cinco anos, suas universidades e grupos de pesquisa são submetidos a um grande processo de avaliação, que define a distribuição de verbas públicas pelo período seguinte. O último deles, o Research Excellence Framework (REF 2014), foi divulgado pelo Hefce em dezembro passado. Cento e cinquenta e quatro universidades submeteram 1.911 itens em 36 campos do conhecimento. Cada item apresentado reuniu um conjunto de trabalhos científicos, estudos de caso, patentes, projetos de pesquisa em curso, informações sobre o desempenho de pesquisadores e indicadores bibliométricos, vinculado a um departamento ou a um grupo de pesquisa, e foi avaliado por painéis de especialistas. A qualidade da produção científica foi responsável por 65% da avaliação, o impacto da pesquisa fora da academia (uma novidade do REF 2014) valeu 20% e o ambiente de pesquisa, 15%. O julgamento concluiu que 30% das universidades do Reino Unido são líderes mundiais, 46% têm nível internacional de excelência, 20% têm reconhecimento internacional e 3% reconhecimento apenas nacional.

DISCREPÂNCIAS

A equipe responsável pelo relatório avaliou os dados do REF 2014 e concluiu que nem sempre indicadores individuais e análise por pares produziam resultados convergentes. Discrepâncias acentuadas foram observadas, por exemplo, no desempenho de pesquisadores em início de carreira. Da mesma forma, a cobertura dos indicadores mostrou-se desigual nos campos do conhecimento, com problemas específicos atingindo em especial o painel de Artes e Humanidades. O relatório recomendou que o modelo atual do REF seja mantido, baseado na avaliação qualitativa feita por especialistas que podem, porém, considerar indicadores selecionados de forma criteriosa. E sugeriu que se aumentem os investimentos em "pesquisa sobre a pesquisa", para aprofundar a compreensão sobre o uso de indicadores. O grupo também instituiu um prêmio às avessas, o Bad Metric Prize, para denunciar os usos inapropriados de indicadores quantitativos. Os primeiros premiados serão conhecidos no ano que vem.

FALTA TRANSPARÊNCIA NA SELEÇÃO DE INDICADORES USADOS POR *RANKINGS* DE UNIVERSIDADES, DIZ RELATÓRIO

Enquanto o relatório era produzido, a comunidade científica britânica chocou-se com uma tragédia relacionada à pressão exercida sobre os pesquisadores pelas métricas. Stefan Grimm, por 10 anos professor de toxicologia da Faculdade de Medicina do Imperial College, suicidou-se aos 51 anos de idade. Estava deprimido com o anúncio de sua demissão e deixou um *e-mail* relatando as sucessivas ameaças que sofreu de seu superior para conseguir um determinado patamar de financiamento para seu laboratório, que ele não conseguiu atingir. O Imperial College anunciou a revisão de seus critérios de avaliação depois do episódio, mencionado na apresentação do relatório.

O documento encomendado pelo Hefce perfila-se com textos recentes que defendem ideias semelhantes. Um deles é o *Manifesto de Leiden* sobre métricas de pesquisa, lançado em setembro de 2014 na 19ª Conferência Internacional de Indicadores em Ciência e Tecnologia, realizada em Leiden, na Holanda. Seus 10 princípios coincidem em grande medida com as recomendações do grupo britânico. Falam, por exemplo, da necessidade de transparência na análise de dados e propõem considerar as diferenças entre

áreas nas práticas de publicação e citação. Outra referência é a *Dora*, sigla para *San Francisco declaration on research assessment*, lançada em dezembro de 2012 em um encontro da American Society for Cell Biology, que faz 18 recomendações para pesquisadores, instituições, agências de fomento e editores científicos. A principal delas propõe eliminar o uso do fator de impacto de revistas como indicador da qualidade de um artigo. Quase 600 instituições científicas e 12,5 mil pesquisadores já assinaram a declaração – o relatório do Hefce sugere que instituições e agências tornem-se signatárias da declaração para orientar publicamente suas práticas no campo da avaliação.

A importância de o Reino Unido adotar esse tipo de postura não é desprezível. "Se a maioria dos países ainda está na infância da discussão sobre avaliação, o Reino Unido está alguns passos adiante, querendo sair da adolescência", observa Sergio Salles-Filho, professor da Universidade Estadual de Campinas (Unicamp) e coordenador do Grupo de Estudos sobre Organização da Pesquisa e da Inovação (Geopi), que avaliou programas da FAPESP. Ele observa que a inclusão de novos parâmetros para avaliar a produção científica também é motivada pela necessidade de sofisticar a avaliação, mensurando aspectos diversos ligados a seu impacto na sociedade. "Em certas áreas, o mais importante não é publicar artigos, mas produzir manuais usados na indústria, promover mudanças nas políticas públicas ou mudar as diretrizes da política econômica. Os processos de avaliação estão se transformando e daqui a 20 anos serão muito diferentes."

CICLO COMPLETO

O Brasil tem avançado na ampliação de critérios de avaliação. "As agências de fomento não se satisfazem mais em saber apenas qual é o impacto específico de um artigo científico e buscam fazer avaliações de ciclo completo, que acumulam informações sobre o trabalho de pesquisadores e os resultados de programas para medir sua contribuição no longo prazo", diz, referindo-se ao esforço da FAPESP e mais recentemente da Financiadora de Estudos e Projetos (Finep) de sistematizar um processo de coleta de dados que siga reunindo informações sobre os resultados de pesquisa ao longo do tempo.

Rogério Mugnaini, professor da Escola de Comunicações e Artes (ECA) da Universidade de São Paulo (USP), está estudando a diversidade de critérios de avaliação dos programas de pós-graduação feita pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), se baseando em todos os documentos propostos pelas áreas do conhecimento desde 1998. Já observou que as áreas recorrem cada vez mais a indicadores de impacto, mesmo que esses parâmetros não sejam valorizados pela cultura daquela disciplina. "Algumas áreas, como a geografia, estão adotando o modelo de avaliação das ciências duras", diz Mugnaini. "Como o volume de títulos a ser avaliado é muito extenso, há uma tendência de adoção de indicadores, desconhecendo suas limitações." Para Salles-Filho, há mais aspectos que precisam ser contemplados na avaliação dos programas de pós-graduação. "Formamos 15 mil doutores por ano no Brasil, mas não sabemos onde estão e o que estão fazendo com o conhecimento e a experiência adquiridos no doutorado, se estão orientando teses, trabalhando no setor público ou na iniciativa privada. Deveríamos ter uma visão dos impactos sociais da formação em pós-graduação", diz. ■

FOTOS LÉO RAMOS