A bibliometric analysis of digestive health research in Canada: “Fair is foul, and foul is fair”

Stephen S Vanner MSc MD FRCPC

In the current issue of The Canadian Journal of Gastroenterology, Tuitt et al (1) (pages 609-614) attempted to evaluate the impact and/or influence of Canadian researchers and their research groups using two measurement tools: the Hirsch factor or index (h-I), and their newly described Influence factor. They suggest that such metrics can provide information to funding agencies, presumably for the purpose of rank ordering, and to institutions to guide staffing and recruitment decisions. Their study focuses on the relative merits of such bibliometric tools for assessing individual researchers and, at the same time, serves to demonstrate the dangers when such metrics are applied without scientific rigour.

From the perspective of assessing ‘groups’, a fundamental flaw of this particular study was that it rested on a database that does not accurately represent the digestive health research community in Canada. First, a number of the defined research ‘groups’ are artificial. Although there are several true groups included in the analysis, there are many sites represented by aggregations of several highly capable solo investigators who do not belong to any such functional research groups. It is, therefore, quite misleading to present the data according to the authors’ artificial research ‘groups’. Second, the selection of 106 investigators evaluated across Canada represents a serious bias, failing to include some high profile investigators, while at the same time, including others who are not part of an active research program. Accurately identifying the functional research groups within Canada and determining the minimum research domain mix necessary to define a digestive diseases researcher is a difficult task, and certainly not one that can be left solely in the purview of data analysts.

Another major flaw of this study is the primary metric, the h-I and, in particular, the manner in which it was used. This simple counting tool was conceived by a physicist, Jorge Hirsch (2). The development of the tool was necessarily empirical. He ‘validated’ the tool initially by constructing mathematical models using several intuitive assumptions, and then examined its performance with a limited number of elite, established (Nobel Laureates) investigators, first in the physical sciences and then with a limited array of biomedical scientists, again with only elite investigators. In this validation exercise, the gold standard was pre-existing scientific status based on very exceptional criteria. Most importantly, the performance of the tool over a broad range of investigators at different stages of career development has never been validated. Moreover, there is considerable active debate among experts concerning the precise utility of this and other similar tools in the evaluation of individual contributions to research (3). As an example, Poynard et al (4) recently showed that the h-I does not predict truth survival in at least one branch of biomedical research. Truth survival is arguably a more robust measure of the ultimate importance of research contributions in any field. In particular, the authors appear to discount the observation that the h-I for investigators in a nonmainstream field will be systematically lower. For example, the investigation of a possible infectious cause for peptic ulcer by Marshall and Warren (5) was not considered to be a mainstream field. His emergence as a paradigm shifter occurred after this discovery. They also appear to ignore the problems created by collaborative research and co-authorship, they pay no heed to the thorny issue of self-citation and the emerging problem of citation networking. Particularly troubling in the study by Tuitt et al is the failure to correct the index according to career stage – as Hirsch himself recommended. This is a substantial problem. The h-I will be strongly affected by the length of the career of the individual being valued. The authors’ decision in this report to limit the valuation to the past 10 years to correct for seniority is mechanistically naive, and represents a substantial bias toward groups that are top heavy with senior investigators. A method of normalizing the valuation measure is crucial.

To refine the valuation process, the authors propose the use of the Influence factor. This factor, a construct of Research Excellence Metrics Ltd (Burlington, Ontario) (6), additively combines the h-I with two terms, neither of which is likely to change the rank orders determined by the h-I. Thus, correlations between the two measures, one of which depends directly on the other, are of little importance.

Even a cursory review by educated readers will identify additional fundamental flaws in this study. The authors have a scientific obligation to describe the nature of the data and to provide some sort of statistical treatment. This is a very difficult problem. The authors calculated the means of h-I’s among the arbitrary sites or ‘groups’. There is no evidence that these values are linearly distributed. Is an investigator working in field A for seven years with an h-I of 16 half as valuable as an investigator working in field B for 30 years with an h-I of 32? It is not appropriate to draw such conclusions based on the evidence. Similarly, drawing conclusions about recruitment and comparisons of basic and clinician scientists is meaningless in the absence of a clearly defined and representative database.

In fairness, the authors did demonstrate some insight into the importance of an accurate database, by acknowledging the problems inherent in the collection of investigator-specific bibliographic information. Unfortunately, they suggest that they controlled for data quality in their study by contacting – post facto – the ‘top’ 30 researchers to affirm the accuracy of the h-I calculation. This is not scientifically acceptable. The entire data base should be verified methodically. In the absence of such confirmation, the authors must use a representative random sample of the study population to check the data – not a biased sample of the ‘top’ 30. This is another violation of fundamental principles of evidence-based research practice.

Beyond the lack of scientific rigour, there are other troublesome aspects of this article. Almost one year before this publication, the main results (and, in some instances, specific details not in the report) appeared online as a promotion piece for Research Excellence Metrics Ltd – a for-profit company. Simultaneously, they also appeared on the website of one of the centres in the study. This is unusual in traditional academic research, and in an era of e-journals, raises challenging questions for print journals. The public dissemination of biased data unfettered by critical appraisal is risky business. The use of these results to draw comparisons among research groups can be divisive, particularly unfortunate in a community that has worked closely to enhance
the research environment and training in our country. If the goal was truly
to examine the impact and influence of digestive health investigators
across Canada, then accurate comparisons with other disciplines or with
foreign research content would have been much more constructive.

There is no question that accountability of research funding is
imperative. Bibliometrics will no doubt be increasingly used as part of
this process, but this study is a compelling example of potential abuses
and the limitations of such an approach. The challenge going forward is
to ensure that these metrics are further validated where needed, used
appropriately and, most importantly, complemented with other meas-
tures to reflect overall scientific merit. Hopefully, scientists in the digest-
ive health community will play an important role in developing an
accurate accountability framework and, in the meantime, discourage
irresponsible bibliometric reporting and those who embrace the results.

REFERENCES
1. Tuitt D, Knight F, Lipman T. A bibliometric analysis of digestive
2. Hirsch JE. An index to quantify an individual’s scientific research
3. Saleem T. The Hirsch index – a play on numbers or a true appraisal
4. Poynard T, Thabut D, Munteanu M, Ratziu V, Benhamou Y,
Deckmyn O. Hirsch index and truth survival in clinical research.
(Accessed on October 11, 2011).