

Statistical methodology for the scientometric study of the growth of medical sciences in India

Ashwani Kumar Mishra* and
Yatan Pal Singh Balhara

Department of Psychiatry and National Drug Dependence
Treatment Centre (NDDTC), Room No. 4096, 4th Floor,
Teaching Block, All India Institute of Medical Sciences, Ansari Nagar,
New Delhi 110 029, India

There is evidence that the growth of medical literature in India is phenomenal. However, the trajectory of this growth requires further study and the findings need to be disseminated. With this in mind the present study attempts to draw inferences on the trajectory of four broad domains of medical sciences in India over the span of 16 years, utilizing the available scientometrics information. The results are indicative of differential growth trajectory in many sub-disciplines of medical sciences. The specialities such as epidemiology, obstetrics and gynaecology, geriatrics and psychiatry and mental health, need to be pursued more seriously.

Keywords: Growth trajectory, medical sciences, scientometrics, statistical methodology.

SCIENTIFIC development is a continuous process attributed to theoretical and applied research conducted by scientists, academicians, professionals and researchers. They continuously strive for excellence in their research domain, by translating the research results as publications in discipline-specific international and national journals. Consequently, scientific productivity and visibility are enhanced globally, regionally and locally.

The science of studying bibliographic databases is termed as scientometrics. More specifically, the science of studying and analysing science is called scientometrics. It embeds information pertaining to the scientific productivity of academic researchers, institutions, organizations, universities and advanced scientific centres. The emerged scientific evidence on these lines highlights the importance of scientometrics databases that are specific to a country and discipline¹, researchers and scientists², and also at the organizational level³.

The last few decades have witnessed exponential growth in scientific output. This has been ascribed to the technological expansion and its accessibility and application of sound scientific methodologies matching with the pace of technological revolution. In the times to come the technological expansion is bound to exhibit greater upsurge and researchers may face challenges in delivering technology based scientific productivity. The scien-

tific productivity possesses multifaceted dimensions, with innovations, inventions, publications, patents being some of them. The major dimension of health also falls into this group, which has greater visibility in the modern era. The indicators signifying its growth are increase in the quality of life and improved prognosis of terminally ill patients⁴, declining maternal mortality rates to some extent⁵ and increase in the healthy life expectancy⁶. However, the medical fraternity should not just be euphoric about such indicators; rather it should introspect periodically and prioritize research areas. Such introspection in addition to understanding the trajectory of progress will also provide ample opportunity to systematically evaluate the contribution of various subcomponents of medical sciences. In the process the precise delineation of domains depicting greater or lower scientific visibility or requiring appropriate attention will also be possible.

In the Indian context scientometric information exists in medical sciences and its related disciplines like genetics⁷, biotechnology⁸ and medical sciences⁹ to mention a few. The detailed study of medical sciences scientometric analysis was performed recently in north India⁹. However, the conclusions drawn from such studies are limited to the areas considered, based on only one subject domain at a time, provide descriptive scientometrics information on scenario of research in medical sciences in India and thus lacked adequate application of statistical methodology.

With these points in mind the present study attempts at disseminating the information on trajectory of four medically related disciplines from an Indian scenario. The data extraction and statistical methodology adopted for drawing inferences form the core aspects of the study.

The data for the present study have been extracted from user-friendly graphical user interface, web-based portal of SCImago¹⁰. The bibliographic details are in tabular format distributed over 16 years (1996–2011). After this the data were retrieved under four broad disciplines/domains of medical sciences, viz. medicine (MED); biochemistry, genetics and molecular biology (BGMB), immunology and microbiology (IM) and health profession (HP). These domains constituted 47, 15, 6 and 9 sub-disciplines respectively, distributed over the same time-span.

A single spreadsheet was formed that contained data for 16 years (1996–2011) from 77 subdisciplines. The fields extracted for the analysis were total number of published documents [doc], percentage of cited documents [cd], uncited documents [ucd], percentage of international [pi], regional [pr] and world [pw] collaborations.

The 16 years' trajectories of the four broad disciplines mentioned above for published documents (Figure 1); percentage of cited documents (Figure 2a), international collaboration (Figure 2b); regional collaboration (Figure 3) and world contribution (Figure 4) were plotted. The error plots (point estimate of mean along with its 95% confidence interval; CI) of each subdiscipline of all

*For correspondence. (e-mail: ashwaniop@gmail.com)

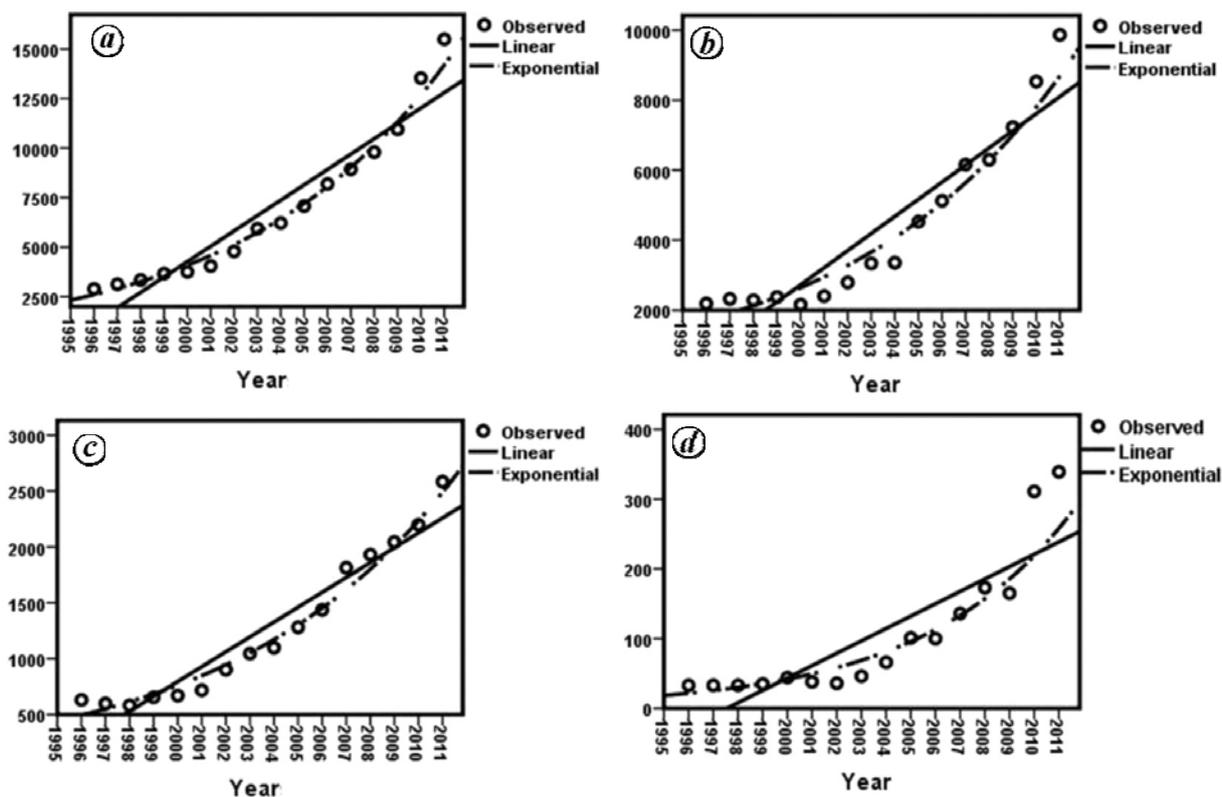


Figure 1a-d. Curve fitting to publications from the domains of medicine (a); biochemistry, genetics and molecular biology (b); immunology and microbiology (c) and health profession (d) over a period of 16 years in linear and exponential form.

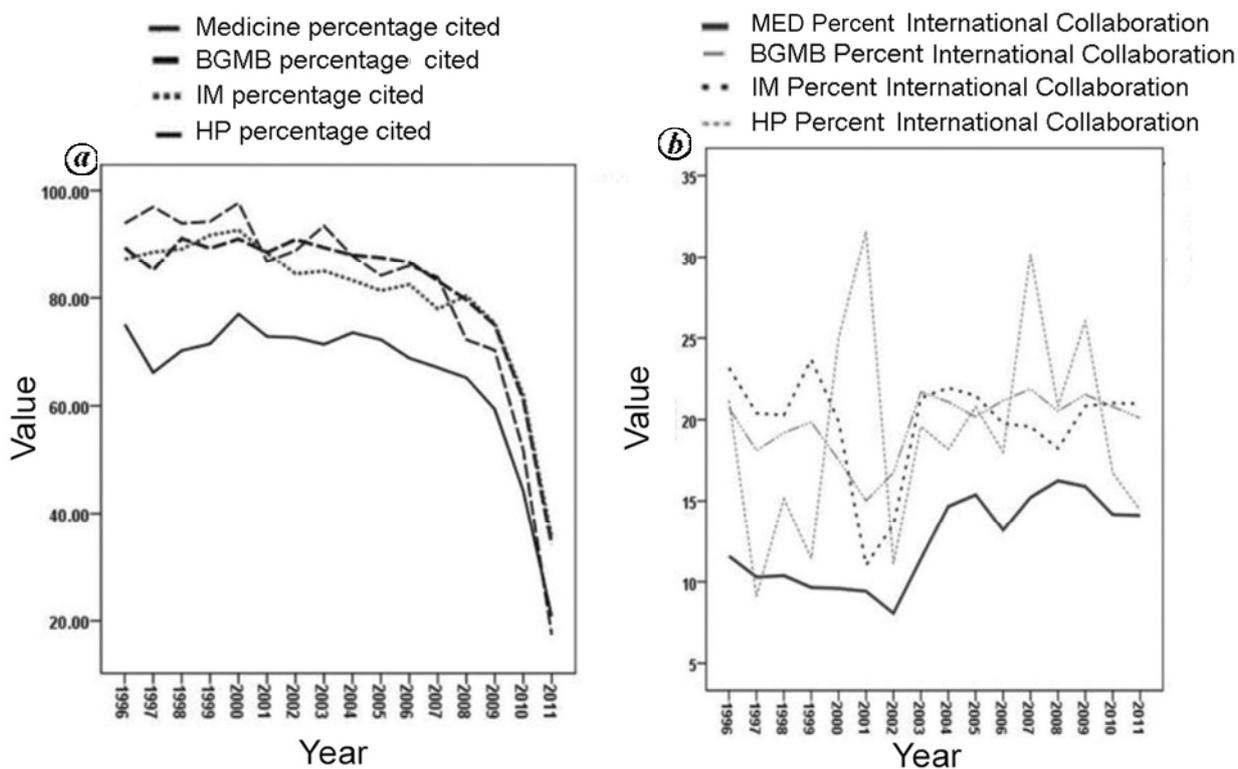


Figure 2a-b. Curve fitting to percentage of cited documents (a) and international collaboration (b) in the four disciplines considered.

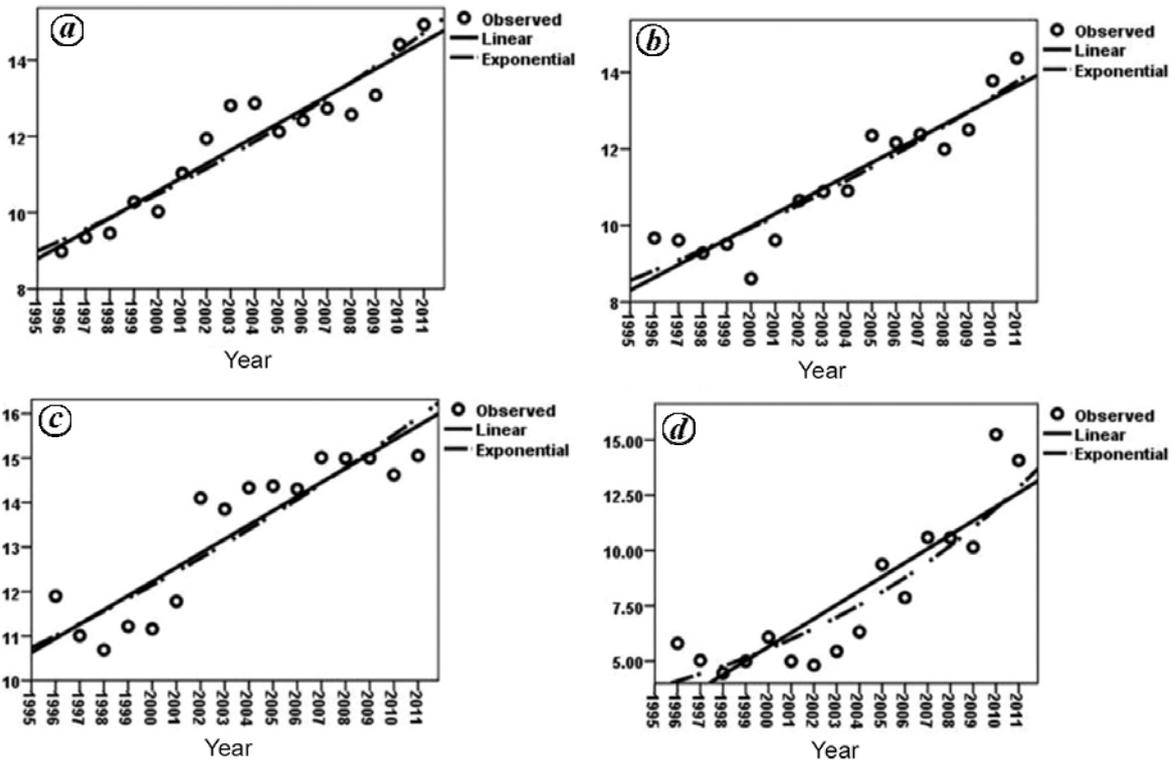


Figure 3 a-d. Curve fitting to percentage of publications having regional collaboration in the four domains of medicine (a); biochemistry, genetics and molecular biology (b); immunology and microbiology (c) and health profession (d) over a period of 16 years in linear and exponential form.

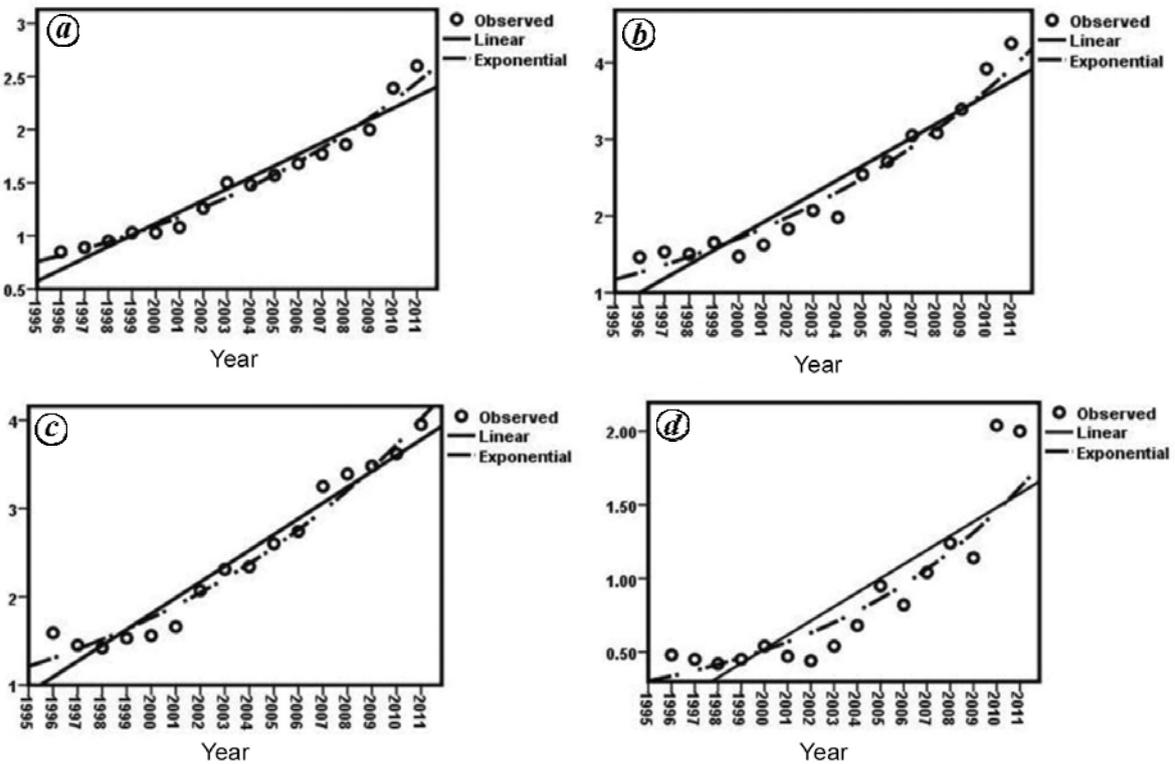


Figure 4 a-d. Curve fitting to percentage of publications having world collaborations in the four domains of medicine (a); biochemistry, genetics and molecular biology (b); immunology and microbiology (c) and health profession (d) over a period of 16 years in linear and exponential form.

major disciplines are shown in Figure 5. The curve fitting methodology was used to fit the trajectory of published documents; percentage of cd, pi, pr and pw. This resulted in estimate of coefficient of determination (R^2) for each disciplines. Statistical analysis was performed using licenced version of SPSS 21.0 software.

The four broader disciplines considered in this study exhibited an exponential growth of scientific publications, in the order of MED, BGMB, IM and HP, compared to the linear model. This was evident by the greater value of coefficient of determination (MED: exp: $R^2 = 0.98$ versus linear: $R^2 = 0.91$; BGMB: exp: $R^2 = 0.93$ versus linear: $R^2 = 0.87$; IM: exp: $R^2 = 0.97$ versus linear: $R^2 = 0.92$; HP: exp: $R^2 = 0.90$ versus linear: $R^2 = 0.74$; Figure 1). The fitting of data for 'cd', depicted decreasing trend and random fluctuation for 'pi', for all major disciplines (Figure 2). The fitting of parameters for 'pr' and 'pw' leads to almost similar pattern for linear and exponential form for pr (Figure 3); MED: exp: $R^2 = 0.90$ versus linear: $R^2 = 0.90$; BGMB: exp: $R^2 = 0.86$ versus linear: $R^2 = 0.86$; IM: exp: $R^2 = 0.78$ versus linear: $R^2 = 0.80$; HP: exp: $R^2 = 0.80$ versus linear: $R^2 = 0.77$ and exponential for 'pw', for category (MED: exp: $R^2 = 0.98$ versus linear: $R^2 = 0.94$; BGMB: exp: $R^2 = 0.94$ versus linear: $R^2 = 0.90$); HP (exp: $R^2 = 0.85$ versus linear: $R^2 = 0.75$) and almost similarly both (exponential and linear) in IM category (exp: $R^2 = 0.95$ versus linear: $R^2 = 0.95$; Figure 4).

Figure 5 depicts mean (95% CI) of published documents for 47, 15, 6 and 10 subdomains of four major disciplines. In MED, for medicine (miscellaneous) estimated mean was maximum and for paediatrics, perinatology and child health lesser than medicine (miscellaneous). For majority of them the mean number of published documents was below 500. The core clinical specialities like medmisc, ppch, neurocl, sur, rnmim, opthal and cvmed published more on average than other equally important clinical categories like obg, gast, uro, psymh, ana, edmet, neph, trans, ger, epi and phys. In the case of BGMB, the highest mean was for bioche. For majority of disciplines pertaining to BGMB the average was less than 500, with only the exception of biotech and bioche (Figure 5 b). In case of IM, the maximum mean was found for ambiotech followed by microb, immune, parasite, virol and immicmisc (Figure 5 c) and for HP, rulra, preceded by hpmisc and cmt (Figure 5 d).

The scientometrics information of medical sciences in India highlights its inadequacy in terms of scope and quality¹¹. An attempt has been made to understand collectively the trajectory of publications with wider spectrum in four medical sciences-related disciplines. The simultaneous consideration of four domains is attributed to the fact that medicine is a multidisciplinary speciality and such consideration offers flexibility for assessing collective growth trajectory.

The findings revealed exponential growth of publications for all the four disciplines. However, cited documents and percentage of international collaboration (except MED) exhibited respectively, decreasing trend and random fluctuations for all major disciplines (Figure 2 a and b). The finding of decreasing trend in percentage of cited documents and international collaboration not matching with volume of growth is corroborated with a recently released report by the Department of Science and Technology¹². The rather better international collaboration of MED can be regarded as promotion of interdisciplinary research in recent times. This needs to be replicated for other areas like BGMG, IM and HP. The curve of percentage of regional and world collaboration among the published documents demonstrated linear and exponential fitting in parallel. The exponential growth signifies large research infrastructure; enhanced capacity building and greater interdisciplinary collaborative research activities. But the decreasing trend in cited documents raises fundamental questions of what is better in medical sciences – quantity or quality? Surely, such trajectories might be interpreted as quality compromise, but it will be rather illogical. The increase in the quantity of published documents can be ascertained to the growing number of journals in recent times¹³ and greater access to open access articles¹⁴. The random fluctuation in percentage of international collaborations might be ascertained to lack of convergence of common research interest among investigators from different countries and constrained availability and accessibility of trained manpower and resources.

For each of the sub-disciplines of broader disciplines, the mean number of documents published along with its 95% CI has been also plotted (in Figure 5 a–d). In 'MED', for medicine (miscellaneous) the mean maximum stood followed by paediatrics, perinatology and child health. For all the remaining 45 fields, the mean was below 1000. The clinical specialities like neurology (clinical), surgery, radiology, ophthalmology, cardiology and cardiovascular medicine in addition to medicine (miscellaneous) and paediatrics, perinatology and child health had greater quantum of publications relative to obstetrics and gynaecology, gastroenterology, urology, psychiatry and mental health, endocrinology, diabetes and metabolism, nephrology, transplantation, geriatrics, epidemiology and physiology (Figure 5 a). In case of BGMB, the highest mean was for biochemistry and biotechnology (Figure 5 b). In case of IM, the maximum mean was for applied microbiology followed by microbiology, immunology, parasitology, virology and immunology and microbiology (Figure 5 c). For health profession, it was radiological and ultrasound technology preceded by health profession (miscellaneous) and complementary and manual therapy (Figure 5 d).

Broadly, the four disciplines among themselves are quite different entities and each category under them has

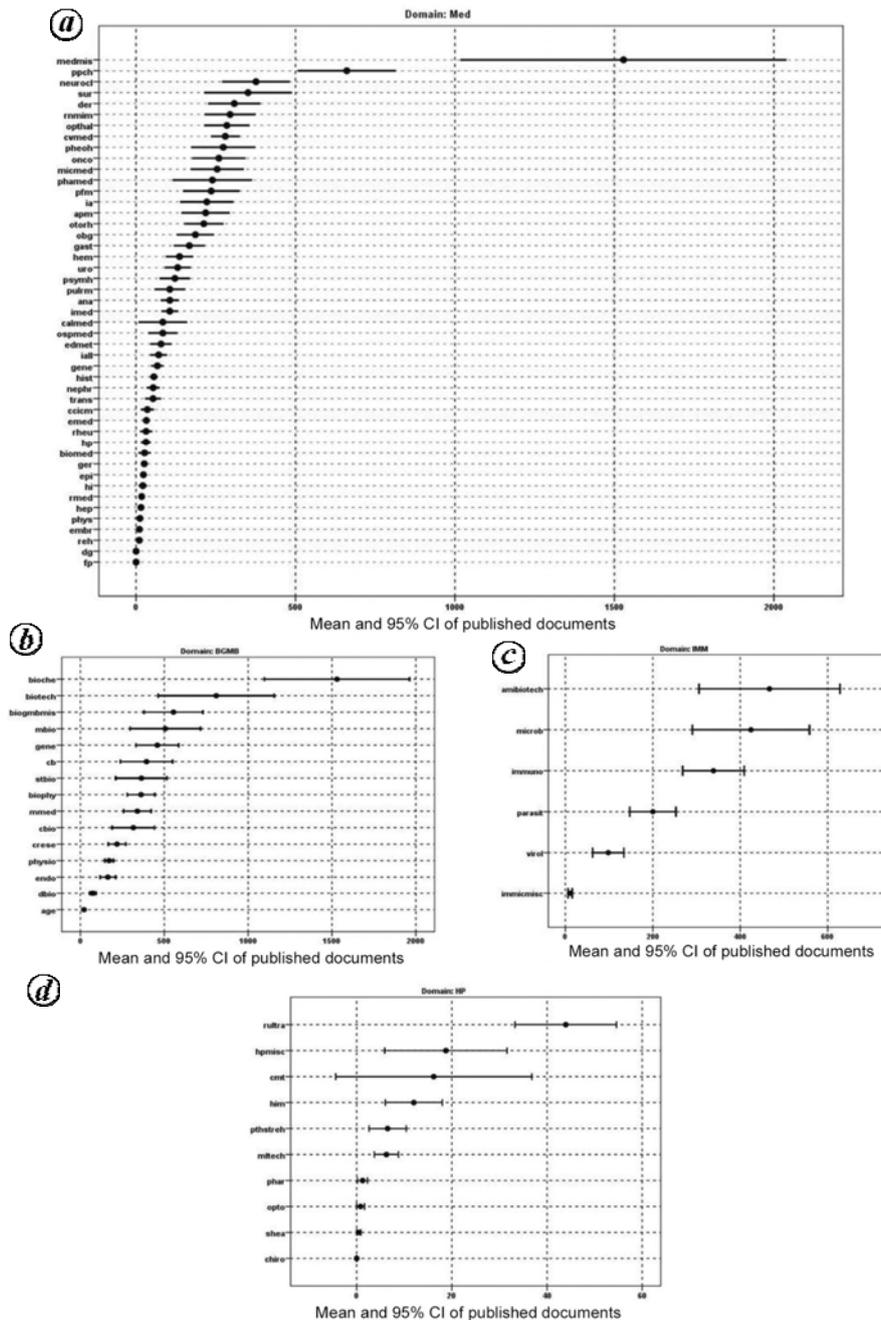


Figure 5 a-d. Error plots. *a*, For 47 categories under medicine. medmis: medicine (miscellaneous); ppch: pediatrics, perinatology and child health; neurocl: neurology (clinical); sur: surgery; der: dermatology; rnmim: radiology, nuclear medicine and imaging; opthal: ophthalmology; cvmed: cardiology and cardiovascular medicine; pheoh: public health, environmental and occupational health; onco: oncology; micmed: microbiology (medical); phamed: pharmacology (medical); pfm: pathology and forensic medicine; id: infectious diseases; apm: anaesthesiology and pain medicine; otorh: otorhinolaryngology; obg: obstetrics and gynecology; gast: gastroenterology; hem: haematology; uro: urology; psymh: psychiatry and mental health; pulrm: pulmonary and respiratory medicine; ana: anatomy; imed: internal medicine; calmed: complementary and alternative medicine; ospmed: orthopaedics and sports medicine; edmet: endocrinology, diabetes and metabolism; iall: immunology and allergy; gene: genetics (clinical); hist: histology; neph: nephrology; trans: transplantation; ccicm: critical care and intensive care medicine; emed: emergency medicine; rheu: rheumatology; hp: health policy; biomed: biochemistry (medical); ger: geriatrics and gerontology; epi: epidemiology; hi: health informatics; rmed: reproductive medicine; hep: hepatology; phys: physiology (medical); embr: embryology; reh: rehabilitation; dg: drug guides; fp: family practice. *b*, For 15 categories under biochemistry, genetics and molecular biology. Bioche: biochemistry; biotech: biotechnology; biogmbmis: biochemistry, genetics and molecular biology (miscellaneous); mbio: molecular biology; genet: genetics; cb: cell biology; stbio: structural biology; biophy: biophysics; mmed: molecular medicine; cbio: clinical biochemistry; crese: cancer research; physio: physiology; endo: endocrinology; dbio: developmental biology; age: aging. *c*, For six categories under immunology and microbiology. ambiotech: applied microbiology; microb: microbiology; immuno: immunology; parasit: parasitology; virol: virology; immimisc: immunology and microbiology (miscellaneous). *d*, For ten categories under health professions. rultra: radiological and ultrasound technology; hpmisc: health profession (miscellaneous); cmt: complementary and manual therapy; pthstreh: physical therapy, sports therapy and rehabilitation; mltech: medical laboratory technology; phar: pharmacy; opto: optometry; shea: speech and hearing; chiro: chiropractics.

unique individual features. Owing to the relative importance of paraclinical/subclinical or clinical/paramedical specialities, the between and within comparisons are not warranted. The greater contribution of medicine compared to the other three broad domains is accounted for by far greater specialities put under one category in SCImago¹⁰. The relatively lesser contribution of many important specialities in MED, like obstetrics and gynaecology, gastroenterology, urology; psychiatry and mental health, endocrinology, diabetes and metabolism, nephrology, transplantation, geriatrics, epidemiology and physiology demands attention.

India possesses a double burden of infectious as well as non-communicable diseases. In view of the changing population dynamics with a greater aged population¹⁵, prevailing epidemiological transition¹⁶, inequalities in life expectancy at birth¹⁷, non-appealing maternal mortality indicators⁵ and upsurge in behavioural emergencies¹⁸ the specialities of epidemiology, obstetrics and gynaecology, geriatrics, and psychiatry and mental health need to be more seriously pursued. These four specialities have been emphasized because they have public health implications. Many other disciplines with less promising trajectories as revealed can also be taken up subsequently.

The present study endeavours in drawing inferences from the scientometrics database of medical sciences through a statistical methodology. Obviously, the study considers a single database and simultaneous consideration of more databases is worth exploring. In addition, creation of databases with incorporation of wider spectrum of quality-related attributes might provide deeper insights into differential growth trajectories as observed. Such attempts help in generating newer hypotheses – what are the causes of differential growth trajectory? Is publication dependent on funding priorities or disease prevalence, and why are the regional and international collaborations not keeping pace with the volume of growth? The answers to these questions cannot be addressed in a single study. Certainly, more studies like the present one can generate newer insights among the scientific community in general and academicians, researchers and policy planners in particular. In a transient economy like India, dissemination of such information might go a long way for planning research activities more effectively.

5. Guin, G. *et al.*, Trends in maternal mortality and impact of Janani Suraksha Yojana (JSY) on maternal mortality ratio in a tertiary referral hospital. *J. Obstet. Gynaecol. India*, 2012, **62**, 307–311.
6. Lau, R. S., Johnson, S. and Kamalanabhan, T. J., Healthy life expectancy in the context of population health and ageing in India. *Asia Pac. J. Public Health*, 2012, **24**, 195–207.
7. Garg, K. C., Kumar, S. and Dutt, B., Scientometric profile of 'genetic and heredity' research in India. *Ann. Libr. Inf. Stud.*, 2010, **57**, 196–206.
8. Patra, S. W. and Chand, P., Biotechnology research profile of India. *Scientometric*, 2005, **63**, 583–597.
9. Gupta, B. M. and Bala, A., A scientometric analysis of Indian research output in medicine during 1999–2008. *J. Nat. Sci., Biol., Med.*, 2011, **2**, 87–100.
10. SCImago, SJR–SCImago Journal & Country Rank, 2007; <http://www.scimagojr.com>, accessed on 10 January 2013
11. Dandona, I., Raban, M. Z. and Guggilla, R. K., Trends of the public health research output from India during 2001–2008. *BMC Med.*, 2009, **7**, 59.
12. Department of Science and Technology, Bibliometric study of India's scientific publication outputs during 2001–10. New Delhi; www.dst.gov.in/whats_new/whats_new12/report.pdf.
13. Satyanarayana, K., Impact factor and other indices to assess science, scientists and scientific journals. *Indian J. Physiol. Pharmacol.*, 2010, **54**, 197–212.
14. Björk, B., Welling, P. and Laakso, M., Open access to the scientific journal literature: situation 2009. *PLOS ONE* 2010, **5**, e11273.
15. Agrawal, S., Deo, J. and Verma, A. K., Geriatric health: need to make it an essential element of primary health care. *Indian J. Public Health*, 2011, **55**, 25–29.
16. Dhillon, P. K., Jeemon, P. and Arora, N. K., Status of epidemiology in the WHO South-East Asia region: burden of disease, determinants of health and epidemiological research, workforce and training capacity. *Int. J. Epidemiol.*, 2012, **41**, 847–860.
17. Goli, S. and Arokiasamy, P., Trends in health and health inequalities among major states of India: assessing progress through convergence models. *Health Econ. Policy Law*, 2013, pp. 1–26; doi: 10.1017/S1744133113000042.
18. Saddichha, S. and Vibha, P., Behavioral emergencies in India: would psychiatric emergency services help? *Prehosp. Disaster Med.*, 2011, **26**, 65–70.

Received 16 July 2013; revised accepted 7 August 2013

1. Mowafi, H. A., Bibliometric analysis of the volume and visibility of Saudi publications in leading anesthesia journals. *Saudi J. Anaesth*, 2012, **6**, 393–397.
2. Horlings, E. and Gurney, T., Search strategies along the academic lifecycle. *Scientometrics*, 2013, **94**, 1137–1160.
3. Hardeman, S., Organization level research in scientometrics: a plea for an explicit pragmatic approach. *Scientometrics*, 2013, **94**, 1175–1194.
4. Madhusudhan, C. *et al.*, Palliative stenting for relief of dysphagia in patients with inoperable esophageal cancer: impact on quality of life. *Dis. Esophagus*, 2009, **22**, 331–336.