AJKD

Correspondence

RESEARCH LETTER

Predicting the Number of US Medical Graduates Entering Adult Nephrology Fellowships Using Search Term Analysis

To the Editor:

There has been a worrisome decrease in the number of applicants to nephrology fellowship training programs. First identified in 2008, the number of nephrology applications has been decreasing annually since 2002.[1] Given the annual increase in kidney disease and projected shortage of nephrologists, this downward trend is concerning. Various studies have suggested that a lack of interest in nephrology is a key factor in this decrease.[1-3] Parker et al[2] have used the percentage of adult nephrology positions filled by US medical graduates as a surrogate marker of nephrology interest, but this measurement is a late indicator. Investigations have suggested that the frequency of internet search term use can serve as a leading and accurate indicator of future events.[4,5] We hypothesize that the frequency of nephrology-specific search term queries can predict the percentage of US medical graduates entering nephrology fellowships.

The frequency of queries using the search term "nephrology" was analyzed from 2004 to mid-September 2011 using Google Insights for Search. The terms "gastroenterology," "oncology," "endocrinology," and "cardiology" were analyzed as controls. The frequency of each search term was restricted to results in the Health Education and Medical Training category. Frequencies were converted into yearly search term interest levels (STILs) by Google Insights for Search. STILs ranged from 0-100, with the latter indicating the greatest interest for the search term. A linear regression model was used to investigate the relationship between average yearly STILs and annual percentage of US medical graduates entering each fellowship listed in the Graduate Medical Education Track Database.[6]

We observed a decrease in yearly STILs for "nephrology," from 51 in 2004 to 31 in 2011, with the greatest monthly STIL occurring in June 2004. The linear regression model showed the following relationship between the STIL for "nephrology" at year t and fraction of fellowship positions filled by US medical graduates at year $t + 1$: Fraction of positions filled by USMGs$_{(t + 1)}$ = 0.00904 × STIL$_{(t)}$ + 0.08167, where USMG is US medical graduate.

This relationship showed a strong correlation during the study period ($r = 0.94$; $P < 0.01$). The regression model also showed a significant relationship for the control specialties (gastroenterology: $r = 0.89$; $P < 0.05$; oncology: $r = 0.84$; $P < 0.05$; endocrinology: $r = 0.81$; $P < 0.05$; and cardiology: $r = 0.83$; $P < 0.05$). Table 1 lists the predicted values of US medical graduates filling fellowships in 2011-2012 and 2012-2013. The model anticipates a continued decrease in US medical graduates entering adult nephrology fellowships.

We find that interest levels for the search term "nephrology" accurately predict the percentage of US medical graduates entering nephrology fellowship programs. The percentage of fellowships filled by US medical graduates is a meaningful but late measure of nephrology interest. An educator would have to wait until 2013 to learn the percentage of adult nephrology fellowship positions filled by US medical graduates who graduated in 2011. Our forecasting model can predict the percentage of fellowship positions filled by US medical graduates 2 years earlier by allowing educators to use STILs from the same calendar year. This reduction would allow more efficient resource allocation for recruitment, quicker assessments of interventions, and timely modifications to efforts designed to increase interest.

Table 1. Observed and Predicted Fractions of Adult Fellowship Positions Filled by US Medical Graduates

Application Year	Calendar Year for STIL	Search Term = Nephrology		Search Term = Gastroenterology		Search Term = Oncology		Search Term = Endocrinology		Search Term = Cardiology	
		Predicted[a]	Observed[b]	Predicted[a]	Observed[b]	Predicted[a]	Observed[b]	Predicted[a]	Observed[b]	Predicted[a]	Observed[b]
2005-2006	2004	0.51 (0.43-0.59)	0.53	0.70 (0.66-0.75)	0.70	0.54 (0.38-0.69)	0.59	0.61 (0.51-0.72)	0.61	0.66 (0.60-0.72)	0.65
2006-2007	2005	0.52 (0.44-0.59)	0.49	0.67 (0.63-0.71)	0.69	0.54 (0.38-0.70)	0.49	0.61 (0.50-0.71)	0.60	0.64 (0.59-0.70)	0.65
2007-2008	2006	0.42 (0.35-0.49)	0.44	0.67 (0.63-0.70)	0.65	0.44 (0.30-0.57)	0.41	0.55 (0.46-0.65)	0.59	0.63 (0.58-0.69)	0.65
2008-2009	2007	0.4 (0.32-0.47)	0.40	0.65 (0.61-0.69)	0.65	0.42 (0.27-0.56)	0.39	0.54 (0.44-0.64)	0.57	0.62 (0.57-0.67)	0.63
2009-2010	2008	0.39 (0.32-0.46)	0.40	0.65 (0.60-0.69)	0.64	0.41 (0.27-0.55)	0.41	0.52 (0.44-0.63)	0.52	0.61 (0.55-0.66)	0.60
2010-2011	2009	0.38 (0.31-0.46)	0.36	0.64 (0.60-0.69)	0.65	0.41 (0.25-0.55)	0.45	0.53 (0.43-0.63)	0.49	0.61 (0.55-0.66)	0.59
2011-2012	2010	0.38[c] (0.31-0.45)	—	0.64[c] (0.60-0.68)	—	0.40[c] (0.25-0.54)	—	0.52[c] (0.42-0.63)	—	0.60[c] (0.54-0.66)	—
2012-2013	2011	0.37[c,d] (0.30-0.44)	—	0.64[c,d] (0.60-0.68)	—	0.39[c,d] (0.24-0.54)	—	0.52[c,d] (0.42-0.63)	—	0.60[c,d] (0.54-0.65)	—

Abbreviation: STIL, search term interest level.
[a] Predicted fraction of positions filled by US medical graduates (95% confidence interval).
[b] Observed fraction of positions filled by US medical graduates.
[c] Predicted fraction of positions filled by US medical graduates that could not be compared with observed data.
[d] Prediction calculated from January to mid-September.

Search term frequency has been linked to future events in a variety of fields. Numerous financial studies have successfully shown a correlation between search term frequency and weekly jobless claims and investor attention.[7,8] In health care, search term frequencies have accurately predicted the rates of methicillin-resistant staphylococcal-induced hospitalizations and the percentage of influenza-like illnesses.[9,10] However, to our knowledge, such correlations have never been shown in the field of medical education.

Whereas our analysis focused on the term "nephrology," US medical graduates may have used alternative search terms. We attempted to address this limitation by filtering the analyses to the Health Education and Medical Training category, thereby analyzing web sites that provided information in education and/or training only. Although one must cautiously interpret the percentage of adult fellowships filled by US medical graduates as a marker of nephrology interest, the former is not subject to survey biases and is a logical outcome of interest in nephrology.

The percentage of US medical graduates filling adult nephrology positions is decreasing and serves as a late marker of interest in nephrology. Our model accurately forecasts US medical graduates entering fellowships from web-based search term analyses. Nephrology educators can use this model to assess the effectiveness of interventions designed to increase the percentage of fellowship positions filled by US medical graduates with a shorter delay than conventional means.

Tejas Desai, MD,[1] Maria Ferris, MD, PhD, MPH[2]
Cynthia Christiano, MD,[1] Xiangming Fang, PhD[1]
[1]East Carolina University Brody School of Medicine,
Greenville, North Carolina
[2]University of North Carolina, Chapel Hill, North Carolina
Corresponding author: desait@ecu.edu

References

1. Lane CA, Healy C, Ho MT, Pearson SA, Brown MA. How to attract a nephrology trainee: quantitative questionnaire results. *Nephrology*. 2008;13:116-123.

2. Parker MG, Ibrahim T, Shaffer R, Rosner MH, Molitoris BA. The future nephrology workforce: will there be one? *Clin J Am Soc Nephrol*. 2011;6:1501-1506.

3. Molitoris B. Nephrology workforce crisis: a wake-up call for all of us. *ASN Kidney News*. 2011;3:7.

4. Kein B. Google search patterns could track MRSA spread. Wired science. http://www.wired.com/wiredscience/2011/05/google-mrsa/. Accessed May 1, 2011.

5. Shimshoni Y, Efron N, Matias Y. On the predictability of search trends 2009. http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en/us/archive/google_trends_predictability.pdf. Accessed May 15, 2011.

6. GMETrack 2011. https://services.aamc.org/gme/admin/login/index.cfm. Accessed September 22, 2011.

7. Choi H, Varian H. Predicting initial claims for unemployment benefits [published online ahead of print July 5, 2009]. *J Finance*. http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en/us/archive/papers/initialclaimsUS.pdf. Accessed June 1, 2011.

8. O'Grady S. How Google can tell the Bank of England what to do next. Belfast Telegraph 2011. http://www.belfasttelegraph.co.uk/lifestyle/technology-gadgets/how-google-can-tell-the-bank-of-england-what-to-do-next-16011174.html. Accessed June 1, 2011.

9. Dukic VM, David MZ, Lauderdale DS. Internet queries and methicillin-resistant *Staphylococcus aureus* surveillance. *Emerg Infect Dis*. 2011;17:1068-1070.

10. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature*. 2009;457:1012-1015.