Too Much To Read

Kendrick Labs scientists have worked with a variety of clients since 1987 to help solve biomedical protein problems**.** To that end, we have become skilled at literature searches to determine scope of projects, and also to find trends — to see which research topics are robust. A philosophical question has lurked continually in the background: **What's the most important biomedical problem?**

The easy answer is cancer, no surprises there. However, examining publication trends reveals an ominous problem of a different nature, the Publication Explosion. It may not be the most important problem, but it's definitely serious, and ignored by many.

What's the evidence? Entering a single-year date range into PubMed without a keyword gives the total number of publications for that year. A plot of these values, papers/year versus year, is shown in Figure 1.

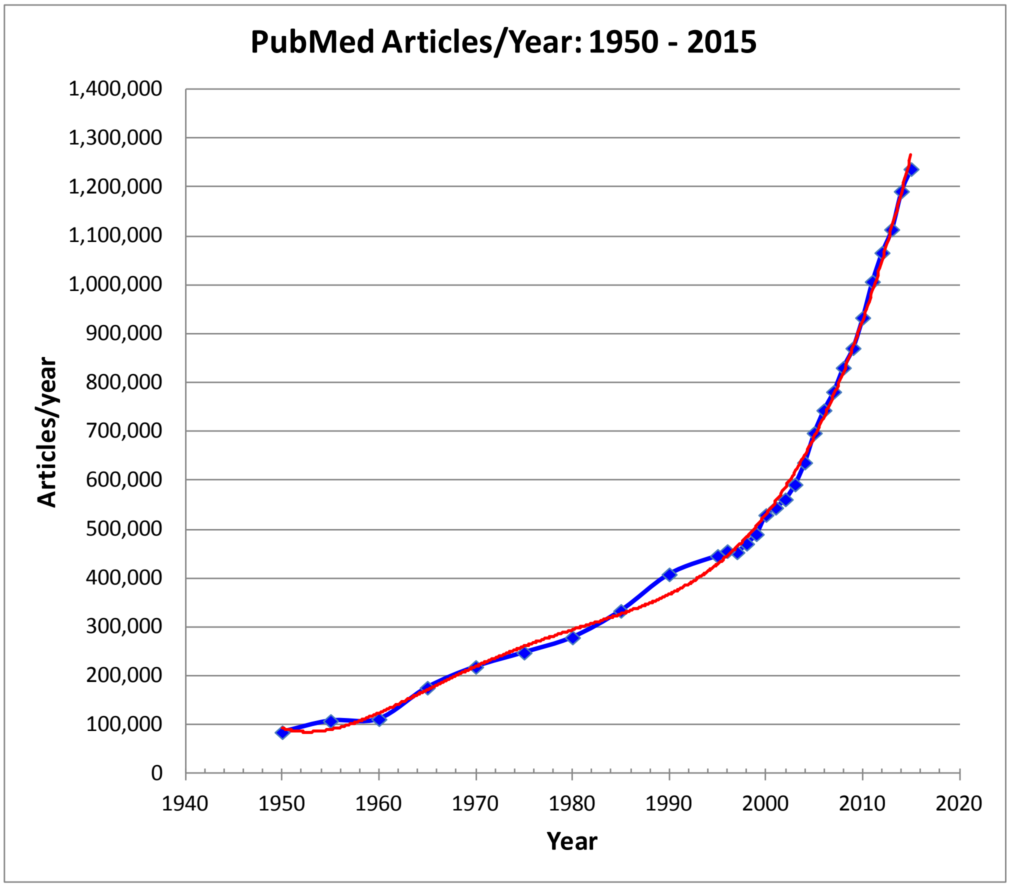

Figure 1. Publication Explosion. The blue line shows Papers per year listed on PubMed when the publication date is entered by year without keywords. The red Excel trend line is a 5th order polynomial.

Everyone knows that the number of papers published per year is large but, **holy smoke, look at the shape of that curve!** Over one million peer-reviewed papers were published in 2011 alone, and the number is rising like a rocket. What's going on — is this a bubble? When will the publication rate level off? The probable answers — not a bubble, not anytime soon — can be deduced by using various countries as keywords for "anywhere in text" with the HighWire search engine. Table 1 shows those results for two recent years.

Rank	"Anywhere in Text" **2009**	# papers	Percent
1	USA U.S.A. "United States"	285,488	32.4
2	UK "United Kingdom"	101,655	11.5
3	Germany	65,503	7.4
4	China	65,348	7.4
5	Japan	60,455	6.9
6	Canada	47,572	5.4
7	France	43,908	5.0
8	Sweden Denmark Norway Finland	41,172	4.7
9	Italy	38,689	4.4
10	Australia "New Zealand"	38,154	4.3
11	Netherlands	28,807	3.3
12	Spain	26,602	3.0
13	India	25,163	2.9
14	Brazil Argentina	22,665	2.6
15	Korea	20,978	2.4
16	Switzerland	19,711	2.2
17	Greece Portugal	13,982	1.6
18	Israel	13,493	1.5
19	Belgium	12,979	1.5
20	Singapore	11,648	1.3
21	Mexico	10,933	1.2
22	Ireland	9,936	1.1
23	Austria	9,493	1.1
24	Poland	8,941	1.0
	2009 Total: 881,973	1,023,275	116.0

Rank	"Anywhere in Text" **2012**	# papers	Percent
1	USA U.S.A. "United States"	315,692	28.4
2	UK "United Kingdom"	125,276	11.3
3	China	98,313	8.9
4	Germany	79,228	7.1
5	Japan	72,217	6.5
6	Canada	57,677	5.2
7	France	51,919	4.7
8	Sweden Denmark Norway Finland	51,298	4.6
9	Australia "New Zealand"	50,064	4.5
10	Italy	47,756	4.3
11	Netherlands	37,126	3.3
12	India	36,024	3.2
13	Spain	34,157	3.1
14	Korea	32,429	2.9
15	Brazil Argentina	30,584	2.8
16	Switzerland	24,114	2.2
17	Greece Portugal	17,262	1.6
18	Belgium	16,013	1.4
19	Israel	15,455	1.4
20	Mexico	12,577	1.1
21	Ireland	12,506	1.1
22	Singapore	12,452	1.1
23	Austria	12,316	1.1
24	Poland	11,061	1.0
	2012 Total: 1,110,037	1,253,516	112.9

Table 1. Number of biomedical publications/year ranked by country for the years 2009 and 2012. Results were obtained from Stanford's HighWire Search Engine by searching "any" of the country key words listed for the given year "anywhere in text" with "include PubMed" checked. Note that HighWire had a funding problem and stopped including PubMed publications on 1/1/13; 2013 and beyond cannot be surveyed. Presumably the summed percentages are over 100 because of collaborations between scientists in different countries.

Table 1 shows that in both 2009 and 2012 the US led in biomedical publications with ~32%and 28% respectively of the total, a lot but not enough to control output. China was ~9% in 2012 and rising. The remaining 63% of 2012 biomedical publications came from all over the world. The pressure to publish is apparently strong everywhere. Number of publications/ year is the universal way to judge productivity of individuals, departments, and universities. In addition, academic scientists are instilled with the idea that all results of all projects must be published, or else those efforts are wasted. There's no clear way to stop the publication explosion, even if anyone wanted to.

What are the consequences of the publication explosion?

Upside: A healthy amount of redundancy is present in the literature. As search engines improve it will become easier to find papers repeating the same experiments from different viewpoints without necessarily citing each other. Results that agree across several labs are robust and reliable. For example, a literature search for data on relationships between mRNA and protein showed a common result from several different labs (see our white paper.)

Downside: There's too much to read**.** PubMed 5-year searches of virtually any targeted keyword yield tens and sometimes hundreds of relevant papers. In a hot field, it can be much more. Searching the broad keyword "lung cancer" in PubMed Title/Abstract from 2011 thru 2015 brings up 37,929 papers. When "EGFR" is added to limit the search, *4020* papers containing both key words in Title/Abstract come up. Adding more keywords limits the number, but increases the risk of missing important papers. Keeping up with biomedical breakthroughs used to be possible by browsing top journals Nature, Science, or Lancet. Then Nature metamorphosed into 36 journals, Science into 4 journals and Lancet into 11 journals. It's too much.

The publication explosion is almost certainly interfering with the reliability of academic research and its conversion to practical applications. John Ioannidis' seminal paper, "Why Most Published Research Findings Are False," [1] and Prinz et al's, "Believe it or not: how much can we rely on published data on potential drug targets?" [2] indicate the negative consequences of a high ratio of writing to reading. The Economist (3) provides a salient review of the problem in the Oct 2013 issue, "How Science goes Wrong" with thought-provoking lead article, "Trouble at the Lab." This problem is so worrisome that NIH is fostering a reproducibility initiative.

What can be done? Well, there's the rub. The literature is a goldmine containing the lost treasures of the recent past, but the numerous small, related, gold nuggets are constantly being lost in an avalanche of papers. One remedy would be for academic scientists to slow down and read more before moving ahead. That's not possible though as everyone knows. Given the low NIH award rate , they must write ~six grants to get one funded, plus enough supporting papers to be competitive. Reading more is untenable.

Alberts et al (4) have written a surprisingly frank article "Rescuing US biomedical research from its systemic flaws" that describes how research has reached its current hypercompetitive state. The four distinguished authors present several recommendations for righting the capsized research ship. One recommendation is that career paths should be broadened for the more than 40,000 postdoctoral fellows in the US biomedical research system. That seems right to us. Perhaps the path can be smoothed for numerous small biotech startup companies that want to apply the wealth of academic knowledge to cure diseases.

Sadly though, scientists who leave academia to start or join a biotech business quickly lose library privileges. To gain access to high impact journals in the Nature, Science, and Lancet groups they must pay, for example, $32 each for Nature Genomics papers; $30 - Science Signaling; $31.50 - Lancet Oncology. This gets prohibitively expensive fast. One work-around is to favor open access journals like the PLoS group but in many cases that's not enough. An obvious solution is for university libraries to sell PubMed access permits to local biotech scientists for some reasonable amount, say $500/year. That money could in turn be used to fund new journal subscriptions. Unfortunately, that's not going to happen. Seems like this is an entropy problem. A prohibitive amount of energy is required to apply voluminous biomedical research information to a practical purpose.

References

1. Ioannidis, J.P., *Why most published research findings are false.* PLoS Med, 2005. **2**(8): p. e124.
2. Prinz, F., T. Schlange, and K. Asadullah, *Believe it or not: how much can we rely on published data on potential drug targets?* Nat Rev Drug Discov, 2011. **10**(9): p. 712.
3. The Economist, Oct 19, 2013 issue. How Science Goes Wrong. (Popular press, not PubMed)
4. Alberts, B., M.W. Kirschner, S. Tilghman, and H. Varmus, Rescuing US biomedical research from its systemic flaws. Proceedings of the National Academy of Sciences of the United States of America, 2014. 111(16): p. 5773-7.