

Is the Force Awakened? Publication Trends in Oncology Big Data as Phase II CancerLinQ is Launched

Hind Rafei^a, Benjamin Viernes^b, Angelike Liappis^a, Dalia A. Mobarek^c

^aDepartment of Medicine, the George Washington University, Washington, DC

^bInformatics, Decision-Enhancement and Analytic Sciences Center (IDEAS 2.0), Salt Lake City, UT

^cDepartment of Medicine, Division of Hematology and Medical Oncology, the George Washington University, Washington, DC

Background

- The **American Society of Clinical Oncology** launched **CancerLinQ** project in 2010.
- CancerLinQ provides **real-time data collection, mining and visualization, clinical decision support, and quality feedback.**
- Creation of a **big data software platform** is currently underway to power the **CancerLinQ** in the **phase II of the project.**
- This would allow for **evidence driven practice** and **rapid learning** for cancer care providers.
- Adequate knowledge** about the **utility of Big Data** to **encourage provider utilization** is needed.
- This is mainly achieved by **increasing the publication trend** in Big Data.
- It is especially important that **publications are in specialized journals to target the right audience.**
- It is also equally important to have an **increased amount of publications in high impact factor (IF) journals.**
- We aimed to **assess trends and quality of Big Data published in Oncology.**

Methods

- A **systematic search of PubMed® for English publications from 2011 to 2015** using **cancer and Big Data query** was conducted.
- Manual review of manuscripts** was performed in order to select appropriate articles that actually **discuss big data in the field of oncology.**
- Data collected included **publication type, study design, cancer subtype, publication year, journal category, sample size** if applicable, **location, VA vs non-VA, first and corresponding author names**, whether it is **funded or not**, having a **federal fund** if applicable, whether any author holds a **Bioinformatics degree**, the **software used**, the **journal name** and its **impact factor.**
- Statistical analysis** included **descriptive analysis** of findings in a **cohort design.**
- The **percentage of publications in each year** was calculated and a **trend of the number of publications** was drawn.
- Journals were categorized** between basic sciences and clinical.
- The **average impact factor of journals** from each year was calculated and the **trend of impact factor** was assessed.
- The **contribution of specialized journals** to publications was also assessed.
- The **US-based versus international contribution** was compared.
- Some of the **software used** were reported.
- The **Percentage of funded research** was calculated.

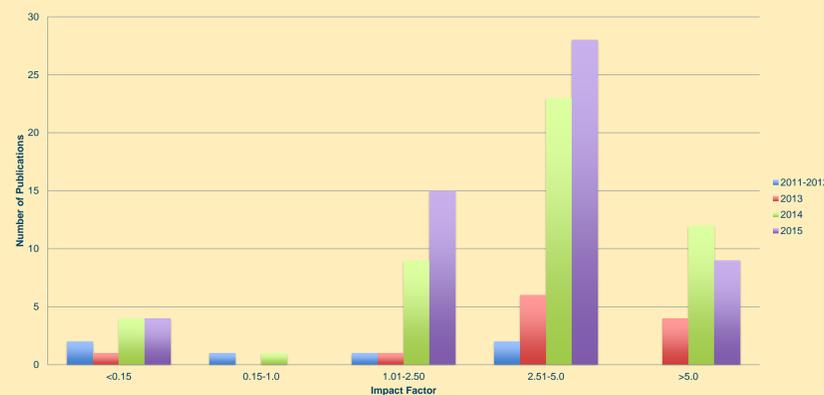


Figure1: Distribution of Publications Among the Different Years and Impact Factors

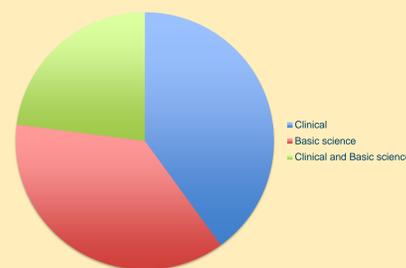


Figure2: Proportions of Publications Research Type

There is an equal distribution of publications among basic science and clinical journals with a small proportion being published in journals that are both basic science and clinically oriented.

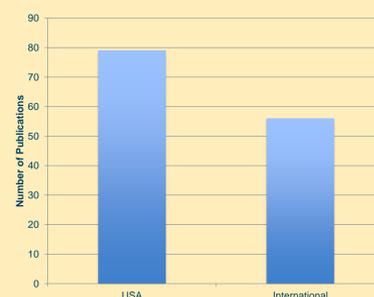


Figure3: Location of Publications

There are more publications based in the United States than in outside countries together.

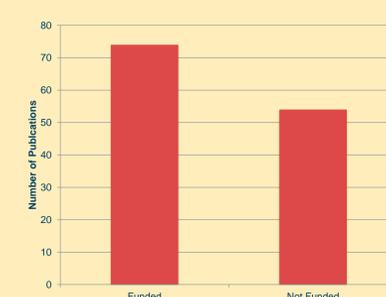


Figure4: Comparison of Funded and Non-Funded Research

More studies are funded worldwide than non-funded.

Results

- We identified **325 publications**
- 135 met inclusion criteria in 105 journals**, of which **36% (n=38) are considered specialized hematology and/or oncology journals.**
- Specialized journals published **29.62% (40/135).**
- Equal distribution of publications was found in clinical and basic science journals;** 54 (37%) and 50 (40%) respectively.
- There was a **trend of increased publications in clinical journals from 2012 to 2015 (16.7% to 42.9%, P = 0.39).**
- Of the available Impact factors (IF)—the **median is 3.234** (range 0.00-41.456).
- 25/125 (20.0%) of available IF is > 5.00** and **12/125 (9.6%) is >10.00** with **no difference in the proportion of IF > 5.00 in clinical versus basic science journals;** 11/51 (21%) versus 11/47 (23) % p = 1.00, respectively.

Impact Factor	2011-12	2013	2014	2015	Total
<0.15	2	1	4	4	11
0.15-1.0	1	0	1	0	2
1.01-2.50	1	1	9	15	26
2.51-5.0	2	6	23	28	59
>5.0	0	4	12	9	25
Total	6	12	49	56	123

Table1: Number of Yearly Publications in Each Impact Factor Category

The total number of publications is consistently increasing from one year to another since 2011. Moreover, there is an increasing number of publications belonging to a higher impact.

Conclusion

- The **need for further publication of studies** addressing Big Data use in furthering oncology research is **being met by the research community in response to the CancerLinQ** as demonstrated by the rapid increase in publications.
- We hypothesize that **this will increase the likelihood of cancer providers using CancerLinQ in the future**
- An **increase in publication in specialized journals and in those with high impact factors is still necessary.**
- Currently, despite the increased trend of publications addressing Big Data in oncology, **less than one-third of these publications are in specialized journals.**