Altmetrics in the biomedical literature: appropriateness of statistical methods
Valeria Scotti, Scientific Documentation Service
Annalisi De Silvestri, Clinical Epidemiology and Biometry Unit
Carmine Tinelli, Clinical Epidemiology and Biometry Unit
Moreno Curti, Scientific Documentation Service
Luigia Scudeller, Clinical Epidemiology and Biometry Unit
All Metrics Team
Fondazione IRCCS Policlinico San Matteo, Pavia, Italy

Background and aims
Alternative metrics have recently gained much attention in the biomedical community, with an increasing number of related publications. Formally, any new metric needs external validation against an accepted “gold-standard”; however, there is not a universally accepted method for validation. Also, like in any new field, the methodology applied in the early stages might be suboptimal. Besides, most alternative metrics tools are currently a sort of “black box” and the underlying algorithms calculating the scores are not made public. We aimed at investigating whether the statistical methods applied to validate the new metrics have been correctly applied in the development literature, with a focus on the biomedical community.

Methods
A systematic review of the literature was performed on 5 August 2015. The search was constructed and carried out by a librarian experienced on Alternative Metrics theme, using Pubmed, ArXiv e-prints archive (http://arxiv.org/) and DPLB database (http://dblp.uni-trier.de/). The following search strategy was used on PubMed: Altmetric*, Altmetrics OR “alternative metrics”. On Arxiv and DPLB only the general term Altmetrics was used, to find similar e-prints abstracts. No language or years restrictions were applied given the novelty of the topic. Eligibility criteria for full-text retrieval were focus on alternative metrics in the abstract or title with a quantitative approach. Selection was performed independently by two researchers (VS and ADS); if at least one researcher suggested full-text retrieval, it was retrieved.

From the selected papers, we extracted the following information: publication type, citation, publication year, scientific aims formally declared, main results presented in abstract, type of statistical methods, overall judgement on appropriateness of statistical methods (on a 0-10 scale, by an experienced statistician).

Results
From ArXiv we retrieved 200 records, from Pubmed 38, and from DPLB 13, for an overall total of 251. Of these, we excluded from screening at the abstract level 185 records focused on social media rather than on alternative metrics, and 24 that were narrative review articles or viewpoints.
We therefore retrieved 42 full text: the publication year of the first report was 2005, followed by a second publication in 2011, 2 in 2012, 6 in 2013 and 16 each in 2014 and 2015. Of the 42 full-text, 8 were abstracts only, 1 was a duplicate publication, 15 were not relevant, 8 were unpublished.

Therefore, only 10 (4.2% of the original sample, and 24% of the full-texts retrieved) were papers formally assessing validity of the new metrics (3 in 2013, 5 in 2014, and 2 in 2015) (see Figure 1). These were published by three different research groups (sometimes overlapping); 6 were published in journals with IF, 4 in journals without IF and 1 in ArXiv repository.

Specific research aims were specified in all 10. Comparison with traditional indicators was performed in 6, while comparison between different altmetrics was done in 4. Only in 4 papers data were highlighted in the abstract. The statistical methods applied were purely descriptive in 2, correlation methods in 5, regression methods in 2, dimensional methods in 3, and meta-analysis of correlation in 1 (NB: total does not add up to 10 because most papers used more than one method). The median quality score was 7.5 (IQR 7-8, min-max range 6-8).

Discussion

Most literature on alternative metrics is represented by introductory papers, descriptive works, or narrative reviews. The first report is 10-year old, and the number of publications increased with time.

In particular, ArXiv is heavily used by researchers who publish articles on altmetrics, while PubMed (which more easily reaches a biomedical audience) is less represented.

Original research on altmetrics represented only a minority of all publications, and was conducted by a small number of research groups, with specific interest in Altmetrics. The statistical methods were generally appropriated, though in 20% of papers only descriptive statistics were used. The overall quality of statistical methods was fair.
In conclusion, in order to support the uptake of altmetric indicators in the biomedical community, for research evaluation or other purposes, in the future the formal validation of alternative metrics need to be expanded. Inclusion of a statistician in any research groups is strongly suggested.