Annotated database and supplementary materials

For the article “A changing landscape in cardiovascular research publication output: bridging the translational gap” in the *Journal of the American College of Cardiology*.

By Gal D, Thijs B, Glänzel W, Sipido KR

March 2018, KU Leuven, Belgium

Table of Contents

Methods - expanded 2

Data sources 2

Journal classification and contribution 2

Flow of communication – definition and clustering methods 3

Statistical analysis and visualizations 3

Results - supplementary data 4

An expanding and diverse community in CV research - distribution of cardiovascular publications 4

Central journals with strong communication 7

Changing communities in CV research 8

Decreasing fragmentation of communities supporting translation 11

Discussion 12

Comparison to previously published bibliometric data on cardiovascular research 12

Limitations of the study 12

References 14

Methods - expanded

Data sources

The dataset was established using a hybrid information retrieval strategy, and the methods and steps undertaken to establish the dataset are as previously described (1, 2). Data retrieval aimed to include all publications (articles, reviews, letters or notes) on the cardiovascular system (heart, blood vessels or pericardium; including all types of research: basic/preclinical, clinical and epidemiological). The dataset includes the reference, abstract, address, and citation data for 804,152 cardiovascular publications published in 5984 separate journals between 1993 and 2013. We obtained the data from Clarivate Analytics Web of Science Core Collection (WoS) through a custom data license held by ECOOM, KU Leuven. We calculated and compared all indicators annually or using three 7-year time periods: 1993–1999, 2000–2006, and 2007–2013.

Journal classification and contribution

All journals in the WoS local database were previously reviewed and assigned unique identification numbers manually in order to combine all journals that changed names that separated or that merged over time.

The most active and most visible journals, which contributed the top 80% of all research output or citations by rank from 1993 to 2013, were included in the final dataset.

Each journal was classified as either an established journal, defined as a journal that was in place before 1993 and that consistently published cardiovascular research throughout the study period; or classified as a new/stopped journal, defined as a journal that was initiated (Volume 1) or was discontinued, or that started/stopped publishing cardiovascular research, during the study time period (1993-2013). The dates were also examined of when Web of Science starts or stops indexing publications, as this varies for each journal and may even have been interrupted within the study time period.

Each journal was also classified into one of five mutually exclusive journal types: Core Cardiovascular, Broader Biomedical Science, Multidisciplinary, Natural Science OR Applied and Social Science journals, based on the Leuven-Budapest journal classification scheme (3).

The number and share of publications per journal, by establishment and journal type were compared. Statistically significant changes in proportions were evaluated using chi-squared, p-value<0.05 unless otherwise stated.

The threshold to consider a journal with a large change in relative share over time was established using the characteristic scales and scores (CSS) method, including only the journal with a change in relative share higher than average in the top category. CSS is used instead of selecting the top 1% or 10%, change since the division between difference in relative share of journals in the top 1% or 10% is often not clear (e.g. journals with a change of 1.001% may have a very similar number of change in publications as those with a change of 0.99%, however they would be placed in two different categories, top 10% vs. top1%). CSS distribute the top change in journals over time into mathematically self-adjusting groups without

arbitrarily separating journals around a set percentage. Therefore, the division between CSS is made at percentages where there is a more clear separation between the change in share of publications (4).

Flow of communication – definition and clustering methods

The standardised flow of citations between individual journals was defined as per Equation 1 (5).

Equation 1. Strength of citation links $AtoB=(Citations\ AtoB)/((RefsA*CitesB)^{½})$

Where Citations AtoB= Citations from JournalA to JournalB;
RefsA=Total number of references* in cardiovascular publications by JournalA;
CitesB=Total number of citations received for cardiovascular publications by JournalB.

The references and the citations were only counted for those received/given up to two full years before and two years after publication, respectively.

We analysed both the complete network, and also undertook iterative testing of the citation links and citation link strength thresholds to reduce the network size to only include strong and very strong links. We then clustered the journals using the most appropriate algorithm available. We selected the networks to further compare based on the most optimal modularity, number of communities and size of communities. We named the journal communities based on the titles of the most informative journals in each community. We tested the statistical significance of the change in the communities of the complete journal network over time using the Infomap bootstrap networks method (6, 7). In addition, we tested the change in significance using manual resampling to create bootstrap networks and applied a standalone significance clustering (8). To disturb the networks we randomly selected 10% of journal citation link pairs in each time period and in scenario one we changed the weight of these links to the mean of all citation links and in scenario two we changed the weights to 0.00000001 in each time period and built 1000 disturbed networks for each scenario. We then applied the standalone significance clustering to these bootstrap networks.

Statistical analysis and visualizations

We made all calculations of indicators in Oracle SQL Developer version 4.0.1.14 and RStudio (9) version 1.0.143. The following R packages were used for calculation and data handling: igraph (10), reshape2 (11), tidyr (12), plyr (13) and dplyr (14). Graphs and network diagrams were produced using ggplot2 (15), gridExtra (16), and scales (17) R packages and the Infomap Map Equation (18). Journal names are depicted using standard abbreviations as per Table S2, available in Excel format at https://lirias.kuleuven.be/handle/123456789/607264 in the file listed as '2_GalD_annotated_database_extended_tables_2018.xlsx'.

Results - supplementary data

An expanding and diverse community in CV research - distribution of cardiovascular publications

Journal categories

The Table S1 below presents the numbers and categories for the summary data presented in Figure 1A in the article "A changing landscape in cardiovascular research publication output: bridging the translational gap" Gal et al., 2018,*Journal of the American College of Cardiology*. The data and further detail can be consulted in Excel format at https://lirias.kuleuven.be/handle/123456789/607264 in the file listed as '2_GalD_annotated_database_extended_tables_2018.xlsx'.

Table S1. Distribution of journals by establishment and journal field category (1993-2013)

Journal field category	Established journals*	Number of Journals	% Journals	Number of Publications	% Publications
Core Cardiovascular	N	47	8,3%	37 247	5,8%
Core Cardiovascular	Y	75	13,3%	278 980	43,2%
Multidisciplinary	N	1	0,2%	4 120	0,6%
Multidisciplinary	Y	5	0,9%	5 505	0,9%
Biomedical Sciences	N	74	13,1%	32 451	5,0%
Biomedical Sciences	Y	355	62,8%	284 983	44,1%
Natural Sciences	Y	5	0,9%	1 605	0,2%
Applied and Social Sciences	Y	3	0,5%	1 572	0,2%
Total		565	100,0%	646 463	100,0%

*Established journals defined as journals started before and consistently publishing cardiovascular research throughout the study period. If N then journal started (Volume 1) or stopped as a journal, or publishing cardiovascular research, in the study time period (1993-2013).

Evolution in journals contributing to cardiovascular research

Of the 565 journals, one-fifth (n=122) either started or stopped publishing cardiovascular research during the study period and time indexed. A full overview of these data by journal can be found in **Table S2. Distribution of publications by journal and time period** (available in Excel format at https://lirias.kuleuven.be/handle/123456789/607264 in the file listed as '2_GalD_annotated_database_extended_tables_2018.xlsx'.)

Journals that started to publish cardiovascular research included 104 journals that were newly launched during the study period and nine journals that started publishing cardiovascular research at least one year after initiation of the journal. Two longstanding journals *Molecular Biology Reports* and *AIDS* published their first cardiovascular publications in this dataset in 2002 and 1997, ten and four years after the beginning of study, respectively. In contrast, only six journals were discontinued before the end of the study and seven additional journals stopped publishing cardiovascular research at least one year before the end of the study. Less than 2% of the journals have left cardiovascular research publishing over time, though several journals changed their names to reflect a more international focus, for example the *British Journal o*f… , or the *Japanese Journal of*.., changing to the *International Journal of*… . Such changes reflect the globalization and expansion of cardiovascular research.

Evolution of participation of individual journals

Figure S1 presents the share of publications per journal in the cardiovascular dataset. Please note that only journals with at least 0.5% share of publications in 2007-2013 are included in the figure.

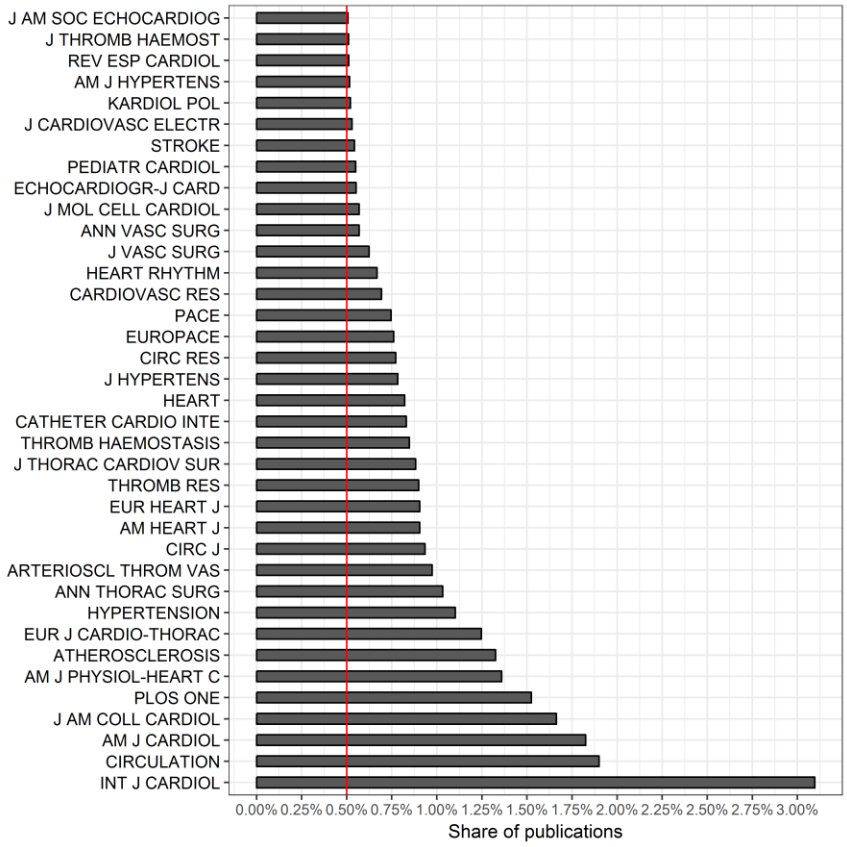


Figure S1. Distribution of publications by journal 2007-2013 (Data sourced from Clarivate Analytics Web of Science Core Collection)

As discussed in the article, we further examined in detail the changes in the publication of cardiovascular papers in leading broader biomedical journals, illustrated here.

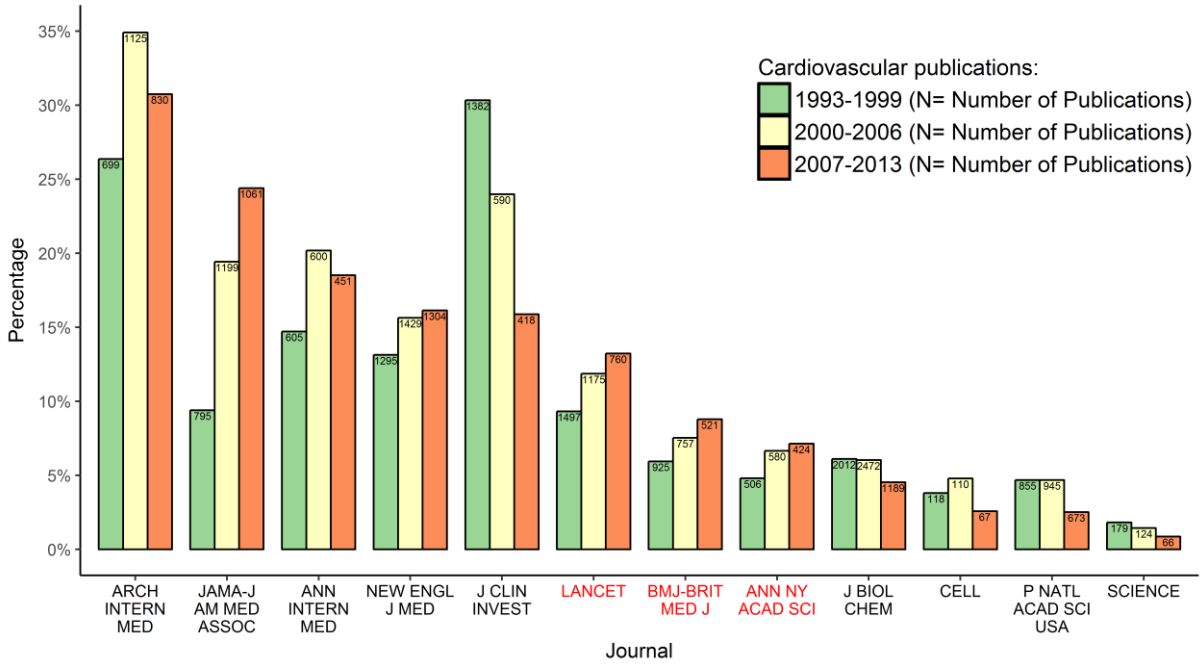


Figure S2. Cardiovascular research in broader biomedical/multi-disciplinary journals. The vertical axis presents the percentage of cardiovascular publications in the journal's publications; the number in the column presents the number of cardiovascular publications. Journal titles highlighted in red present journals that increased their share of cardiovascular publications while decreasing their total publications over time. (Data source:Clarivate Analytics WoS.

Central journals with strong communication

When considering the communication and transfer of cardiovascular research knowledge between all journals, the total number of observed citation links to the cardiovascular publications grew two and a half times between the 1993-1999 and 2007-2013 time periods, while the number of journals citing cardiovascular research grew 1.7 times (**Table S3. Communication flow of cardiovascular research between journals** - available in Excel format at https://lirias.kuleuven.be/handle/123456789/607264 in the file listed as '2_GalD_annotated_database_extended_tables_2018.xlsx'). However, these journal citation links were extensively spread out, with half of the links representing only one or two citations between individual journals during a seven year time period, and with only 13% of links between journals having a greater than average number of citations (6-8 citations on average). We found that the citations received came from a wide range of journals (Table S3). Half of all journals were cited at least once by more than 298 journals in 1993-1999, increasing to 505 journals in 2007-2013. At the extreme, the top 25 journals in 1993-1999 were cited by over 1000 journals, with 16 of these top 25 journals remaining in the top 25 cited journals in 2007-2013.

The central journals with the highest number of strong links to other journals (**Table S4. Central journals with strong communication of cardiovascular research** - available in Excel format at https://lirias.kuleuven.be/handle/123456789/607264 in the file listed as '2_GalD_annotated_database_extended_tables_2018.xlsx') tend to be the journals with the highest number of citations received, with *Circulation* and the Journal of the *American College of Cardiology* leading. However, the strength is a normalized indicator, accounting for the number of publications and number of citations and thus truly indicates the relative strength between both journals. The network therefore brings information on impact of a journal beyond and above other citation-based metrics such as the journal Impact Factor (IF), as illustrated e.g. by the position in Figure 2 of the *American Journal of Physiology-Heart and Circulatory Physiology* as an important hub of communication.

Changing communities in CV research

The flow of communication through citations can identify networks and communities as presented in Figure 2 in the main article.

Visualization map and underlying data

The complete map data presented in Figure 2 in the main manuscript are available by time period in the online annotated database files at https://lirias.kuleuven.be/handle/123456789/607264 , including files GalD_Journal_communities_1993_to_1999.map (data Figure 2A), GalD_Journal_communities_2000_to_2006.map (data not illustrated) and GalD_Journal_communities_2007_to_2013.map (data Figure 2B).

To explore the data interactively online one needs to access the MapEquation program. To view the community level maps you will need to download supplementary files labeled *.map to your computer, and then go to http://www.mapequation.org/apps/MapGenerator.html. Load one map file at a time by clicking on 'load map' and select 'open file' to locate and select one file from your computer, such as file GalD_Journal_communities_1993_to_1999.map, to view the journal communities from 1993-1999.

Once the map is open in the box on the left hand size, place your mouse cursor over the community circle to view the communication flow values. By double clicking on each community circle, a list of the journals that is included in the community will pop up in a new box. You can also increase the size of the map by clicking and dragging the bottom right corner of the map box. Then to add a second time period map click on the 'add network' button and repeat above with a new file. When more than one map is uploaded you can view the alluvial plot on the far right hand side of the screen by scrolling the bar at the bottom of the screen.

It is possible to view the more detailed directed links between all journals in each community shown in panels C-F of Figure 2, using another feature of Mapequation called the network navigator http://www.mapequation.org/apps/NetworkNavigator.html. The files to be

uploaded have the extension *.ftree and are also located on the annotated database webpage at https://lirias.kuleuven.be/handle/123456789/607264. Download all files to your computer then load one file at a time by clicking on 'load map' and select 'open file' to locate and select one file from your computer, such as file GalD_crossreferences_1993_to_1999.ftree, to view the journal communities from 1993-1999. At this level, the communities are to be identified by the leading journal within the community, i.e. the one with the highest communication flow. In this particular example, the clinical group is called 'Circulation' (data underlying Figure 3C) and the basic/preclinical groups is called 'J Biol Chem' (data underlying Figure 3E). The names further down the list are the groups in the circle of Figure 2A, not illustrated in detail in the main manuscript. Once the community level map is open in the network view, click on the name of one journal community on the left hand side to view the list of journals in this community, as well as, to see the most influential journals in the network view box. To adjust the number of journals and links viewed, click on the 'advanced controls' button  in the top right hand side of the network view box. Then when the advanced settings panel opens on the right, tick the box in front of the 'modules' control and then increase the number of journals(modules) and links to be viewed. It is best to slowly increase the values (and not to exceed 300 journals and 1000 links) as it takes time to view a large number of journals and links. The network view box can also be made larger by clicking and dragging the bottom right corner of the network view box.

The authors are available for further guidance and tutorial.

It is also possible to consult the data in Excel format. **Table S5. Cardiovascular Journal Communities and Communication Flow** is available in Excel format at https://lirias.kuleuven.be/handle/123456789/607264 in the file listed as '2_GalD_annotated_database_extended_tables_2018.xlsx'.

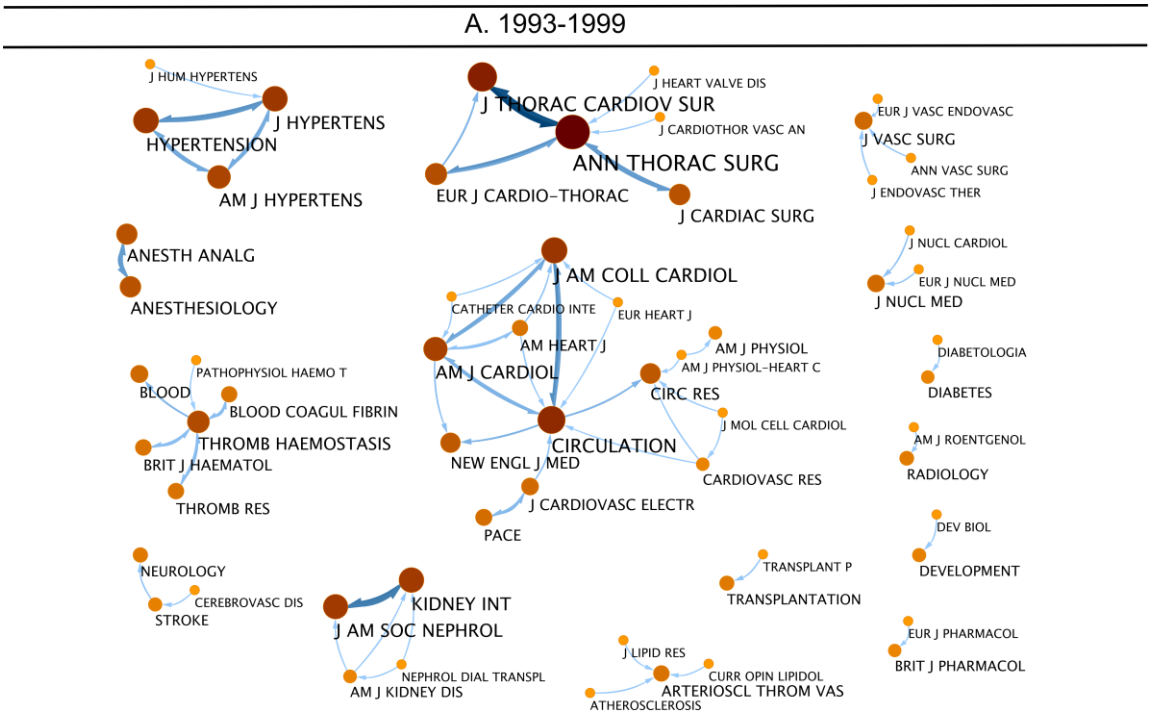
Strong & very strong citation links

We also examined the network structures and citation flow based only on strong and very strong intercommunication between journals. In the first time period few differences were noted between the clinical communities identified when using all citation links compared to only using the strong citation links between journals (**Table S6. Strong citation links between all journals -** available in Excel format at https://lirias.kuleuven.be/handle/123456789/607264 in the file listed as '2_GalD_annotated_database_extended_tables_2018.xlsx'.)

However, the general basic/preclinical journal community divides into six separate communities. In contrast, in the most recent time period, the basic/preclinical community only divides into two communities, while the number of clinical communities with more than one journal doubles.

When only using the very strong links between journals, we found fewer links between journal communities, along with missing communities because they did not contain any very strong links (**Figure S3**). Overall, the strongest links in each community were predominantly

between or with American journals, or with a non-cardiovascular top journal in each discipline. In 2007-2013 some subspecialty communities are visible: imaging, thoracic cardiac surgery, vascular surgery, interventional; cardiology and arrhythmias.



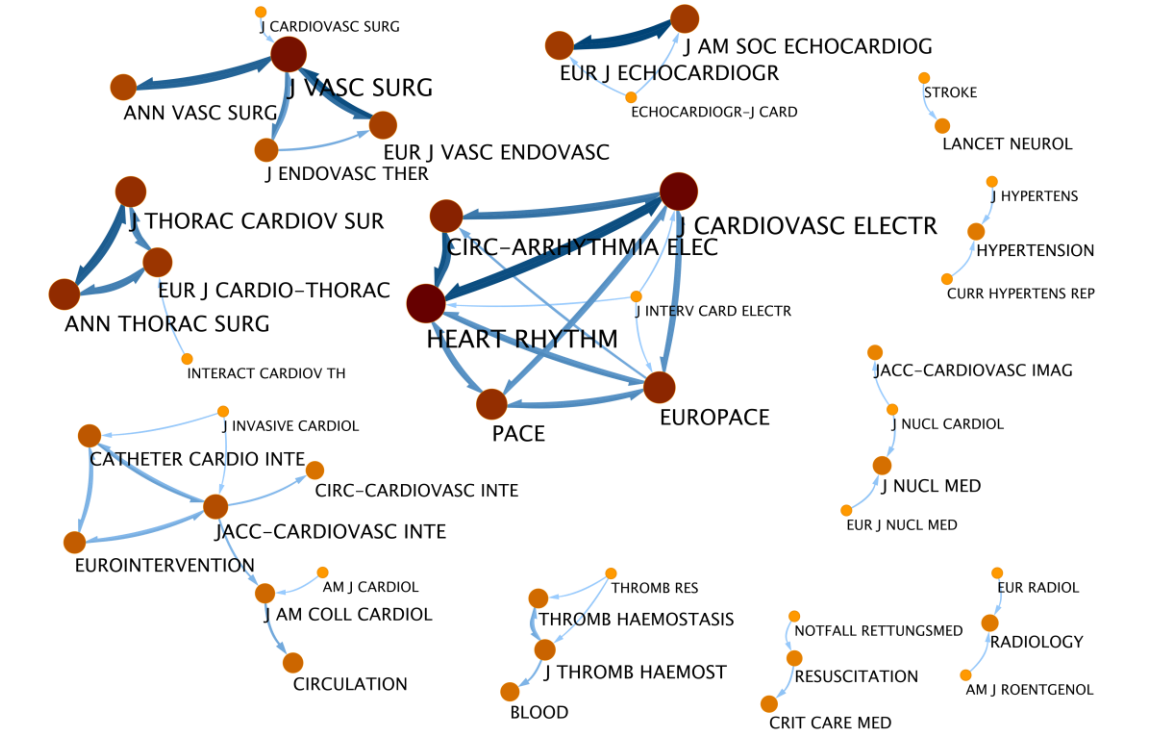


Figure S3. Journals with VERY strong inter-communication
Comparison of the flow of communication of all very strong standardized citation links between all journals in the first and last time period. Arrows point in the direction of the flow of citations- from citing to cited journals. The lines represent the strength of the information flow between journals increasing from light grey to dark blue. The circles represent the share of information flow of each journal relative to others increasing from light orange to dark red. The circles are positioned based on the Force Kamada-Kawai algorithm ensuring the length of the lines are homogenous and that the lines are crossed as little as possible. (Data sourced from Clarivate Analytics Web of Science Core Collection)

Changes in output from the different communities

Underlying the change in size of the communities can be a change in the output from the different communities. The data shown in Table S6 below illustrate the different growth rate in the communities.

Table S7. Growth in journals and publications in clinical and basic/preclinical cardiovascular journal communities

	Journals publishing cardiovascular research			Manuscripts in cardiovascular journals		
	Basic/preclinical	Clinical	Total	Basic/preclinical	Clinical	Total
1993-1999	135	115	250	43569	62318	105887
2007-2013	147	177	324	53990	115295	169285
Increase over time	12	62	74	10421	52977	63398

Decreasing fragmentation of communities supporting translation

We further assessed the stability of the networks over time and examined whether the allocation of journals in each journal community was statistically significant or not using re-iterative testing and bootstrap methods. The alluvial plot resulting from this analysis indicates decreased fragmentation of flow of communication between clinical and basic/preclinical journals over time as presented in Figure 3 in the article.

To further test the robustness of the infomap mapequation community structures we also re-ran the network algorithm based on two scenarios: 1) including all citation links to the clinical and basic/preclinical journals only, 2) only including the citation links to and from the clinical and basic/preclinical journals.

Overall these did not affect all network/community structures, the large basic/preclinical and clinical communities were similar, with a smaller number of clinical-related communities not

identified earlier. However in scenario 1) the clinical and basic/preclinical communities merged into one community in the third time period..

As discussed in the article, one possible reason for decreased fragmentation is the changing culture in referencing with increasing cross-referencing and consequent increase in number of references in publications. This evolution is illustrated in **Table S8. Growth in number of references in cardiovascular publications** - available in Excel format at https://lirias.kuleuven.be/handle/123456789/607264 in the file listed as '2_GalD_annotated_database_extended_tables_2018.xlsx'.

Discussion

Comparison to previously published bibliometric data on cardiovascular research

Few publications have focused on the evolution of cardiovascular research and when compared to this study(19, 20), the two publications to note included more limited datasets—of shorter duration, and with limited coverage of journals. Along with differences in the datasets reported, different bibliometric strategies and methods were applied in both studies.

Baldridge et al presented the change over time in cardiovascular output in five general and three specialty journals and found that the number and share of cardiovascular publications from 1999 to 2008 only increased in the European Heart Journal. Whereas, the publication volume remained steady for the Journal of the American College of Cardiology (JACC) and decreased for Circulation, Annals of Internal Medicine, BMJ, Circulation, JAMA, Lancet and NEJM. (19) With the inclusion of data from 2007 to 2013 in our study, we observed different trends with a generalized decrease in the number of cardiovascular publications in all journals, except for in JACC. In addition, the relative share of cardiovascular publications in the non-disciplinary journals increased over time in all except for the Annals of Internal Medicine.

In the article by Jones et al, network analysis was used to characterize the citations from the articles published in 75 top cardiovascular journals in 2009 (20). By applying journal level assignment of types of research, including basic, clinical, clinical-mix and clinical observation, they presented the journals in three main groups based on the citation links between journals. Their main result is in line with our data, showing evidence of translation through overlap and strong inter-connections between basic and clinical research.

Limitations of the study

This study is presenting data from the 565 journals that have the highest output or citations to cardiovascular publications over the study period, however, the journals not included that publish less cardiovascular research and are cited less frequently may be producing research that is of equal or greater value to health and society.

The present study comes too early to evaluate the impact of open access initiatives on growth in cardiovascular publications. *PLOS ONE* was founded in 2006 as part of the 2002 PLOS initiative as an independent non-profit organization for faster, easier and open access to science. With a different editorial evaluation policy based on data accuracy, rather than novelty, *PLOS ONE* has accepted many papers and taken a dominant place in the PubMed database. Other publishers have followed suit with *Scientific Reports* launched by the Nature Publishing Group in 2011, possibly reaching a similar volume as *PLOS ONE* in 2017. Other journals created open access online-only journals in 'second-tier' to their prime journals, such as the *Journal of the American Heart Association* (2012) and *JCI Insights* (2016).

The limitations and variability of clustering items using network algorithms are well described (21, 22). Since we included journals that published a relatively small number of cardiovascular publications in each time period, it meant that many journals had low or no information flow and so tended to cluster around the main journals they linked to in each community. This means small changes in the underlying sub-structures are difficult to detect. However, the Infomap Mapequation is a robust method to show sub-structures because of the method used to cluster items, presenting the data in the best fitting hierarchical network (22). The additional possibility to test and visualise the significance of any changes in the clustering is also a strength of the method (6).

References

1. Gal D, Sipido KR, Glänzel W. Using Bibliometrics-aided Retrieval to Delineate the Field of Cardiovascular Research. Proceedings of ISSI 2015 Istanbul 2015. Available at: http://issi2015.org/en/Proceedings-of-ISSI-2015.html.

2. Gal D, Glänzel W, Sipido KR. Mapping cross-border collaboration and communication in cardiovascular research from 1992 to 2012. Eur Heart J 2017;38:1249–1258.

3. Glänzel W, Schubert A. A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. Scientometrics 2003;56:357–367.

4. Glänzel W, Thijs B, Debackere K. The application of citation-based performance classes to the disciplinary and multidisciplinary assessment in national comparison and institutional research assessment. Scientometrics 2014;101:939–952.

5. Zhang L, Janssens F, Liang L, Glänzel W. Journal cross-citation analysis for validation and improvement of journal-based subject classification in bibliometric research. Scientometrics 2010;82:687–706.

6. Rosvall M, Bergstrom CT. Mapping Change in Large Networks Rapallo F, editor. PLOS ONE 2010;5:e8694.

7. Mirshahvalad A, Beauchesne OH, Archambault É, Rosvall M. Resampling Effects on Significance Analysis of Network Clustering and Ranking Lambiotte R, editor. PLOS ONE 2013;8:e53943.

8. Rosvall M. Standalone significance clustering. MapEquation project Significance clustering 2017. Available at: https://bitbucket.org/mapequation/significance-clustering/src. Accessed October 5, 2017.

9. RStudio Team. RStudio: Integrated Development Environment for R. Boston, MA: RStudio, Inc.; 2015. Available at: http://www.rstudio.com/.

10. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal 2006;Complex Systems:1695.

11. Wickham H. Reshaping Data with the reshape Package. J Stat Softw 2007;21:1–20.

12. Wickham H. tidyr: Easily Tidy Data with `spread()` and `gather()` Functions. 2016. Available at: https://CRAN.R-project.org/package=tidyr.

13. Wickham H. The Split-Apply-Combine Strategy for Data Analysis. J Stat Softw 2011;40:1–29.

14. Wickham H, Francois R. dplyr: A Grammar of Data Manipulation. 2015. Available at: https://CRAN.R-project.org/package=dplyr.

15. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2009. Available at: http://had.co.nz/ggplot2/book.

16. Auguie B. gridExtra: Miscellaneous Functions for “Grid” Graphics. 2015. Available at: https://CRAN.R-project.org/package=gridExtra.

17. Wickham H. scales: Scale Functions for Visualization. 2015. Available at: https://CRAN.R-project.org/package=scales.

18. Bohlin L, Edler D, Lancichinetti A, Rosvall M. Community Detection and Visualization of Networks with the Map Equation Framework. In: Ding Y, Rousseau R, Wolfram D, editors. Measuring Scholarly Impact. Cham: Springer International Publishing, 2014:3–34. Available at: http://link.springer.com/10.1007/978-3-319-10377-8_1. Accessed January 27, 2017.

19. Baldridge AS, Huffman MD, Bloomfield GS, Prabhakaran D. Footprint and imprint: an ecologic time-trend analysis of cardiovascular publications in general and specialty journals. Global Heart 2014;9:263–269.e2.

20. Jones DS, Cambrosio A, Mogoutov A. Detection and characterization of translational research in cancer and cardiovascular medicine. J Transl Med 2011;9:57.

21. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech-Theory E 2008;2008:P10008.

22. Rosvall M, Bergstrom CT. Multilevel Compression of Random Walks on Networks Reveals Hierarchical Organization in Large Integrated Systems Rapallo F, editor. PLOS ONE 2011;6:e18209.