Title

How industry influences collaborative research communities: a network analysis

Authors

[1]Adam G. Dunn, PhD

[1]Blanca Gallego, PhD

[1]Enrico Coiera, PhD, MB BS

Affiliations

[1]Centre for Health Informatics, Australian Institute of Health Innovation, University of New South Wales, Sydney, NSW, 2052 Australia

Corresponding author information

Adam Dunn, Centre for Health Informatics, University of New South Wales, SYDNEY NSW 2052, Australia. Telephone: +61(2) 9385 8699, Fax: +61(2) 9385 8692, Email: a.dunn@unsw.edu.au.

Funding Information

This work was supported by National Health and Medical Research Council (NHMRC) Program Grant 568612. The funding body did not have a role in the research. No authors have any competing interests. An ethics statement was not required for this work.

Article information

Word count: 2908 in the manuscript text, 197 in the abstract.

Pages: 23

Figures: 3

Tables: 3

Objective: To measure the relative influence that industry authors have within collaborative research communities.

Study Design and Setting: Using 22 commonly-prescribed drugs, 6711 randomised controlled trials (RCTs) and 28104 authors, 22 collaboration networks were constructed and analysed. The directly industry-affiliated authors (DIA) were identified in the networks according to their published affiliations. Measures of influence (network centrality) and impact (citations) were determined for every author. Network-level measures characterising community structure and collaborative preference were used to indicate further characterise the groups.

Results: Six percent (1741 of 28104) of authors listed a direct affiliation with the manufacturer of a drug evaluated in the RCT. These authors received significantly more citations ($p<0.05$ in 19 networks) and were significantly more central in the networks ($p<0.05$ in 20 networks). The networks show that DIA authors tend to have greater reach in the networks, and collaborate more often with non-DIA authors despite a preference towards their own group. Potential confounders include publication bias, trial sizes and conclusions.

Conclusions: Industry-based authors are more central in their networks and are deeply embedded within highly-connected drug research communities. As a consequence, they have the potential to influence information flow in the production of evidence.

Keywords: Randomised controlled trials; Research translation; Collaboration; Network analysis; Community Structure.

Running title: Networks of research collaboration and influence

Manuscript word count: 2901

Introduction

Industry and non-industry funded trials contribute differently to the evidence bases for pharmacological interventions. Industry-funded trials that reach publication are less likely to include active comparisons, are of different sizes and more often report outcomes that favour the intervention being tested (1-3). Publications also differ in terms of how closely the conclusions reflect the results (4). However, industry influence is not binary, and authors from across the spectrum of industry influence interact via direct collaborations in trials and trial publications. In the present study, patterns of co-authorship in the production of evidence are examined for a sample of common drugs. The study involves comparing directly industry affiliated (DIA) authors to other authors to quantify their relative positions in the research community, and to determine if they have the potential to disproportionately influence the production of an evidence base.

Network analyses exploit a set of relatively recently-devised metircs that quantify the structure of relationships between large numbers of interacting entities e.g. proteins, people or network switches (5-7). Large decentralised networks such as protein networks, social networks and the Internet are characterised by non-linear relationships in which some entities and some connections are more important than others. The overall structure of a network may be random, hierarchical, scale-free or jellyfish-like (8-10), and the structure may be used to suggest properties of the system such as efficiency or vulnerability (11-13), or be used to predict behaviour (14). Network metrics that quantify differences between individual entities within a network often indicate the relative importance of one entity in comparison to others. Both network-level and individual-level metrics are used in the present study to determine if the structure of research collaboration networks and the position of DIA authors in that structure allows for a disproportionate level of influence in the construction of an evidence base for approved drugs.

Collaboration networks may be used to reveal the ways in which researchers cooperate and assemble collaborative groups (15-18). These networks are constructed by linking authors that publish articles together, creating a single, measurable view of a research community. In a network, an individual's

centrality measures the influence associated with an individual's position within a network (19, 20). *Betweenness centrality* is a specific measure of the power that an individual has in shaping information flow in a network (5, 21, 22). Betweenness centrality is a common method for measuring influence in networks of co-authorship (23).

This study is an examination of collaboration networks associated with randomised controlled trials (RCTs) for a cross-section of drugs that are widely prescribed in primary care. The study involved measuring (i) the direct affiliations of authors listed in each publication (but not their financial disclosures); (ii) the centrality of authors in the collaboration networks for each of the 22 drugs; (iii) the number of citations received by each of the authors aggregated across their publications; and (iv) the structure and composition of the 22 communities. The aim of the study was to measure the influence of directly industry-affiliated (DIA) authors in comparison to their counterparts in these networks, and examine how differences between the two groups may benefit or hinder the construction of a trustworthy evidence base.

Methods

The sample comprised 22 drugs that are widely prescribed and recently developed (95% or more of the RCTs were published since 1996). For each drug, the set of published RCTs were retrieved (PubMed via Scopus, and using EMBASE synonyms). For each published RCT, the number of citations, authors and their direct affiliations were then extracted by automated analysis of the text. Further details about the selection process and the collection of publications are given in the Supplementary Material. Since the vast majority of RCTs are published after a medicine has been approved for marketing, these publications generally represent the accumulation of clinical evidence during the post-marketing phase. The total number of publications was 6711 and involved 28104 authors.

Authors

The set of authors was extracted from the articles associated with each drug, using up to two initials and a family name as is common in existing studies of co-authorship networks (16, 23). An author is considered to be directly affiliated with industry if that author lists an affiliation matching the name of the company producing or marketing the medicine in at least one publication about that medicine. This approach is more conservative than previous studies investigating industry support using manual searches (24-28), because it does not include any further manual investigation of financial disclosures. Indeed, the overall proportion of DIA authors in the set of published RCTs is 6.19%, so the group of authors under scrutiny represents the most certain end of the industry spectrum.

The automated classification of DIA authors was confirmed to be accurate by comparing the automated classification with results of two blinded manual searches over a random sample of 200 authors ($\kappa = 0.936$ and $\kappa = 1.00$, see Supplementary Material for details). An author's individual citation count is given by the sum of citations (all citations collected by Scopus to December 2010) for each RCT article to which the author contributed.

Collaboration networks

The aggregated set of authors contributing to the RCT literature of an individual drug was used to generate the network of co-authorship and thus defines the community of researchers involved in the production of evidence for a given drug. The networks were constructed by creating connections between any two authors that co-author a publication on at least one occasion and this defines the structure of a collaboration network (Fig. 1).

Insert Figure 1 approximately here

Given that some relationships between authors are stronger than others, connections in the collaboration network were weighted by an inter-author *distance* (23). The distance is defined as follows:

$$d_{ij} = \left(\sum_k \frac{\delta_i^k \delta_j^k}{n_k - 1}\right)^{-1} \quad (1)$$

where *i* and *j* are distinct authors, δ_i^k is 1 if the author *i* contributed to paper *k* and 0 otherwise, n_k is the number of authors on paper *k* and single-author papers are excluded from the summation (15).

Influence within the network of collaboration was estimated using degree centrality and betweenness centrality, common proxies for influence (5). *Degree centrality* for an author is given by the sum of the weights of each of the connections associated with the author. Betweenness centrality is determined by counting the number of times an author appears along the shortest path between any two other nodes (see Supplementary Material). Thus, betweenness centrality is a way of measuring the effect that removing an author would have on the connectivity of the network. Non-parametric tests were performed to compare the DIA authors to their counterparts across the three metrics (degree centrality, betweenness centrality and citation counts), and across the 22 drugs.

In a randomly-constructed network, where authors are not influenced by geography, past experiences, variation in productivity and organisational hierarchy, each author would be expected to have a similar degree centrality (tightly distributed around the network density) and betweenness centrality. The network would look quite similar when viewed from the perspective of any author and each author would feature in a similar number of shortest paths. However in real networks, contextual differences in organisation structure, geography and chronology lead to variations that affect the connectivity in the network, the flow of information, the structure of communities and the disparity of influence amongst groups.

Community structure and collaborative preference

Community structure detection uses the shape of a network to identify the optimal sub-groupings within a network to reveal how individuals fall into natural communities within which they form stronger or more frequent connections (29-31). A network grouping assigns authors to exactly one group (note that some recent methods consider overlapping communities) (32). The network grouping that produces the maximum *modularity* is defined as the best approximation of community structure for a given network. The algorithm used for finding the maximum modularity (and thus the optimal grouping) in each network is provided in the Supplementary Material. The results of a community

structure analysis are used along with collaborative preference (below) to indicate the composition of research sub-communities in the population of authors.

Collaborative preference is the tendency for individuals in a network to form connections with others of similar characteristics, and is commonly known as homophily (33). In the networks we examine, authors are characterised only by their association with a company, so collaborative preference is determined by the proportional rate at which DIA authors choose to collaborate with other DIA authors compared to non-DIA authors. The collaborative preference is given as a percentile where values close to 0% (across group) or 100% (within group) indicate collaborative preference. Details for constructing the percentile from the homophily are given in the Supplementary Material (34, 35).

Results

Network-level characterisation

The networks derived from 22 recently-developed and widely prescribed drugs, included an average of 1278 authors (from 424 for meloxicam to 3582 for atorvastatin), of which 81 were DIA authors (from 8 for clopidogrel to 316 for olanzapine). Each author had an average of 1.39 co-authors, indicating a relatively low network density, which was consistent across all networks. The networks tended to have dense cores of co-authorship involving between 10% and 60% of the authors, with the remaining authors distributed across a set of small, disconnected groups involving only one to several papers (Fig. 2). The network level characteristics for each of the 22 networks are given in Table 1.

Higher proportions of DIA authors tend to be found in denser networks ($\beta=1.31\times10^{-2}$ percentage points per extra co-author, CI 1.29×10^{-3} to 2.74×10^{-2}, $R^2 = 0.153$, p=0.0721). However, the result is not statistically significant.

Insert Figure 2 approximately here

Insert Table 1 approximately here

Disparity in influence and impact

Kruskal-Wallis tests comparing individual-level network metrics for DIA authors with their counterparts in each network demonstrated significant differences between the two groups, which were consistent across the set of drugs (Table 2). In a majority of networks, DIA authors tend to be significantly more central (by both degree and betweenness) and received a greater number of citations. For degree centrality, the exceptions were clopidogrel, latanoprost and rabeprazole and for betweenness centrality, the exceptions were clopidogrel and latanoprost. For disparity in citations, exceptions in which DIA authors did not receive a significantly greater number of citations were clopidogrel, glimepiride and latanoprost. These exceptions appear to be associated with the community having a small overall number of DIA authors.

Insert Table 2 approximately here

Community structure and collaboration preference

In the 22 networks, the mean modularity within the largest core component is 0.680 (from 0.456 in ezetimibe to 0.824 in candesartan). Since a modularity value over 0.3 is typically considered evidence of community structure, all networks may be thus described as having internal communities in the largest connected section of the network.

The proportion of collaborations DIA authors make with other DIA authors ranges between 5% and 63% (Table 3). However, analysis of the networks shows that all networks produce a collaborative preference of 100% indicating that DIA authors consistently prefer to collaborate within their own group, relative to opportunity (Fig. 3).

Insert Table 3 approximately here

Insert Figure 3 approximately here

Taken together, the results indicate that DIA authors occupy positions in which they are directly connected to a larger number of other authors and positions that are associated with the mediation of information between otherwise disparate groups. DIA authors tend to receive a greater number of

citations and preferentially collaborate within their own group despite inhabiting communities that also include their non-DIA peers.

Discussion

Centralities and citation rates

Since centrality metrics are measures of influence, the disparity of centrality values across the two groups indicates the existence of a significant disparity in influence. The results reveal a disproportionate level of influence that favours DIA authors, both in terms of their immediate influence (degree centrality) and their influence as mediators of information (betweenness centrality). In particular, positions of high betweenness centrality indicate positions in which the author is responsible for connecting groups that are otherwise poorly-connected – positions that are important in facilitating information flow in a collaborative environment. As indicated in Table 2, this disparity was found in 19 of the 22 collaboration networks.

Citations measure the impact of an individual publication, and when aggregated, are used to measure the impact of an author's publications, relative to his or her peers (36). A disparity in the number of citations received by authors across distinct groups suggests that one group has a greater impact within the evidence base for a drug. As indicated in Table 2, this disparity significantly favoured DIA authors in 18 of the 22 collaboration networks. Given the consistent and significant differences, the results suggest that DIA authors have a disproportionate impact within the evidence base.

Communities and collaborative preference

The community structure analysis demonstrates that the majority of collaboration networks comprise multiple communities, which are each characterised by strong internal connections and weaker connections to other communities within the network. This type of structure is common to many different types of naturally-formed networks in which the connections are formed chronologically and by preferential attachment (37-39).

Since the majority of collaborations involving DIA authors are across the industry interface, the results suggest that DIA authors collaborate across the industry interface with relative ease. Despite this, the results also show that DIA authors preferentially collaborate within their own group, relative to opportunity. Together, the results suggest that DIA authors are embedded in communities that are well-connected and span the industry-academia interface.

Limitations

The definition of DIA authors is restrictive in the sense that a large number of other authors may be financially tied to the company associated with the drug in question, and trial funding is not explicitly included in the analysis. Given the challenges in defining other forms of affiliation and financial disclosure in a large scale examination of the literature, a conservative measure was chosen. The results therefore underestimate the scope of company influence as a whole, instead concentrating on authors whose interests are most closely aligned to that of the company marketing the drug.

The process of determining individual authors from disparate published RCTs is imperfect because authors are determined by a family name and one or two listed initials. This means that authors listed differently on different articles may be counted as two authors, and authors with exactly the same surname and initials would be counted as one. This is consistent with the approaches taken by Barabási et al. (16) and Newman (18), who showed that the error was found to be at the level of a few percent. Following these examples, it is assumed that the effect is unlikely to have a major impact on the pattern of results since it likely affects both groups similarly.

The distance metric simplifies complex relationships between co-authors by making the assumption that two authors on a paper with many more authors are likely to be less close than two authors that publish a paper with no other authors. The method is limited insofar as it does not account for other effects that might influence collaboration and the eventual impact, such as geography, proximity or special relationships between first and senior authors (40-42).

A significant proportion of clinical trials are never published, and there is a tendency for industry-funded clinical trials to be published less often than otherwise funded trials (3). This publication bias has a specific effect on the reliability of the networks we constructed – that the degree centrality of authors involved in industry-funded trials may be under-estimated in the network. Under the assumption that DIA authors tend to be involved in industry-funded trials, the consequence is a potential under-representation of DIA authors (and their degree centrality) in the network, specifically, when the networks are compared to their participation in trials. Thus, DIA authors may be more central than is represented in the published literature, further increasing the significant differences demonstrated in the results.

Implications

Research occurs in diverse settings from industry to academia, and sometimes collaborations span the boundary between the two. It should not be assumed industry's influence on the evidence base stops at the industry boundary. In the current environment, 36% of published evidence is funded by industry and that proportion is increasing (43). Typical examinations of industry bias tend to directly sum the contributions of industry and non-industry, ignoring the importance collaborations between the groups. As a consequence of their central positions and "reach" within collaborations in wider community, DIA authors may be able to take advantage of their disproportionate level of influence over the construction of the evidence base. In combination with the known biases affecting what industry-affiliated authors publish (1, 44), the effects of their influence may have as yet unmeasured implications for what is known about the safety or efficacy of approved drugs. For example, when a meta-analysis indicated a safety problem for rofecoxib, industry-affiliated researchers responded by authoring articles claiming safety and this may have contributed to the slow withdrawal (45, 46). With other clear examples of unsafe or ineffective drugs entering the market and then leaving too slowly (24, 47-49), it is therefore important to be able to quantify the ways in which the research agendas of various groups differentially contribute to how the safety and efficacy of drugs are perceived as a consequence.

Conclusions

The present study is the first to measure influence across groups in collaborative research networks for clinical trial research, using co-authorship to define professional relationships. The networks are characterised by the disparity between DIA authors and their other peers. The analysis shows that DIA authors tend to collaborate across the industry interface, prefer to collaborate within their group, have the potential to exert more influence over information flow and thus may have greater impact on the evidence base.

Acknowledgements

The authors thank Dr Tatjana Zrimec (PhD) and Dr Frank Lin (MB BS, PhD) for comments and assistance in evaluating the analysis and acknowledge the support of the National Health and Medical Research Council (NHMRC) Program Grant 568612.

References

1. Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of Registered and Published Primary Outcomes in Randomized Controlled Trials. JAMA. 2009 September 2, 2009;302(9):977-84.
2. Hochman M, McCormick D. Characteristics of Published Comparative Effectiveness Studies of Medications. JAMA. March 10, 2010;303(10):951-8.
3. Bourgeois FT, Murthy S, Mandl KD. Outcome Reporting Among Drug Trials Registered in ClinicalTrials.gov. Annals of Internal Medicine. August 3, 2010;153(3):158-66.
4. Yank V, Rennie D, Bero LA. Financial ties and concordance between results and conclusions in meta-analyses: retrospective cohort study. BMJ. 2007 December 8, 2007;335(7631):1202-5.
5. Borgatti SP, Mehra A, Brass DJ, Labianca G. Network Analysis in the Social Sciences. Science. 2009 February 13, 2009;323(5916):892-5.
6. Goh K-I, Oh E, Jeong H, Kahng B, Kim D. Classification of scale-free networks. Proceedings of the National Academy of Sciences of the United States of America. 2002 October 1, 2002;99(20):12583-8.
7. Barabasi A-L. Scale-Free Networks: A Decade and Beyond. Science. 2009 July 24, 2009;325(5939):412-3.
8. Ravasz E, Barabási A-L. Hierarchical organization in complex networks. Physical Review E. 2003;67(2):026112.
9. Barabási A-L, Albert R. Emergence of Scaling in Random Networks. Science. 1999 October 15, 1999;286(5439):509-12.
10. Leskovec J, Chakrabarti D, Kleinberg J, Faloutsos C, Ghahramani Z. Kronecker Graphs: An Approach to Modeling Networks. The Journal of Machine Learning Research. 2010 3 January;11:985-1042.
11. Albert R, Jeong H, Barabasi A-L. The diameter of the world wide web. Nature. 1999;401:130-1.
12. Schneider CM, Moreira AA, Andrade JS, Havlin S, Herrmann HJ. Mitigation of malicious attacks on networks. Proceedings of the National Academy of Sciences. 2011 March 8;108(10):3838-41.

13. Albert R, Jeong H, Barabasi A-L. Error and attack tolerance of complex networks. Nature. 2000;406(6794):378-82.
14. Vespignani A. Predicting the Behavior of Techno-Social Systems. Science. 2009 July 24, 2009;325(5939):425-8.
15. Newman MEJ. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Physical Review E. 2001;64(1):016132.
16. Barabási AL, Jeong H, Néda Z, Ravasz E, Schubert A, Vicsek T. Evolution of the social network of scientific collaborations. Physica A: Statistical Mechanics and its Applications. 2002;311(3-4):590-614.
17. Moody J. The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999. American Sociological Review. 2004;69(2):213-38.
18. Newman M. The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences of the United States of America. 2001;98:404 - 9.
19. Freeman LC. Centrality in social networks conceptual clarification. Social Networks. 1978;1(3):215-39.
20. Borgatti SP. Centrality and network flow. Social Networks. 2005;27(1):55-71.
21. Borgatti SP, Everett MG. A Graph-theoretic perspective on centrality. Social Networks. 2006;28(4):466-84.
22. Freeman LC. A set of measures of centrality based on betweenness. Sociometry. 1977;40(1):35-41.
23. Newman M. Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Sciences. 2004;101:5200 - 5.
24. Wang AT, McCoy CP, Murad MH, Montori VM. Association between industry affiliation and position on cardiovascular risk with rosiglitazone: cross sectional systematic review. BMJ. March 18, 2010;340:c1344.
25. Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of Funding and Conclusions in Randomized Drug Trials: A Reflection of Treatment Effect or Adverse Events? JAMA. 2003 August 20, 2003;290(7):921-8.

26. Kjaergard L, Als-Nielsen B. Association between competing interests and authors' conclusions: epidemiological study of randomised clinical trials published in the BMJ. BMJ. 2002 August 3, 2002;325(7358):249.
27. Lundh A, Barbateskovic M, Hróbjartsson A, Gøtzsche PC. Conflicts of Interest at Medical Journals: The Influence of Industry-Supported Randomised Trials on Journal Impact Factors and Revenue - Cohort Study. PLoS Med.7(10):e1000354.
28. Bhandari M, Busse JW, Jackowski D, Montori VM, Schunemann H, Sprague S, et al. Association between industry funding and statistically significant pro-industry findings in medical and surgical randomized trials. CMAJ. 2004 February 17, 2004;170(4):477-80.
29. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. Physical Review E. 2004;69(2):026113.
30. Girvan M, Newman MEJ. Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America. 2002 June 11, 2002;99(12):7821-6.
31. Newman MEJ. Detecting community structure in networks. The European Physical Journal B - Condensed Matter and Complex Systems. 2004;38(2):321-30.
32. Ahn Y-Y, Bagrow JP, Lehmann S. Link communities reveal multiscale complexity in networks. Nature.466(7307):761-4.
33. McPherson M, Smith-Lovin L, Cook JM. Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology. 2001;27(1):415-44.
34. Dunn AG, Westbrook JI. Interpreting social network metrics in healthcare organisations: A review and guide to validating small networks. Social Science & Medicine. 2011;72(7):1064-8.
35. Rizos EC, Salanti G, Kontoyiannis DP, Ioannidis JPA. Homophily and co-occurrence patterns shape randomized trials agendas: illustration in antifungal agents. Journal of Clinical Epidemiology.In Press, Corrected Proof.
36. Evans JA. Electronic Publication and the Narrowing of Science and Scholarship. Science. 2008 July 18, 2008;321(5887):395-9.

37. Newman MEJ. Clustering and preferential attachment in growing networks. Physical Review E. 2001;64(2):025102.
38. Vázquez A. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. Physical Review E. 2003;67(5):056104.
39. Wagner CS, Leydesdorff L. Network structure, self-organization, and the growth of international collaboration in science. Research Policy. 2005;34(10):1608-18.
40. Lee K, Brownstein JS, Mills RG, Kohane IS. Does Collocation Inform the Impact of Collaboration? PLoS ONE.5(12):e14279.
41. Yousefi-Nooraie R, Akbari-Kamrani M, Hanneman R, Etemadi A. Association between co-authorship network and scientific productivity and impact indicators in academic medical research centers: A case study in Iran. Health Research Policy and Systems. 2008;6(1):9.
42. Jones BF, Wuchty S, Uzzi B. Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science. Science. 2008 November 21, 2008;322(5905):1259-62.
43. Buchkowsky SS, Jewesson PJ. Industry Sponsorship and Authorship of Clinical Trials Over 20 Years. Ann Pharmacother. 2004 April 1, 2004;38(4):579-85.
44. Lee K, Bacchetti P, Sim I. Publication of Clinical Trials Supporting Successful New Drug Applications: A Literature Analysis. PLoS Med. 2008;5(9):e191.
45. Jüni P, Nartey L, Reichenbach S, Sterchi R, Dieppe PA, Egger M. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. The Lancet. 2004 2004/12/10/;364(9450):2021-9.
46. Topol EJ. Failing the Public Health - Rofecoxib, Merck, and the FDA. N Engl J Med. 2004;351(17):1707-9.
47. Moynihan R. Rosiglitazone, marketing, and medical science. BMJ. April 7, 2010;340:c1848.
48. Hayes DF. Bevacizumab Treatment for Solid Tumors. JAMA: The Journal of the American Medical Association. February 2, 2011;305(5):506-8.
49. Brody H, Light DW. The Inverse Benefit Law: How Drug Marketing Undermines Patient Safety and Public Health. Am J Public Health. March 1, 2011;101(3):399-404.

Figures

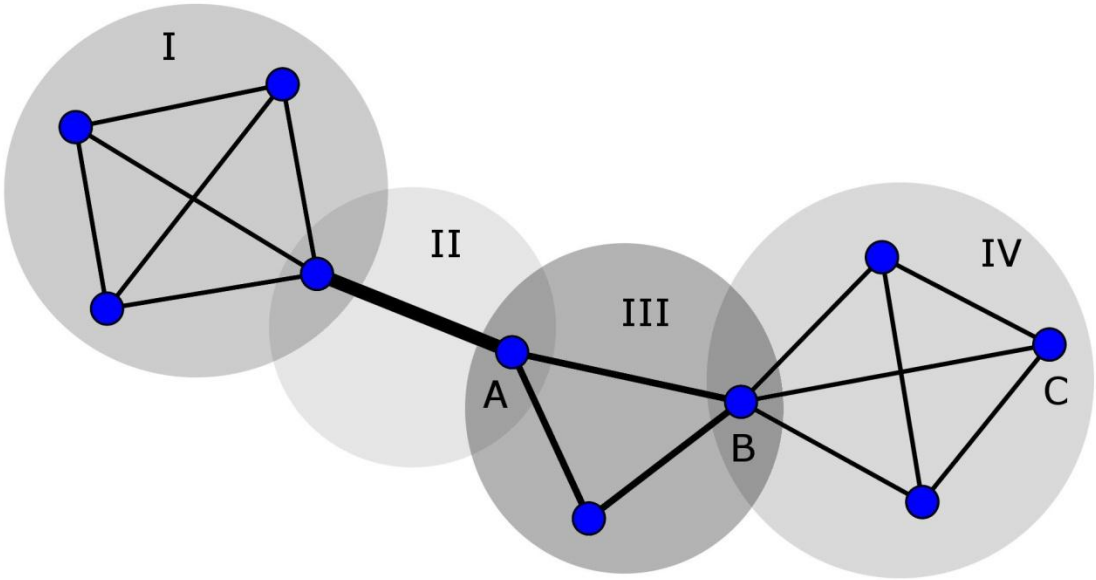

Figure 1. An example research collaboration network including four publications (I to IV), ten authors, and connections based on weighted co-authorship (thicker lines indicate shorter distance). Author A has the highest betweenness centrality (featuring in forty shortest paths), B has a high weighted degree centrality (collaborating in two papers with five authors), and C is one of seven peripheral author nodes.

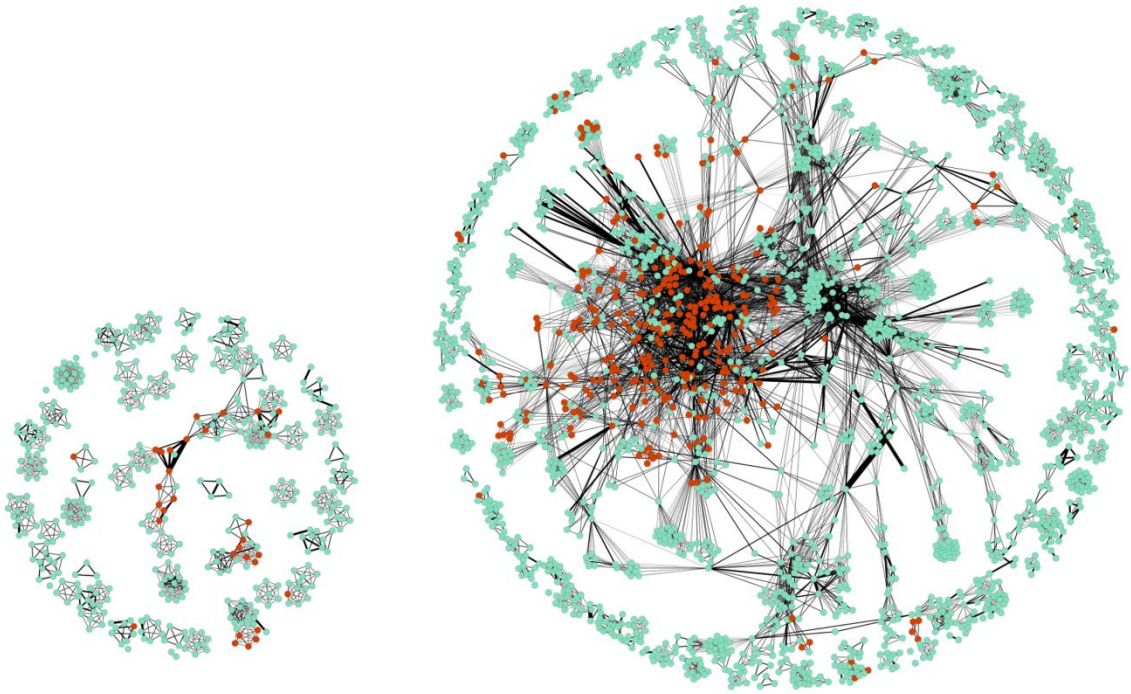

Figure 2. Two examples of the research collaboration networks indicate the diversity within the group of 22. The mirtazapine network (left) is small, relatively sparse with a large proportion of small disconnected groups in the periphery, and features an average proportion of DIA authors. Conversely, the olanzapine network (right) is larger, has a larger, denser core and features a high proportion of directly industry affiliated authors, predominantly in the core.

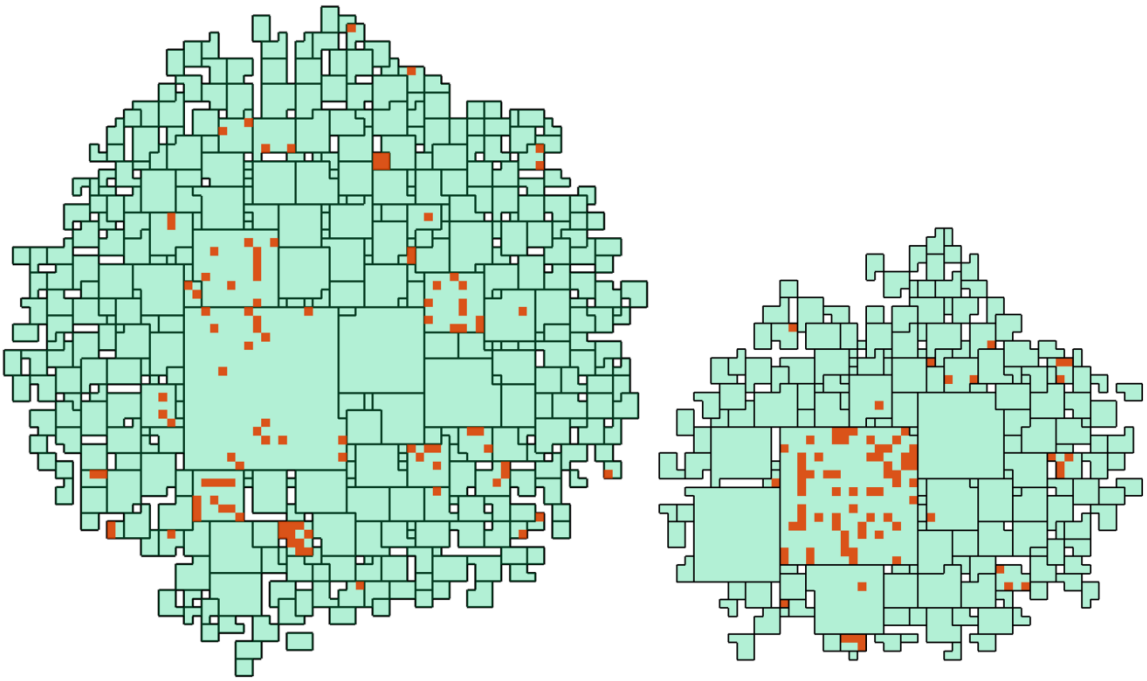

Figure 3. Community structure and community membership of directly industry-affiliated authors for atorvastatin (left) and rosiglitazone (right) constructed using all authors in the two networks. Directly industry-affiliated authors (dark squares) tend to inhabit communities in which non-company authors (light squares) are also present. Communities are delineated by black lines and include authors in the core and the periphery of the networks (each peripheral component is a community).

Table 1. Network-level information about the 22 networks that are defined by co-authorship of RCTs.

Drug Name (INN*)	Company or Companies	Network size (authors)	DIA§ authors (%)	Average number of co-authors	Authors in the core‖ (%)
alendronic acid	Merck	1646	131 (7.96)	1.57	57.8
atorvastatin	Pfizer	3582	103 (2.88)	1.44	37.3
candesartan	AstraZeneca	1776	56 (3.15)	1.48	61.8
celecoxib	Pfizer	1758	61 (3.47)	1.25	35.8
clopidogrel	Sanofi-Aventis and BMS†	2723	8 (0.29)	1.53	58.5
escitalopram	Lundbeck & Forest	573	57 (9.95)	1.47	36.5
esomeprazole	AstraZeneca	895	115 (12.8)	1.54	51.3
ezetimibe	Merck & Schering-Plough	746	132 (17.7)	1.55	60.9
glimepiride	Sanofi-Aventis	440	13 (2.95)	1.12	9.09
irbesartan	Sanofi-Aventis & BMS†	951	69 (7.26)	1.44	38.3
latanoprost	Pfizer	1116	12 (1.08)	1.38	60.4
meloxicam	Boehringer Ingelheim	424	31 (7.31)	1.16	24.3
mirtazapine	Organon	463	31 (6.70)	1.16	19.4
olanzapine	Eli Lilly	2210	316 (14.3)	1.68	61.2
rabeprazole	Eisai & Janssen	1059	40 (3.78)	1.40	36.7
risedronic acid	P&G‡ & Sanofi-Aventis	684	65 (9.50)	1.43	51.6
rofecoxib	Merck	1314	127 (9.67)	1.24	35.7
rosiglitazone	GlaxoSmithKline	1688	86 (5.09)	1.29	39.6
rosuvastatin	AstraZeneca	1136	93 (8.19)	1.39	46.8
telmisartan	Boehringer Ingelheim	986	55 (5.58)	1.36	37.4
tiotropium	Boehringer Ingelheim	689	62 (9.00)	1.36	63.3
venlafaxine	Wyeth	1245	109 (8.76)	1.38	38.8

*International Non-proprietary Name; †Bristol-Myers Squibb, ‡Procter and Gamble; §Directly industry-affiliated; ‖The core is the largest connected component in the network.

Table 2. Measures of disparity between directly industry-affiliated and other authors in 22 networks defined by co-authorship of RCTs.

Drug Name (INN*)	Network size (authors)	Number of DIA† authors	MR‡ for Log-Betweenness centrality	MR‡ for Degree centrality	MR‡ for Log-Citations
alendronic acid	1646	131	**(804, 1054)***	**(799, 1107)***	**(794, 1169)***
atorvastatin	3582	103	**(1782, 2102)***	**(1778, 2232)***	**(1780, 2177)***
candesartan	1776	56	**(879, 1188)***	**(876, 1264)***	**(878, 1223)***
celecoxib	1758	61	**(874, 1028)***	**(875, 1017)**	**(873, 1070)***
clopidogrel	2723	8	(1362, 1456)	(1362, 1485)	(1362, 1485)
escitalopram	573	57	**(280, 353)***	**(278, 367)***	**(270, 440)***
esomeprazole	895	115	**(431, 565)***	**(425, 602)***	**(428, 586)***
ezetimibe	746	132	**(353, 471)***	**(337, 542)***	**(355, 461)***
glimepiride	440	13	**(220, 253)**	**(218, 302)**	**(218, 304)**
irbesartan	951	69	**(469, 566)***	**(470, 556)***	**(471, 542)**
latanoprost	1116	12	(558, 619)	(557, 651)	*(561, 339)*
meloxicam	424	31	**(208, 264)***	**(208, 264)**	**(209, 257)**
mirtazapine	463	31	**(228, 282)***	**(227, 304)***	**(228, 295)***
olanzapine	2210	316	**(1062, 1366)***	**(1056, 1402)***	**(1063, 1361)***
rabeprazole	1059	40	**(526, 619)***	(526, 621)	**(525, 654)***
risedronic acid	684	65	**(334, 427)***	**(336, 406)***	**(329, 476)***
rofecoxib	1314	127	**(636, 863)***	**(638, 840)***	**(633, 883)***
rosiglitazone	1688	86	**(834, 1039)***	**(833, 1056)***	**(834, 1044)***
rosuvastatin	1136	93	**(551, 766)***	**(551, 763)***	**(550, 775)***
telmisartan	986	55	**(484, 651)***	**(483, 679)***	**(476, 788)***
tiotropium	689	62	**(339, 410)***	**(337, 427)***	(341, 388)
venlafaxine	1245	109	**(609, 770)***	**(609, 771)***	**(613, 729)***

*International Non-proprietary Name; †Directly industry-affiliated; ‡Mean ranks (non-DIA, DIA), where a higher MR indicates a higher median value; Bold: $p<0.05$, Asterisk: $p<0.01$, Italics: $p<0.05$ favouring higher values amongst non-DIA authors.

Table 3. Measures of community membership and the mixing of directly industry-affiliated authors within the collaboration networks.

Drug Name (INN*)	Network size (authors)	Core size (authors)	Number of communities in core	Modularity in the core[†]	Collaborative preference[‡] (% of connections)
alendronic acid	1646	951	118	0.480	36.2
atorvastatin	3582	1336	77	0.751	24.2
candesartan	1776	1097	14	0.824	21.5
celecoxib	1758	630	47	0.770	28.2
clopidogrel	2723	1594	133	0.655	5.4
escitalopram	573	209	10	0.688	43.9
esomeprazole	895	459	23	0.683	47.4
ezetimibe	746	454	161	0.456	52.6
glimepiride	440	40	3	0.784	29.6
irbesartan	951	364	15	0.784	50.7
latanoprost	1116	674	27	0.822	13.0
meloxicam	424	103	4	0.601	31.7
mirtazapine	463	90	8	0.741	51.3
olanzapine	2210	1352	78	0.656	63.4
rabeprazole	1059	389	13	0.785	24.5
risedronic acid	684	353	12	0.635	39.8
rofecoxib	1314	469	18	0.691	42.9
rosiglitazone	1688	668	22	0.685	28.3
rosuvastatin	1136	532	12	0.715	28.2
telmisartan	986	369	153	0.506	23.1
tiotropium	689	436	10	0.735	23.0
venlafaxine	1245	483	14	0.729	47.5

*INN: International Non-proprietary Name; [†]All values above 0.3 indicate the presence of community structure. [‡]Connetions between DIA authors as a percentage of all connections involving DIA authors, all are significantly higher than for random simulated networks.

Appendix: Details of methodology

This supplementary material includes specific details about the methodology for (A) drug selection and literature search; (B) verification of our classification for authors' direct industry affiliations; and methods for calculating (C) betweenness centrality, (D) community structure, and (E) collaborative preference. The descriptions are intended to complement the main text and provide complete information for the purpose of replication.

A. Drug selection and literature search

We chose high-volume, recently developed drugs. The 22 drugs were selected as follows:

1. Drugs featuring amongst the top fifty (by prescription volume) in Australia between the years 2003 and 2009 were retrieved from yearly reports (1). Of the 115 drugs represented in these lists (since many featured in reports for more than one year), we selected the top 50 by volume.
2. For each drug, Scopus (which also covers PubMed) is searched to obtain all articles for which the name of the drug (by International Non-proprietary Name) or any of its synonyms (by EMBASE subject headings) was used in the title or the abstract (2). Amongst these, only articles in which "randomised controlled trial" was used as an index term were used.
3. Any drug for which 5% or more randomised controlled trials (RCTs) were published prior to 1996 was excluded. From the top 50 drugs by volume, 28 were excluded in this step. The number of RCTs published for the resulting 22 drugs was 6711.

The resulting 22 drugs (see the main text for details) represent those drugs that were prescribed in the order of 10^5 to 10^7 times per year in Australia in the period 2003 to 2009, and for which the vast majority of RCTs were published after the beginning of 1996.

We found that the literature search method described in Step 2 above is an appropriate compromise between sensitivity and specificity. This approach minimises the number of false positives because

there are many spurious articles that do not examine a drug as part of an RCT but still mention the drug by name. There is a chance for false negatives in the case where an RCT evaluates a drug, but the term or its synonyms are not used in the title or abstract, or the abstract is not accessible.

Each of the 22 drugs is also associated with one or more pharmaceutical companies responsible for the manufacture and marketing of the drug. Note that these companies are inferred from US Food and Drug Administration approvals and augmented by partnerships listed on commercial websites.

B. Verification of affiliation

Authors' affiliations are extracted as free text from the retrieved citations, and acknowledgements and declarations of funding sources are not examined. Each author may include more than one affiliation on an article, and may contribute to more than one article about a given drug. We use an automated text search to determine if the affiliation of an author in a publication matches the name of the company associated with the medicine. If the name of the company producing the medicine under trial is found in any of an author's affiliations (implying that the author is in some way employed by the company), then that author is classified as company-affiliated. The possibility of inaccuracy arises when affiliations are listed without attribution to individuals, or when affiliations are missing or incomplete. One author (AGD) and an independent examiner performed a blinded classification of a random sample of 200 authors, which produced a strong agreement between the two humans ($\kappa = 0.936$) and between human and computer ($\kappa = 0.936$ and $\kappa = 1.00$ for the author and independent examiner respectively).

C. Betweenness centrality

Weighted betweenness centrality (3-5), as it is applied to collaboration networks (6), uses the shortest distances between all pairs to determine the importance of each connection (and author) in the flow of information within the network. Intuitively, one can imagine that betweenness centrality is a way of measuring the effect that removing a collaborative link (or an author and his or her links) would have

on the connectivity of the network. The betweenness centrality algorithm for a weighted collaboration network is as follows:

1. Calculate the shortest paths between all pairs of individual authors using the distances associated with each connection.
2. Assign a value of one to each connection and each author along the shortest path (excluding the two authors on the ends of the path). If there is more than one unique shortest path, assign a proportional value (half for two paths, one-third for three paths and so on) to that path or section of a path.
3. Aggregate all the values along each connection and define this value as the betweenness centrality of that connection.
4. Aggregate all the values associated with each author and define this as the betweenness centrality of that author.

D. Community structure analysis

The community structure of a network may be defined as the optimal grouping of a network into a set of connected sub-groups, each of which is better connected internally than it is externally. For any given network, the grouping that produces the largest *modularity* is defined to be the best approximation of the community structure. Modularity takes a value between zero and one, and any network with a modularity value above 0.3 is said to exhibit a community structure (7). Modularity is defined as follows:

$$Q = \frac{1}{2m}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(c_i c_j)$$

where *m* is the total sum of the weights in the network (weights are the inverse of the distance), A_{ij} is the adjacency matrix form of the network, k_i is the sum of the weights for connections emanating directly from node *i*, and $\delta(c_i c_j)$ is 1 if the nodes *i* and *j* are in the same partition and 0 otherwise (7).

One popular algorithm for detecting the optimal grouping uses betweenness centrality to define inter-community connections and uses them to partition the network. Community structure analysis may be applied to weighted networks using a similar approach to that used for unweighted networks (8-10). The method proceeds with the following steps:

1. The weighted betweenness centrality of every connection is calculated without considering the distances (all non-zero connections are counted as a connection).
2. The betweenness for every connection is multiplied by its distance to produce a score.
3. The connection with the highest resulting score is removed from the network and the process is repeated. Once all connections are removed, the result is a completely disconnected network in which every author forms his or her own community.
4. Starting with the completely disconnected set of authors, the connections removed in the previous step are added sequentially in the reverse order to produce a new network.
5. The modularity (see below) is calculated after each added connection by assuming each of the components in the new network is a separate community.

The algorithm ends when there are no new connections to be added. The grouping with the maximum modularity is defined to be the optimal community structure.

E. Collaborative preference

A network metric measuring collaborative preference is used to determine if nodes in a network have a preference for connecting with other nodes that share the same characteristics (11, 12). In a collaboration network where authors belong to one of two groups (either DIA authors or non-DIA authors), preferential collaboration undertaken by DIA authors may be defined as the sum of the number of connections between DIA authors, divided by the total number of connections involving DIA authors.

Thus, the collaborative preference metric is a proportion of connections and takes a value between zero and one. Since the density and the number of authors in each group may confound the result, we account for this by comparing the proportion against a simulated random baseline, as has been

previously applied for homophily and other network metrics in small networks (13, 14). The method works by locating the collaborative preference along the distribution of random networks – if the value is significantly different from the simulated networks it provides evidence of a network that is constructed under collaborative preference in one of two directions

References

1. Department of Health and Ageing. Reports for the twelve months to 31 December 2003 to the twelve months to 30 June 2009, PBS Expenditure and Prescription. [04/12/2009; cited 2011 January]; Available from: http://www.health.gov.au/internet/main/publishing.nsf/Content/pbs_expenditure_prescriptions-copy1.

2. Elsevier. Ovid Database Guide, EMBASE: Excerpta Medica Database Guide. [21/08/2007; cited 2011 January]; Available from: http://www.ovid.com/site/products/ovidguide/embase.htm.

3. Barrat A, Barthélemy M, Vespignani A. Modeling the evolution of weighted networks. Physical Review E. 2004;70(6):066149.

4. Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A, Parisi G. The Architecture of Complex Weighted Networks. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(11):3747-52.

5. Barthélemy M, Barrat A, Pastor-Satorras R, Vespignani A. Characterization and modeling of weighted networks. Physica A: Statistical Mechanics and its Applications. 2005;346(1-2):34-43.

6. Newman MEJ. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Physical Review E. 2001;64(1):016132.

7. Newman MEJ. Analysis of weighted networks. Physical Review E. 2004;70(5):056131.

8. Girvan M, Newman MEJ. Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America. 2002 June 11, 2002;99(12):7821-6.

9. Newman MEJ. Detecting community structure in networks. The European Physical Journal B - Condensed Matter and Complex Systems. 2004;38(2):321-30.

10. Newman MEJ. Modularity and community structure in networks. Proceedings of the National Academy of Sciences. 2006 June 6, 2006;103(23):8577-82.

11. Ibarra H. Homophily and Differential Returns: Sex Differences in Network Structure and Access in an Advertising Firm. Administrative Science Quarterly. 1992;37(3):422-47.

12. McPherson M, Smith-Lovin L, Cook JM. Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology. 2001;27(1):415-44.

13. Dunn AG, Westbrook JI. Interpreting social network metrics in healthcare organisations: A review and guide to validating small networks. Social Science & Medicine. 2011;In Press, Corrected Proof.

14. Rizos EC, Salanti G, Kontoyiannis DP, Ioannidis JPA. Homophily and co-occurrence patterns shape randomized trials agendas: illustration in antifungal agents. Journal of Clinical Epidemiology.In Press, Corrected Proof.