Undercounting the gift givers: Issues when tallying acknowledgments in life sciences research.

Katherine W. McCain
College of Computing & Informatics
Drexel University
mccainkw@drexel.edu

SUBMITTED FOR PRESENTATION

BACKGROUND. Citations and acknowledgments are two different (but likely interrelated) ways of giving credit for the input of other researchers in the author's own research. Citation data demonstrate the impact/utility/influence of prior published work (the author's or someone else's contributions). Acknowledgments deal with the many other ways that living researchers and funding organizations can provide support for scholarly research—in the former case by providing research materials, unpublished protocols or data, various support services, or by reading versions of a manuscript or engaging in useful discussion as the research progresses (see, e.g. Cronin & Franks, 2006; McCain, 1991, and a thorough review of the field by Desrochers, Paul-Hus & Pecoskie, forthcoming in *JASIST*). Most, if not all, of the prior work on acknowledgments patterns has focused on the specific section of the research article where authors explicitly thank individuals and agencies for supporting the reported research in some fashion. This is most generally located at the end of the text proper and most commonly labeled as "Acknowledgments" but can also be found in footnotes at the bottom of the first page, the last (unlabeled) paragraph of the paper, or at the end of the bibliography (*Science* is a good example of the last). Focusing on the formal Acknowledgments section works well for data collection in many fields but may miss data in life sciences research, where donors are thanked for providing research materials and other useful inputs in the section most commonly labeled "Materials and Methods." This section can be found in different locations such as following the introduction, following the discussion or conclusions, or embedded in extensive figure captions. Another challenge in tallying personal acknowledgments is name disambiguation because of the use of nicknames (e.g., Chuck/Charles, Bob/Robert, Bill/William, Janni/Christiane), alternative spellings (Stephen/Steven, Philip/Phillip), surname misspellings, and variation in the use of initials for Asian personal names. Both search space and allonym control may be particularly important in automated text extraction (see, e.g., Khabsa, Koppman, & Giles, 2012). Finally, simple name tallies can hide the degree to which the person thanked may play more than one role in supporting the research process (and, indeed, whether the recipient is even part of the research community). This workshop paper explores these issues through an in-depth acknowledgments analysis of the literature of zebrafish research, 1980-2004.

PURPOSE. (1) To look at trends in acknowledgment patterns over time in a focused life sciences literature; (2) To examine the range of resources, services, etc. for which individuals were thanked in early zebrafish research. (3) To assess the impact of including or ignoring personal acknowledgment statements in "Materials & Methods" sections of articles reporting original scholarly research.

METHODS. To provide a focus for this study and increase the likelihood of multiple acknowledgments over time, data were collected from papers originally used for a study of collaboration patterns in zebrafish research 1980-2004 (McCain, 2013). All papers examined and retained for this analysis met the following criteria: English language, accessible in full text, original research report (not review, meeting abstract, letter to the editor, etc.), focus on zebrafish as main topic with no more than one other organism discussed. Personal acknowledgment text was extracted from "standard" acknowledgments sections (ACK) and from Materials & Methods (M&M) or similar sections of all available papers. Papers were coded by the location of personal acknowledgments and degree of name overlap between M&M and ACK sections. Individual acknowledgment statements were assigned to a category & location based on an acknowledgments classification scheme adapted from McCain, 2015. Allonyms were disambiguated (to the extent possible) using authors' publications, the ZFIN database of zebrafish researchers, and personal/laboratory websites.

PRELIMINARY FINDINGS. Table 1 shows the rise in zebrafish publications (original research) and the distribution of personal acknowledgments within ACK and M&M categories across the four time periods studied (1980-1989, 1990-1994, 1995-1999, 2000-2004). There were 2590 papers with at least one personal acknowledgment. Six hundred twenty four(22%) papers included at least one name outside of the formal acknowledgments text. Most M&M acknowledgment text used variants of "was a (kind) gift of/from, was (kindly) provided by, courtesy of, were/was obtained from." Very few papers thanked researchers only in the M&M section; most using the M&M section also posted thanks in the ACK section. Table 2 presents summary data showing the kinds of research support for which other researchers were thanked. In each of the 4 time periods (data not shown), the distribution of individual personal names was noticeably skewed with a relatively few names occurring frequently and long tails of names mentioned only once. Overall, roughly 7800 names appeared at least once (some as allonyms—surnames with initials that could not be assigned a clear fuller name). (Note to reviewers—more detailed data would be presented in the workshop). As an example of the issues in bypassing the M&M section, Table 3 illustrates personal acknowledgment tallies for 6 highly-acknowledged authors in 1995-1999 (including known allonyms and proportion of acknowledgments found only in M&M) as well as marked differences in the range of things for which the various donors are thanked. These preliminary results suggest that, while ignoring the M&M section will still capture a large percentage of the personal acknowledgments (Table 1), it can distort the ranking of researchers who are primarily sources of research materials (e.g. Grunwald, Riggleman in Table 3). High acknowledgment tallies do not necessarily point to zebrafish researchers—some highly ranked names may be skilled technical support staff or researchers in other fields whose only role is to serve as a source for specific research reagents. Additionally, automated text mining for personal names in M&M sections would need to distinguish between a named thanked person (associated with a thanking phrase) and a name/date reference to prior published work (where gift fish/library/antibody, protocol, etc.) were first described.

REFERENCES

Cronin, B. & Franks, S. 2006. Trading Cultures: Resource Mobilization and Service Rendering in the Life Sciences as Revealed in the Journal Article's Paratext. *Journal of the American Society for Information Science & Technology,* 57(4): 1909-1918

Desrochers, N. Paul-Hus, A. and Pecoskie, J. (forthcoming). Five decades of gratitude: A meta-synthesis of acknowledgements research. *Journal of the Association for Information Science and Technology.*

Khabsa, M., Koppman, S. & Giles, C. L. 2012 "Towards Building and Analyzing a Social Network of Acknowledgments in Scientific and Academic Documents." In: International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction SBP 2012: Social Computing, Behavioral - Cultural Modeling and Prediction pp 357-364. Lecture Notes in Computer Science v.7227.

McCain, K.W. 1991. Communication, competition and secrecy—the production and dissemination of research-related information in genetics. *Science, Technology & Human Values, 16*(4): 491-516.

McCain, K.W. 2013. Charting the rise of the zebrafish as a model organism: Persistent co-author networks, 1980-2004 Presented at METRICS 2013: Workshop on Informetric and Scientometric Research, Montreal, Canada, November 2, 2013

McCain, K.W. 2015. Collaboration patterns in model organism research: Co-authorship, acknowledgment, and the starlet sea anemone (*Nematostella vectensis.)* Presented at METRICS 2015: Workshop on Informetric and Scientometric Research, St. Louis, MO, 7 Nov 2015.

Table 1: Distribution of personal acknowledgments in reports of original research

LOCATION OF PERSONAL ACKNOWLEDGMENTS IN PAPERS	1980-89	1990-94	1995-99	2000-04	TOTAL	Percentage total
M&M only	0	2	13	38	53	2%
M&M & ACKs--no overlap	6	25	99	159	289	11%
M&M & ACKs--partial overlap	0	6	198	78	282	11%
M&M & ACKs—complete overlap	1	36	4	219	260	10%
ACKs only	99	128	409	1070	1706	66%
TOTAL PAPERS WITH ACKNOWLEDGMENTS	**106**	**197**	**723**	**1564**	**2590**	
No personal acknowledgments	44	25	69	204	342	
TOTAL	**150**	**222**	**792**	**1769**	**2933**	

Table 2: Types of acknowledgments: summary data 1980-2004

ACKNOWLEDGMENT CATEGORIES	Papers with names in M&M only	Papers with names in both M&M and ACK	Papers with names in ACK only	Total papers
1a Experimental animals	84	128	312	524
1b Providing materials	235	483	699	1417
1c Equipment & facilities	1	0	232	233
1d Unpublished data	15	11	229	255
1e Unpublished software	13	5	16	34
1f Genome information	1	1	13	15
2a Unpublished results	2	3	105	110
3a Specific information or suggestion	4	3	439	446
3b Comments on manuscript	0	0	1339	1339
3c Advice & discussion	0	0	967	967
3d Special thanks	0	0	58	58
3e Peer support	0	0	303	303
4a Specific analyses	3	3	818	824
4c Animal husbandry	0	0	511	511
4d Technical assistance	0	0	818	818
4e Collecting animals	0	0	3	3
5a Manuscript production	0	0	132	132
5b Graphics & images	0	1	218	219
5c English editing	0	0	22	22
6a Administrative support	0	0	9	9
Other/not classifiable	0	0	23	23

Table 3: Selected Acknowledgment Recipient Profiles: 1995-1999

Recipient Name	M&M only	Allonyms	Acknowledgment categories
David J. Grunwald (lab director)	22/93	D. Grunwald, D.J. Grunwald, David Grunwald, David Jonah Grunwald	1b Providing materials (M&A) – primarily cDNA library (with Helde, Riggleman)* 3a Specific information or suggestion (A) 3c Advice & discussion (A)
Charles B. Kimmel (lab director)	6/88	C. Kimmel, C.B. Kimmel, Chuck Kimmel Charles Kimmel	1a Experimental animals (M & A) 1b Providing materials (A) 1d Unpublished data (A) 2a Unpublished results (A) 3a Specific information or suggestion (A) 3b Comments on manuscript (A) 3c Advice & discussion (A) 3e Peer support (A)
Bob Riggleman (post-doc, Grunwald lab)—has left the field	18/55	B. Riggleman, R. Riggleman, Robert Riggleman	1b Providing materials (M&A) – primarily cDNA library (with Helde, Grunwald), twist probe * 1d Unpublished data (A) 2a Unpublished results (A) 3a Specific information or suggestion (A)
Stefan Schulte-Merker (PhD student MPIE/Post-doc)	1/55	S. Shulte-Merkur [sic], S. Schulte-Merker	1a Experimental animals (M,A) 1b Providing materials (M, A)* 1d Unpublished data (A) 3a Specific information or suggestion (A) 3b Comments on manuscript (A) Max-Planck-Institut für Entwicklungsbiologie 3c Advice & discussion (A)
Marnie E. Halpern (post-doc U of Oregon, Staff Scientist, Carnegie Institute of Washington)	2/32	M. Halpern, M.E. Halpern, Marnie Halpern	1a Experimental animals (A)* 1b Providing materials (A) 1d Unpublished data (A) 2a Unpublished results (A) 3b Comments on manuscript (A) 3c Advice & discussion (A)
Ruth BreMiller (Research Associate, IoN, University of Oregon)	1/27	R. BreMiller, R. Bremiller	1b Providing materials (M&A) 4a Specific analyses (A)* 4d Technical assistance (A)*

* Most common acknowledgment category