Lemke, Steffen; Mazarakis, Athanasios; Peters, Isabella

Book Chapter — Published Version

Path model of the interplay between the promotion and the received attention of research articles

Suggested Citation: Lemke, Steffen; Mazarakis, Athanasios; Peters, Isabella (2023) : Path model of the interplay between the promotion and the received attention of research articles, In: Broer, Irene et al. (Ed.): The Science-Media Interface. On the Relation Between Internal and External Science Communication, ISBN 9783110776546, De Gruyter, Berlin, pp. 117-138,
https://doi.org/10.1515/9783110776546-005

Steffen Lemke, Athanasios Mazarakis, and Isabella Peters

5. Path model of the interplay between the promotion and the received attention of research articles

Abstract: Existing systematic analyses of the associations between the visibility that research articles receive within different formats of external science communication (e.g., press releases, embargo e-mails, or news stories) and their later impact metrics are mostly restricted to specific case studies, despite these studies' recurring findings of substantial potential effects. This study aims to give a consolidating and more comprehensive perspective on the interplay between research articles' promotion within press releases and embargo e-mails, their publishing journal's prestige, as well as their received attention in mainstream media, on Twitter, and their academic impact as proxied by citations. To achieve this goal, we use the path analysis method to specifiy models that manifest a range of hypotheses motivated by literature and theory on the relationships that may exist between these variables. We estimate and evaluate our models based on a dataset of 67,581 research articles, which we construct through a combination of empirical data from Web of Science, Altmetric.com, EurekAlert!, and Science Media Center Germany. The resulting model confirms the conformity of the hypotheses we derived from past literature with the large set of empirical observations within our sample. More specifically, our findings highlight the considerable associations between promotion in external science communication and the attention research articles can be expected to receive on both mainstream and social media. The strongest correlations in the model exist between mainstream media mentions and both embargo e-mail promotion (normalized path coefficient of 0.605) and press release promotion (normalized path coefficient of 0.568).

Keywords: press releases, embargo e-mails, citations, altmetrics, external science communication, path analysis

1 Introduction

Citation analysis – i.e., the analysis of citations to scientific publications as indicators for the latter's academic influence – assumes that the likelihood of an ar-

https://doi.org/10.1515/9783110776546-005

ticle getting cited correlates with relevant inherent qualities, e.g., its scientific rigor, its novelty, or the significance of its findings. While certain studies argue that citations can indeed be useful predictors for such inherent qualities (see Aksnes et al., 2019 for an in-depth discussion), scientometric research has revealed numerous factors without apparent relation to quality that also affect an article's likelihood of being cited. Being informed of their existence and effects is crucial to assess individual citation analysis applications' validity and potential limitations.

In their review of literature examining factors that affect citations, Tahamtan et al. (2016) divide 28 such factors into the categories of paper related, journal related, and author related factors, albeit acknowledging that this selection is not exhaustive. As the three categories identified by Tahamtan et al. (2016) indicate, a large portion of the research focuses on how respective research findings are communicated within the academic sphere, e.g., how decisions regarding publication formats or publication venues affect citations. These decisions can, for the largest part, be considered aspects of internal science communication or scholarly communication, i.e., the communication of research findings primarily targeted at an academic audience. However, another, less analyzed set of issues affecting citations results from how research is featured and processed in channels of external science communication, i.e., in media targeted at stakeholders outside the scientific community.

Considering the scientific journal article as the exemplary unit of observation, such processing by public media may begin even before said article's publication: many scholarly journals regularly disseminate advance information on upcoming issues to science journalists in an arrangement known as an embargo (Kiernan, 1997). In short, this arrangement provides registered journalists with early access to unpublished research findings in exchange for their pledge not to pass on that information before a specified embargo date has passed. The embargo system serves both involved parties well: for the science journalists, embargoed information allows them to timely prepare their coverage on new findings while also providing a certain assurance that other journalistic outlets' respective stories will not leapfrog their own – provided those other outlets do not break the embargo date, of course. For scholarly publishers, the embargo system provides an opportunity to highlight specific publications and topics to the media as well as a certain control over the respective coverage's timing (Kiernan, 2014). As the embargo information given to journalists usually requires prior registration and thus is not openly accessible, it remains difficult to assess how this specific form of promotion affects the attention individual journal articles later receive, let alone their probability of getting cited.

Other types of interventions in external science communication that serve the purpose of directing attention to certain articles are less difficult to track. One of the most common tools used by press officers of scholarly publishers, journals, or research institutes to promote selected publications is the press release, described by Carver (2014, p. 2) as "essentially a short news article written in a journalistic style that explains a newly published scientific result in a common and not too specialized language." While research on the relationship between press release promotion and articles' later citations is rather scarce, some case studies indicate a positive association between the two (Chapman et al., 2007; Fanelli, 2013; Lemke, 2020), although Fanelli (2013) found that this association becomes negligible when controlling for the media coverage of the respective articles..

On the topic of media coverage, a further body of case studies examined how mentions in newspapers affect scientific publications' later citations (Dumas-Mallet et al., 2020; Fanelli, 2013; Kiernan, 2003; Phillips et al., 1991). Phillips et al. (1991) found articles from the *New England Journal of Medicine* (NEJM) to receive significantly more citations if these had been featured in the *New York Times* (NYT) – although this advantage could not be detected for NEJM articles that had been chosen for coverage in NYT issues that in the end were not disseminated due to the NYT being on strike. This finding backs up what is called the publicity hypothesis, which attributes articles' increase in citations associated with press coverage to the concomitant increase in visibility. The earmark hypothesis, on the other hand, explains higher citation counts for press-covered research articles with the assumption that journalistic agents merely apply similar criteria in their decisions on which research to cover as researchers do when deciding which research to cite (Phillips et al., 1991). Kiernan (2003) added to the work by Phillips et al. (1991) by additionally regarding how coverage from 24 daily newspapers and several evening broadcasts of major U.S.-television networks affected citation rates. The author found that the NYT's influence on citation rates is not unique, as NYT coverage did not correlate significantly with citation counts once the author controlled for coverage by other newspapers and television networks. In his study of the association between newspaper coverage and citations, Fanelli (2013) also found regional effects to play a substantial role, as the apparent positive effect of newspaper coverage on citations was stronger for English media than for Italian media, which primarily only affected authors from Italy. More recently, in their analysis of the citation advantages of 162 biomedical association studies reported in newspapers from six specific countries, Dumas-Mallet et al. (2020) found the strength of the observed effects to depend on the influence of the covering newspaper as well as on the number of published press articles. Moreover, they found the positive correlation between newspaper coverage and citation counts to be most significant for research articles published in journals with lower impact factors.

As the examples of embargo e-mails, press releases, newspapers, and television broadcasts illustrate, the landscape of sources of external science communication with potential effects on research impact is vast, heterogeneous, and at times opaque. Several current developments add to this intricacy: ongoing professionalization of institutes' own research communication, as well as increasing commitments to Open Science and the "third mission" of higher education (Berghaeuser & Hoelscher, 2020), eradicate formerly existing boundaries between scientists, journalists, and public audiences and lead to the establishment of new tools and formats of science communication, many of which enable more direct, bidirectional exchanges of and about research content (Liang et al., 2014). In this vein, an additional layer of complexity results from the increasing digitalization of journalistic media, the advent of social media, and the continually blurring line between these two spheres. In a reciprocal give and take, news stories are posted to and might evoke public discussions on social media platforms like Facebook or Twitter, while journalists also use these platforms to gather news in the first place (Hermida, 2012). Although altmetrics (Priem et al., 2010) provide flexible technical means to track the attention individual research publications attract on various online domains, modelling the relationship between the attention received, for instance, on Twitter, mentions received in the news, and academic citations remains complicated because of such chicken-or-egg dilemmas.

1.1 Research aim and model

While numerous previous studies put spotlights on specific types of research promotion in external science communication and the subsequent impact metrics of the respective research articles, what is lacking are more comprehensive models that provide entry points for understanding the interdependencies that work between the various interventions made to increase the publicity of research and the attention observable in different spheres of media and academia as a whole. The present analysis represents a step towards closing that research gap. It aims to consolidate the findings from previous case studies, assess their conformity with large sets of empirical data, and extend them by widening the focus to, in this context, under-analyzed forms of research promotion (in the form of embargo e-mails) as well as research metrics (in the form of altmetrics). The findings quantify the degree to which measures of external science communication potentially affect research metrics and provide first steps towards making these degrees comparable, also with a factor from the domain of scholarly communication – journal reputation – whose significance for metrics is well documented (Tahamtan et al., 2016). More substantiated knowledge of such dependencies of research metrics from external factors is necessary to develop more precise descriptions of their meaning, their caveats, and their limitations, enabling more substantiated assessments of their validity within different use cases, e.g., research evaluation. Therefore, our research question can be framed to determine the direct and indirect effects of journal prestige and mentions of scientific publications in various media on later citations. Thus, this study aims to inform the development of a more profound theory on the meaning of various research metrics and to shed light on the complex relationships of attention distribution and channeling that surround and impact research metrics, in order to ultimately increase their usefulness in practice.

Methodologically, this study aims to disentangle the complexity of relationships between various formats of science communication and research impact by formulating and testing path models comprised of several variables that capture particularly relevant manifestations of research articles' impact and uptake in external science communication. We derive hypotheses from literature about the interplay between the attention that research articles receive within press releases, embargo e-mails, mainstream media, social media, their citations within the academic sphere, and their publication venue's prestige and apply the method of path analysis (Regorz, 2021; Streiner, 2005) to see to which degree these hypotheses can be confirmed through the testing of models based on a large set of empirical data. We choose the path analysis method because of suspected multi-level interdependencies between the different indicators to be included in our models and rely on its implementation from the R-package lavaan.[1] As an extension of the statistical method of multiple regression, path analysis allows us to test more complex models in which certain variables simultaneously affect and are affected by others.

The model we start our analysis with (Figure 1) is motivated by past empirical research and theory, and is based on six such variables, which we explain in detail after listing them:

– research articles' numbers of mentions in mainstream news media (En1);
– their numbers of mentions on Twitter, as a prototypical example of a social media platform that is broadly used in academic contexts (En2; Tahamtan & Bornmann, 2020);
– their numbers of (academic) citations (En3);

1 https://cran.r-project.org/web/packages/lavaan/index.html

- the (binary) variable of whether the articles have been featured in a publisher's embargo e-mail (Ex1);
- the (binary) variable of whether the articles have been promoted in a press release (Ex2);
- the prestige of their publishing journal, proxied by the median number of citations received by articles within said journal during the three years preceding the observed article's publication (Ex3). We use median-based impact factors instead of the more commonly seen mean-based impact factors to at least partly account for problems resulting from the latter's skewed distributions; see Kiesslich et al. (2021) for a detailed discussion of this issue.

Our starting model will be fitted to a large set of empirical data, then evaluated for its fit, and afterwards, if appropriate, respecified and reevaluated until no substantial enhancements appear possible with the data at hand. This process of model estimation serves two purposes: first, it serves as a test of whether the range of hypotheses about the interplay between certain events of external science communication and article metrics drawn from past research conforms with a large set of empirical observations; and second, it shall provide a comprehensive view on how the hypothesized interactions compare to each other.

Figure 1 shows the model we start this analysis with. We assume mentions in embargo e-mails and press releases to be exogenous variables, as in almost all cases these events will happen before or very shortly after the promoted article's publication – making it implausible to assume that respective press officers' or editors' decisions could be affected by any of the endogenous variables in our model, which all accumulate later. As a third exogenous variable, we include the publishing journal's prestige, as a large body of research has found this to be a crucial factor for an article's expected citations (see the review by Tahamtan et al., 2016 for an overview of such studies), which therefore cannot be omitted in an endeavor of convincingly modeling the attention articles receive in academic and media spheres. Furthermore, it seems reasonable to assume that also in science journalism there is a common awareness of what the most prestigious journals within a respective covered field of research are – perhaps even of journal impact factors as indicators for such prestige. This makes it likely that journal prestige would be an important variable to explain articles' expected media presence as well (which is also suggested by results from previous case studies such as Dumas-Mallet et al., 2020).

To briefly illustrate the effects assumed in our model with a fictitious example: imagine we have an article that was published in 2018 in the *New England Journal of Medicine* (NEJM), an indubitably highly prestigious journal in its field (into our model's estimation, this prestige would in this case enter as the median of citations that NEJM-articles received over the years of 2015 to 2017, as our article was published in 2018). Based on past studies (see below), we believe this prestige to have a positive effect on the likelihoods with which journalists will choose said article for their stories, users will post about it on social media, and researchers will cite it in their works. The article's presence on mainstream and social media will likely be higher if the article also received promotion by means of an embargo e-mail or a press release – and the media presence itself will have a positive effect on the likelihood of the article being cited within academic publications as well.

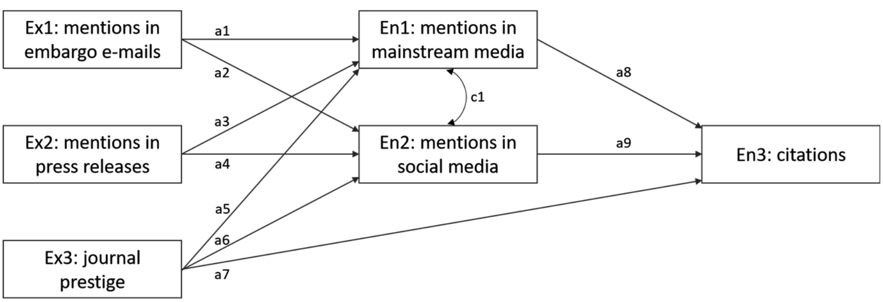

Figure 1: Path model with three exogenous variables (Ex1-3), three endogenous variables (En1-3), nine assumed effects between these (a1-9), and one assumed covariance (c1).

The reasoning behind the supposed effects in our model (a1-9) is based on positive associations found in past empirical studies:

- Such suggesting that a research article that is featured within an embargo e-mail will be more likely to be mentioned in mainstream and social media (a1 and a2) have been found in a case-control study by Lemke et al. (2022). Similarly, the case studies by Fanelli (2013), Lemke (2020), and Stryker (2002) all found positive associations between articles being promoted in press releases and their later mainstream media mentions (a3). Regarding indicators for online attention on the other hand, the findings by Chapman et al. (2007) and Lemke (2020) suggest that press release promotion is associated with an increase of these as well (a4).
- The positive correlation between journal prestige (*a7*), most commonly represented by journal impact factors, and citations has been confirmed by numerous studies (see Tahamtan et al., 2016). Correlations between journal prestige and certain altmetrics (a5 and a6) have, for instance, been found by

Li & Thelwall (2012) and Thoma et al. (2015). Studies on the positive correlation between mentions in news media and future citations (a8) have already been discussed in detail in this article's introduction (Dumas-Mallet et al., 2020; Fanelli, 2013; Kiernan, 2003; Phillips et al., 1991). The (varying) potential of social media-based altmetrics to predict later citations (a9) is proposed by another rich body of studies (see, for instance, Konkiel et al., 2016; Priem et al., 2012; Thelwall & Nevill, 2018).

While cases are imaginable in which high numbers of citations for an article precede it being mentioned in news media or on social media, we for now consider this a less likely case due to findings that indicate that most attention around research articles in social and news media usually happens soon after their publication (Brainard, 2022; Waltman et al., 2021), while the majority of a scientific article's citations typically occur two to seven years later (Schloegl & Gorraiz, 2010). Similarly, a positive direct effect of having a press release on expected subsequent citations is conceivable. However, the findings by Fanelli (2013) indicate that this effect likely is already covered to a large extent by the combined effect of positive associations between press release promotion and mainstream media mentions (a3) and the positive correlation between mentions in news media and future citations (a8). Finally, regarding the presumably complex reciprocal relationship of social media content spawning news content and vice versa – as for example supported by correlations found by Haustein et al. (2015) or Lemke et al. (2022) – we do not assume a unidirectional causal effect, but model the relationship as a covariance (c1) instead.

We use maximum likelihood estimation with robust standard errors and a Satorra-Bentler scaled test statistic (Satorra & Bentler, 2010) for model estimation, as citation and altmetrics data are usually not normally distributed. For the evaluation of the models' global fit, we also consult the (robust) comparative fit index (CFI), root mean squared error of approximation (RMSEA), and standardized root mean squared residual (SRMR) and apply established cutoff criteria proposed by Hu & Bentler (1999). To evaluate models' local fit and identify potential improvements, we calculate modification indices with a cutoff value of 10 (Regorz, 2021). All statistical analyses are performed in R (R Core Team, 2020).

2 Data sampling

To analyze the relationships between mentions of research articles in external science communication and their performance regarding citations and altmetrics, we start with data obtained from *EurekAlert!*[2]. EurekAlert! is a platform for the distribution of research-related press and news releases that was set up by the American Association for the Advancement of Science in 1996. The platform enables publishers, universities, research institutes, corporations, government agencies, and eligible organizations that engage in scientific research to disseminate press releases to journalists and the public against payment of submission fees. With over 14,000 registered journalists from more than 90 countries, EurekAlert! has become the most prominent multilingual platform of its kind (Vrieze, 2018). Or, as Vrieze (2018) put it: right now, EurekAlert! has become for science press and news releases "what *Google* is for searching and *Amazon* for online shopping."

We focus our analysis on research articles published between 2016 and 2018 to balance the timeliness of the research analyzed in our study with the ability to obtain meaningful citation windows. Thus, the starting point for our dataset is data on 84,194 press releases provided to us by EurekAlert! that were published on the platform between 2016 and 2018. Most of these (79,419, 94.33%) are categorized by EurekAlert! as type "research." In addition, EurekAlert! press releases on new research articles frequently contain a DOI link to the respective article – this enables us to extract 41,937 DOIs. Of these identifiers, 34,055 refer to a valid Web of Science record with publication type "journal," document type of either "article" or "review," and a publication year between 2016 and 2018. These 34,055 DOIs form the starting point for our bibliometric analysis.

As a next step, we enrich this data with information about the presence of research publications in publishers' embargo e-mails to journalists. For this, we rely on data from the Science Media Center Germany (SMC). The SMC is an editorially independent non-profit institution with the mission of supporting journalists in reporting on science-related topics. As one service contributing to this mission, the SMC regularly sends out comments by scientific experts on new research findings that are still under embargo. To be able to provide this service, the SMC aims to monitor as many scholarly journals that send out embargo e-mails as possible. Since 2016, the SMC has accumulated 2,638 ingoing e-mails identified as

2 https://www.eurekalert.org/

embargo e-mails from scholarly publishers. Each of these embargo e-mails contains information about one or more upcoming articles from one or more journals belonging to the same publisher.

As we rely on DOIs to track articles' citations and altmetrics, we query the SMC's e-mail archive for embargo e-mails containing references to scientific articles via DOI. This led to 953 articles with Web of Science records of document type "article" or "review," publication type "journal," and publication year 2016, 2017, or 2018 that also have been promoted in an embargo e-mail between 2016 and 2018 with reference to a DOI. Merging these articles with our dataset of 34,055 articles promoted on EurekAlert! enlarges our sample to 34,413 articles that received promotion in an embargo e-mail and/or a press release on EurekAlert!. Table 1 shows to which extent the two regarded types of promotional activities overlap within our dataset. The two events are not statistically independent from each other (Fisher's exact test, two-sided, $p < .001$).

Table 1: Contingency table of promotion in embargo e-mails and on EurekAlert! for research articles in our sample.

		Featured in embargo e-mail?	
		No	Yes
Featured on EurekAlert!?	No	0	358
	Yes	33,460	595

We also added "control group articles" to our dataset – articles which, to our knowledge, did not receive any promotion in EurekAlert! press releases or publishers' embargo e-mails, but otherwise should have been published under comparable circumstances as our "treatment group articles." The addition of a control group enables us to assess the effects of external promotion via such comparison. To do so, we match every article from the treatment group (i.e., the group of articles that received the "treatment" of getting promoted within an embargo e-mail, an EurekAlert! press release or both) to one random article from Web of Science that was published in the same publication year and with the same ISSN, but which is not part of the treatment or control group yet. We again restrict our matching to the Web of Science document types "article" and "review" to avoid matching research articles with, for instance, editorials, letters to the editor, etc.

For articles from multidisciplinary journals (e.g., *PLOS ONE*, *Nature*, *Science*), this procedure might lead to suboptimal matchings of articles from domains with highly heterogeneous citing or publication behaviors – it would, for instance, be questionable to match a biomedical article from *PLOS ONE* with a sociology-related article from *PLOS ONE*. For articles published in journals classified as multidisciplinary in Web of Science (24.65% of our sample), we therefore apply an additional step, inspired by the matching methodology described in Fraser et al. (2020). To do so, we reclassify these articles and all potential control group articles' disciplines based on the most frequently cited Web of Science subject categories among their references and subsequently use concurrence among these new classifications as an additional matching criterion for articles from multidisciplinary journals.

It should be noted that for a relatively small number of articles from our sample (3.62%), our control group matching procedure does not return a valid match fulfilling the criteria explained above. Thus, a total of 67,581 unique publications serves as our dataset for model estimation.

For bibliometric analyses we use data provided by the *Competence Centre for Bibliometrics* (CCB). The CCB administers databases based on Web of Science, which are updated annually. The bibliographic and citation data used in this study therefore reflects the status of Web of Science from April 2020. The altmetric data used in this study (i.e., articles' numbers of mentions in mainstream media and on Twitter) was obtained via the API of Altmetric.com in November 2021.

3 Results

Before model estimation, we briefly examine some of the articles' metadata to achieve an understanding of our dataset's composition. In total, 3,419 individual journals are represented within our dataset, the most frequently represented journals being *Nature Communications* (5.38% of all articles), *Scientific Reports* (4.26%), *PNAS* (3.96%), *PLOS ONE* (2.62%), and *Nature* (2.57%). The frequency of the remaining journals follows in a long tail distribution, with most journals (1,450) only represented by a single article within our treatment and control group. Applying traditional Web of Science subject categories, a total of 241 different categories are represented in our data. Table 2 shows the ten most strongly represented journals and Web of Science subject categories from our sample and their respective shares. The outstandingly high share of the category Multidisciplinary Sciences (15.32%) appears to back up what the examination of most commonly represented journals had also shown – namely, that prominent multidisciplinary journals like *Nature*, *PNAS*, or *PLOS ONE* account for a large share of the research that gets featured in press releases and/or embargo e-mails. Also, a look at the most heavily represented categories suggests a particularly substantial representation of research dealing with biomedical subjects.

Table 2: Most frequent journals and Web of Science subject categories among our sample.

Journal	Freq.	Subject Category	Freq.
Nature Communications	5.38%	Multidisciplinary Sciences	15.32%
Scientific Reports	4.26%	Cell Biology	4.95%
PNAS	3.96%	Biochemistry & Molecular Biology	3.97%
PLOS ONE	2.62%	Neurosciences	2.89%
Nature	2.57%	Materials Science, Multidisciplinary	2.54%
Science	2.24%	Chemistry, Multidisciplinary	2.41%
Science Advances	1.33%	Biology	2.35%
eLife	1.14%	Ecology	2.33%
Cell Reports	1.11%	Medicine, General & Internal	2.19%
Physical Review Letters	1.05%	Environmental Sciences	2.03%

Four of the six variables in our model are metrically scaled: articles' citation counts, numbers of mentions in tweets, numbers of mentions in mainstream media (MSM mentions), and the median of citations articles within the respective journal received during the past three years (abbreviated as JCM or journal citation median). Table 3 shows descriptive statistics of our sample regarding these variables.

Table 3: Descriptive statistics of metric variables within the model.

	Citations	Tweet mentions	MSM mentions	JCM
Minimum	0	0	0	0
Mean	26.92	39.25	8.70	12.37
Median	12	7	2	8
Maximum	13,715	9,951	533	876
Standard deviation	106.98	150.13	20.51	14.10

According to our research aim and method, we estimate a path model based on our publication data and the specification outlined in Figure 1. To account for the considerable differences in their variances, the four metric variables included in our model (citations, tweet mentions, mainstream media mentions, and journal citation median) were all standardized via z-transformation (subtraction of mean and division by standard deviation for each observation) before model estimation.

Table 4: Global fit indices for our first model.

	CFI	RMSEA	SRMR
Model 1	1.000	0.010	0.003

Table 5: Global fit indices for our second model.

	CFI	RMSEA	SRMR
Model 2	1.000	0.001	0.001

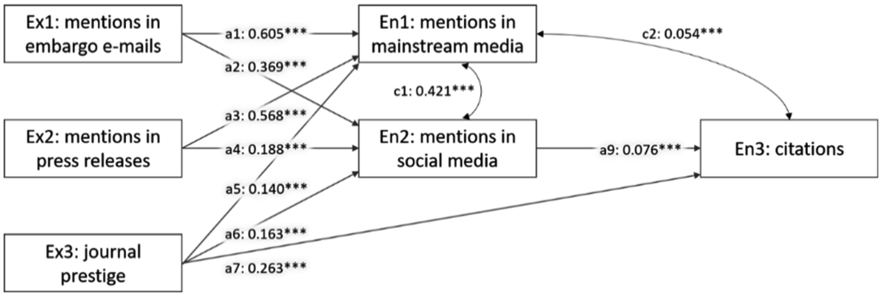

Figure 2: Final path model with path coefficients. *** indicate $p < 0.001$.

We evaluate our model's global fit before we consult the estimated model's coefficients. As a result, the χ^2 test for exact model fit is significant ($\chi^2 = 11.503$, df = 2, p = .003). The comparative fit index (CFI), root mean squared error of approximation (RMSEA), and standardized root mean squared residual (SRMR) for our model are shown in Table 4.

Applying cutoff criteria recommended by Hu & Bentler (1999), the CFI > 0.95, RMSEA < 0.06, and SRMR < 0.08 all indicate an already relatively good fit between our hypothesized model and observed data. The significance of the χ^2 test is undesirable, so we are interested in whether further substantial improvements to the model are possible. Using the lavaan package's function modindices, we calculate modification indices to see if and how the change of existing or addition of further effects to our model could increase its fit with the data. The suggested model change associated with the highest expected model improvement is the replacement of the unidirectional effect a8 by a covariance – so the abandonment of the assumption that mainstream media mentions have a mostly unidirectional effect on citations in favor of a model in which no clear causal direction between citations and media mentions is assumed. As such a non-unidirectional relation might make sense from a theoretical standpoint as well, we respecify the model accordingly and again assess its global fit through a χ^2 test ($\chi^2 = 2.161$, df = 2, p = 0.339) and the calculation of fit indices shown in Table 5.

Table 6: Final path model's parameter estimates.

	Estimate	Standard error	*z*-value	p
a1	0.605	0.062	9.773	<0.001
a2	0.369	0.061	6.075	<0.001
a3	0.568	0.007	77.265	<0.001
a4	0.188	0.008	24.859	<0.001
a5	0.140	0.010	14.608	<0.001
a6	0.163	0.011	15.248	<0.001
a7	0.263	0.012	22.336	<0.001
a9	0.076	0.014	5.385	<0.001
b1	0.421	0.036	11.699	<0.001
b2	0.054	0.015	3.661	<0.001

Both χ^2 test and fit indices indicate an improvement in global model fit compared to the first model. As another iteration of modification index calculation does not reveal any further model changes that would be linked to substantial expected improvements (applying our cutoff value of 10 for modification indices), we continue with the interpretation of the model's coefficients. Figure 2 shows the resulting model along with its path coefficients and covariances, and Table 6 provides additional statistics. Table 7 shows R^2-values for the three endogenous variables within the model.

Table 7: Endogenous model variables' R^2-values.

	Citations	Tweet mentions	MSM mentions
R^2	0.081	0.037	0.104

4 Discussion

The obtained fit indices and significances of effects suggest a good fit between the model we hypothesized based on findings from past case studies (Chapman et al., 2007; Dumas-Mallet et al., 2020; Fanelli, 2013; Kiernan, 2003; Konkiel et al., 2016; Lemke, 2020; Lemke et al., 2022; Li & Thelwall, 2012; Phillips et al., 1991; Priem et al., 2012; Stryker, 2002; Tahamtan et al., 2016; Thelwall & Nevill, 2018; Thoma et al., 2015) and the actual publication data at hand. Thus, the first stated purpose of our analysis, which was to test whether our hypotheses about the interactions between events of external science communication and article metrics drawn from past studies conform with a large set of empirical observations, can be considered as fulfilled.

Regarding the effects of exogenous variables on various media mentions, our final model suggests that promotion with embargo e-mails seems to affect articles' expected mainstream media presence to a slightly larger (0.605) but similarly high degree as promotion with a press release (0.568). Furthermore, considering social media presence, the comparatively higher strength of embargo e-mail promotion as a predictor for later mentions becomes even more apparent (0.369, opposed to 0.188 from press releases). As it might seem counterintuitive that embargo e-mails would to a greater extent contribute directly to an article's visibility on various media than the more openly accessible format of a press release, these findings might also suggest that embargo e-mail promotion (as tracked by the Science Media Center Germany) is – compared to press release promotion – reserved for even more elite research publications, which due to innate qualities not represented in our model (e.g., particular originality, societal value, or some form of provocativeness, to just name a few possibilities) will likely attract more media attention on their own. If the selection of articles for embargo e-mail promotion indeed typically follows more rigorous criteria than the selection for press releases, this finding might also be considered a hint towards the validity of the earmark hypothesis as suggested by Phillips et al. (1991). However, further research on publishers', journals', and institutions' criteria behind the selection of research articles for both these forms of promotion would be necessary to solidify this hypothesis.

Our third exogenous variable, journal prestige, proves to be a significant predictor for citations (0.263), mainstream media mentions (0.140), and Twitter mentions alike (0.163) – the direct effect on citations, however, is most substantial. What might be considered a surprising finding are the fairly weak relationships between both mainstream and social media mentions and citations (0.054 and 0.076, respectively). Possibly, controlling for journal prestige (which, as we have seen, is also a strong predictor for mentions in both forms of media) already accounts for most of the citation advantage expected from increased media presence; this interpretation is in contrast to findings by Dumas-Mallet et al. (2020) though, who for their biomedical sample found the expected citations of articles from lower impact journals to benefit particularly strongly from media mentions. Another interesting and perhaps surprising insight concerns the observation that led to our model respecification, namely that a model without the assumed directional effect of mainstream media mentions on citations fits the empirical data better than our initial model, where this effect was present. This finding might hint at a more bidirectional relationship between academic impact and media attention than past case studies suggested, which in line with Phillips et al.'s (1991) publicity hypothesis often focused on how media exposure might increase citations without much consideration of the opposite phenomenon of outstanding research evoking media attention.

Our findings add to existing case studies on associations between media coverage of research and said research's impact by taking a more comprehensive perspective than past studies, which mostly focused on fewer variables, and by analyzing a large sample of articles from a wide variety of journals. Also, to the best of our knowledge, there has neither been a comparative analysis of the effects of embargo e-mails and press releases before, nor an application of path analysis in a large-scale bibliometric analysis like ours. For the interpretation of the results, readers should however keep in mind that our sampling approach started with articles featured on either EurekAlert! or within an embargo e-mail tracked by the SMC Germany and that our study therefore remains a case study whose generalizability might be limited by its sample. Moreover, with its limited number of variables considered, our study can only serve as a starting point for disentangling the complex relationships and effects between the systems of science communication and academic reputation.

Our study comes with some further limitations. First, it is virtually impossible to prove that "control group articles" did not ever receive any kind of external promotion similar to a press release or an embargo e-mail that was not tracked by our data sources. However, as both observed kinds of promotion will most likely be the exception rather than the rule among randomly sampled articles as done in this study, we consider it unlikely that this limitation would have impaired our results significantly.

A second limitation results from our reliance on DOIs to identify references on EurekAlert! and in the SMC's database. While the high amount of DOIs found within the two data sources (which on EurekAlert! has been increasing over the years, see also Chapter 1 in this book) suggests that they are a common way of referencing articles within them, our DOI-based approach means that our data likely underestimates the total shares of articles receiving promotion within press releases or embargo e-mails.

Third, as is the case for many bibliometric and altmetric analyses, results should be interpreted with the data sources used to obtain metrics in this study in mind, as with our reliance on these sources we also inherit some of their limitations; e.g., the limitation of Web of Science to publications indexed by it, or the limited transparency of what Altmetric.com tracks as mainstream media mentions and what it does not (for a recent assessment of Altmetric.com's news mention data see also Fleerackers et al., 2022). The representation of subject fields encountered in this study (see Table 2) suggests that the promotional formats of science press releases and embargo e-mails topically are dominated by natural sciences and, more specifically, biomedical subjects. The extent to which our model also applies to other, less publicly visible fields of research, e.g., the humanities, is an aspect that future research should investigate more deeply. Moreover, the national foci of EurekAlert! (likely towards sources from the United States, see also Bowman & Hassan (2019)) as well as the SMC Germany (towards press materials sent to journalists based in Germany) might have an influence on our results, which should be assessed more precisely in the future.

Overall, our findings support the existence of statistically significant associations between the promotion of research in science PR (i.e., embargo e-mails and press releases) and impact metrics that past case studies had found for individual parts of our model's components and smaller, more restricted samples of scientific articles (Dumas-Mallet et al., 2020; Fanelli, 2013; Kiernan, 2003; Lemke, 2020; Lemke et al., 2022; Phillips et al., 1991). It thus underlines the importance of further, more in-depth research on the selection criteria with which PR officers and science journalists decide on which research to cover (see, for instance, Badenschier & Wormer, 2012; Broer, 2020), as these criteria might diverge substantially from the characteristics that metrics-based indicators are supposed to reflect when used for evaluative purposes. Also, although not at the center of our inquiry, our look at the discipline- and journal-wise composition of our sample of articles promoted in either press releases or embargo e-mails indicates that large shares of them were published in multidisciplinary mega-journals and cover subjects from life sciences, which would confirm observations of media coverage of science made in previous studies (e.g., Elmer et al., 2008; Lemke et al., 2021).

It is important to emphasize that the effects observed in our model cannot readily be assumed to prove causal relationships between articles' external promotion and increased metrics. While past case studies provide convincing arguments for the significance of such causal, publicity-related effects (the results of the case-control study by Phillips et al. (1991), the region-specific differences in effect size found by Fanelli (2013), and the link between numbers of press mentions and citations found by Dumas-Mallet et al. (2020) can all be considered such arguments), other, non-causal phenomena might explain the associations shown by the model as well. Such phenomena are, for instance, backed up by Weingart's (1998) theory of the medialization of science, which would explain an increasing convergence between the criteria with which press offices select publications for promotion and the criteria with which researchers choose research subjects (which therefore could also be expected to experience overall rises in impact metrics like citations). To clarify the precise degree to which such causal and non-causal links explain the associations seen in our model, additional qualitative investigations of the selection mechanisms behind press releases, embargo e-mails, and different metric events will be necessary (see also Lemke, 2022). What the model derived in our study can provide, however, is a quantification of the potential magnitude to which promotional measures and impact metrics are linked to each other.

5 Conclusions

We specified a path model of the interplay between two prevalent measures of external science communication, journal prestige, presence in mainstream and on social media, as well as citations, and tested the model against a large set of empirical data from Web of Science, Altmetric.com, EurekAlert!, and SMC Germany. The empirical results confirmed the significance of the effects assumed in the model and signaled substantial associations between the three exogenous variables and articles' expected later impact in both media and academia. In particular, the results highlighted the considerable potential effects of embargo e-mails and press releases on (social) media attention and of journal prestige on citations.

Future research should work on providing a more detailed picture of the criteria that affect an individual research article's likelihood of receiving certain forms of promotion in external science communication, as well as investigate which article properties (for instance, considering topics, authors, or publication venues) influence respective promotional activities' effectiveness with regard to impact generation. Furthermore, as citation norms and behaviors as well as the ways media discusses research findings can vary considerably between fields, a worthwhile continuation of this analysis could lie in the specification, estimation, and comparison of discipline-specific models of the interdependencies examined in this study.

Finally, it should be noted that we focused our analysis on six variables that we deemed particularly relevant and that were generated via literature research to explain the phenomena we aimed to explore. However, there certainly are more factors that might affect the likelihood of research articles being promoted in external science communication, featured on various media platforms, or cited within other publications, that could, in principle, be included in an analysis like ours. In fact, we would argue that within a context as intricate as the system of science communication, no matter which number of factors a model considers, there will probably always be more one could add. Nevertheless, the incorporation of further variables would be another promising path for future research that aims to build upon this work to take – for instance, including a parametrization of authors' prestige might lead to new valuable insights on the interplay between promotion and received attention of research articles, to name just one example. Linked to this, a potential methodological continuation of this research would be the utilization of structural equation modeling, where even factors as abstract as "pure luck" could be incorporated into the model.

6 Acknowledgments

We thank Sophie Rotgeri for her crucial support regarding the provision of data on embargo e-mails by Science Media Center Germany and on press releases from EurekAlert!, as well as Nicholas Fraser for his input regarding the control group matching algorithm for articles from multidisciplinary journals. Moreover, we thank the SMC, EurekAlert!, and the Competence Centre for Bibliometrics for providing us with access to their databases. This study is part of the research project MeWiKo funded by the German Federal Ministry of Education and Research (grant number 01PU17018).

7 References

Aksnes, D.W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open, 9*(1), 2158244019829575. https://doi.org/10.1177/2158244019829575

Badenschier, F., & Wormer, H. (2012). Issue selection in science journalism: Towards a special theory of news values for science news? In S. Rödder, M. Franzen, & P. Weingart (Hrsg.), *The sciences' media connection – public communication and its repercussions* (S. 59-85). Springer Netherlands. https://doi.org/10.1007/978-94-007-2085-5_4

Berghaeuser, H., & Hoelscher, M. (2020). Reinventing the third mission of higher education in Germany: Political frameworks and universities' reactions. *Tertiary Education and Management, 26*(1), 57-76. https://doi.org/10.1007/s11233-019-09030-3

Bowman, T.D., & Hassan, S.-U. (2019). *Science news and altmetrics: Looking at EurekAlert!* 5. http://altmetrics.org/wp-content/uploads/2019/10/Bowman_altmetrics19_paper_6.pdf

Brainard, J. (2022). Riding the Twitter wave. *Science*. https://doi.org/10.1126/science.abq1541

Broer, I. (2020). Rapid reaction: Ethnographic insights into the Science Media Center and its response to the COVID-19 outbreak. *Journal of Science Communication, 19*(5), A08. https://doi.org/10.22323/2.19050208

Carver, R.B. (2014). Public communication from research institutes: Is it science communication or public relations? *Journal of Science Communication, 13*(3), C01. https://doi.org/10.22323/2.13030301

Chapman, S., Nguyen, T.N., & White, C. (2007). Press-released papers are more downloaded and cited. *Tobacco Control, 16*(1), 71-71. https://doi.org/10.1136/tc.2006.019034

Dumas-Mallet, E., Garenne, A., Boraud, T., & Gonon, F. (2020). Does newspapers coverage influence the citations count of scientific publications? An analysis of biomedical studies. *Scientometrics, 123*(1), 413-427. https://doi.org/10.1007/s11192-020-03380-1

Elmer, C., Badenschier, F., & Wormer, H. (2008). Science for everybody? How the coverage of research issues in German newspapers has increased dramatically. *Journalism & Mass Communication Quarterly, 85*(4), 878-893. https://doi.org/10.1177/107769900808500410

Fanelli, D. (2013). Any publicity is better than none: Newspaper coverage increases citations, in the UK more than in Italy. *Scientometrics, 95*(3), 1167-1177. https://doi.org/10.1007/s11192-012-0925-0

Fleerackers, A., Nehring, L., Maggio, L.A., Enkhbayar, A., Moorhead, L., & Alperin, J.P. (2022). *Identifying science in the news: An assessment of the precision and recall of Altmetric.com news mention data*. Scientometrics, *127*, 6109-6123. https://doi.org/10.1007/s11192-022-04510-7

Fraser, N., Momeni, F., Mayr, P., & Peters, I. (2020). The relationship between bioRxiv preprints, citations and altmetrics. *Quantitative Science Studies, 1*(2), 618-638. https://doi.org/10.1162/qss_a_00043

Haustein, S., Costas, R., & Larivière, V. (2015). Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns. *PLOS ONE, 10*(3), e0120495. https://doi.org/10.1371/journal.pone.0120495

Hermida, A. (2012). Social journalism: Exploring how social media is shaping journalism. In *The handbook of global online journalism* (p. 309-328). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118313978.ch17

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Kiernan, V. (1997). Ingelfinger, embargoes, and other controls on the dissemination of science news. *Science Communication, 18*(4). https://doi.org/10.1177/1075547097018004002

Kiernan, V. (2003). Diffusion of news about research. *Science Communication, 25*(1), 3-13. https://doi.org/10.1177/1075547003255297

Kiernan, V. (2014). Public relations practices at medical journals. *Learned Publishing, 27*(1), 5-13. https://doi.org/10.1087/20140102

Kiesslich, T., Beyreis, M., Zimmermann, G., & Traweger, A. (2021). Citation inequality and the journal impact factor: Median, mean, (does it) matter? *Scientometrics, 126*(2), 1249-1269. https://doi.org/10.1007/s11192-020-03812-y

Konkiel, S., Madjarevic, N., & Lightfoot, A. (2016). *Altmetrics for librarians: 100+ tips, tricks, and examples*. https://doi.org/10.6084/m9.figshare.3749838.v2

Lemke, S. (2020). *The effect of press releases on promoted articles' citations and altmetrics*. Metrics 2020: ASIS&T Virtual Workshop on Informetrics and Scientometrics Research. https://doi.org/10.5281/zenodo.4351360

Lemke, S. (2022). *An assessment of impact metrics' potential as research indicators based on their perception, usage, and dependencies from external science communication* [Doctoral thesis, CAU Kiel University]. https://macau.uni-kiel.de/receive/macau_mods_00003250

Lemke, S., Brede, M., Rotgeri, S., & Peters, I. (2022). Research articles promoted in embargo e-mails receive higher citations and altmetrics. *Scientometrics, 127*, 75-97. https://doi.org/10.1007/s11192-021-04217-1

Lemke, S., Sakmann, J., Brede, M., & Peters, I. (2021). Exploring the relationship between qualities of press releases to research articles and the articles' impact. In *Proceedings of the 2021 International Conference on Scientometrics and Informetrics*, 639-644.

Li, X., & Thelwall, M. (2012). F1000, Mendeley and traditional bibliometric indicators. In *Proceedings of the 17th international conference on science and technology indicators*, 451-551.

Liang, X., Su, L. Y.-F., Yeo, S.K., Scheufele, D.A., Brossard, D., Xenos, M., Nealey, P., & Corley, E.A. (2014). Building buzz: (Scientists) communicating science in new media environments. *Journalism & Mass Communication Quarterly, 91*(4), 772-791. https://doi.org/10.1177/1077699014550092

Phillips, D.P., Kanter, E.J., Bednarczyk, B., & Tastad, P.L. (1991). Importance of the lay press in the transmission of medical knowledge to the scientific community. *New England Journal of Medicine, 325*(16), 1180-1183. https://doi.org/10.1056/NEJM199110173251620

Priem, J., Piwowar, H.A., & Hemminger, B.M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. *ArXiv:1203.4745 [Cs]*. http://arxiv.org/abs/1203.4745

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010, Oktober 26). *Altmetrics: A manifesto* [Altmetrics]. http://altmetrics.org/manifesto/

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Regorz, A. (2021). *Einführung in die Pfadanalyse mit R/lavaan*.

Satorra, A., & Bentler, P.M. (2010). Ensuring positiveness of the scaled difference Chi-square test statistic. *Psychometrika, 75*(2), 243-248. https://doi.org/10.1007/s11336-009-9135-y

Schlögl, C., & Gorraiz, J. (2010). Comparison of citation and usage indicators: The case of oncology journals. *Scientometrics, 82*(3), 567-580. https://doi.org/10.1007/s11192-010-0172-1

Streiner, D.L. (2005). Finding our way: An introduction to path analysis. *The Canadian Journal of Psychiatry, 50*(2), 115-122. https://doi.org/10.1177/070674370505000207

Stryker, J.E. (2002). Reporting medical information: Effects of press releases and newsworthiness on medical journal articles' visibility in the news media. *Preventive Medicine, 35*(5), 519-530. https://doi.org/10.1006/pmed.2002.1102

Tahamtan, I., Afshar, A.S., & Ahamdzadeh, K. (2016). Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics, 107*(3), 1195-1225. https://doi.org/10.1007/s11192-016-1889-2

Tahamtan, I., & Bornmann, L. (2020). Altmetrics and societal impact measurements: Match or mismatch? A literature review. *El Profesional de La Información, 29*(1). https://doi.org/10.3145/epi.2020.ene.02

Thelwall, M., & Nevill, T. (2018). Could scientists use Altmetric.com scores to predict longer term citation counts? *Journal of Informetrics, 12*(1), 237-248. https://doi.org/10.1016/j.joi.2018.01.008

Thoma, B., Sanders, J.L., Lin, M., Paterson, Q.S., Steeg, J., & Chan, T M. (2015). The social media index: Measuring the impact of emergency medicine and critical care websites. *Western Journal of Emergency Medicine, 16*(2), 242-249. https://doi.org/10.5811/westjem.2015.1.24860

Vrieze, J. de. (2018, September 27). EurekAlert! Has spoiled science news. Here's how we can fix it. *VWN - Vereniging voor Wetenschapsjournalistiek en - communicatie Nederland*. https://www.vwn.nu/?p=2225/

Waltman, L., Pinfield, S., Rzayeva, N., Oliveira Henriques, S., Fang, Z., Brumberg, J., Greaves, S., Hurst, P., Collings, A., Heinrichs, A., Lindsay, N., MacCallum, C.J., Morgan, D., Sansone, S.-A., & Swaminathan, S. (2021). *Scholarly communication in times of crisis: The response of the scholarly communication system to the COVID-19 pandemic* [Report]. Research on Research Institute. https://doi.org/10.6084/m9.figshare.17125394.v1

Weingart, P. (1998). Science and the media. *Research Policy, 27*(8), 869-879. https://doi.org/10.1016/S0048-7333(98)00096-1