Article

Content and Phrasing in Titles of Original Research and Review Articles in 2015: Range of Practice in Four Clinical Journals

Mary Ellen Kerans [1],*, Anne Murray [2] and Sergi Sabatè [3]

[1] Freelance authors' editor and translator, Carrer Indústria 331, àtic 2a, 08027 Barcelona, Spain

[2] Freelance authors' editor and translator, Carrer Major 17, 43422 Barberà de la Conca, Spain;
annetrans@gmail.com

[3] Department of Anesthesiology, Fundacio Puigvert, Carrer Cartagena 340-350, 08025 Barcelona, Spain;
sergisabate@gmail.com

* Correspondence: mekerans@gmail.com; Tel.: +34-934-080-997

Abstract: Reporting guidelines for clinical research designs emerged in the mid-1990s and have influenced various aspects of research articles, including titles, which have also been subject to changing uses with the growth of electronic database searching and efforts to reduce bias in literature searches. We aimed (1) to learn more about titles in clinical medicine today and (2) to develop an efficient, reliable way to study titles over time and on the fly—for quick application by authors, manuscript editors, translators and instructors. We compared content and form in titles from two general medical journals—the *New England Journal of Medicine* (NEJM) and the *British Medical Journal*—and two anesthesiology journals (the *European Journal of Anaesthesiology* and *Anesthesiology*); we also analyzed the inter-rater reliability of our coding. Significant content differences were found in the frequencies of mentions of methods, results (between general and subspecialty titles), and geographic setting; phrasing differences were found in the prevalence of full-sentence and compound titles (and their punctuation). NEJM titles were significantly shorter, and this journal differed consistently on several features. We conclude that authors must learn to efficiently survey titles for form and content patterns when preparing manuscripts to submit to unfamiliar journals or on resubmitting to a new journal after rejection.

Keywords: academic writing; search strategies; reading strategies; titles; English as an additional language (EAL); English for research publication purposes (ERPP); English for specific purposes (ESP)

1. Introduction

Today's clinical researchers are understandably convinced of the importance of title content and form in English. Modern readers often evaluate these short texts in isolation when doing literature searches in PubMed and other electronic indexes, and sometimes titles are all that busy readers look at before discarding an article. Whereas potential users of information (for clinical application or citing) once saw titles and abstracts together when browsing their preferred journals in print, readers now use time-saving titles-first-or-only search strategies online. A 2001 survey at one medical school reported that faculty members (two-thirds of whom received their last degree between 1970 and 1990) still relied on their own print subscriptions for 70% of their reading [1]. In contrast, today's medical researchers are encouraged to obtain and update their reference lists more systematically by also doing exhaustive online searches in several indexes and avoiding language filters [2]. A titles-first approach has been shown to be more efficient than a traditional titles-plus-abstract approach for such searches and to yield the same reference results [3]; consequently, even large groups dedicated to systematic reviewing

may now advise that screening the title "and/*or* abstract" (our emphasis) will allow many articles to be quickly rejected (e.g., [4]).

A title's potential effect on an article's survival probably begins much earlier, however. At the *British Medical Journal* (BMJ), the in-house screening process (before peer review) in 2008 was reported to begin with a handling editor's initial assessment of "interest" based on reading the abstract and skimming the paper (average time, 7 min) [5]. Eight years later, we suspect many editors may spend even less time on this initial step—given that the criterion is only interest—and are tempted to take the same titles-first approach recommended to systematic reviewers [3,4]. This trend toward making a first judgment based on a title has been facilitated by the reporting guidelines for specific research designs that began to emerge about 20 years ago. The first guideline to appear (1996) addressed how to report a randomized controlled trial (the CONSORT statement—for Consolidated Standards of Reporting [6]); it stipulated that research design should be mentioned in the "title *or* abstract". The uptake of CONSORT recommendations accelerated with open access to the 2001 revision [7], and more reporting guidelines began appearing for a larger range of clinical studies—qualitative research, diagnostic and prognostic studies, and many more. All made mention of CONSORT's earlier title-*or*-abstract formula for naming the research design. However, both the newest guidelines for reporting systematic reviews and meta-analyses (2009) [8] and the most recent CONSORT revision (2010) [9] now advise making research design explicit in the title itself. Certainly, complying with this advice will make it easier for both editors and readers to screen articles with these short, content-packed texts—some of which are so awkwardly long they have been called mini-abstracts [10].

We hypothesized that as the twentieth anniversary of CONSORT approaches, reporting guidelines have taken root in most clinical disciplines and practice has become fairly stable, so that a study of current titling practices across journals will provide information that will be useful for some time to come. Many besides authors have a stake in understanding the range of practices and possible differences in journal preferences. Stakeholders include authors' pre-submission manuscript editors—especially where authors are users of English as an additional language (EAL)—as well as translators working for authors or publishers. Accurate information about titling patterns could also be valuable for instructors of English for research publication purposes (ERPP) and university instructors of English for specific purposes (ESP). Even ESP instructors who work with undergraduates or non-researching clinicians are concerned with titles, as the critical reading of search results pages is the necessary starting point for further critical reading [11] for an evidence-based practice of medicine.

To date, the objectives and scope of the few studies on clinical research article (RA) titles have varied greatly. Editor-physicians who do research into publication practices are also users of information from published articles and have mainly focused on title content [10,12–16] rather than detailed language features. This group is particularly concerned with whether research design is included or not (Table 1). Some have also looked at other content and phrasing features potentially associated with citing, notably length [14–16]. Their findings for that question have been contradictory ([14,15] *vs.* [16]), suggesting a need for continued research. It is possible that discrepancies like this stem from how they compiled their title corpora, especially the decision to mix journals publishing clinical research with others focusing on bench research [15,16]. Applied linguists, in contrast, have logically looked at various surface structures besides length in the interest of informing ERPP and ESP instructors, who must teach forms (Table 2). We found that many linguistics studies compared titles in several disciplines, so the table summarizes the only three that looked at clinical RAs in corpora that did not predate the introduction of the CONSORT statement in 1996. Like the editor-physicians, the linguists took different approaches to building a corpus. One looked only at *New England Journal of Medicine* (NEJM) titles [17], one used a convenience sample of veterinary journals to extract and analyze examples for an ESP course [18], and one compared titles in various disciplines [19], including the "biological sciences"—a category that encompasses bench sciences, for which reporting guidelines development started later [20]. Thus, the literature we reviewed could only provide us with a starting point, suggesting which variables were relevant if we wanted to

be able to compare our findings to others'. Besides variety in corpus development, methodological or reporting problems have made interpretation difficult. For example, one linguistics study that compared disciplines [19] did not test for statistical significance of differences, although it is difficult to know if studies in fields with small samples have the statistical power for valid comparisons in any case.

After reviewing this literature, we set out with two objectives for this study: (1) to learn more about the previously studied features of titles in clinical medicine 20 years after reporting guidelines began to influence practice and (2) to develop an efficient, reliable way to study titles over time and on the fly—for quick application by authors, manuscript editors, translators and instructors. We limited our attention to two highly respected general medical journals and two anesthesiology journals, one of which ranks at the top of this subspecialty and the other considerably lower. We chose to contrast a subspecialty with general medicine to test the following hypothesis: although patterns in high-ranked general medical journals—such as the often-studied NEJM [10,12,15,17]—are assumed to reflect a gold standard of language usage that can safely be followed, the more tightly knit discourse communities of subspecialties may use different patterns. If that is the case, authors and those who support them need to know this because most researchers target subspecialty journals throughout their career; in our experience even authors doing increasingly common multidisciplinary research (between clinical disciplines, or with engineers, for example) target a journal in one of the co-authors' subspecialties. We compiled a series of content and phrasing questions that could begin to provide information about the current state of titling and potentially form the basis for an efficient method for studying different clinical disciplines over time or for checking practice in unfamiliar target journals. We were able to use this question set to identify titling differences between general medicine and anesthesiology and between individual journals (notably the NEJM and the others). In particular, we found there seems to be differential uptake of reporting guidelines' advice to include mention of research design in titles. Differences in form were also detected.

Table 1. Editor-physicians' reports on English language original research article (RA) titling: content, length and citing associations.

Author (date)	Aim	Method/corpus	Significant Findings	
			Content	Length
Goodman *et al.* (2001) [10]	"To determine whether the titles included information about methods, results, and conclusions"	Journals: 4 top general medical journals—NEJM, JAMA (both US); *Lancet*, BMJ (both UK) No. of titles: 420 Type: original RAs Time: July–December 1995	Methods mention: 29%–40%; none named less powerful study designs (e.g., "case series") Results mention: 16%–21% "Topic only": 40%	—
		Editors were interviewed about title policies.	Only BMJ had a titles policy and peer-review protocol (favoring methods mention).	
Siegel *et al.* (2006) [12]	To "report the information content of titles of articles" and "compare the results obtained for titles in the same journals in 1995" (*i.e.*, [10])	Journals: NEJM, JAMA, *Lancet*, BMJ No. of titles: 417 Type: original RAs Time: July–December 2001	Methods mention: 96% (2001) in BMJ, up from 49% (1995); others tended to go up (JAMA) or down (*Lancet*, NEJM) non-significantly Results mention: down in all	—
Ubriani *et al.* (2007) [13]	"To assess the frequency of ... type of study design" in titles (*or* abstracts) "using standard key words"	Journals: 3 top dermatology journals—*British Journal of Dermatology, Archives of Dermatology, Journal of the American Academy of Dermatology* No. of titles: 95 Type: original RAs Time: 12 months (2004–2005)	Methods mention[1]: 50.6%–79.6% (title *or* abstract)	—
Jacques and Sebire (2010) [14]	To find "specific features of journal titles [that] may be related to citation rates"	Journals: *Lancet*, BMJ, *Journal of Clinical Pathology* No. of titles: 150 (25 most-cited titles *vs.* 25 least-cited in each journal Type: RAs ("primary" + meta-analyses) and reviews Time: 2005	More cites: titles with methods mention or acronym [2] Fewer cites: geographic location mention	More cites: longer titles (median, 18 words *vs.* 9 words for least-cited) More cites: 2 parts (70%) with colon
Habibzadeh and Yadollahie (2010) [15]	"To investigate the correlation between title length and number of citations"	Journals: non-random "arbitrarily chosen" sample of 22 with IFs; clinical and non-clinical science No. of titles: 9031 Type: all	—	More cites: longer titles (association is stronger for higher IF journals) US titles longer than UK titles
Paiva *et al.* (2012) [16]	"To evaluate some features of titles ... from OA journals and to assess the possible impact of ... features on predicting ... views and ... citations"	Journals: OA journals (7 from PLoS journals, 12 from BMC); clinical and non-clinical science No. of titles: 423 Type: RAs Time: October 2008	More cites: results mention (*vs.* methods mention) Fewer cites (non-significant trend): geographic location mention and questions	More cites: short titles

Abbreviations: BMC = Biomed Central; BMJ = *British Medical Journal*; IF = impact factor; JAMA = *Journal of the American Medical Association*; NEJM = *New England Journal of Medicine*; PLoS = Public Library of Science; RA = research article; [1] Ubriani *et al.* [13] studied mentions in the title or the abstract; the percentages for titles only would probably have been lower, but we include this reference because it is the only one for a specific subspecialty; [2] The variable "acronym" studied by editor-physicians refers to ones used to name a study group to help make the group's ongoing research better known.

Table 2. Applied linguistics researchers' reports on features of English language medical RA and review titles

Author (Date)	Aim	Method/Corpus	Structure
Soler (2007) [19]	"To examine ... recurrent structural features of titles in two ... genres ... in the biological ... and social sciences"	Journals: *American Journal of Cardiology*, *Journal of Hepatology* (for RAs); *Arteriosclerosis, Thrombosis and Vascular Biology, Archives of Internal Medicine,* and *Transfusion Medicine Reviews* (for reviews) No. of titles: 15 RAs, 15 reviews Type: RAs and reviews Time: 1996–2002	Phrasal [1] titles: RAs, 72%; reviews, 46% Compound titles: RAs, 12%; reviews, 40% (Soler's Table 5, p. 95) Declaratives: RAs, 16%; reviews, 0% (Soler's Table 4, p. 95) Questions: RAs, 1%; reviews, 6% Length: RAs, 15.5 words; reviews, 10.7 words. Biological science titles were longer than clinical titles.
Wang and Bai (2007) [17]	"To reveal the syntactic structures that English native speakers frequently use [to write concise titles]"	Journal: NEJM No. of titles: 417 Type: RAs and reviews Time: 2003–2005	Phrasal titles: 99%, nominals (4 gerund phrases) Heads: 2 heads, 21.8%; ⩾3, 3.6% Between heads: *and*, 73.3%; *vs.*, 22.2%; *or*, 4.4% Length: shorter titles over the 3 years (2003–2005)
Cianflone (2010) [18]	"To discuss the format of titles (noun phrase, compound, full-sentence declarative or question) in a small corpus of veterinary medicine RAs"	Journals: *Preventive Veterinary Medicine, The Veterinary Journal, Journal of Feline Medicine and Surgery* and *Journal of Comparative Pathology* (current IFs 2.2, 1.8, 1.7 and 1.1) No. of titles: 63 Type: RAs Time: 2010	All title patterns (phrasal, compound, declaratives and questions) were seen.[2] All examples of compound sentences show methods were in the second part.[1]

Abbreviations: IF = impact factor; RA = research article; [1] *i.e.*, not sentences; [2] No descriptive statistics were compiled [18].

2. Materials and Methods

2.1. Corpus: Journals, Article Types, Time Frame, and Size

We gathered the titles of original RAs (including systematic reviews and meta-analyses) and so-called "expert opinion" reviews (*i.e.*, not systematic ones). Original RAs are highly prized because they provide the raw data to guide clinical decisions about risk and treatment. Their information is synthesized in systematic reviews, which share the RA genre structure (introduction, methods, results and discussion format). Expert review articles are still published in some journals today, in spite of their low status in the level-of-evidence hierarchy [21], because clinicians value the critical insights they provide. These reviews are usually invited submissions, and the opportunity to write them comes once a clinical researcher has gained the respect of peers. We anticipated finding few reviews [19] but nonetheless included this genre to make our work comparable to previous studies [14,15,17,19].

Two high-ranked general medical journals, the NEJM and BMJ, were logical choices. Clinicians admire the NEJM because it is the oldest continuously published journal and its impact factor (IF) is the highest in clinical medicine (54.42 in 2014). The BMJ (IF 16.38) has been a leader in the development and promotion of reporting guidelines, and this journal's practices influence many smaller journals. For the subspecialty journals, we chose top-ranked *Anesthesiology* (2014 IF 6.17) and the lower-ranked *European Journal of Anaesthesiology* (EJA) (2014 IF 3.01).

We began by copying titles from the first issue of the NEJM in 2015 into a spreadsheet and continued until the 19 March issue (61 titles). Special issues in the time frame, if any, were excluded from all the journal corpora because of our concern that the preferences of a special editor or the concentration on a single topic might skew the titling patterns somehow. Each title occupied a row on the spreadsheet and codes (categorical values) or real numbers (continuous variables) were entered into columns corresponding to specific content or form questions (Table 3). For example, one column was used to code article type immediately: 1, original RA; 2, review; 3, special articles or non-editorial opinion articles with features of an expert review (collected in an effort to increase the number of reviews). Most questions referred to categorical variables: Q1, all Q2s, Q3a, Q3c, and Q3e). (Supplement S1 gives more detailed coding instructions, definitions, and examples).

Table 3. Questions about title content and form.

Type	Question	
Methods mention	Q1	Is research design mentioned? If yes, in first or second position?
Other content	Q2a	Are results mentioned?
	Q2b	Is clinical context indicated?
	Q2c	Is a geographic location indicated?
	Q2d	Is a treatment named?
	Q2e	Is a patient population named?
Surface structure	Q3a	Full sentence? Question or declarative?
	Q3b	Number of parts?
	Q3c	Punctuation between parts?
	Q3d	Number of heads in phrasal titles?
	Q3e	Link (e.g., *and*, *but*, *vs.*, *or*) between heads?

After one author (M.E.K.) coded all the NEJM titles for all questions, the spreadsheet was sent to the statistician (co-author S.S.), who compiled preliminary statistics on the frequencies observed for that journal so we could confirm that we seemed to have enough observations of varying content and phrasing patterns to detect differences between journals. We sought journal corpora that would be large enough but not unnecessarily large. In the absence of a well-established procedure for analyzing minimum size and statistical power for this type of study *a priori*, we looked at the preliminary statistics for the trial journal to confirm that we had a range of possible codes for most questions and that some

extreme frequencies (approaching 0% or 100%) were also present. This suggested that we were already seeing a pattern. Such a pattern, we reasoned, meant that even if we sampled a very large number of NEJM issues to produce a few examples of content or phrasing features not already present in this corpus (such as full-sentence titles for the NEJM), the percentage of such features would still probably approach zero. This reasoning is similar to the calculation of types-to-token ratios [22] to indicate that a corpus is large enough to guide phrasing choices (syntax and collocations) for specialized translators and manuscript editors even though it might not have the full range of terminology present in the specialty [23].

Thus, given that the preliminary statistics for the NEJM were satisfactory, we stopped sampling that journal at 61 titles. We then went on to collect similar numbers for the other journals, starting with the first issue of 2015. For the BMJ we gathered 76 titles in total, stopping with the 22 February issue; for *Anesthesiology* (*n* = 68) and the EJA (*n* = 64) we sampled until the May and September issues of 2015, respectively. In total, we studied 269 titles.

2.2. Rationale for the Research Questions and the Coding System

As translator-editors (M.E.K. and A.M.) and additional-language authors (M.E.K. and S.S.) we have struggles with titles that have competing points of focus and might need to meet word limits and comply with current reporting recommendations. As a result, we sought guidance on the placement of research-design information (methods) in particular, and on how to punctuate titles with more than two parts (when more than the widely used colon would be needed). We describe the rationale for each question below and give a brief account of coding. Supplement S1 gives more detailed instructions, definitions, and examples.

Q1 concerned whether or not research design (*i.e.*, methods) was specified in the title. This content feature was a main focus of the editor-physician research in response to the appearance of reporting guidelines to support evidence-based medicine, as explained in the introduction. Methods mention is considered necessary for transparency (regarding the level of evidence a clinician will be reading about) even outside the world of titles-only searching for systematic reviews because clinician readers are said to have difficulty identifying this aspect of an RA [13]. To guide our own title composing, translating or editing process, we were also interested in whether methods appeared at the end (usually in a two-part title) or not, so the codes were 0, no methods mention; 1, methods mention in first position (the start); and 2, methods mention later (end of a single-part title or in a later part of a multi-part title).

Q2a through Q2e dealt with further content. One research group observed weighty percentages (35%–48%) of so-called topic-only titles in four prestigious general medical journals when they looked only at topic-only, methods and results mentions and whether a study group was named by an acronym in 1995 [10] and 2001 [12]. We decided to look further to attempt to describe title richness—including that of the apparently lowly topic-only titles—by tallying titles mentioning not only results (Q2a) but also clinical context (Q2b), geographic setting (Q2c), treatment (Q2d), and patient population (Q2e). Results mention was of interest because it has been linked to higher citation in one study [16]; in addition, some journals promote or require it in instructions for authors, and we would like to know if this inclusion is becoming a trend. Clinical context, treatment, and patient population were chosen because they are used as search parameters and are considered when articles are screened for use because they affect generalizability. Geographic setting was chosen because we often see EAL authors include the fact that their research is local in titles and have to defend the generalizability of their work to peer reviewers. In addition, two studies have found that mention of geographic location is linked to low citation [14,16].

Q3a through Q3e were concerned with a series of phrase and sentence features studied by linguists [17–19] and were coded as follows in this study: full sentence (0, no; 1, yes—a question; and 2, yes—an affirmation); number of parts (continuous variable); punctuation between parts (1, colon; 2, period; 3, dash; 4, semicolon; 5, comma; 6, parentheses; 7, question mark); number of heads in multi-head phrasal titles (⩾2 noun or –*ing* form heads) recorded as a continuous variable; and link

between heads (1, *and*; 2, *or*; 3, *versus* or *vs.*; 4, *with* as a conjunction; 5, a comma; 6, other). Our separate coding of questions and declarative sentences reflected the intention of analyzing how often options other than a declarative sentence are used for titles specifying results. We also noted that one study linked question titles to fewer cites [16]; we therefore wanted this information in our database for descriptive statistics at this time and possibly to study citing in the future. We coded the other features partly to extend the work of Wang and Bai [17] to journals other than the NEJM, and partly to look into the punctuation of compound titles with more than two parts, which have proliferated since the inclusion of database (study group) acronyms became popular. Such titles are thought to make a research group better known and one group found a link to higher citation [14].

Q4 on length was calculated automatically using a formula embedded in the spreadsheet. Given that we also had noted the contradictory findings of editor-physician researchers on the relation between length and cites ([14,16] *vs.* [15]), and the trend toward shorter titles in the NEJM between 2003 and 2005 [17], we wished to have this descriptive statistic on record for comparison with others' findings or for future use in studying citing.

2.3. Coding Standardization and Quality Checks

The familiar difficulty of categorical coding comes into play whether university instructors are grading student writing, peer reviewers are evaluating manuscripts, systematic reviewers are assessing the quality of published evidence, or clinicians are assessing prognosis. Thus, all studies based on coding raise questions about accuracy related to human judgment, fatigue or distractions, drift over time, and the intrinsic vagueness of some criteria. We applied a system using two coders, which has been shown to improve the accuracy of results over those of single-observer coding [24]. The two coders (M.E.K. and A.M.) are language professionals (editors, translators, instructors) who worked independently and then discussed and resolved discrepancies (following analysis of inter-rater agreement), sometimes with the help of S.S., who is an anesthesiologist and could offer an insider's perspective. For example, at the final consensus stage he confirmed our impression that in the title "Anaesthetic and ICU management of aneurysmal subarachnoid haemorrhage: A survey of European practice," the first word is referring to a clinical department, not a treatment. In the title "Sleep disorders among French anaesthesiologists and intensivists working in public hospitals: A self-reported electronic survey," he confirmed that the persons named constitute a patient population. At other times, the coders simply felt a third opinion would help. For example in the title "Cost-Effectiveness of Hypertension Therapy According to 2014 Guidelines" we all agreed that cost-effectiveness named a type of analysis, a methods mention in this case.

The coders also included an early norming step. While the preliminary statistics for the trial journal (the NEJM) were being compiled to assess corpus size, the second observer coded titles for two issues, blinded to the codes of the first observer and equipped only with the first draft of coding instructions. After consensus on discrepancies, the instructions were improved and each coder worked on the rest of the titles for all journals independently. They could ask each others' opinions on general concepts in the instructions but could not discuss specific titles during this process so that inter-rater agreement could be assessed. The data were also checked at this point for missing or anomalous data for resolution by the coders.

Once agreement statistics were on record, differences were discussed and final consensus reached on the coding. Sometimes it was necessary to look at an abstract for confirmation. The final database was then used for descriptive statistics and comparisons.

2.4. Statistical Analysis

Frequency data for content mentions and structural features were expressed as absolute number of observations and percentages of the total number of relevant titles for each journal and overall. Length measures were expressed as the mean and 95% confidence interval (CI) to give an indication of

the precision of the observed mean in relation to the population mean, which would fall somewhere between the limits of the CI.

We used the Fisher-Freeman-Holton exact test to compare categorical variables (content and feature frequencies) and, after verifying normal distribution of observations, we used analysis of variance to compare mean lengths. Differences were considered significant at a level of $p < 0.05$ and are expressed to four decimal places unless they approach 0 or 1, in which case approximate values (e.g., <0.0001) are given. The program StatsDirect version 3 (Stats Direct Ltd., UK) was used.

To estimate inter-rater reliability between two coders we used Cohen's κ statistic, which is an estimate of level of agreement adjusted for accidental coincidence in coding. κ is expressed on a scale from −1 (systematic disagreement) to 1 (perfect agreement), on which 0 represents random agreement. We expected high agreement (>0.80) [25] on most questions. Besides using κ to support the reliability of the method, we wished to know if it estimated our subjective assessment of the relative difficulty of coding and the amount of training or norming required for coders in the future.

3. Results and Discussion

3.1. Main Findings and Limitations

The important title content differences we detected between journals affected mainly methods and results mentions, although there were also differences between some journals for mention of geographic location. Differences in phrasing features were found in the prevalence of full sentences and multi-part titles (and in their punctuation), and in length. Some differences were found between a single journal and all the others (the NEJM *vs.* others on most features, but the EJA *vs.* others on length), and some were between the general medical journals and the subspecialty ones.

The main limitation of this study is that when our title corpus was divided by genre (original RAs *vs.* reviews), we found we had insufficient statistical power to make comparisons between journals for review article titles by themselves; we therefore analyzed between-journal differences only for the combined RA and review corpus for most questions. (Reviews did not present this problem for methods and results mentions—Q1 and Q2a—because that content is only relevant to RAs). Only BMJ sampling produced a substantial number of review titles during the period, partly because we collected titles of extended "review-like" opinion articles from this journal. None of the other journals published such opinion articles during the sampling period and the chosen subspecialty journals published very few reviews. This genre was distributed as follows in the journals: NEJM, 7 (11.5% of that journal's total); BMJ, 21 (27.6%; $p = 0.0001$ *vs.* every other journal); *Anesthesiology*, 5 (7.4%); and the EJA, 2 (3.1%). To be able to detect between-journal differences in review titling (by obtaining at least 21–23 titles), we would have needed to sample about 36 of the NEJM's weekly issues (we sampled 12) and at least two years' worth of the subspecialty journals' monthly issues. That approach would have created subspecialty review article corpora that were not chronologically comparable to their corresponding RA corpora. Given this situation, and since the invited review article might be termed a genre for mature writers in medicine—invitations come from the best journals only to some authors and only in the latter half of their careers in our experience—we think that between-journal comparisons for reviews will be of interest to only a few of the stakeholders our research addresses. An ESP instructor interested in studying review titles, possibly for critical reading lessons or to contrast with systematic review titles, would be best advised to go directly to review journals to look for examples. Material for classroom instruction on reviews is unlikely to be needed by ERPP writing instructors working with novices.

3.2. Methods and Results Mentions in Original RA Titles

Methods and results mentions were relevant only to original RA titles. Significant differences between some journals were found in both (Table 4).

Table 4. Methods and results mentions in original RA titles in four clinical medicine journals [1].

Content	Journals										*p* Value [2]
	NEJM (*n* = 53)		BMJ (*n* = 55)		*Anesthesiology* (*n* = 62)		EJA (*n* = 62)		All (*N* = 232)		
	n	(%)	*n*	(%)	*n*	(%)	*n*	(%)	*n*	(%)	
Methods (Q1)											
No	47	(88.7)	1	(1.8)	26	(41.9)	2	(3.2)	76	(32.8)	*p* < 0.0001
Yes, first position [3]	6	(11.3)	0	(0)	11	(17.7)	1	(1.6)	18	(7.8)	
Yes, second position [3]	0	(0)	54	(98.2)	25	(40.4)	59	(95.2)	138	(59.4)	
Results (Q2a)											
No	52	(98.1)	55	(100.0)	44	(71.0)	53	(85.5)	204	(87.9)	*p* < 0.0001
Yes	1	(1.9)	0	(0)	18	(29.0)	9	(14.5)	28	(12.1)	

Abbreviations: BMJ = *British Medical Journal*; EJA = *European Journal of Anaesthesiology*; NEJM = *New England Journal of Medicine*; Q = question; RA = research article; [1] Data are number (percentage). The denominator for each journal is the total number of original RAs extracted, shown at the head of each journal column; [2] Fisher-Freeman-Holton exact test. Frequencies were compared between all journals. The *p* values in the column indicate that both features (methods and results mentions) showed significant differences, though not necessarily between all journals. For methods mention, the EJA and BMJ were similar to each other (*p* > 0.9999). For results mention, the NEJM and BMJ were similar (*p* = 0.409), as were the two anesthesia journals to each other (*p* = 0.0805). The general medicine *vs.* subspecialty differences, however, were significant; [3] First position refers to placement in the first part of a ⩾2-part title or at the beginning of a single-part phrasal or sentence title. Second position refers to placement in the second (or later part) of a multi-part title or at the end of a single-part or sentence title.

We expected a very high prevalence of methods mention across the board in 2015, nearly 20 years after reporting guidelines began to appear, and five years after the CONSORT statement specified that study design should be in the title; however, the prevalence was low in the NEJM, at around 11%, and modest in *Anesthesiology*, at just under 60% (Table 4). Only the BMJ and EJA had high CONSORT compliance (approaching 100%). Differences were significant between journals (Table 4) in both aspects coded for this question (presence/absence and first/second positioning) except between the EJA and the BMJ (*p* > 0.999). In the EJA methods were nearly always specified in second position in two-part titles and in the BMJ they always were; in contrast, NEJM titles always used the first position for this information. We think our findings confirm that specific journals or small families of journals have fairly tight discourse practices and show that preferences are not always driven by reporting guidelines. Nor can our findings be predicted by IF-based rank, since the BMJ and EJA are both highly guideline compliant. In contrast, unstated preferences as observed previously [10] seem to explain the convergence of NEJM authorial style and/or the imposition of style by copy editors or other shapers [26] of the forms we see in published articles. We also note that *Anesthesiology* is tolerant of a range of practice.

The practical effect of not mentioning methods can be illustrated by pairing examples of titles for the same study designs found in different journals. The first example in each pair below is explicit about the methods used. The second is a topic only [10,12] title that requires the insider reader to notice the attribute or outcome under study, infer the required study design, and assume the particular journal would not have published the paper if the proper method had not been used. We have underlined the methods-explicit phrasing and/or the outcome words that would trigger realization of method for the insider reader:

1. Two observational studies:

    Intraoperative temperature monitoring with zero heat flux technology (3M SpotOn sensor) in comparison with sublingual and nasopharyngeal temperature: <u>an observational study</u> (EJA)

    Intraoperative core temperature <u>patterns</u>, transfusion <u>requirement</u>, and <u>hospital duration</u> in patients warmed with forced air (*Anesthesiology*)

2. Two randomized controlled trials:

> Impact of a behavioural sleep intervention on symptoms and sleep in children with attention deficit hyperactivity disorder, and parental mental health: <u>randomised controlled trial</u> (BMJ)

> Efficacy of a tetravalent dengue vaccine in children in Latin America (NEJM)

Both *Anesthesiology* and the NEJM seem confident that target readers will intuit the study designs accurately, assuming, for example that readers know the NEJM would not publish a study on vaccine efficacy if the researchers had not used the only method that can demonstrate that a treatment has that attribute. Likewise, the intended readers would be expected to realize that the *Anesthesiology* title above would have to be an observational study of associations because commonly recorded data rather than interventions were being compared. These two top-ranked journals also seem confident that searchers interested in the topics named will find and not discard the studies they publish—perhaps a privilege of rank. The rate of non-mention of methods we observed in the NEJM (88.7%) was even higher than the level observed (80%) in that journal's 2001 titles [12], whereas non-mention in the BMJ fell from 51% in 1995 to nearly 0% in 2015. This high-ranked journal's commitment to promoting evidence-based medicine and reporting guidelines is well established, so their near full compliance is unsurprising, but the interest of the other journals in good reporting standards surely cannot be called into question; in fact, their instructions to authors refer to the same reporting guidelines the BMJ's and EJA's titles comply with. It is perhaps tempting to wonder whether we are seeing a US–UK difference of opinion about readers' needs, but to answer that question many more journals would have to be studied. Our hypothesis is that different US and UK attitudes to readers' needs are unlikely, consistent with an earlier study that showed the British *Lancet*'s methods-mention percentage (34%) to be more similar to the rates of the NEJM (31%) and the *Journal of the American Medical Association* (38%) in 1995 than to the BMJ's rate of 49% [12]. The same study found that only the BMJ had increased methods mention significantly by 2001 while all three of the other journals' rates had tended to hold steady, going only slightly up or down (non-significant differences). Our interpretation is also consistent with the finding of Goodman *et al.* [10] that the BMJ was the only journal they studied that had a specific policy to promote methods mention. For the moment, we think our findings have two main messages for authors and their supporters: (1) even journals that direct authors to reporting guidelines, especially the CONSORT statement, require different degrees of compliance with respect to titling, so authors should check practice by looking at tables of contents; and (2) just as authors reformat their abstracts before submitting a rejected manuscript to a different journal, they should also reassess their titles.

In general clinical medicine, results seem to be rarely mentioned in titles (nearly 90% of titles in our corpus omitted results), although they appear fairly often in titles in the subspecialty we studied; the general-*vs.*-specialist difference was significant (Table 4). Both anesthesiology journals had substantial proportions of results-mention titles (EJA, nearly 15%; *Anesthesiology*, nearly 30%), whereas the BMJ had none and the NEJM only one (<2%). Our findings are consistent with the decreasing frequency of results mention between 1995 [10] and 2001 [12]. According to the cited studies, the BMJ's rate of results-mention titles fell from 16% to 0% in six years and according to our data was still at 0% after another 14 years. The NEJM's fell from 20% in 1995 to 4% in 2001 and now approaches zero. In contrast, Jaime Sisó [27] found that results-mention titles continued to increase in developmental biology, a non-clinical discipline, between 1990 and 2001. She points out that editors and researchers were debating the advisability of what she called "conclusive" titles (for results mention) for clinical trials in the late 1980s and early 1990s. For example, some results-mention titles in the BMJ in 1996 were found to give conclusions that were later called into question or had overreached the actual findings of the trials [28] (in other words, misrepresenting the level of evidence obtained) and were later contradicted, yet the results claimed early on remain frozen in the titles forever—creating a potential problem for medical practitioners searching the literature. The relative rarity of results mention in general medicine titles today may have come about for this reason. However, we note that some clinical subspecialties (anesthesiology is an example) publish a fair amount of basic research—reports of animal

models or other non-clinical designs that do not specifically target readerships of general practitioners or even specialist clinicians. In such studies, conclusions may be more straightforward, given that they involve direct observations *in vitro* or in the tissues of sacrificed animals, for example, rather than indirect findings like reported symptoms or even biological markers that clinicians study. We think that the interest of some researcher-readers in these studies in the anesthesiology journals probably accounts for the editors' tolerance for results-mention titles in the discourse of the subspecialty overall, even when such titles are for clinical studies like the following two full-sentence titles from our corpus:

> Anaphylaxis is more common with rocuronium and succinylcholine than with atracurium *Anesthesiology*, based on a retrospective cohort study

> Postoperative delirium is an independent risk factor for posttraumatic stress disorder in the elderly patient: a prospective observational study (EJA)

The observation of Paiva *et al.* [16] that results-mention RA titles tended to attract more cites increases interest in continuing to use Q2a in future research. It is possible, however, that the association they observed may have been an artifact of studying a corpus of titles from all journals belonging to two open-access publishing houses with a mix of basic and clinical research journals (see their Table 1, p. 510). That mix might also explain why they saw a high rate of results mention (over 40%).

The message we take from these findings is that results-mention titles may be inadvisable for strictly clinical research, except in cases where a subspecialty journal's instructions say they require one. Clinical researchers should reflect carefully about whether they want to emphasize results rather than method and other content, and if opting for results, we would suggest they consider phrasal titles as well as the more obvious present-tense sentences. Even when a journal asks for results to be highlighted in titles, authors should look at four to five tables of contents to confirm compliance with this rather surprising preference for clinical research and to see how results mention is phrased. A brief *post-hoc* check of one such journal (*Clinical Gastroenterology and Hepatology*) showed that in 2015 fewer than half the 198 RA titles ($n = 92$, or 46.5%) stated results in a way that was unequivocally clear to us, and almost two-thirds ($n = 60$) of those clear titles were full sentences in the present tense. Thus, a third of the clear titles used phrases with combinations like "improved [an outcome]... with [a treatment]." We thought 12 more would probably imply a result to an insider. Inferring a result can be risky, however. When we viewed the abstracts to help us interpret a few titles, we found that some seemed to imply a positive result of a treatment to a reader expecting to see results in this particular journal (e.g., "efficacy of..."), whereas the abstract showed that results were negative or inconclusive. All told, the number of vague or outright non-compliant titles was not trivial, and we think these data suggest a strategy that authors with negative (*i.e.*, "no-difference") findings can use to increase the likelihood that a submission will survive in-house screening for "interest" given that bias in favor of positive (significant-difference) findings is still well established [29].

3.3. Other Content Mentions in RA and Review Titles

Slight variations in other title content features were non-significant between journals except for mention of a geographic area (Table 5, Q2c). Specific places were named in a few titles (ranging from about 10%–15%) in all journals except *Anesthesiology*, which had none ($p = 0.0012$ *vs.* all the other journals). Watching this content item is important because two studies have linked geographic mention to lower citation rates: one looked at two open access journal groups (Public Library of Science and Biomed Central) [16] and the other at two general medical journals (the *Lancet* and the BMJ) and a single subspecialty journal belonging to the BMJ group [14]. We are unable to comment on citing associations at this time, but in general it seems reasonable for authors to title their work with the details that represent their research accurately regardless of citing associations, which may vary in relation to content features over time because of changes in the intrinsic interest of topics themselves. For example, Jacques and Sebire [14] found that the presence of the word *children* in a title

was associated with fewer cites [14], yet we would hardly expect a pediatrician or any author to avoid using it if relevant. Nevertheless, we think geography mention may be worth discussing with EAL authors because it is clearly not always necessary according to our data and has the potential to lower some readers' perception of generalizability. Anecdotally, we can mention that we have seen reviewers ask EAL authors to defend the international relevance of studies done in non-Anglophone settings (e.g., with questions about the effects of presurgical medication differences, unfamiliar health care systems, *etc.*), so we now anticipate this problem when preparing manuscripts. In any case, it seems unnecessary to highlight location in the titles of most studies, and journals do not seem to expect it: 85%–100% of the titles in our corpus, depending on journal, did not include this information.

Table 5. Additional content in original RA and review titles in four clinical medicine journals [1].

Content	Journals										p Value
	NEJM (*n* = 61)		BMJ (*n* = 76)		*Anesthesiology* (*n* = 68)		EJA (*n* = 64)		All (*N* = 269)		
Clinical context (Q2b)											
No	6	(9.8)	15	(19.7)	15	(22.1)	11	(17.2)	47	(17.5)	*p* = 0.2774
Yes	55	(90.2)	61	(80.3)	53	(77.9)	53	(82.8)	222	(82.5)	
Geography (Q2c)											
No	52	(85.2)	64	(84.2)	68	(100.0)[†]	58	(90.6)	242	(90.0)	*p*= 0.0012 [2]
Yes	9	(14.8)	12	(15.8)	0	(0.0)	6	(9.4)	27	(10.0)	
Treatment (Q2d)											
No	26	(42.6)	48	(63.2)	32	(47.1)	37	(57.8)	143	(53.2)	*p* = 0.0616
Yes	35	(57.4)	28	(36.8)	36	(52.9)	27	(42.2)	126	(46.8)	
Patient type (Q2e)											
No	45	(73.8)	55	(72.4)	54	(79.4)	54	(84.4)	208	(77.3)	*p* = 0.3195
Yes	16	(26.2)	21	(27.6)	14	(20.6)	10	(15.6)	61	(22.7)	

Abbreviations: BMJ = *British Medical Journal*; EJA = *European Journal of Anaesthesiology*; NEJM = *New England Journal of Medicine*; [1] Data are number (percentage). The denominator for each journal is the total number of original RAs extracted, shown at the head of each journal's column; [2] Significant difference (*p* < 0.05, Fisher-Freeman-Holton exact test). The significance for geography mention was attributable to *Anesthesiology vs.* every other journal.

Although title content differences between original RAs and reviews could not be analyzed between journals, we did see one significant difference when the two genres were analyzed for the entire corpus: treatments were mentioned less often in reviews (18.9%) than in original RAs (18.9% *vs.* 51.3%, respectively; *p* = 0.0003). However, we think this finding is a mere accident of topic range in journals at the time we compiled the corpus. If titles are sampled when a very effective new therapy has emerged, treatments might well be named in review titles. For example, when the efficacy of the biologic drug propranolol for severe infantile hemangioma was casually discovered about 8 years ago [30], a spate of retrospective case series and expert reviews appeared first in the literature, after which prospective studies were designed; if we had sampled titles in general medicine, dermatology or pediatrics around 2009–2014, we would probably have seen many review titles mentioning that treatment.

3.4. Phrasing Patterns in RA and Review Titles

Several surface features relevant to how titles are phrased or look on the page differed between journals (Table 6).

Table 6. Formal features of original RA and review titles in four clinical medicine journals [1].

Features	Journals										*p* Value
	NEJM (*n* = 61)		BMJ (*n* = 76)		*Anesthesiology* (*n* = 68)		EJA (*n* = 64)		All (*N* = 269)		
Full sentence (Q3a)											
No	61	(100.0) [2]	73	(96.1)	56	(82.4)	54	(84.4)	244	(90.7)	
Yes, question	0	(0.0)	3	(3.9)	0	(0.0)	1	(1.6)	4	(1.5)	$p < 0.0001$ [2]
Yes, affirmation	0	(0.0)	0	(0.0)	12	(17.6) [2]	9	(14.0)	21	(7.8)	
No. of parts (Q3b)											
1	54	(88.5) [2]	19	(25.0)	45	(66.2)	3	(4.7)	121	(45.0)	
2	6	(9.9)	57	(75.0)	23	(33.8)	60	(93.7)	146	(54.3)	$p < 0.0001$ [2]
3	1	(1.6)	0	(0.0)	0	(0.0)	1	(1.6)	2	(0.7)	
Punctuation, multi-part titles (Q3c) [3]											
Colon	2/7	(28.6)	57	(100.0) [2]	23	(100.0)	60/61	(98.4)	142/148	(95.9)	
Colon + colon	0	(0.0)	0	(0.0)	0	(0.0)	1/61	(1.6)	1/148	(0.7)	$p = 0.0001$ [2]
Colon + dash	1/7	(14.3)	0	(0.0)	0	(0.0)	0	(0.0)	1/148	(0.7)	
Dash	4/7	(57.1)	0	(0.0)	0	(0.0)	0	(0.0)	4/148	(2.7)	
No. of heads in multi-head phrases (Q3d) [4]											
2	15/18	(83.3)	34/37	(91.9)	4/5	(80.0)	10/11	(90.9)	63/71	(88.7)	
3	3/18	(16.7)	2/37	(5.4)	1/5	(20.0)	1/11	(9.1)	7/71	(9.9)	$p = 0.5802$
4	0	(0.0)	1/37	(2.7)	0	(0.0)	0	(0.0)	1/71	(1.4)	
Link between heads (Q3e) [4]											
and	15/22	(68.2)	35/41	(85.4)	5/6	(83.3)	9/12	(75.0)	64/81	(79.0)	
vs. (or *versus*)	4/22	(18.2)	3/41	(7.3)	0	(0.0)	3/12	(25.0)	10/81	(12.4)	$p = 0.3202$
comma	3/22	(13.6)	3/41	(7.3)	1/6	(16.7)	0	(0.0)	7/81	(8.6)	

Abbreviations: BMJ = *British Medical Journal*; EJA = *European Journal of Anaesthesiology*; NEJM = *New England Journal of Medicine*; [1] Unless otherwise specified, data are number (percentage). The denominator for each journal is the total number of RAs and reviews extracted, shown at the head of each journal's column; [2] Significant differences ($p < 0.05$, Fisher-Freeman-Holton exact test). The significant P values for the following features were attributable to the following comparisons: full sentence (Q3a), NEJM *vs.* the subspecialty journals ($p = 0.0093$); multi-part titles and punctuation between parts (Q3b and Q3c), NEJM *vs.* every other journal ($p < 0.0001$, all comparisons); [3] Punctuation information (Q3c) is only relevant for multi-part titles. Thus, the denominators shown after the frequency specify the number of such titles for each journal; [4] Number of heads and linking-word information (Q3d and Q3e) are only relevant for multi-head phrasal titles. Thus, the denominators shown after the frequency specify the number of such titles (or title parts) for each journal. When a multi-part title had multiple heads in one or more parts, each part was looked at and quantified separately. See supplementary file S1 for more detailed information on coding.

One difference was in the prevalence of full-sentence titles (Q3a), which was low overall at 9.3% of all titles (Table 6). The NEJM—which published no such titles in the period sampled—differed significantly on that feature, but only *vs.* the two subspecialty journals (*p* = 0.0003 *vs. Anesthesiology*; *p* = 0.0020 *vs.* the EJA). Our observations for this question confirm the stability of the NEJM's title style that Wang and Bai [17] described for 2003–2005, when 99% of titles were phrases (neither sentence affirmations nor questions). The general-*vs.*-subspecialty difference we detected for this title form is unsurprising given the significantly higher prevalence of results-mention titles in our anesthesiology corpus (Section 3.2, Table 4). A present-tense sentence, which is the clearest way to express a conclusion about a result, accounted for 67.9% of results-mention titles in our corpus. Our observations are also consistent with Soler's [19] finding of more full sentences (declaratives, usually in the present tense) in biology RA titles (51%) than medical ones (16%). Biology RAs would be more similar to reports of *in vitro* or animal models in medicine, and when we checked their prevalence in our corpus *a posteriori*, we found that the NEJM and BMJ had none whereas the two anesthesiology journals had 17 (almost 13% of their total) and that 8 of them were full sentences. Sentence titles were also used to convey the results of strictly clinical research (in humans) in the subspecialty, possibly reflecting a style influence of bench science titles in *Anesthesiology* and the EJA. Finally, we saw only one sentence title in the past tense, but that single example of using tense to tie a result to a specific study rather than generalizing suggests a strategy authors can use to hedge a claim, consistent with the advisability of being cautious when mentioning negative or inconclusive results in titles as we discussed in Section 3.2.

The form-function tie between sentences and results in the subspecialty drew our attention to how phrasal titles expressed results in the corpus because such titles might represent another strategy for hedging a claim. Given that phrasal titles are less salient than sentences (because they are more common), they could mention results when required to comply with journal instructions without calling attention to themselves. In the following examples of results-mention phrasal titles, the words relevant to claims have been underlined:

Improved overall survival in melanoma with combined dabrafenib and trametinib (NEJM)

Effective reversal of edoxaban-associated bleeding with four-factor prothrombin complex concentrate in a rabbit model of acute hemorrhage (*Anesthesiology*).

p.L1612P, a novel voltage-gated sodium channel nav1.7 mutation inducing a cold sensitive paroxysmal extreme pain disorder (*Anesthesiology*).

It is interesting to note that while *effective reversal* clearly conveys a positive finding in the second title, beginning a title with *efficacy of* would be vague about results: a title with that beginning would merely suggest a clinical trial. Positive-sounding noun phrases were also used as points of departure (e.g., *discovery of*, practice *improvements based on...*, *improvement in...*). The third example above, with a postnominal modifier is stylistically subtle because it calls attention to the hot topic (a genetics study) first and names the positive result last: the authors have found a gene responsible for inducing the named condition. It is clearly possible to mention results without emphasizing them with a stylistically salient full sentence, something ERPP instructors, manuscript editors and translators might want to bear in mind.

Our findings for Q3a also confirm the extremely low prevalence of questions (<2%, Table 6). The BMJ's three sentence titles (a prevalence statistically similar to the NEJM's, *p* = 0.2535) were all questions and all were titles of reviews. Soler [19] (p. 95) also saw question titles used more often for reviews (6%) than for RAs (1%). Cianflone [18] reported that questions (indeed, all formal patterns) were present in a convenience sample of veterinary journal titles, but no rates were mentioned. Opinions on the merits of questions in titles may come from many quarters, including reading exposure in national-language journals, mentors, or advice books. For example, Day [31], whose *How to Write and Publish a Scientific Paper* has been updated and translated many times and whose opinions are often handed down behind the scenes of writing, disliked questions (p. 20). Given that we have seen

that these titles are rare but tolerated in clinical medicine, we see no reason to recommend they be avoided, especially for review titles given the three questions harvested from the prestigious BMJ.

A between-journal difference was also detected for the frequency of multi-part titles (Q3b), which Day [31] (p. 20) also disliked, and their punctuation (Q3c). The difference is attributable to the markedly different titling pattern in the NEJM (Table 6). All journals included some multi-part titles, but in different proportions, ranging from over 95% in the EJA, for example, to only slightly more than 10% in the NEJM. Reviews were less often headed by two-part titles than original RAs (73% *vs.* 40.5%, respectively, in the sample overall; *p* < 0.0001). That observation differs from the percent distribution Soler [19] found in a medical title corpus collected from 1996 to 2002, when 40% of reviews and 12% of original RAs had two parts. Table 6 shows the percentages of each journal's multi-part titles that were punctuated with the default choice (a colon) (Q3c), using the findings for Q3b as the denominator. This data presentation reveals that only the two general medical journals used another option, the dash. The difference was statistically significant only for the NEJM (where a dash was used in 5 out of 7 multi-part titles and a colon in only 3 of the 7) in comparison with every other journal. This information is probably useful for EAL authors and their language supporters who struggle to manage titles with three or more parts. Once the usual first-choice punctuation (a colon) has been used, based on our observations we would recommend using a dash to set off third or fourth parts rather than calquing punctuation use from other languages. Our findings also suggest that authors and their supporters should explore phrasing options to reduce three- or four-part tiles to two or three.

It is interesting that the frequencies of multi-head phrasal titles (Q3d) differed so much in our four journals, ranging from a low of 7.4% for *Anesthesiology* (5 out of the total of 68 titles) to a high of 48.7% for the BMJ (37/76) (*p* < 0.0001). For the NEJM and EJA, on the other hand, the differences in proportions, at 29.5% (18/61) and 17.2% (11/64), respectively, were non-significant (*p* = 0.1379). Table 6 shows the proportions of two-, three- and four-head phrases in each journal's set of multi-head phrasal titles. The frequency of multi-head phrasal titles in our full corpus (26.4%) was similar to the 25.4% Wang and Bai [17] saw on average for comparable titles in the NEJM between 2003 and 2005 even though their observations for individual years showed an unusually high rate of 62% for 2004. The overall similarity between our study and theirs suggests to us that these prevalences offer a faithful picture of noun-phrase use in medicine overall and possibly even prevalences in English in general.

For readers unfamiliar with parsing heads in nominal phrases, we show five examples from the corpus in which we have underlined the heads:

Two heads:

1. Blood glucose <u>concentration</u> and <u>risk</u> of pancreatic cancer: systematic <u>review</u> and dose-response <u>meta-analysis</u> (BMJ, RA, a two-part title with two heads in each part).
2. Sodium <u>bicarbonate</u> and renal <u>function</u> after cardiac surgery: a prospectively planned individual patient <u>meta-analysis</u> (NEJM, RA, a two-part title with two heads in the first part and one in the second).
3. Best <u>position</u> and <u>depth</u> of anaesthesia for laryngeal mask airway removal in children: a randomised controlled <u>trial</u> (EJA, RA; a two-part title with two heads in the first part and one in the second).

Three heads:

4. <u>Disorders</u> of fluids and electrolytes: <u>disorders</u> of plasma sodium—<u>causes</u>, <u>consequences</u>, and <u>correction</u> (NEJM, review article, a three-part title; the first two parts have one head each, the third has three heads).
5. Intraoperative core temperature <u>patterns</u>, transfusion <u>requirement</u>, and hospital <u>duration</u> in patients warmed with forced air (*Anesthesiology*, RA, one part with three heads).

A noun (or alternatively in our study, an *–ing* form [see supplement S1 for coding instructions and examples]), is a head, or nucleus, of a phrase only if it is not a prenominal modifier or part of a prepositional phrase or relative clause post-modifying a head that precedes it. Although it is true that we counted noun heads and *–ing*-form heads as one phrasal type in our coding, *–ing* form heads were a minority (3 out of the 18 phrasal titles [16%] with two or three heads in the NEJM, 1 out of the 37 [2.3%] in the BMJ, none of the 5 [0%] in *Anesthesiology*, and 1 of the 11 [9%] in the EJA). Therefore, the general patterns we report would be unaffected.

3.5. Length

Our findings for 2015 were consistent with previously reported lengths in two aspects. First, review titles were significantly shorter than RA titles on average: reviews, 7.9 (95% CI, 6.7–9.0) words; RAs, 15.0 (95% CI, 14.3–15.7) words; $p < 0.0001$. Soler [19] reported lengths of 10.7 and 15.5 words for medical reviews and RAs, respectively. Mean lengths of both article types combined in our study varied considerably between journals, and with the exception of the BMJ and Anesthesiology, whose lengths were in the middle range and almost identical, the differences were significant (Figure 1). The exact means were as follows: NEJM, 8.6 (95% CI, 8.04–9.1) words; BMJ, 14.5 (95% CI, 12.9–16.1) words; *Anesthesiology*, 14.6 (95% CI, 13.4–15.7) words; and EJA, 18.1 (95% CI, 16.9–19.3) words. The narrow CIs for length confirm that the mean lengths we report are not very far off the true means for all titles in these journals. The editor-physicians who studied associations between length and citing did not show mean lengths for their samples for comparison, although one group reported a median length of 12 words (interquartile range, 6 words) in 2010 and an actual range of 1–43 words [15] in a very large corpus (over 9000 titles from 22 journals) in which length clearly varied greatly. Finally, we can confirm a trend toward even shorter titles in the NEJM over time. At 8.6 words, the average 2015 length we saw in the NEJM was down from the 9.9 words Wang and Bai [17] reported for 2005, which was already lower than the 11.6 words they saw for 2003 titles.
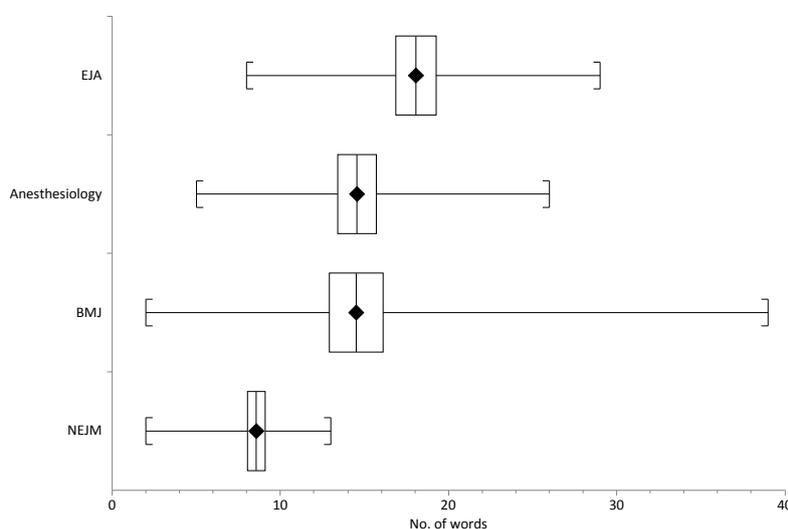

Figure 1. Differences between the NEJM and all other journals and the EJA and all other journals were significant ($p < 0.0001$); BMJ and *Anesthesiology* lengths were similar ($p = 0.957$). The diamonds show the means, the boxes show the 95% confidence intervals, and the bracketed lines show the actual ranges of lengths in each journal. The narrow confidence intervals indicate the observed means are probably close to the true means.

Our finding that average lengths were roughly the same overall in 2015 as lengths seen in previous studies [17,19] surprised us given that reporting guidelines call for more specificity about methods than in the past. Given the variability in our study and [15], however, and the apparent trend toward

shorter NEJM titles since 2001, we were curious about whether skillfully written or edited short titles can be as rich in content as long ones. A *post-hoc* analysis of our results suggests that they apparently cannot be. We first checked the distribution of "yes" codes on content items (Q1 through Q2e) in the NEJM and BMJ to find a logical cut point. Nearly three-quarters of the NEJM titles (73.6%) mentioned two or fewer content items, whereas over half (56.4%) of the BMJ titles mentioned three or more. Applying that cut point ($\geqslant$3 "yes" codes) as a rough richness criterion, we saw that 62.9% of titles in the EJA (longest in the corpus) were rich; in contrast, the NEJM had the smallest proportion (26.4%) of rich titles ($p = 0.0001$ *vs.* the EJA). It seems to us that ERPP instructors in clinical disciplines should be guiding EAL and other authors toward longer, richer titles (without mention of geography necessarily) in compliance with reporting guidelines—unless they are submitting to the NEJM, where a shorter phrasal title on a submission may suggest a share in the discourse community's self-confidence.

Length and richness are particularly important features to watch at this time, not only because some editor-physicians are hoping to discover a correlation of length with citing [14–16], but also because journals are increasingly leaving practice up to authors' discretion. Instructions to authors today may give advice on title content, remind authors of title functions, or recommend attributes like conciseness, but they generally just ask that the appropriate reporting guidelines be followed. When preparing this paper, we found no mention of mandatory title word limits on the websites of any of the journals we studied, although the EJA's includes a link to a 2011 editorial advising authors [35]. That editorial includes familiar opinions—"ideally...brief and concise" (p. 819)—but immediately moves on to discuss completeness of description and mention that the CONSORT statement encourages authors to include research design in the title, advice that was new in 2011. The editorial analyzes past titles and discusses why it is wise to increase length—from 10 to 23 words in one example. Journals may of course request title changes during review, a practice Goodman *et al.* [10] reported as new in 2001, citing examples of "implicit policies" (p. 76) that editors established but did not make public and that could be assumed to affect the impressions made on discourse community gatekeepers. Such implicit policies are precisely the reason why ERPP instructors train authors to do their own genre research when preparing manuscripts [32,33] and why translators and manuscript editors are beginning to do so too [34].

Two studies have shown a correlation between long titles and higher citation [14,15]. The single study showing more cites for shorter titles used an open access journal corpus with a mixture of clinical and basic science journals [16]; the authors did not stratify the corpus by journal type. We think this discrepancy supports our position that authors and their language supporters need to undertake their own genre analysis when preparing manuscripts.

3.6. Ease and Reliability of Coding: Inter-Rater Agreement

The κ statistics (Table 7) indicated that coding agreement was almost perfect or substantial (>0.80 or >0.60 to 0.80, respectively) [25] for 9 out of 11 questions (82%). One of the two questions with moderate agreement (>0.40 to 0.60) was one we found difficult to code (Q2b, clinical context), especially in the subspecialty journals; we therefore worked harder on it, discussing criteria several times during the process, probably improving agreement. Clinical context refers to a patient's disease or condition, or perhaps a period of care (e.g., postoperative recovery) that states the situation in which the research findings will be used. (See supplementary file S1 for further information on this question). We coded Q2b fairly intuitively for the general medical journals, but often hesitated in the subspecialty. At the final stage M.E.K. and A.M. consulted the specialist (S.S.) to check some interpretations and learned that anesthesiologists also have to explain their contexts to the members of hospital research ethics review boards when seeking approval for studies. We suspect this question will be difficult in any subspecialty. In contrast, that Q3e on links between heads also showed only moderate agreement is surprising because the question involves noticing a simple surface feature (words like and, or, and *vs.*) in the specified context between heads of noun phrases. We can only attribute it to coding fatigue because every title must be parsed carefully. We will probably eliminate Q3e in the future, however,

because Wang and Bai [17] have covered the topic exhaustively for the NEJM, and our observations produced no new insights about phrasal clinical research titles in that or other journals.

Table 7. Inter-rater coding agreement.

Type	Question		κ Statistic	Interpretation [1]
Methods mention	Q1	Is research design mentioned? If yes, in first or second position?	0.89	Almost perfect
Other content	Q2a	Are results mentioned?	0.80	Substantial
	Q2b	Is clinical context indicated?	0.52	Moderate
	Q2c	Is a geographic location indicated?	0.87	Almost perfect
	Q2d	Is a treatment named?	0.65	Substantial
	Q2e	Is a patient population named?	0.66	Substantial
Surface structure	Q3a	Full sentence? Question or declarative?	0.77	Substantial
	Q3b	Number of parts?	0.93	Almost perfect
	Q3c	Punctuation between parts?	0.93	Almost perfect
	Q3d	Number of heads?	0.74	Substantial
	Q3e	Link between heads?	0.41	Moderate

[1] Interpretation follows [23]. The κ statistic is expressed on a scale from −1 to 1: a negative value indicates systematic disagreement, 0 represents random agreement, and a positive value indicates increasing degrees of agreement.

Finally, we notice that several questions that achieved high κ statistics were ones that sometimes felt difficult for us and required us to slow down (e.g., Q3d). We therefore feel that inter-rater agreement should be analyzed as part of any study based on human coding as a quality-control measure. From this analysis, however, we realize that this statistic cannot predict which questions will need special training or periodic norming in the future: rather, all questions will.

4. Conclusions

4.1. Medical Titles After 20 Years of Reporting Guidelines

Twenty years after reporting guidelines for clinical research designs began to emerge, we saw that only two of the journals we studied insist on methods-mention RA titles (the BMJ and EJA) and that one (*Anesthesiology*) publishes a large percentage of such titles. The NEJM seems to be quite different in its titling preferences—having published fewer methods-mention titles in 2015 than in 2001 [12]; this may change if more authors begin to follow the latest revision of the CONSORT statement [9] or as other reporting guideline revisions emerge. Furthermore, the NEJM placed methods information at the beginning when it was provided, whereas the BMJ and EJA titles placed this information in a second part of most titles when it was included. *Anesthesiology* used the second position more often but not exclusively.

Results-mention titles of clinical RAs were practically nil in the two general medical journals but were seen in the two subspecialty journals, possibly because the subspecialty we studied publishes a fair amount of animal and other laboratory research and familiarity with the form may have encouraged its adoption for clinical titles. Alternatively, the general medical journals may have internal policies in place. We were able to identify interesting phrasal alternatives to full sentences in the present tense for use in results-mention titles if authors want to hedge their claim slightly by making their title more similar to other phrasal RA titles in clinical medicine.

The only other content mention that differed in frequency was geographic location, which was relatively rare overall (10%) and which may not always be necessary. We would recommend authors consider carefully before they include geographic location in a title.

Length, which is no longer mentioned in the instructions to authors of the journals we studied, differed significantly between them. The NEJM was once again very different from the others; whereas

their titles have become shorter over time, the BMJ's have become longer. The EJA's are remarkably and significantly longer.

Other differences in formal features were the frequency of full sentences and titles with two or more parts (and their punctuation). Yet again, the NEJM was different from the others (no sentence titles, fewer multi-part titles, and more frequent use of dashes in them).

4.2. Researching Medical Title Content and Phrasing Systematically or on the Fly

Inter-rater agreement was high in this study, suggesting that with due diligence in training and norming of coders, researchers can reuse the coding method and question set to examine other journals. Sampling regular issues (not special issues with guest editors) to harvest around 70 titles should be adequate to reflect a journal's practice.

Although researchers interested in comparison with the literature may wish to include all our questions, we mention that questions Q3d and Q3e on heads and the words used to join them were included mainly to extend the work of Wang and Bai [17] beyond the NEJM. We found these two questions difficult to code, and the similarity we saw between our four journals and the results of their previous exhaustive examination suggests that further research is not a priority unless some form-function hypothesis emerges. To provide information on the usage of linking words (Q3e) for novice translators or instructors for example, it may be sufficient to use a concordancer to look at the frequency and placement of useful but less frequent links between heads (notably *or* and *versus* or *vs.*) in a corpus of titles. Even checking a corpus of complete RA articles should give sufficient information.

Our observations on the prevalence of sentence titles and results-mention titles in anesthesiology, which publishes a considerable number of RAs reporting bench science, suggest that future research should differentiate between *in vitro* or animal model research designs and strictly clinical research when coding Q1. Interactions between results-mention and sentence titling could then be confirmed without resort to *post-hoc* calculations. This change in the coding method might also facilitate the testing of our hypothesis that phrasal results titles can hedge claims made about less robust findings.

We conclude that per-journal sample sizes like those we used in this study will be adequate in future research that does not compare review and RA titles between journals. Another research design should be chosen if review article titles are important to know about. In support of this conclusion about sample size, we emphasize the fact that we were able to discern journal-specific patterns for content and phrasing, to be confirmed through further research, and that the 95% CIs for the length variable were very narrow.

Given the differences we detected, including ones that cannot be explained by explicit instructions to authors, we encourage authors and supporters of research writing to look at several target-journal tables of contents (but not guest-edited special issues) to detect differences in features. We would particularly recommend that authors and their supporters reassess titles when resubmitting to another journal after rejection.

For such on-the-fly research, we recommend looking mainly at where research design is most often mentioned (our Q1). Since most titles with two parts end by specifying research design, checking Q1 will simultaneously reveal whether multi-part titles (Q3b) predominate (as in the BMJ and EJA) or not (as in the NEJM) or are both present in substantial numbers (as in *Anesthesiology*). Rough length should also be noticed, as journal preferences seem to be implicit at this time, but exact counts are not needed for practice-based research. If a target journal requires results mention, we would want to notice how authors tend to provide this information—with sentences or phrases—and indeed, if they comply or not. Again, exact numbers are not needed. Other content that does not seem to vary between journals need not be checked specifically. Nor is there a need to look for geography mentions (our Q2c), we think, since its inclusion is a strategic decision related to the need to underline limited generalizability or to imply the intrinsic interest of a location at a specific point in time (e.g., in ebola research).

Supplementary Materials: The following is available online at www.mdpi.com/2304-6775/4/2/11/s1: File: S1: System for coding content and surface features of article titles in clinical medicine.

Author Contributions: M.E.K. and S.S. conceived the study and designed the coding and statistical analysis. M.E.K. and A.M. refined and carried out the coding. S.S. analyzed the data and advised on how to present it. M.E.K. drafted the paper, in consultation with A.M. and S.S., both of whom approved the submitted manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BMC	Biomed Central
BMJ	*British Medical Journal*
CONSORT	Consolidated Standards of Reporting
EAL	English as an additional language
EJA	*European Journal of Anaesthesiology*
ERPP	English for research publication purposes
ESP	English for specific purposes
IF	impact factor
JAMA	*Journal of the American Medical Association*
NEJM	*New England Journal of Medicine*
RA	research article
PLoS	Public Library of Science

References

1. Tenopir, C.; King, D.W.; Bush, A. Medical faculty's use of print and electronic journals: Changes over time and in comparison with scientists. *Med. Libr. Assoc.* **2004**, *92*, 233–241.
2. Khan, K.S.; Kunz, R.; Kleijnen, J.; Antes, G. Five steps to conducting a systematic review. *J. R. Soc. Med.* **2003**, *96*, 118–121. [CrossRef] [PubMed]
3. Mateen, F.J.; Oh, J.; Tergas, A.I.; Bhayani, N.H.; Kamdar, B.B. Titles *versus* titles and abstracts for initial screening of articles for systematic reviews. *Clin. Epidemiol.* **2013**, *5*, 89–95. [CrossRef] [PubMed]
4. Systematic reviews and meta-analyses: A step-by-step guide. Available online: http://www.webcitation. org/6eXMRFyg7 (accessed on 14 January 2016).
5. Langdon-Neuner, E. Hangings at the bmj: What editors discuss when deciding to accept or reject research papers. *Write Stuff* **2008**, *17*, 84–85.
6. Begg, C.; Cho, M.; Eastwood, S.; Horton, R.; Moher, D.; Olkin, I.; Pitkin, R.; Rennie, D.; Schulz, K.F.; Simel, D.; *et al.* Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* **1996**, *276*, 637–639. [CrossRef] [PubMed]
7. Moher, D.; Schulz, K.F.; Altman, D.G.; for the CONSORT Group. The CONSORT Statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann. Intern. Med.* **2001**, *134*, 657–662. [CrossRef] [PubMed]
8. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; the PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Med.* **2009**. [CrossRef] [PubMed]
9. Schulz, K.F.; Altman, D.G.; Moher, D.; for the CONSORT Group. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *J. Clin. Epidemiol.* **2010**. [CrossRef] [PubMed]
10. Goodman, R.A.; Thacker, S.B.; Siegel, P.Z. What's in a title? A descriptive study of article titles in peer-reviewed medical journals. *Sci. Editor* **2001**, *24*, 75–78.
11. Labassi, T. Reading titles of empirical research papers. *Read. Matrix* **2009**, *9*, 166–174.
12. Siegel, P.Z.; Thacker, S.B.; Goodman, R.A.; Gillespie, C. Titles of articles in peer-reviewed journals lack information on study design: A structured review of contributions to four leading medical journals, 1995 and 2001. *Sci. Editor* **2006**, *29*, 183–185.
13. Ubriani, R.; Smith, N.; Katz, K.A. Reporting of study design in titles and abstracts of articles published in clinically oriented dermatology journals. *Br. J. Dermatol.* **2007**, *156*, 557–559. [CrossRef] [PubMed]

14. Jacques, T.S.; Sebire, N.J. The impact of article titles on citation hits: An analysis of general and specialist medical journals. *J. R. Soc. Med.* **2010**, *1*. [CrossRef] [PubMed]

15. Habibzadeh, F.; Yadollahie, M. Are shorter article titles more attractive for citations? Cross-sectional study of 22 scientific journals. *Croat. Med. J.* **2010**, *51*, 165–170. [CrossRef] [PubMed]

16. Paiva, C.E.; da Silveira Nogueira Lima, J.P.; Paiva, B.S. Articles with short titles describing the results are cited more often. *Clinics* **2012**, *67*, 509–513. [CrossRef]

17. Wang, Y.; Bai, Y. A corpus-based syntactic study of medical research article titles. *System* **2007**, *35*, 388–399. [CrossRef]

18. Soler, V. Writing titles in science: An exploratory study. *Engl. Specif. Purp.* **2007**, *26*, 90–102. [CrossRef]

19. Cianflone, E. Scientific titles in veterinary medicine research papers. *Engl. Specif. Purp. World* **2010**, *9*, 1–8.

20. Taylor, C.F.; Field, D.; Sansone, S.-A.; Aerts, J.; Apweiler, R.; Ashburner, M.; Ball, C.A.; Binz, P.-A.; Bogue, M.; Booth, T.; *et al.* Promoting coherent minimum reporting guidelines for biological and biomedical investigations: The MIBBI project. *Nat. Biotechnol.* **2008**, *26*, 889–896. [CrossRef] [PubMed]

21. Oxford Centre for Evidence-based Medicine. Available online: http://www.webcitation.org/6eYfjBQ1i (accessed on 15 January 2016).

22. Corpas Pastor, G.; Seghiri, M. Specialized corpora for translators: A quantitative method to determine representativeness. *Transl. J.* **2007**, *11*. Available online: http://www.webcitation.org/6gDtLCZWu (accessed on 23 March 2016).

23. Maher, A.; Waller, S.; Kerans, M.E. Acquiring or enhancing a translation specialism: The monolingual corpus-guided approach. *J. Spec. Transl.* **2008**, *10*, 56–75. Available online: http://www.webcitation. org/6giYF6iGb (accessed on 12 April 2016).

24. Buscemi, N.; Hartling, L.; Vandermeer, B.; Tjosvold, L.; Klassen, T. Single data extraction generated more errors than double data extraction in systematic reviews. *J. Clin. Epidemiol.* **2006**, *59*, 697–703. [CrossRef] [PubMed]

25. Viera, A.J.; Garrett, A.M. Understanding interobserver agreement: The kappa statistic. *Fam Med.* **2005**, *37*, 360–363. [PubMed]

26. Burrough-Boenisch, J. Shapers of published NNS research articles. *J. Second. Lang Writ.* **2002**, *12*, 223–243. [CrossRef]

27. Jaime Sisó, M. Titles or headlines? Anticipating conclusions in biomedical research article titles as a persuasive journalistic strategy to attract busy readers. *Miscelánea: J. Engl. Am. Stud.* **2009**, *39*, 29–54.

28. Goodman, N.W. Survey of active verbs in the titles of clinical trial reports. *BMJ* **2000**, *320*, 914–915. [CrossRef] [PubMed]

29. Song, F.; Parekh, S.; Hooper, L.; Loke, Y.K.; Ryder, J.; Sutton, A.J.; Hing, C.; Kwok, C.S.; Pang, C.; Harvey, I. Dissemination and publication of research findings: An updated review of related biases. *Health Technol. Assess.* **2010**, *14*. Available online: http://www.webcitation.org/6giYZiKZN (accessed on 28 March 2016). [CrossRef] [PubMed]

30. Léauté-Labrèze, C.; Dumas de la Roque, E.; Hubiche, T.; Boralevi, F.; Thambo, J.B.; Taïeb, A. Propranolol for severe hemangiomas of infancy. *N. Engl. J. Med.* **2008**, *358*, 2649–2651. [CrossRef] [PubMed]

31. Day, R.A. *How to Write and Publish a Scientific Paper*, 5th ed.; Oryx Press: Phoenix, AZ, USA, 1998.

32. Burgess, S.; Pallant, A. Teaching academic writing in Europe: Multilingual and multicultural contexts. In *Supporting Research Writing: Roles and Challenges in Multilingual Settings*; Matarese, V., Ed.; Chandos Publishing: Oxford, UK, 2013; pp. 1–14.

33. Burgess, S.; Cargill, M. Using genre analysis and corpus linguistics to teach research article writing. In *Supporting Research Writing: Roles and Challenges in Multilingual Settings*; Matarese, V., Ed.; Chandos Publishing: Oxford, UK, 2013; pp. 55–71.

34. Lounds, A. Abstracts and introductions: Genre analysis for editors and translators of research articles. Mediterranean Editors and Translators (workshop abstract). Available online: http://www.webcitation. org/6ebp6ZD7M (accessed on 17 January 2016).

35. Elia, N.; Tramèr, R. How to write a good title. *Eur. J. Anaesthesiol.* **2011**, *28*, 819–820. [CrossRef] [PubMed]