



Searching for The Holy Grail How valuable are metrics in medical writing?

by Wendy Kingdom

Estimates of the number of copies of *The Da Vinci Code* sold worldwide since its publication in 2003 range from 15 million to 38 million [1]. It is unfortunate that we do not have accurate figures because the number of copies of a book sold is an objective metric. Therefore, it should tell us whether or not Dan Brown is a good author and worth our investment of buying a copy of his book. Alternatively, we could view the star rating metric of this book on the Amazon website and use this for our decision. Sadly, although *The Da Vinci Code* is clearly a 'best seller', it gets only three and a half stars out of five in terms of readers' opinions. Crikey! How much must a 5-star author earn?

The question of applying metrics to medical writing arises periodically because sponsors have a need to evaluate a service before they invest in it. There is a corresponding commercial need by Contract Research Organisations (CROs) to demonstrate their ability to meet potential customers' requirements. However, by definition, metrics have to be things that you can measure, weigh, count or otherwise attach a number to. They are not intelligent, i.e. they are unable to adapt to varying situations and are unable to take account of influencing factors. On the other hand, good medical writing is all about taking the best approach according to the target audience, the data, the purpose of the work etc., which you simply cannot count. Therefore, the logical thing to do is to forget about metrics in medical writing and get back to work. Sadly, this is not the way of the world and if there is a commercial need for metrics, sooner or later we will all be judged by them.

If we really need to identify metrics for medical writing, we can have some fun brainstorming medical writing output and attaching numbers, e.g., the number of pages in a document, the weight of the document, the number of words on a page, and so on. However, if we want to try and be sensible, we might consider the following:

- The number of documents written (i.e. experience).
- Average time to produce a document.
- The number of papers published.
- The number of citations of a paper.
- The number of quality control (QC) findings.

The problem is that even these relatively sensible metrics are all seriously flawed. For example, the number of a type of document written, i.e. experience, tells you no more and no less than how many. It doesn't say anything about how clear, interesting, sensible, suitable, thorough, or consistent

the documents were, nor does it say anything about any writer's potential when they are presented with new challenges. It also doesn't tell you whether most of a company's experience was held by one person who has since gone to work elsewhere.

Time to produce a document might give some indication of efficiency, but this metric is influenced by the experience of the medical writer, the complexity of the document, the number of source documents, the volume of data, and the clarity of the source documents. Assuming that a shorter time is a better time in metric terms, does this mean that an inexperienced writer is a bad writer? A journey of a thousand miles begins with a single step (Confucius). Even the most experienced writers started somewhere. Alternatively, if fast is the most important variable, a document can be divided into small parts, each of which can be written by different writers in parallel. The output will be rapid but the document is likely to be disjointed. As soon as we start to ask questions about the variables that can influence any metric, the list of flaws rapidly outweighs any merits.

The number of papers submitted for publication and that are actually published might, theoretically, give some indication of the quality of writing. However, a medical writer needs little experience of writing papers to recognise that many would-be authors are simply deluded about the value of their research. It's tempting to write the letter of rejection from *The Lancet* yourself and not waste any time on the work in the middle, but that is not sporting, nor is it good customer service, so we do our best with what we have. It is no surprise, therefore, to see that poor quality

writing is not included in The World Health Organization's top 10 reasons for rejecting a paper (see Box).

Good writing cannot be quantified any more than you can measure good acting, good painting, or beauty.

Despite the obvious flaws in assessing value or ability by counting the number of papers published, this metric has been in operation in academic institutions for many

years. Now, researchers rush to publish work early so that they can meet deadlines for their appraisals, heads of departments put their names on papers irrespective of their true involvement in the work, and there is an inevitable consequent reduction in scientific depth per paper published. Was the gain worth the loss? Not in my opinion.

Another consequence of judging academic researchers by the number of publications is that there has been an explosion in the volume of publications, and the number of journals needed to accommodate all of the papers. This means

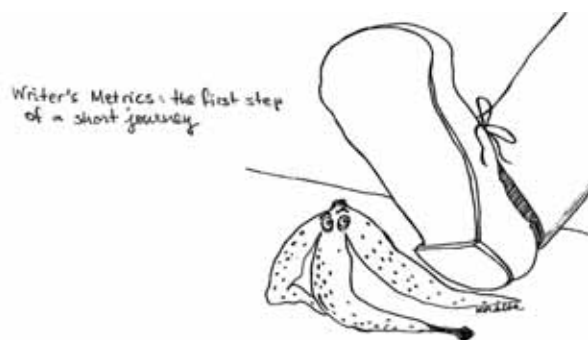
Searching for The Holy Grail...

that we now have a bewildering array of journals to choose from when we are interested in a particular area of research. So now we need a metric that helps us to judge the worth of the publications. What can we count? We can count the number of times a publication is cited. Problem solved?

The number of citations is such an important metric that it has a name: bibliometrics. Well, citations are good up to a point but all sorts of games can be played to increase impact scores, e.g., taking one study, producing five papers from it, publishing the papers in series and citing every published paper in the series. There is no metric for evaluating whether one paper would have been more informative to the scientific community than the five separate papers so we don't try. It's too difficult so we don't do it, especially when we already have something that we can count.

Turning now to regulatory writing, there is a real danger that the target audience for any document will cease being the reader (who isn't involved in the metric process) and will become the metric itself. If writing for the public is judged on a readability score, the writer can select from a panoply of pithy words and keep the readability score low. Of course the reader might not be able to understand the text as well as if it had been a lengthier but detailed explanation using longer more precise words, but nobody is counting that so it will cease to matter.

A metric that can appear at first glance to be a relatively sensible measure of quality is the number of QC findings. However, this is not an intelligent metric in that it cannot incorporate factors such as complexity of the study, quality of the source information, time pressures, etc. In essence, QC is a consistency check whereas a good medical writer will make adaptations according to the target audience, the data, etc. Alistair Reeves points out to us how the terminology used in the Medical Dictionary for Regulatory Activities (MedDRA) has been creeping into



pharma-company language [2]. He tells us that he was required to write, 'One patient developed hallucination auditory'. I think that Alistair's example illustrates my point very nicely. The clinical study report and the MedDRA dictionary have different audiences and different purposes. Coding is used to build a computerised safety database across studies and to compile lists, whereas a clinical study report is intended to be read by a human being who wants to know what happened to the study subjects. In Alistair's example, the audience has ceased being the reader of the report and has become a disembodied concept called consistency. By applying an unintelligent metric to intelligent work we must inevitably dumb down the output. Consistency, though unquestionably very important, is not always the most important thing under all circumstances.

Furthermore, if QC findings are used as metrics, the dilemma described above becomes a nightmare. Should the medical writer use the standard medical term in the text and risk a QC finding of 'inconsistency' with the source data, or use the MedDRA term and risk appearing to be an idiot? Yet if the same medical writer used the same data to write a paper for publication, there is no question that the medical term should be used—same medical writer, same data, same metric; different score.

Yet things can get even worse. As soon as we start to use QC as a metric, the value of the QC process will become corrupted and distorted. QC is part of the overall quality assurance umbrella that is intended to assure the ethics and quality standards of a study. It should be a positive, team-building process whereby project teams work together to produce high quality documents. If QC is used as a metric, project teams will break down into groups of individuals unwillingly participating in a point-scoring war. Every finding will be fought over, political battles will be won and lost, and resentments burrowed deeply. This is bad for morale, bad for medical writing, bad for quality assurance and bad for clinical research.

The truth is that good writing cannot be quantified any more than you can measure good acting, good painting, or beauty. If we accept the premise that it is possible to apply metrics to medical writing, before long, computer programs will be written that will conjure up a composite score that is intended to quantify our writing. No doubt the score will be given a technical-sounding name so that it is not obviously analogous to an Amazon star rating but we

>>>

Top 10 reasons for rejecting a manuscript

1. Content of the paper not suitable for an international journal of public health.
2. Design of the study not appropriate for the question asked.
3. Lack of novelty and or timeliness.
4. Lack of either or both ethical committee approval and informed consent.
5. Lack of an appropriate search strategy.
6. Conclusions not justified by the results.
7. Lack of a feedback step in descriptions of audit.
8. Insufficient sample size.
9. Lack of a clear message to the public health community.
10. Secondary analyses of demographic surveys or simple prevalence studies that are difficult to generalise.

<http://www.who.int/bulletin/contributors/rejection/en/index.html> (World Health Organization)

>>> Searching for The Holy Grail...

will be befuddled into believing that the score has some value. Or worse, our managers and clients will believe that the score has value and will measure us against it. As soon as we know that we are being judged by a metric, we will adapt our writing to improve our scores irrespective of the needs of the audience. Like trying to pick up jelly with your fingers, you can do it but, in the attempt, you change its substance.

My fear is that we are heading towards valuing what we measure because we can measure it, not because the metric is intrinsically valuable. In reality, we decide whether or not to read a book based on previous experience of that author, recommendations by our friends, and reviews.

Sponsors select CROs and freelancers based on previous experience, recommendations, and interviews. Nobody has devised a meaningful metric for medical writing yet because there isn't one. We must accept that and move on.

Wendy Kingdom

*Wendy Kingdom Limited
Somerset, UK
info@wendykingdom.com
www.wendykingdom.com*

References:

1. <http://answers.google.com/answers/threadview?id=610195>
2. Alistair Reeves. MedDRA 'preferred' terms. TWS 2007;16(2):49.

A Comment not a Counter

When I was approached by the editor, Elise, about writing an article on the pros of medical writing metrics as a counter argument to Wendy's article, my first reaction was "no problem!" After all, when I worked at a Contract Research Organisation (CRO), I was often asked to supply 'numbers' for potential clients. The 'numbers' were used in response to a client's 'Request for Information' (RFI). Questions such as the following were regularly posed:

- "How many and what type of documents have you prepared in the last 2 years?"
- "How many weeks does it take you from Last Case Report Form in house to Final Clinical Study Report (CSR)?" That one had a veritable minefield of variables attached to it so when possible we responded with the manageable 'Number of weeks from Final Data Listings to Draft CSR'.

Yes, I could certainly come up with 'numbers'! For example, I readily knew how many CSRs had been issued by our CRO's global writing group in a defined period, although I used to worry that this number may have been viewed 'out of context' ... I needed the reviewer to be aware that the number of available medical writers had varied during that time, as had their level of experience and, most importantly, for each CSR there had been factors beyond the writer's control such as the highly variable quality of the data and the varied experience of clinicians, statisticians, and other contributors to the reports.

Looking back more carefully, I recalled my concerns about collating these numbers, frequently asking myself "just how valuable are the metrics we supplied?" Working for a CRO at the time, and later as a freelancer, however, I believed that if you wanted a chance of getting a particular job, then you just had to live with the metric requests, and supply responses with assumptions or annotations where possible.

When I read Wendy's article, I found myself agreeing with everything she wrote. So, how could I produce a counter argument? As Wendy stated "Nobody has devised a meaningful metric for medical writing yet because there isn't one. We must accept that and move on." With nothing more to add, I was going to decline Elise's kind offer when I happened to mention Wendy's article to a colleague.

"Don't abandon them (metrics) ... they must be of some use? Figure out how to apply them" they said but didn't offer any suggestions. I also don't have an answer but their plea did make me wonder about the people who use metrics. So, I spoke to a few colleagues at different pharmaceutical companies that outsource complete trials or medical writing-related tasks such as protocol and CSR preparation. From my, admittedly limited, survey, it appears that there are two approaches regarding metrics.

Some companies simply ignore metrics in relation to outsourcing, with decisions made on the basis of personal experience or recommendation from a trusted colleague. This appears to be the route taken when a clinical team makes the decisions; selectors go with gut reactions, and the chemistry between individual people and groups.

Metrics are of interest, however, where there is an 'Outsourcing Department' or a formal outsourcing procedure. Detailed RFIs including requests for metrics are issued to potential service providers. Responses to the metric-related queries appear to be used to assure the outsourcer that the potential providers are at least in the right 'ball park', that processes are being followed, and pre-defined standards are being met. In addition, for some companies, gathering metrics from potential providers appears to serve another purpose—showing senior management that a 'fair selection' process is in place.

For writers providing responses to metric requests, this can be a way to get a 'foot through the door' with a new client. The challenge, however, is not only to come up with the numbers but also to provide context, possibly in the form of annotations if the RFI allows. The explanations need to be sufficiently clear such that the outsourcer appreciates that metrics provided by others are not necessarily directly comparable. Of course, ensuring the outsourcer realises this is not something fully in the writer's control.

A final comment: the outsourcing manager who said they were reluctant to throw out metrics completely was also the same person who said they "were slightly suspicious of numbers" ... so, for now, I'm staying in Wendy's camp!

Barbara Grossman

barbara.grossman@hawkeye-medical.co.uk