

Datum isn't; data are

The views expressed in this editorial are those of the author and do not necessarily reflect the position of the publisher.

Selection of manuscripts for publication in a scientific journal is a subjective process. Safeguards, such as using multiple independent reviewers, review templates and conflict of interest guidelines, try to make the process fair. It is clear to editors that some reviewers' judgments are guided by whether they just liked a manuscript or not. The editor's job is to divine if the reasons for the opinion are valid. For some manuscripts, the description "poorly written" weighs heavily in the mind of the reviewer. Use of the term "data" as a singular noun is often seen as a marker of either scientific naïveté or poor writing. I suspect that failure to use "data" as a plural noun has, on occasion, resulted in dismissal of an otherwise acceptable manuscript. Hopefully not in *C7S*.

Acceptance of the singular form of the word "data" has increased in all fields except academic medicine. The justification cited in medicine is that the word "data," the Latin plural of "datum," refers to multiple facts or observations. In ordinary speech where use of the plural form of data can be cumbersome and pompous, the singular noun is much more acceptable. Journalists use only the singular form, regardless of whether their reports are written or spoken. Other branches of science, such as computing and engineering, have abandoned use of "data" as a plural noun. A search of Google Books shows that the use of plural form of "data," which once outnumbered the singular form by a factor of 4, has been reduced to equality in the last 2 decades. However, a similar search of PubMed shows the plural form in academic medicine to have remained at 3 times the use of the singular form. Therefore, academic medicine, alone among all the sciences, has stuck firm to the requirement for the plural form of "data." Curiously, the term "datum" is almost never used and appears to be replaced by compound nouns, such as "data point." So, is the word "data" singular or plural, and does it matter?

We cannot ask the Romans because they did use not the words "datum" or "data" as we do in science. The past participle of the verb to give, the words may have been applied as adjectives meaning "given." For example, "data" was written on a letter to signify that it was given to the recipient. However, "data" in this example is the feminine form of the word to agree with the word "epistola" rather than the plural. Interestingly, this use

led to the modern word "date" and its use in letters. Since the seventeenth century logicians have applied the word "data" to elements of a premise upon which conclusions are deduced. Here, "given" implies that further proof is not required. Since the premise should consist of facts if a conclusion is to be sound, the origin of the modern scientific term "data" to mean facts or observations is explained. Many writers credit Pullen¹ with the earliest use of the modern scientific meaning, but a quick perusal of his book, which is available online, reveals that he used "mathematical data" as the title to a chapter containing common trigonometry formulas, the equivalent to the logicians' given premise. Others credit the Computation Laboratory at Harvard University for the first modern use of the term "data" when they described, in 1946, the cards used to feed the computer "with empirical or other data." However, these writers, and subsequent computer scientists, used "data" to mean the functions programmed by the cards. The introduction of computers into science and medicine in the 1960s resulted in a gradual shift of this meaning to its modern form in science that describes the observations themselves rather than the methods used to analyze them.

Therefore, "data," a very recently coined term in academic medicine, has evolved as required by writers for the purposes of clear communication. The belief that "data" must be treated as a plural noun is not part of that evolution. It is worth remembering the origin and fate of a similar word, "agenda." It was also coined in the seventeenth century as a noun to describe a collection of items with no "singular form." This time the word was derived from the Latin verb *agere* (to do) and had a theological meaning of things to do. Its use evolved in the Nineteenth century to mean items of business at a meeting. A collective noun, it is always treated with the singular tense. Like "datum," its singular form, "agendum," feels contrived.

There are many situations where "data" is meant as a collective noun to describe the body of observations of an experiment, and use of the singular tense is more natural. There are other situations where "data" describe multiple discrete observations and the plural feels more appropriate. This editor's advice is to use the phrase that feels most like your spoken voice. Our promise is

that use of the singular or plural tense with the word “data” will have no impact on the evaluation of a manuscript for *CJS*.

Vivian C. McAlister, MD

Coeditor, *Canadian Journal of Surgery*

Acknowledgements: The idea for this editorial, for which the author is solely responsible, grew out of enlightening conversations with Prof. Paul Ridgway, Dublin, Ireland.

Competing interests: None declared.

DOI: 10.1503/cjs.009316

References

1. Pullen WWF. Engineering and mathematical data. The Scientific Publishing Company, Manchester, UK; 1899. Available: <https://archive.org/stream/engineeringtable00pullrich#page/n4/mode/1up> (accessed 2016 July 9).

De l'usage du pluriel (ou non) du mot « data »

Les opinions exprimées dans cet éditorial sont celles de l'auteur et ne représentent pas nécessairement celles de l'éditeur.

La sélection des manuscrits pour publication dans une revue scientifique est un processus subjectif. Des mesures sont prises pour assurer un certain contrôle et rendre la démarche aussi équitable que possible, notamment le recours à de multiples examinateurs indépendants, l'imposition de grilles d'examen et la mise en place de lignes directrices sur les conflits d'intérêts. Les rédacteurs voient bien que le jugement de certains examinateurs est guidé par leur sympathie (ou leur antipathie) pour un manuscrit. Le travail de l'éditeur consiste à discerner le bien-fondé des opinions. Dans certains cas, la description « mal écrit » pèse lourd dans l'esprit de l'examineur. L'utilisation (en anglais) du mot « data » comme nom commun singulier est souvent perçue comme signe soit de naïveté scientifique ou de faute de grammaire. Or, je soupçonne que le défaut d'attribution du pluriel au mot « data » a eu comme résultat, à l'occasion, le refus d'un manuscrit par ailleurs acceptable. Pas dans le *JCC*, espérons-le.

L'usage du mot « data » au singulier s'est étendu à tous les domaines, sauf aux milieux universitaires de la médecine. En médecine, on justifie ce refus par le fait que le mot « data », pluriel du latin « datum », renvoie à de multiples faits ou observations. Dans le langage ordinaire, où l'utilisation de la forme plurielle de « data » peut paraître lourde et pompeuse, le nom au singulier est beaucoup plus acceptable. Les journalistes utilisent uniquement la forme singulière, à l'écrit ou à l'oral. D'autres branches de la science, notamment l'informatique et le génie, ont abandonné l'usage de « data » comme nom pluriel. Une recherche dans Google Books révèle que « data » pluriel, autrefois 4 fois plus fréquent que « data » singulier, a évolué au cours des 2 dernières décennies jusqu'à atteindre une fréquence d'utilisation

égale. Une recherche semblable dans PubMed révèle toutefois que, dans les écrits savants en médecine, la forme plurielle est 3 fois plus utilisée que la forme singulière. Ainsi, la médecine, seule parmi toutes les sciences, est demeurée ferme dans son utilisation de la forme plurielle de « data ». Curieusement, le mot « datum » n'est presque jamais utilisé et semble être remplacé par des expressions telles que « data point ». Alors, le mot « data » est-il singulier ou pluriel, et cela a-t-il de l'importance?

Nous ne pouvons pas consulter les Romains, puisqu'ils n'utilisaient pas les mots « datum » et « data » comme nous le faisons aujourd'hui en sciences. Pour le participe passé du verbe « to give » (donner), ils utilisaient peut-être des adjectifs signifiant « given » (donné), par exemple, le mot « data » écrit sur une lettre pour indiquer qu'elle était donnée au récipiendaire. On note toutefois que le mot « data » dans cet exemple est au féminin, car il doit s'accorder avec « epistola » plutôt qu'avec le pluriel. Il est intéressant de voir que cet usage a donné lieu au mot contemporain « date » et à son usage dans la correspondance. Depuis le dix-septième siècle, les logiciens ont appliqué le mot « data » aux éléments d'une prémisse de laquelle des conclusions sont déduites. Ici, le mot « given » sous-entend qu'aucune autre preuve n'est requise. Puisqu'une prémisse doit être constituée de faits pour que la conclusion soit valide, cela explique l'origine de l'acceptation scientifique moderne de « data », soit faits ou observations. Beaucoup d'auteurs attribuent à Pullen¹ la première utilisation du mot dans sa signification scientifique moderne, mais un coup d'œil à ce livre, que l'on peut consulter en ligne, révèle qu'il a utilisé l'expression « mathematical data » comme titre d'un chapitre portant sur les