Computers in Biology and Medicine 43 (2013) 962–965

Contents lists available at SciVerse ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Letter to the Editor

A Lack of statistical pitfalls in the comparison of multivariate causality measures for effective causality

A B S T R A C T

In a 2011 paper, Wu et al. Comp. Biol. Med. 41 (2011) 1132–1141, compared the performance of several standard causal connectivity measures including Granger Causality (GC) using both simulated data sets and real magnetoencephalography data. Parameters for the causal connectivity measures were obtained using the Dynamic Autoregressive Neuromagnetic Causal Imaging (DANCI) algorithm. In a letter, Dr. Florin and Dr. Pfeifer Comp. Biol. Med. 43 (2013) 131–134, outline four shortcomings of Wu et al. Comp. Biol. Med. 41 (2011) 1132–1141, study. We provide counterarguments for the appropriateness of our approach and demonstrate how, despite any shortcomings, the Wu et al. Comp. Biol. Med. 41 (2011) 1132–1141 study provides an important and valid analysis of these various causal connectivity methods. In particular, none of the findings are consistent with limitation of the dynamic autoregressive neuromagnetic causal imaging (DANCI) algorithm and/or Granger causality (GC) method described by Frye and Wu Comp. Biol. Med. 41 (2011) 1118–1131. In fact, many of the limitations raised by Florin and Dr. Dr. Pfeifer illustrate the significant advantage of the DANCI algorithm and GC method for the analysis of causal connectivity.

1. Introduction

I thank Dr. Florin and Dr. Pfeifer [1] for their interest in the manuscript that I co-authored with Wu [2]. I would have to agree with Dr. Florin and Dr. Pfeifer's sentiment; it is important to evaluate the performance of the methodology used for connectivity analysis. This sentiment is exactly what motivated the Wu et al. [2] paper and the companion paper in the same issue by Frye and Wu [3]. In Wu et al. [2] certain practical limitations had to be made to allow the comparison of the various causality methods. However, these limitations were not significant and do not invalidate the results. Most importantly, none of the findings are consistent with limitation of the dynamic autoregressive neuromagnetic causal imaging (DANCI) algorithm and/or Granger causality (GC) method described by Frye and Wu [3]. Below I provide a point-by-point explanation for the suggested issues. As the reader can see, within the limitations of Wu et al. [2], the conclusions are sound.

2. Methodological concerns

2.1. Significance computation

Dr. Florin and Dr. Pfeifer [1] are concerned that the procedure used to compare the accuracy of the causality methods does not adequate represent the true accuracy of the DANCI algorithm and GC method. Dr. Florin and Dr. Pfeifer point out that the counting method was used to compare the various causality methods and further points out the reason why this method for comparison was used as well as its limitation. Specifically, Dr. Florin and Dr. Pfeifer correctly points out that only the GC measure provides a known statistical distribution that can be used to set a predetermined significance threshold in order to determine which connections are significant whereas the other causality measures do not provide well described statistical distributions thereby limiting the ability to select a comparable statistical threshold across causality methods. Indeed, Dr. Florin and Dr. Pfeifer provides a good argument for adopting the DANCI algorithm and GC method as it provides a rational for determining which connections are significant and which are not significant whereas the other causality measures leave the procedure for determining a threshold open to speculation and involve the application of various untested methods.

However, Dr. Florin and Dr. Pfeifer argue that the 100% accuracy found for the GC measure using the counting method is unrealistic if the known GC statistical distribution was used to set to a threshold based on a level of statistical significance. Dr. Florin and Dr. Pfeifer points to the fact that statistical tests are based on a preset α probability which, by definition, sets the level of falsely rejecting the null hypothesis. In addition, Dr. Florin and Dr. Pfeifer go on to say that this risk increases with the number of channels involved because often the number of actual connections does not increase as fast as the number of potential connections. In a footnote Dr. Florin and Dr. Pfeifer discuss the fact that the only way to achieve 100% accuracy is to set the α probability to zero combined with an effect size large enough to prevent a Type 2 error from occurring. However, Dr. Florin and Dr. Pfeifer then go on to say that in practical terms this is not possible as α and β probabilities are inversely related.

While Dr. Florin and Dr. Pfeifer is correct that α and β probabilities are inversely related to each other, this relationship is not absolute, rather the relative relationship between the α and

DOI of original article: http://dx.doi.org/10.1016/j.compbiomed.2012.11.009

0010-4825/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2013.02.001

β probabilities depends on the effect size and the sample size. First we will consider the effect size which is essentially the difference between the two distribution's means divided by the standard deviation. With a large effect size or a large sample size, both the α and β probabilities can be very small (near zero). Below I illustrate this graphically with the help of G-Power version 3.1.3 (Düsseldorf, Germany).

Fig. 1A is a plot of the null distribution (H_0) in red on left side of the figure and the alternative distribution (H_1) in blue on the right side of the figure. For this illustration we will use z distributions with standard deviations of 1. H_0 has a mean of 0 while H_1 has a mean of 3. In this example the α probability is 0.05. We can see that the β probability is essentially the proportion of the H_1 distribution that falls to the left of the critical z (which is set by the α probability). In this example the β probability is 0.09. Fig. 1B shows the same distributions with the α probability changed to 0.01 which moves the critical z higher and increases the β probability to 0.25. So we see that α and β probabilities are indeed related for this specific set of distributions. But now let us change the characteristics of the distributions. Fig. 1C demonstrates the same distribution parameters with an increase in the effect size. The effect size can be increased by increasing the difference between means or reducing the standard deviation. In this case we will double the difference between the means to 6. An α probability of 0.01 results in a β probability of 0.0001. Now, let us double the difference between the means again to 12 (Fig. 1D). In this case an α probability of 0.01 results in a β probability that is essentially zero (below the limit of the software). In Fig. 1E we can even decrease the α probability to 0.0000001 (the smallest α probability for the software) and the β probability is still essentially zero. So, in conclusion, although α and β probabilities are not really zero, if the effect size is large enough, the α and β probabilities can be very close to zero and thus you can differentiate the two distributions (in this case H_0 and H_1) from each other with perfect accuracy in practicality. Although it could be argued that such an effect size is not possible, in actuality this depends on the lower limit of the magnitude of the connectivity that is wished to be detected and the signal-to-noise ratio of the recording instrument. However, even if one rejects this argument, there is an alternative factor (i.e., sample size) which can improve the ability to differentiate the two distributions (in this case H_0 and H_1).

In neurophysiological measurements such as magnetoencephalography (MEG) and electroencephalography (EEG), sample size can be an advantage. For example, many MEG recordings are performed at 500 Hz or above, providing sufficient data points for which to base connectivity calculations (at least when looking at medium to high frequencies). In fact, in the analyses of MEG signals during the 500 ms pre-stimulus period of a language task using the short-window approach we were able to obtain approximately 39,300 samples per participant [4,5]. Now let us see how large sample sizes translate to differentiating two distributions (H_0 and H_1). Fig. 2A demonstrates a *t*-distributed sample with the H_0 mean set to 0 and the H_1 mean set to 1 with a standard deviation of 1. In this example the sample size is five for both groups and the α probability is 0.05. This results in a β of 0.58. Now we can increase the sample size by 10 fold to have 50 samples for each distribution. With the α probability at 0.05, the β probability is calculated as 0.0005 (Fig. 2B). If we reduce the α probability to 0.01, the β probability is calculated as 0.005 (Fig. 2C). Now if we increase the number of samples by 10 fold again to 500 for each distribution, we find that an α probability to 0.01 produces a β probability that is essentially zero (Fig. 2D). Now by decreasing the α probability to 0.0000001 (the smallest alpha for the software) and the beta is still essentially zero (Fig. 2E).

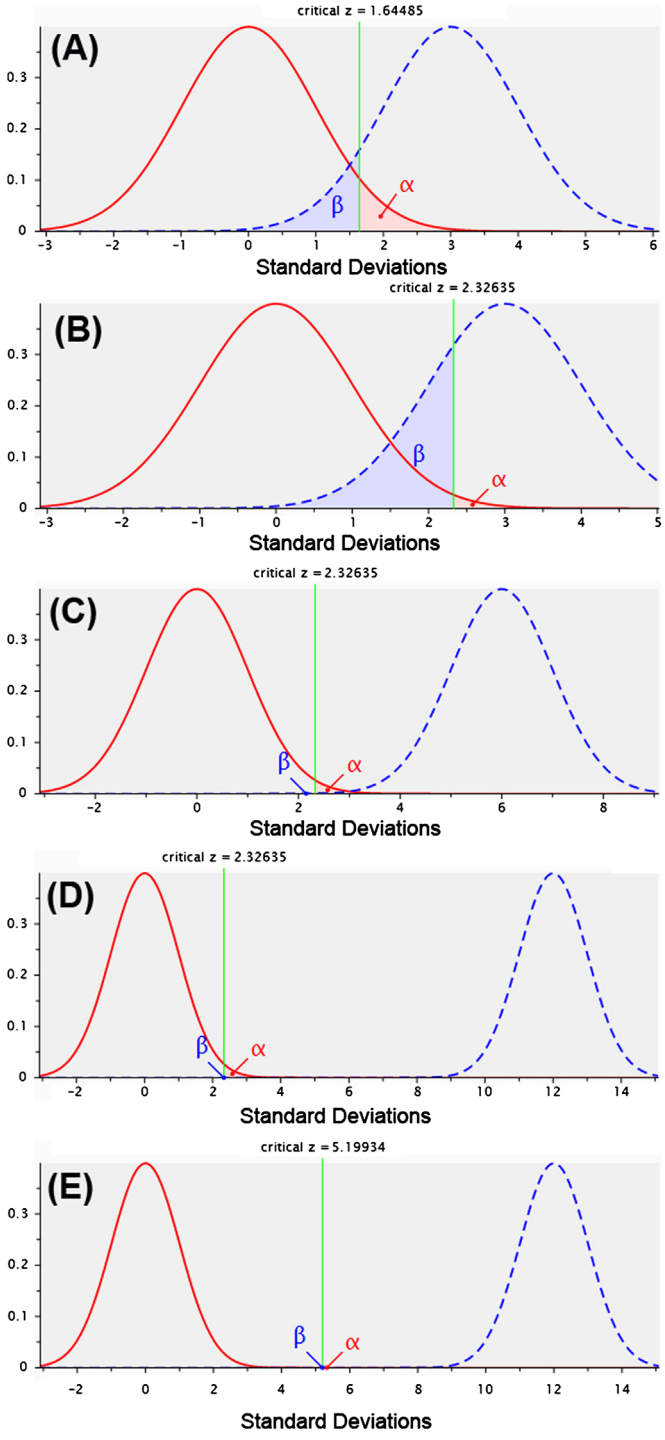

Fig. 1. Statistical distribution overlap depends on effect size. In this example *Z* distributions with standard deviation of 1 are used to illustrate this effect and H_0 (red distribution on left side of figure) has a mean of 0. (A) H_1 (blue distribution on right side of figure) has a mean of 3 and α probability is 0.05. (B) H_1 has a mean of 3 and α probability is set to 0.01. (C) H_1 has a mean of 6 and α probability is set to 0.01. (D) H_1 has a mean of 6 and α probability is set to 0.01. (D) H_1 has a mean of 12 and α probability is set to 0.01. (E) H_1 has a mean of 12 and α probability is set to 0.0000001. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

It could also be argued that the effect size is just not big enough even with a large sample size to separate the distributions. However if we take some examples from our application of MEG data we can see that this is just not the case. In our studies we used an α probability of 0.0001 and approximately 39,300