Journal of Traditional and Complementary Medicine 8 (2018) 1–3

Contents lists available at ScienceDirect

Journal of Traditional and Complementary Medicine

journal homepage: http://www.elsevier.com/locate/jtcme

Perspective

Homeopathic *Arnica* from Boiron and post-operative bleeding in mastectomized women in Milan: Statistical flaws and bias to be addressed

Salvatore Chirumbolo [a, *], Geir Bjørklund [b]

[a] Department of Neurological and Movement Sciences, University of Verona, Italy
[b] CONEM-Council for Nutritional and Environmental Medicine, Mo i Rana, Norway

ARTICLE INFO

Article history:
Received 8 March 2017
Accepted 2 May 2017
Available online 19 May 2017

Keywords:
Arnica
Homeopathy
Breast cancer
Mastectomy
Statistics
Bias

In a recent paper, Sorrentino et al., reported that a homeopathic preparation of an extract from *Arnica montana* L, known as *Arnica* 1000 K, significantly reduced the amount of post-operative blood and seroma (improving bleeding) in female subjects who underwent an unilateral total mastectomy.[1] This double blind randomized controlled trial versus placebo was performed throughout two years-long period, from 2012 to 2014, and about 53 patients (26 cases and 27 controls) with an age range 20–75 years and hospitalized for unilateral total mastectomy, were enrolled.[1] In their study, the authors stated that patients underwent also a standardized protocol of treatment for surgical procedures including an antithrombotic prophylaxis with 4000 UI heparin 12 h before surgery.[1] Patients who followed a randomization blinded protocol received a 30% v/v ethanol preparation of *Arnica* 1000 K (five drops sublingual) or a placebo (i.e. simply 30% v/v ethanol) from 1 day before surgery to the fourth day following surgery, which was performed by the same surgical team.[1] Those patients treated with *Arnica* 1000 K and collected in the group who excluded subjects who violated the protocol (PP group), showed less blood leakage and seroma (serum from surgical drainages) release from post-operatory drainages, reporting a better bleeding process.[1]

Arnica 1000 K is a korsakovian homeopathic dilution, made into 30% v/v ethanol, which according to the authors should not contain any toxic compound, as the starting material is very highly diluted.[1] Starting from a 0.04% sesquiterpene lactones (SLs) amount, which following previous reports might be about 1.0 mM SLs,[2] *Arnica* 1000 K is absolutely far away from a chemical ponderal dosage, as the dilution is made by emptying completely the vial with the previous content, refilling it with only 30% v/v ethanol, shaking strongly, emptying it again, refilling with the same solvent, shaking, emptying and so following this iterative process 1000 times, reaching the last step in a refilling with the alcoholic solvent.[1] The method is practically similar to the one used by Boiron, which prepared *Arnica* 1000 K and placebos for the study, for its *Oscillococcinum* 200K.[3]

In this perspective it is rather difficult to achieve a molar mass evaluation of SLs in the 30% ethanol solvent, as at least theoretically each emptying procedure might leave less than 10% of the total volume, according to the authors[1] and our calculations should reach a value that exceeds the Avogadro-Loschmidt's threshold already following only 7–8 passages, thus leading to the consideration that *Arnica* 1000 K does not contain any SL molecule but simply solvent. Actually, if considering that *Arnica* 1000 K is practically a control (placebo) from a chemical point of view, any resulting value at $p < 0.05$ through a Student's t-test, might lead to false positives if variances in the distribution are biased and inhomogeneously dispersed. The authors did not succeeded in retrieving statistically significant differences with the exception of a single $p < 0.03$ in the third treatment model in the PP group.[1] The statistical re-calculation of the data is shown in Table 1. In the *Arnica* ITT group (26 subjects) means followed a normal distribution (Anderson–Darling test $p = 0.394$, Shapiro–Wilk test $p = 0.307$, Cramer–von Mises test $p = 0.430$), as like as placebos (Anderson–Darling test $p = 0.136$, Shapiro–Wilk test $p = 0.134$, Cramer–von Mises test $p = 0.137$) and apparently their two sided paired t-test gave a value of $p = 0.04331$, with $t = -2.34989$.[1] Yet, a Grubb's test for potential outliers, if applied, would turn this significance to a p value = 0.07422.

* Corresponding author. Department of Neurological and Movement Sciences, University of Verona, Strada Le Grazie 9, 37134 Verona, Italy. Fax: +39 0458027403.
E-mail address: salvatore.chirumbolo@univr.it (S. Chirumbolo).

Peer review under responsibility of The Center for Food and Biomolecules, National Taiwan University.

http://dx.doi.org/10.1016/j.jtcme.2017.05.001

Table 1
Statistics reappraisal of data reported in Ref 1.

Test	Data	Statistics	Samples	p
Kolmogorov–Smirnov	ITT group	0.4 Mean 1 = 62.194; Mean 2 = 60.418 Diff = 1.776	Variability	0.87302
	PP group	0.6 Mean 1 = 41.886; Mean 2 = 54.092 Diff = −12.206		0.35714
	ITT group	0.2 Mean 1 = 77.424; Mean 2 = 79.602 Diff = −2.378	Mean	1
	PP group	0.4 Mean 1 = 68.82; Mean 2 = 87.562 Diff = −18.742		0.87302
	ITT + PP	0.4 Mean 1 = 52.04; Mean 2 = 57.255 Diff = −5.215	Variability	0.41752
		0.3 Mean 1 = 73.122; Mean 2 = 83.682 Diff = −10.56	Mean	0.78693
Wilcoxon-Mann Whitney (for data <30)	ITT + PP	18 Variable 1 = 18.55758; Variable 2 = 18.28493	Variability	0.375
	ITT + PP	7 Variable 1 = 38.34219; Variable 2 = 48.05253	Mean	**0.03711**
	ITT + PP (1)	7 Variability 1 = 30.79461; Variability 2 = 37.26469		0.07422

(1) If considering Grubbb's test resulting false outlier $p > 0.3$.
Bold numbers are significant values at $p < 0.05$.

Distribution of patients in the ITT group appeared biased. As a matter of fact, the authors did not retrieve any positive effect of *Arnica* 1000 K on blood and serum amount in drainages in the intention-to-treat (ITT) population. They reported differences only in the per-protocol (PP) population and exclusively if considering the model (number 3) including type of treatment, the amount of blood and/or seroma collected on the day of surgery, and patient's weight.[1] Nevertheless, weight distribution in the placebo (a) and *Arnica* (b) group, respectively, in the ITT population, was rather different, as the delta values (most probably IC_{95}) for weight were 80 kg (a) and 28 kg (b). Distributions, which have comparable medians (a = 60.40; b = 62.90) and means (a = 64.19; b = 66.05), should suggest that, taking into account these differences, homoscedasticity was most probably different in the two groups, resulting in a possible bias in the evaluation of *Arnica* effect because of the existence of size effects and due to the distribution of patients weights and vascular pressure in the placebo respect to the *Arnica* group. Bartlett's test confirmed this inhomogeneous variance distribution between cases (*Arnica* treated) and controls (placebos) ($p < 0.0001$, $\chi^2 = 113.03173$). Difference in the variances within the ITT group is a possible confounder.[4]

This fundamental bias might be generated from the less powerful t-test in a relatively small sized sample. Actually, the authors used Student's t-test for their statistics.[1] While in a frequentist statistical perspective a t-test might be exceptionally adopted for samples <30 subjects (as the number used for either cases or controls in the reported study), though with caution for the possible occurrence Type II errors, the use of small sample sizes, from a Bayesian point of view, should result in inflated effect sizes and false positives.[5–8] A very simple reason shedding a light on this assessment may be addressed. An investigator, particularly if performing a double blinded experimental setting, cannot know whether an H_0 null hypothesis is true or false. The probability that a value is true or false could be estimated from the existing literature on the topic, but, at least for homeopathy and *Arnica* 1000 K, many technical and cultural prejudices lead to the utmost consideration that the dilution "cannot" work (for its chemical absurdity) and that whatsoever is considered as a "positive" signal, would be potentially considered as the confirmation of the excellent quality of the experimental setting, rather than a true effect, the reliability of which is exclusively due to the experimental setting and performance. Therefore the statistical power of the test is particularly stringent in giving the best positive predictive value. If we suppose a 50% of the null hypothesis as true and 50% as false, with D = 1, if the sample size is small, the statistical power is low. With the lowest value of 2 (sample size), the probability for a true positive is 4.7% and for a false positive is 2.5%, i.e. the probability that this statistics results in giving a significant finding as a true effect is only 65%, while for a sample size of 100, the positive predictive value is 95%, so suggesting that for a sample <30, a statistically significant result is clearly more likely to be a false positive.[9–11]

Therefore, in this circumstance, the positive value in the PP population in the model 3 of treatment with *Arnica* 1000 K, might be biased and in this sense should need reproducibility. The authors concluded that *Arnica* 1000 K gave noticeable differences, though small, and attributed this failure to the extremely high dilution of the *Arnica* extract in the *Arnica* korsakovian preparation. This conclusion might apply to any *Arnica* dilution exceeding the Avogadro-Loschmidt threshold, yet.[12,13]

In conclusion, the paper by Sorrentino et al., should need some reappraisal, in order to assess the experimental reproducibility by checking and revising possible effects from confounders and bias.

Conflict of interest

The authors state they have no conflicts of interest.

References

1. Sorrentino L, Piraneo S, Riggio E, et al. Is there a role for homeopathy in breast cancer surgery? A first randomized clinical trial on treatment with Arnica Montana to reduce post-operative seroma and bleeding in patients undergoing total mastectomy. *J Intercult Ethnopharmacol*. 2017 Jan 3;6(1):1–8.
2. Marzotto M, Bonafini C, Olioso D, et al. Arnica Montana stimulates extracellular matrix gene expression in a macrophage cell line differentiated to wound-healing phenotype. *PLoS One*. 2016 Nov 10;11(11):e0166340.

3. Chirumbolo S. Oscillococcinum®: misunderstanding or biased interest? *Eur J Intern Med*. 2014 Mar;25(3):e35–e36.
4. Veith FJ. How can good randomized controlled trials in leading journals be so misinterpreted? *J Vasc Surg*. 2013 Feb;57(2 suppl):3S–7S.
5. Ingre M. Why small low-powered studies are worse than large high-powered studies and how to protect against "trivial" findings in research: comment on Friston (2012). *Neuroimage*. 2013 Nov 1;81:496–498.
6. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005 Aug;2(8):e124.
7. Ioannidis JP. Effect of formal statistical significance on the credibility of observational associations. *Am J Epidemiol*. 2008 Aug 15;168(4):374–383. discussion 384–90.
8. Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology*. 2008 Sep;19(5):640–648.
9. Shrier I. Power, reliability, and heterogeneous results. *PLoS Med*. 2005 Nov;2(11):e386. author reply e398.
10. Goodman S, Greenland S. Why most published research findings are false: problems in the analysis. *PLoS Med*. 2007 Apr;4(4):e168.
11. Ioannidis JP. Why most published research findings are false: author's reply to Goodman and Greenland. *PLoS Med*. 2007 Jun;4(6):e215.
12. Olioso D, Marzotto M, Bonafini C, Brizzi M, Bellavite P. Arnica Montana effects on gene expression in a human macrophage cell line. Evaluation by quantitative real-time PCR. *Homeopathy*. 2016 May;105(2):131–147.
13. Chirumbolo S, Bjørklund G. Commentary: Arnica Montana effects on gene expression in a human macrophage cell line: evaluation by quantitative real-time PCR. *Front Immunol*. 2016 Sep 8;7:280.