

Big-data Clinical Trial (BCT): the third talk

Stephen D. Wang¹, Yaxing Shen^{1,2}

¹AME Publishing Company, HK, China; ²Department of Thoracic Surgery, Zhongshan Hospital, Fudan University, Shanghai 200032, China

Correspondence to: Stephen D. Wang. Founder & CEO, AME Publishing Company & Pioneer Bioscience Publishing Company, FLAT A, 9/F Gold Shine Tower, 346-348 Queen's Road Central, Sheung Wan, Hong Kong, China. Email: swang@amegroups.com; Yaxing Shen, M.D. Department of Thoracic Surgery, Zhongshan Hospital, Fudan University, Shanghai 200032, China. Email: shen.yaxing@zs-hospital.sh.cn; shenyx@amegroups.com.

Submitted Jul 01, 2015. Accepted for publication Jul 30, 2015

doi: 10.3978/j.issn.2072-1439.2015.08.07

View this article at: <http://dx.doi.org/10.3978/j.issn.2072-1439.2015.08.07>

In the previous talk, the randomized control trial (RCT) showed less possibility to keep its major role in the future clinical studies (1). Meanwhile, the concept of “Big-data Clinical Trial (BCT)” was discussed and redefined in the following articles (2,3). In this article, we will continue the topic on BCT from the perspective of “data structure”. We hope BCT will have an adequate arena to take over RCT and win the “relay race” in the upcoming new era.

In “Big Data”, we have 4Vs: *Volume*, *Velocity*, *Variety*, *Veracity*, and there could be three requirements for the clinical data in BCT. Firstly, the *Volume* of data should be “big” enough, even it is mined from a single case: the data would cover the whole process of health care including the diagnosis (e.g., genome-wide findings instead of selected markers), treatment (e.g., real-time surgical videos instead of intra-operative still pictures), and results (e.g., life-long instead of 3 or 5 year follow-ups); secondly, the *Velocity* of the data refreshment would be “high” enough to realize the real-world evidence studies; thirdly, the data would have *Variety* in structure, and new parameters would be continuously introduced to the study, resulting in a “ladder-shaped” data structure (Figure 1). The expanding parameters of the study stand for the increasing human knowledge on diseases; even the information is temporally compromised in *Veracity*.

Currently, most clinical research articles are based on the statistical analyses on the “rectangle-shaped” data. In Figure 1, for instance, the first column represents the numbered individuals (1, 2, 3...), and the rows shows various parameters including sex and age (the same parameters apply each individual, as shown in rectangles 1, 2, and 3 in

No.	C1	C2	C3	...	TM1	TM2	TM3	TM4	TM5	...
1	XX	XX	XX							
2	XX	XX	XX							
3	XX	XX	XX		XX	XX	XX			
4	XX	XX	XX		XX	XX	XX			
5	XX	XX	XX		XX	XX	XX			
6	XX	XX	XX		XX	XX	XX	XX	XX	XX
7	XX	XX	XX		XX	XX	XX	XX	XX	XX
8	XX	XX	XX		XX	XX	XX	XX	XX	XX
9	XX	XX	XX		XX	XX	XX	XX	XX	XX
...										

Figure 1 The expansion of parameters and the “Ladder-shape” data structure.

Figure 1). Suppose we carry out statistical analysis on data in rectangle 3 (Figure 2), data in areas A and B will not be included. The question is: Should data in area A and/or area B be included for analyses?

As was suggested in a recent TED talk by Kenneth Cukier: three additional parameters were discovered in determine the malignancy from breast specimens by data scientist, but not the pathologist! In BCT, the continuous expansion of parameters would probably summon the beast called “*Machine Learning*”, and the future of medicine would welcome the collaboration between physician and data scientist.

All in all, there will be long journey before the human beings realized the “Precision Medicine” from BCT. Long as it is, soon it will be.

No.	C1	C2	C3	...	TM1	TM2	TM3	TM4	TM5	...
1	XX	XX	XX							
2	XX	XX	XX							
3	XX	XX	XX		XX	XX	XX			
4	XX	XX	XX		XX	XX	XX			
5	XX	XX	XX		XX	XX	XX			
6	XX	XX	XX		XX	XX	XX	XX	XX	XX
7	XX	XX	XX		XX	XX	XX	XX	XX	XX
8	XX	XX	XX		XX	XX	XX	XX	XX	XX
9	XX	XX	XX		XX	XX	XX	XX	XX	XX

Figure 2 The statistical analyses of “rectangle-shaped” data.

Acknowledgements

Supported by the American Association for Thoracic

Surgery’s Everts A. Graham Memorial Traveling Fellowship; and the National Natural Science Foundation of China (Grant No. 81400681).

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Wang DY. RCT and “relay race”. Available online: <http://kysj.amegroups.com/articles/1144>
2. Wang SD. Opportunities and challenges of clinical research in the big-data era: from RCT to BCT. *J Thorac Dis* 2013;5:721-3.
3. Wang SD, Shen Y. Redefining big-data clinical trial (BCT). *Ann Transl Med* 2014;2:96.

Cite this article as: Wang SD, Shen Y. Big-data Clinical Trial (BCT): the third talk. *J Thorac Dis* 2015;7(8):E243-E244. doi: 10.3978/j.issn.2072-1439.2015.08.07