Construct Validation in Secondary Data: A Guideline for Medical Data Mining

Mirza Rizwan Sajid

Faculty of Industrial Sciences & Technology, Universiti Malaysia Pahang, 26300 Gambang, Malaysia & Department of Statistics, University of Gujrat, Pakistan

Email: mirzarizwansajid@gmail.com

Dr. Noryanti Muhammad

Faculty of Industrial Sciences & Technology, Universiti Malaysia Pahang, 26300 Gambang, Malaysia

Email: noryanti@ump.edu.my

Dr. Roslinazairimah Zakaria

Faculty of Industrial Sciences & Technology, Universiti Malaysia Pahang, 26300 Gambang, Malaysia

Email: roslinazairimah@ump.edu.my

Abstract:

Construct validation is an important step in the formulation of theory and its testing. Any information infers from the construct and its related hypothesis, without the validation, might be misleading the decision. In the literature, the construct validation guidelines for primary data are available but the exponential usage of secondary data mining approaches in recent times has created its need for secondary data users. However, it is a less addressed domain, specifically for medical sciences due to underdeveloped theoretical foundations of the field. Usually, an assessment of the validity of the secondary data is not evaluated because primary users already underwent the process especially in the tool development phase. Further, secondary data users are more concerned with data mining that is a data-driven approach rather than based on theoretical knowledge of the field. Therefore, if researchers want to explore the hidden structures of data and try some different combinations of items which were not tested by primary users, then they should validate this newly explored group of items for the creation of knowledge of the particular field. In this paper, a guideline of the construct validation process for secondary data users with its practical issues especially validity and reliability coefficients is discussed. This paper concludes that the construct validation based on secondary data needs special attention of data mining experts and medical researchers while designing the studies to maximize the benefits of the secondary data.

ACKNOWLEDGEMENT