UNIVERSITY OF TURKU
Department of Computing, Faculty of Technology

Jouni Kettunen: Regression Discontinuity Design for AI-assisted histopathology data

Master of Science in Technology Thesis, 83 p., 16 app. p.
Digital Health
Master's Degree Programme in Information and Communication Technology
March 2023

The aim of the thesis was to review use of quasi-experimental regression discontinuity design method in histopathological data-analysis and to assess the technical possibility of using the method with algorithmic data. Regression discontinuity design is a rarity in medical research and little used in histopathological data-analysis. The data produced in histopathological analysis relies on cut-offs when reviewing eligibility to diagnose/treatment and fulfils the main assumption of continuity. Also, data around the divisive cut-off can be proven to be randomly divided. Proof-of-concept for regression discontinuity design was done using digitalized breast tissue slides (n = 129) analyzed in Aiforia Cloud. The observations were divided by using a proliferation protein Ki-67's positivity ratio as a cut-off; $< 14\%$ were in moderate group (pathologist n = 37, algorithm n = 47) and $\geq 14\%$ positives (pathologist n = 92, algorithm n = 82). Outcome was chosen to be severity of the decease (CFR%), derived from five-year survival. Thesis used a pre-made API to fetch algorithmic results to R, where the data handling and analysis were done. For realistic study one needs thousands of observations, thus used dataset with n = 129 subjects was too small, also there were only nine observations in outcome. For the reasons, thesis can't give any clinically relevant judgments. Technically it was possible to make a pipeline from Aiforia Cloud to R and proceed to regression discontinuity design. In the future one might consider repeating this study with enough observations and more robust outcome.

Keywords: regression discontinuity design, quasi-experiment, Ki-67, causal effect, breast cancer