Causal Discovery on Health-related Quality of Life of cancer patients

Maria Ganopoulou[1], Efstratios Kontopoulos[2], Konstantinos Fokianos[3], Lefteris Angelis[4], Ioannis Kotsianidis[1], and Theodoros Moysiadis[1,*]

[1] *Department of Hematology, University Hospital of Alexandroupolis, Democritus University of Thrace Medical School, 68100 Alexandroupolis, Greece*
[2] *Foodpairing NV, Oktrooiplein 1, Box 401, 9000 Gent, Belgium*
[3] *Department of Mathematics & Statistics, University of Cyprus, 1678 Nicosia, Cyprus*
[4] *School of Informatics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece*

Abstract

The management of cancer patients increasingly includes Health-related Quality of Life (HRQoL) as a crucial endpoint. Physical, psychological, lifestyle, and social aspects expressed via responses to HRQoL questionnaires offer valuable insights for patient care. However, a still unexplored field is the identification and understanding of causal relationships among the questions involved. This study evaluates the capability of detecting cause-effect relationships in this context, applying causal structure-learning algorithms to simulated data. Different data configurations are examined, encompassing the number of hypothetical questions in an HRQoL questionnaire, the quantity of cause-effect relationships, and the number of participants involved. Exploring this issue holds potential significance in shaping the design and/or selection of HRQoL questionnaires, accounting for limitations in sample size and intuition regarding the underlying causal structure. Uncovering cause-effect relationships can contribute to enhanced management and improved HRQoL for cancer patients.

Keywords

Causal discovery, directed acyclic graphs, HRQoL questionnaire

1. Introduction

Health-related quality of life (HRQoL) is emerging as an important aspect in managing cancer patients. Physical, psychological, lifestyle and social parameters may provide guidance on how to manage the individual patient. Several HRQoL questionnaires are available (e.g. [1], [2]) that assess the patients' quality of life. An issue that has not yet been investigated is to detect and assess the causal relationships among the corresponding questions. Unveiling the underlying cause-effect relationships and building on this knowledge may aid to improve the management, and the HRQoL of cancer patients. For example, assuming that worrying for health issues in the future is the cause, and feeling depressed is the effect, alleviating these worries (e.g., by informing/educating patients) could result in helping the patients to feel less depressed.

This study assessed the ability to detect cause-effect relationships within this context, by employing two causal structure-learning algorithms, based on simulated data. To this end, different data setups were considered, involving the number of hypothetical questions within an HRQoL questionnaire, the number of cause-effect relationships, and the number of participants. This study may be beneficial when considering causal discovery in HRQoL questionnaire related research, by providing valuable insights regarding the selection and/or the design of an HRQoL questionnaire related to sample size limitations, questionnaire size and the importance of scientific intuition related to the underlying causal structure.

SWAT4HCLS 2024: The 15th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences, February 26–29, 2024, Leiden, The Netherlands

tmoysiad@affil.duth.gr (T. Moysiadis); *Corresponding author
0000-0001-6892-1242 (M. Ganopoulou); 0000-0002-4435-2521 (E. Kontopoulos); 0000-0002-0051-711X (K. Fokianos); 0000-0002-6677-4039 (L. Angelis); 0000-0001-7668-0186 (I. Kotsianidis); 0000-0003-1673-196X (T. Moysiadis)

2. Method

Questions within HRQoL questionnaires typically contain 3 to 5 answers that represent increasing health burden related to the question at hand. Six different directed acyclic graphs (DAGs) were considered, based on which random data were generated. The DAGs differed in the number of questions/nodes involved (5 to 19), and the number of direct relationships/edges among them (5 to 17). For each DAG, the parameters were custom fitted and 1000 samples were generated for different numbers n of simulated participants. For each sample, two different constraint-based structure learning algorithms were used to estimate the completed partially DAG, namely PC-Stable [3], a modern implementation of the standard PC algorithm, and Interleaved Incremental Association [4]. Two metrics were used to evaluate the algorithms, the structural Hamming distance (SHD) between the true and the estimated DAG, and a relative SHD, defined as the SHD divided by the number of the true DAG edges. The mean value of these metrics was computed across the 1000 samples, for each DAG and each n.

3. Results and Discussion

The algorithms exhibited similar performance and were both inefficient for $n<200$. On the other hand, larger values of n (≥ 500), and, even better $n \geq 1{,}000$, resulted in a satisfactory performance in both cases. Moreover, as expected, the algorithms performed better for simpler DAGs.

We are now in the process of deploying semantic technologies for representing the responses to the HRQoL questionnaires (e.g., based on the works presented in ([5], [6]), but, more importantly, the detected cause-effect relationships among the questions. Having the results in the form of an RDF semantic knowledge graph will facilitate interoperability with potential third-party stakeholders towards developing extensions of mutual interest across research studies and healthcare institutions. Moreover, capitalizing on those semantic artefacts will contribute to knowledge discovery from HRQoL data within the context of developing decision support systems that leverage semantic technologies in the context of cancer patient management.

Acknowledgements

The research project was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "2nd Call for H.F.R.I. Research Projects to support Post-Doctoral Researchers" (Project Number: 553).

References

[1] N. Aaronson, et al., The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology, Journal of the National Cancer Institute (1993) vol. 85, no. 5, 365-376.

[2] D. Cella, et al., The Functional Assessment of Cancer Therapy scale: development and validation of the general measure, J Clin Oncol (1993) vol. 11, no. 3, 570-579.

[3] D. Colombo and M. Maathuis, Order-independent constraint-based causal structure learning, J. Mach. Learn. Res. (2014) vol. 15, no. 1, 3741-3782.

[4] S. Yaramakala and D. Margaritis, Speculative Markov blanket discovery for optimal feature selection, in: Fifth IEEE International Conference on Data Mining (ICDM'05), 2005.

[5] D. Cella and R. Hays, A Patient Reported Outcome Ontology: Conceptual Issues and Challenges Addressed by the Patient-Reported Outcomes Measurement Information System®(PROMIS®), Patient Related Outcome Measures (2022) 189-197.

[6] P. Fabry, et al., QUESTO–An ontology for questionnaires, in: 11th International Conference on Biomedical Ontologies (ICBO 2020), 2020.