Causal Modeling Between Factors on Quality of Life in Cancer Patients Using S3C-Latent Algorithm

Yohani Setiya Rafika Nur 1), Ridho Rahmadi 2), Christantie Effendy 3)*

1)Master Program in Informatics, Department of Informatics, Universitas Islam Indonesia, Indonesia
Jl. Kaliurang, Kec. Ngemplak, Kabupaten Sleman, Yogyakarta
1)18917225@students.uii.ac.id, 2)ridho.rahmadi@uii.ac.id

2)Department of Medical Surgical Nursing, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Indonesia
Jl. Farnako, Kec. Depok, Kabupaten Sleman, Yogyakarta
3)christantie@ugm.ac.id

Abstract

**Background:** Cancer patients can experience both physical and non-physical problems such as psychosocial, spiritual, and emotional problems, which impact the quality of life. Previous studies on quality of life mostly have employed multivariate analyses. To our knowledge, no studies have focused yet on the underlying causal relationship between factors representing the quality of life of cancer patients, which is very important when attempting to improve the quality of life.

**Objective:** The study aims to model the causal relationships between the factors that represent cancer and quality of life.

**Methods:** This study uses the S3C-Latent method to estimate the causal model relationships between the factors. The S3C-Latent method combines Structural Equation Model (SEM), a multi objective optimization method, and the stability selection approach, to estimate a stable and parsimonious causal model.

**Results:** There are nine causal relations that have been found, i.e., from physical to global health with a reliability score of 0.73, to performance status with a reliability score of 1, from emotional to global health with a reliability score of 0.71, to performance status with a reliability score of 0.82, from nausea, loss of appetite, dyspnea, insomnia, loss of appetite and from constipation to performance status with reliability scores of 0.76; 1; 0.61; 0.76; 0.72; 0.70, respectively. Moreover, this study found that 15 associations (strong relation where the causal direction cannot be determined from the data alone) between factors with reliability scores range from 0.65 to 1.

**Conclusion:** The estimated model is consistent with the results shown in previous studies. The model is expected to provide evidence-based recommendation for health care providers in designing strategies to increase cancer patients’ life quality. For future research, we suggest studies to include more variables in the model to capture a broader view to the problem.

**Keywords:** Cancer, Causal Modeling, Quality of Life, S3C-Latent

**Article history:** Received 28 February 2021, first decision 15 March 2021, accepted 1 April 2021, available online 28 April 2021

I. INTRODUCTION

Cancer is a serious problem around the world [1]. Cancer occurs due to abnormal cell growth in the tissues [2]. The prevalence of death caused by cancer in the province of Special Region (DI) Yogyakarta (DIY henceforth) based on doctor's diagnoses was 4.9% [3]. Cancer patients can experience both physical and non-physical problems like psychosocial, spiritual and emotional problems that impact a person’s quality of life [1]. Quality of life is an individual’s assessment of the physical, mental, social, emotional conditions, health conditions and risks posed by cancer therapies. Cancer patients’ health status can be maintained by measuring their quality of life before evaluating the disease effects, treatment required and what effects have a significant impact on the decrease in quality of life [5][6]. One of the instruments to determine the quality of life in cancer patients is the Organization for Research and Treatment of Cancer Quality of Life Questionnaire-Core 15-Palliative Care (EORTC QLQ C15-PAL instrument) [5]. The EORTC QLQ C15-PAL has 10 factors consisting of two factors on functions (physical and emotional), seven factors on symptoms, and one factor on global health status/QoL.

* Corresponding author
Previous studies have been conducted to determine the correlations between quality of life factors in patients with breast and thyroid cancer [7–9], and in breast cancer patients who had undergone a chemotherapy [9]. However, those previous studies have not answered the fundamental question about the causal relationships [10]. Understanding such causal relationships will help us better understand the problem and thus design better interventions such as therapies or medications which could increase the life quality.

The purpose of this study is to estimate causal relations among variables influencing cancer patients’ quality of life. In this study we will use Stable Specification Search for Cross-Sectional Data with Latent Variable (S3C-Latent) to estimate the causal model. S3C-Latent is designed to estimate causal structures among factors, focusing on the model stability and complexity [11]. The causal model will be widely published on a website called Shiny. This is a free web platform from the R package that aims to build interactive web applications easily using the R programming language [12]. The website will display the causal relationship between factors on quality of life, and its also expected to serve as an evidence-based recommendation for healthcare providers in making decisions and designing treatments to resolve quality of life problems.

II. LITERATURE REVIEW

Cancer patients may experience different negative consequences, including physical and psychological conditions that can impact quality of life. Research has discussed specific correlation between factors that influence quality of life (QoL). Angraini et al. [6] have discussed the correlation between QoL factors among breast cancer patients. The data were collected from a survey with 34 patients. The results showed that the symptoms complained about were fatigue (38.9%) and nausea (80.8%). Respondents’ demography (age, occupation, level of education, body mass index, and cancer stage) were not related to QoL. Overall, the QoL levels of these patients were good.

Other studies involving breast cancer patients conducted by Juwita et al. [8] aim to determine the impact of demographic characteristics (age, level of education), treatments (duration of diagnosis, type of chemotherapy) and clinical conditions of the breast cancer patients. The study's results showed that demographic factors did not have a close correlation, but clinical conditions did.

A study on thyroid cancer patients’ QoL Aryanata et al. [7] examined the QoL of 100 thyroid cancer patients after a total thyroidectomy. The study considered respondents’ demographic characteristics and the affecting factors. The results showed that the quality of life of the participants after total thyroidectomy is lower than the general population. Age negatively affects physical functions, cognitive functions and global health; and positively affects dyspnea, loss of appetite and constipation. Gender affects cognitive functions and global health. The level of education and family income also have a strong influence on QoL, whereas time interval since the diagnosis has a marginal effect.

Wulandari et al. [9] have examined breast cancer patients to find out the QoL after undergoing a chemotherapy treatment. There were 53 female respondents aged between 20-70 in the study. The results showed that patients with breast cancer who had undergone chemotherapy had a good QoL. Cognitive factors and other people’s perceptions of the patient’s condition result in good QoL.

Considering all of the review studies discussed above, it could be concluded that there are correlations between factors that influence QoL. However, a study focusing on the causal relationships of those factors is limited and has not focused on understanding what decreases cancer patients’ QoL.

III. METHODS

A. Data set

The data set used in this study from the FK-KMK Universitas Gadjah Mada. The data were collected between 2017 and 2018. The respondents were all cancer patients at the Dr. Sardjito Hospital Yogyakarta and Prof. Dr. Margono Soekarjo Hospital, Purwokerto, with a total number of respondents of 214. In this study using 8 factors from EORTC C15-PAL consisting of 11 items, i.e., 1) physical (3 items); 2) emotional (2 items); 3) nausea and vomiting (1 item); 4) dyspnea (1 item); 5) insomnia (1 item); 6) loss of appetite (1 item); 7) constipation (1 item); 8) global health status/QoL (1 item); and 10) performance status of respondent characteristics.

B. Research Stages

The present study is conducted with following stages, i.e., literature study, data pre-processing, causal modeling, evaluation, and dissemination (see Fig. 1).
The research stage begins with a literature study to review the findings of previous research [7–10] related to cancer patients’ quality of life. The second stage is data pre-processing, which aims to check the data set if there are missing values, as well as to check the data distribution. At this stage, the problem of missing values can be done by cleaning using code `NewData <- Data[complete.cases(Data),].`

The third step is to apply S3C-Latent to the prepared data set. The model computations are performed parallel on a computer server with an R package named StablespecImptLatent. The computer server has 80 Cores, 250GB RAM, 4 GPUs, with a Jupyter GUI. The output of this stage is a causal model between factors of QoL.

The fourth stage is to evaluate the model by asking opinion from experts in the field, via a designated questionnaire. The final stage is dissemination, where we develop a Shiny website to better visualize and widely publish the estimated causal model. A Shiny web is a web platform developed through R language, which enables direct compilation of R scripts.

### C. Stable Specification Search for Cross-sectional Data

S3C-Latent originates from the Stable Specification Search for Cross-sectional Data (S3C) which focuses on the stable and simple causal estimation [13]. It is a causal method which consists of two phases, namely search and visualization (see Fig. 2). The pseudocode of S3C method (see Algorithm 1). The search phase consists of an inner loop which optimizes model estimation and an outer loop which repeat the inner loop over different data subsets to obtain stable estimation.

Algorithm 1 shows the pseudocode of S3C method. In Line 1, `D` is described as a data set and `C` as initial knowledge. Lines 3-18 represent the process of the outer loop. Line 4 sets `T` as the subset of data from `D` with size `|D|/2`. In Lines 6-16, the inner loop process starts by running variable `I` to get the `pareto` front, then Lines 7-12 create a population `P` that takes the previous population in size `N` randomly using crowding distance sorting. The causal model `X ← Y` is represented by a binary vector `{0,1}`. The population is set on variable `P` to `Q` by using a binary tournament selection, one-point crossover and mutation with binary representation. The crossover with one-point takes two models from the `Mpool` and exchanges the data in the middle of the crossover point.

In Line 14, the variables `P` and `Q` are combined and the result (`P` and `Q`) is set in variable `F` by using fast non-dominated sort. Variable `F` is set with updates `pareto` in line 15. Line 17 is the result of the inner loop iteration and stored in the variable `H`. After iterating on `J` as the value of the inner loop iteration, `pareto` front (`j`) is set in variable `H`. Lines 19 to 22 change `DAG` (variable `H`) to `CPDAG` and is stored in variable `G` by using `consDAG2Cpdag`, and is then calculated to obtain the stability graphs of edge and causal path.
D. Stable Specification Search for Cross-Sectional Data with Latent Variable

S3C-Latent is a development method of Stable Specification Search (S3C). S3C-Latent specifically models the causal relationship between the observed variables and uses the Structural Equation Model (SEM) on the latent variables. Latent variable is referred to as a factor, and the variable observed is referred to as an indicator. SEM with latent variables consists of structural models and measurement models [11]. The equation is as follows:

\[ \eta = B\eta + \Gamma\xi + \zeta \]  

When \( \eta \) is an vector of size \( m \times 1 \) on the variables effect (endogenous), \( \xi \) an vector of size \( n \times 1 \) on the variables cause (exogenous), \( \zeta \) is an vector of size \( m \times 1 \) interference on \( \eta \), and \( \Gamma \) is matrix coefficients of \( m \times n \) between \( \xi \) [11]. The measurement model using equations (2) and (3) as follows:

\[ x = \Lambda_x\xi + \delta \]  
\[ y = \Lambda_y\eta + \epsilon \]  

The \( r \times n \) matrix \( \Lambda_x \) and the \( q \times m \) matrix \( \Lambda_y \) are structural coefficients associated with latent and indicator variables, the \( r \times 1 \) of vector \( \delta \), and the \( q \times 1 \) of vector \( \epsilon \) are errors on indicators. The \( r \times r \) matrix \( \Theta_\delta \) and the \( q \times q \) matrix \( \Theta_\epsilon \) are the covariance matrix of \( \delta \) and \( \epsilon \). Equation (4) represents the breakdown of the model-implied covariance matrix.

\[ \Sigma(\theta) = \begin{bmatrix} \Sigma_{yy}(\theta) & \Sigma_{yx}(\theta) \\ \Sigma_{xy}(\theta) & \Sigma_{xx}(\theta) \end{bmatrix} \]

\[ \Sigma_{yy}(\theta) = \Lambda_y (I - B)^{-1} (\Gamma\Phi\Gamma' + \Psi) [(I - B)^{-1}]' \Lambda_y' + \Theta_\epsilon, \]  
\[ \Sigma_{xy}(\theta) = \Lambda_x \Phi \Gamma' [(I - B)^{-1}]' \Lambda_y' \]  
\[ \Sigma_{xx}(\theta) = \Lambda_x \Phi \Lambda_x + \Theta_\delta \]  

Furthermore, the SEM model with latent must be identified before evaluated. The identification attempts to determine if the parameters of the model can be estimated. The identification conditions are detailed in [11]. After the identification conditions fulfilled, parameter \( \theta \) can be estimated by minimizing the maximum likelihood function that reads

\[ \hat{\theta} = \arg\min_{\theta} F_{M1}(\theta), \]  
\[ F_{M1}(\theta) = \log|\Sigma(\theta)| + \text{Tr} [S\Sigma^{-1}(\theta)] - \log|S| - p \]
Algorithm 2 shows the pseudocode of the S3C-Latent. Line 1 describes variable \( D \) as a dataset, \( C \) as initial knowledge, \( \Lambda \) as a matrix containing factors. Lines 2 to 13 are to ensure the model identification conditions are fulfilled. Line 14 runs S3C-Latent on the data set \( D \) with information of latent variables from \( L \), satisfying any constraint in \( C \), and fulfilling the model identification conditions in \( I \). More detail about the pseudocode can be found in [11].

### IV. RESULTS

The demographic characteristics of the study subjects were described based on the gender, age, type of cancer and performance status with the 214 respondents. TABLE 1 shows that the majority of respondents in this study were female (72.5 %), mostly aged 51-60 years (39.7 %), with breast cancer (30.4 %), and performance status level 1 (42.5 %).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Indicator</th>
<th>Frequency (N)</th>
<th>Percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genders</td>
<td>Male</td>
<td>162</td>
<td>75.7</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>52</td>
<td>24.3</td>
</tr>
<tr>
<td>Ages</td>
<td>&lt; 30</td>
<td>8</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>30-40</td>
<td>32</td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td>41-50</td>
<td>61</td>
<td>28.5</td>
</tr>
<tr>
<td></td>
<td>51-60</td>
<td>85</td>
<td>39.7</td>
</tr>
<tr>
<td></td>
<td>&gt;60</td>
<td>28</td>
<td>13.1</td>
</tr>
<tr>
<td>Type of Cancers</td>
<td>Breast/Mammae</td>
<td>65</td>
<td>30.4</td>
</tr>
<tr>
<td></td>
<td>Cervix</td>
<td>26</td>
<td>12.1</td>
</tr>
<tr>
<td></td>
<td>Ovarium</td>
<td>18</td>
<td>8.4</td>
</tr>
<tr>
<td></td>
<td>Colli/Tiroid</td>
<td>4</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>74</td>
<td>34.6</td>
</tr>
<tr>
<td></td>
<td>Missing value</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>Performance status/ECOG</td>
<td>0</td>
<td>31</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>91</td>
<td>42.5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>34</td>
<td>15.9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>30</td>
<td>14.0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>20</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td>Missing value</td>
<td>8</td>
<td>3.7</td>
</tr>
</tbody>
</table>

Technically, before the computation is performed, we check the missing value data set using the code `NewData <- Data[complete.cases(Data),]`, and set the computation parameters, which includes data subset number (S), iteration number (I) and model population (P), crossover rate (C) and mutation rate (M) [11]. The parameters are set: \( S = 200 \), \( I = 50 \), \( P = 170 \), \( C = 0.45 \), and \( M = 0.01 \). We also added constraints or prior knowledge that global health and
performance status do not cause anything. Fig. 3 shows (a) causal path stability and (b) edge stability with \( \pi_{sel} = 0.6 \) and \( \pi_{bic} = 25 \). Details on how to read these graphs can be found in our study [13].

Fig. 3 shows the aggregated causal path stability and the edge stability graphs. Fig. 4 shows the stability graph for each pair of variables. In Fig. 4, the blue line (---) indicates the edge stability; the green line (----) indicates the causal path stability of length 1 and the red line (-----) indicates the causal path stability of any length [14]. Based on Fig. 3 and Fig. 4, there are nine causal relations which are considered stable and parsimonious. That is, from physical, emotional, nausea, dyspnea, loss of appetite, constipation, and insomnia to performance status, and from physical and emotional to global health status. There are 15 associations between some variables, i.e., between physical and emotional conditions, physical and insomnia, physical and constipation, emotional and insomnia, loss of appetite and emotional, nausea-vomiting and emotional, loss of appetite and nausea, nausea-vomiting and dyspnea, dyspnea and physical, dyspnea and emotional, insomnia and constipation, loss of appetite and insomnia, loss of appetite and constipation, loss of appetite and physical, constipation and dyspnea.
Fig. 4 The stability graphs with the descriptions of variable as follows: var1 physical, var2 emotional, var3 nausea-vomiting, var4 dyspnea, var5 insomnia, var6 loss of appetite, var7 constipation, var8 global health status/QoL, var9 performance status

Based on the stability graphs on Fig. 3 and Fig. 4, Fig. 5 visualizes the relevant edge and causal paths. Each relation has a reliability score, which shows the likelihood/probability of the selection, i.e., the higher the score, the greater
the confidence on these corresponding relations. It is obtained from the highest selection probability of the edge stability in the relevant area [11]. The reliability scores range from 0 to 1.

![Diagram](image.png)

Fig. 5. Visualization of causal models of quality of life using the S3C-Latent method

![Shiny website](image.png)

Fig. 6. Shiny website displaying the quality of life
The arrows indicate causal relations and dashed lines indicate associations. We then developed a Shiny website to better disseminate the estimated model, including its prior computation process and some useful information. The look of the Shiny webs is shown in Fig. 6 and can be accessed at https://kualitaslidup.shinyapps.io/Website/. As part of the evaluation, we sent questionnaire to correspondents with background knowledge relevant to the fields. The questionnaire is designated for asking their opinion. The result of the questionnaire is included in the Shiny web.

V. DISCUSSION

The results indicate that both physical and emotional cause global health status. This finding is in line with the study conducted by Pradana et al. [15] and in other studies, it was found that mental health disorders, such as depression, and the physical disability can impact global health [16][17].

Moreover, physical, emotional, nausea, dyspnea, insomnia, loss of appetite, and constipation were found to influence performance status. These causal relations are corroborated with the studies conducted by Laird et al. [18] and Madulara [19] which show that worse physical condition can affect daily activities, reduced ambulation problems, and that loss of appetite also worsen the performance status.

In addition, there are associations between pairs of variables (relations among variables which causal direction cannot be determined from the data alone), i.e., between physical and emotional, physical and insomnia, physical and constipation, insomnia and constipation, loss of appetite and constipation, and loss of appetite and physical. These associations are consistent with the study by Alatas et al. [20]. The study found that patients with physical activity limitations can trigger the symptoms of emotional and constipation. Studies by Asyf et al. [21] and Suartimingsih et al. [22] found that the symptoms of insomnia include thinking disorders, increased blood pressure, and unhealthy lifestyles can also influence constipation. Moreover, loss of appetite and lack of activity can lead to constipation [23][24].

The association between emotional conditions and insomnia is consistent with the study by Susanti [25] which found that patients with poor sleep patterns have a relationship with the occurrence of distress and experience an increase in thinking load. The association between insomnia and loss of appetite is corroborated in the studies conducted by Julian and Kurniawan [26] and Vicario [27], which explained that the need for sleep and food are interconnected, and that a poor diet will lead to insomnia. The associations between loss of appetite and emotion, and between nausea and loss of appetite are in line with the observations of Ambarwati [23] and Rumastika and Surarso [28]. The studies explained that loss of appetite could be triggered by depression, which impacts the sufferer's emotion. Furthermore, chemotherapy-induced nausea and vomiting can also cause a loss of appetite.

The association between dyspnea and nausea-vomiting is consistent with the studies of Juwita et al.[8] and Rustanti [29]. The studies explained that the symptoms of nausea-vomiting causes complications in the respiratory tract due to cytostatic, which can affect the role of neuroanatomy and abdominal muscles. Furthermore, the association between constipation and dyspnea was strengthened with a statement by Estri et al. [30], which stated that constipation could cause abdominal distension, which can block the diaphragm and improve breathing performance.

VI. CONCLUSIONS

In this study, we applied a causal method called S3C-Latent to a data set consisting of information regarding the QoL of 214 cancer patients. We aim to estimate the underlying causal mechanisms between factors indicating the QoL of cancer patients. The findings of this study, that is, both the estimated causal relations and associations are in general supported by those of relevant previous and by opinion of experts in the field. The obtained model is projected to be a scientific reference for those in the field, e.g., doctors, nurse, researchers, etc., who focus on the QoL of cancer patients. This study also aims to show an alternative modeling in the clinical domain, that is a causal model, which can be useful in better understanding the problem and later propose the solution. Future research will benefit from including other demographical information that could lead to a broader view of the problem. However, it is not in the scope of the present study; for future work, we suggest to include some other demographic information.

Author Contributions: Yohani Setiya Rajika Nur: Conceptualization, Methodology, Software, Formal analysis, Writing - Original Draft. Ridho Rahmadi: Methodology, Software, Project administration, Funding acquisition, Supervision, Writing - Review & Editing. Christantie Effendy: Validation, Investigation, Supervision, Writing - Review & Editing.

Funding: This research received no specific grant from any funding agency.

Conflicts of Interest: The authors declare no conflict of interest.
REFERENCES


[27] B. A. L. Vicario, “Penanganan Insomnia Menggunakan Terapi Akupunktur Titik Shenmen (Hi 7),” Sanyinjiao (Sp 6), Dan Zusani (St 36), Dan Herbal Bunga Chamomile (Matricaria Chamomillla L.) 2020.

