Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume xx, Number xx, month Year
DOI : https://doi.org/10.33395/sinkron.xxx.xxx
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Causal Analysis of Stunting Determinants Using the Peter-Clark and Greedy Equivalence Search Algorithms

Nurhaeka Tou[1)*], Putri Mentari Endraswari[2)], Syafiranur Itizam[3)]
[1)2)3)]Teknologi Informasi, Fakultas Sains dan Teknik, Universitas Bangka Belitung, Indonesia
[1)]nurhaeka@ubb.ac.id, [2]putrimentari@ubb.ac.id, [3)]syafiranur.sn@email.com

Submitted :Nov 02, 2025 | **Accepted** : Nov 22, 2025 | **Published** : Jan 03, 2026

Abstract: Child stunting remains a major public health challenge, reflecting the long-term effects of inadequate nutrition, limited maternal education, and restricted access to health services. However, most existing studies rely on correlational analysis, leaving the underlying causal mechanisms insufficiently understood. This gap limits the development of effective interventions, as policymakers require evidence on how determinants interact causally. To address this issue, this study applies two causal discovery algorithms Greedy Equivalence Search (GES) and Peter Clark (PC) to identify causal relationships among eight key determinants of stunting using secondary data from the West Bangka District Health Office (2024). The variables include anthropometric indicators, maternal characteristics, and environmental conditions. Causal assumptions such as causal sufficiency, acyclicity, and faithfulness were imposed to ensure identifiability of the resulting graphs. Model performance was evaluated using Directed Density (DD) and Causal Density (CD) metrics. GES generated a parsimonious causal structure highlighting maternal education, posyandu visits, and exclusive breastfeeding as dominant causal candidates affecting height-for-age (TB/U) and weight-for-age (BB/U). In contrast, the PC algorithm produced a more complete and dense structure, achieving DD = 1.0 and CD = 0.12, compared with GES (DD = 0.80; CD = 0.10). These results indicate that PC is more exploratory in mapping complex causal interactions, while GES offers a simpler and more conservative model. Collectively, the findings demonstrate that combining score-based and constraint-based discovery approaches yields complementary insights into the mechanisms driving stunting.

Keywords: Stunting; Causal Analysis; Peter-Clark Algorithm; Greedy Equivalence Search; Causal Inference

INTRODUCTION

Stunting is a chronic growth disorder in toddlers characterized by height that falls below the age-standard due to prolonged malnutrition (Efrizal, 2020). According to Indonesian Minister of Health Regulation No. 2 of 2020, stunting is defined based on the height-for-age (H/A) index with a score of less than -2 SD (Khairani et al., 2020). Its impact not only affects physical growth but also brain development, endurance, and productivity in adulthood (Khoerul Ummah, 2024). This condition poses a serious threat to the quality of human resources and national economic development (Herliansyah et al., 2021).

Indonesia still faces a high prevalence of stunting. According to UNICEF and the World Bank (2020), Indonesia ranks 115th out of 151 countries. Basic Health Research (Riskesdas) data show that the prevalence of stunting reached 37.2% in 2013 and decreased to 27.7% in 2019 (Widia Pebrianti et al., 2024). In the Bangka Belitung Islands Province, the prevalence dropped to 18.5% in 2022, but increased again to 20.6% in 2023. West Bangka Regency was recorded as the region with the highest stunting rate, reaching 33% in 2013 and still recording 822 cases by the end of 2024 (Julianti & Elni, 2021). This condition demonstrates the need for comprehensive treatment of the factors causing stunting.

Previous studies have widely examined determinants such as nutrition, parenting, maternal education, sanitation, and socioeconomic status. However, most studies rely on correlational, regression-based, or predictive models, which *cannot* capture directional or cause–effect relationships between variables. This creates a significant

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume xx, Number xx, month Year
DOI : https://doi.org/10.33395/sinkron.xxx.xxx
e-ISSN : 2541-2019
p-ISSN : 2541-044X

research gap because policy interventions require knowledge of *which* factors causally influence stunting, *how* they interact, and *which* variables act as root causes rather than merely correlating with outcomes (Rahmadi, 2019)**.**

Recent advances in causal discovery provide methodological alternatives for identifying causal structures directly from observational data (Rahmadi et al., 2017)Algorithms such as Peter Clark (PC) and Greedy Equivalence Search (GES) can infer causal graphs by examining conditional independencies, uncovering hidden causal pathways and enabling structural interpretation beyond prediction. Although these methods are widely applied in fields such as epidemiology, neuroscience, and social sciences (2022–2025), their use remains limited in stunting research in Indonesia (Rahmadi et al., 2018).

Most stunting studies in Indonesia still depend on correlation or prediction models and have not attempted to reconstruct a causal graph of stunting determinants using regional health datasets. No prior research has applied causal discovery techniques to identify the causal structure of stunting factors in West Bangka or the broader Bangka Belitung region.

This study addresses this gap by: (1) applying the Peter Clark (PC) Algorithm and Greedy Equivalence Search (GES) to West Bangka District Health Office data; (2) reconstructing causal graphs among determinants of stunting; (3) evaluating structural accuracy using Degree of Dependence (DD) and Conditional Dependence (CD); (4) comparing the causal structures generated by PC and GES; and (5) interpreting the causal implications to support evidence-based stunting prevention policies (Tou et al., 2021)(Nur et al., 2021)**.**.

The causal discovery approach in this study relies on key assumptions, including causal sufficiency, acyclicity, and conditional independence patterns encoded within the data. Under these assumptions, PC and GES algorithms can model plausible causal pathways that cannot be identified using traditional statistical approaches.

METHOD

This study uses the R programming language to analyze stunting data. Specifically, the pcalg package is used to implement the Peter Clark (PC) and Greedy Equivalence Search (GES) algorithms. Several relevant and continuous variables are deliberately selected to illustrate this approach. Because the primary objective of this study is to introduce causal modelling as an alternative analytical framework, the implementation is designed to be simple and illustrative, rather than comparative. The PC and GES algorithms are applied to a case study of stunting in toddlers, demonstrating their potential application in public health data analysis.

Stunting Data

This study used secondary data obtained from the West Bangka Regency Health Office. The data used were stunting data in toddlers in 2024. The analysis focused on eight main factors, namely height for age (TB/U), weight for age (BB/U), maternal education, exclusive breastfeeding, frequency of visits to Posyandu, immunization status, and smoking habits of family members at home. The final sample includes N = [**226**] toddlers after data cleaning. Categorical variables were encoded as factors, and continuous variables were standardized prior to analysis.

The pcalg package is used to implement the Peter-Clark (PC) and Greedy Equivalence Search (GES) algorithms. Because the data contain a mix of numeric and categorical variables, the PC algorithm is run using the conditional independence test for mixed data (gaussCItest for continuous variables and discreteCItest for categorical variables, depending on variable type). The GES algorithm is applied using the standard score-based approach with the Gaussian BIC scoring function, following common practice in mixed-data causal discovery when continuous variables dominate the structure. To maintain transparency, the following parameters are specified: PC algorithm: alpha = 0.05, stable skeleton search enabled. GES algorithm: maximum degree unrestricted, scoring function = BIC.

The primary objective of this study is to introduce causal modeling as an alternative analytical framework using a simple and illustrative case. The PC and GES procedures are applied to the stunting dataset to demonstrate their potential application in public health data analysis.

Software and Tools

This study employs the R programming language (version R 4.1.2) for causal discovery analysis. The implementation uses the pcalg package, including the PC algorithm and Greedy Equivalence Search (GES). The GaussL0penObsScore is applied as the scoring method for GES, while the PC algorithm uses a significance threshold of $\alpha = 0.05$ for conditional independence testing. All analyses were conducted using R Studio on a standard computing environment. The conceptual explanation of the GES method and PC-Algorithm is as follows:

Greedy Equivalence Search (GES) Algorithm

The causal algorithm is a score-based causal inference method used to compare and evaluate models according to their goodness of fit. In causal analysis, the Greedy Equivalence Search (GES) algorithm is employed

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume xx, Number xx, month Year
DOI : https://doi.org/10.33395/sinkron.xxx.xxx
e-ISSN : 2541-2019
p-ISSN : 2541-044X

to construct a directed causal structure represented as a graph. During each iteration, the GES algorithm selects modifications that improve the causal model based on a statistical scoring criterion, such as the Bayesian score. The conceptual framework of the GES algorithm is illustrated in Fig 1.

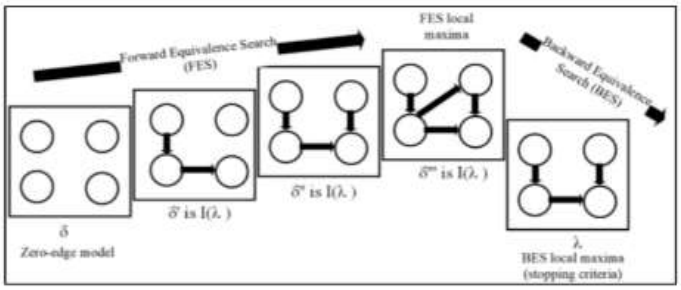

Fig 1. GES Method Concept

GES was configured using the GaussL0penObsScore with standardized continuous variables and numerically encoded categorical variables. The algorithm was run without structural priors to allow full exploration of plausible causal pathways in the stunting dataset. FES added edges that improved model fit, while BES removed redundant edges to obtain a parsimonious graph. GES is advantageous for this setting because it can reveal layered causal chains (e.g., maternal education → feeding practices → TB/U → BB/U), which are difficult to capture with conventional regression. Fig 2 - Fig 4 present the steps used in this study.

Algorithm 2
Add Operation (Forward Phase)

Input *A* dataset *D* and a CPDAG G
Output *A* CPDAG *H*
Let *M* be a set of CPDAGs obtainable form *G* by ***adding*** one edge
Let *K* be a CPDAG from *M* with the highest score *S(D,K)*

If *S(D,G) < S(D,K)* ***then***
H ← K
else
H ← G
else if
return **H**

Fig 2. Algorithm 2 of the GES method

Fig 2 illustrates the process of adding an edge to refine the structure of a Completed Partially Directed Acyclic Graph (CPDAG). The edge addition procedure begins with an initial graph G, from which the algorithm generates a new set of graphs M by progressively adding edges. Among these candidate graphs, the subset K with the highest score, denoted as S(D, K), is selected to evaluate its fit with the dataset D. If the score of graph K exceeds that of graph G, the current graph H is updated to K; otherwise, H remains equivalent to G. This iterative process aims to refine the graph structure by introducing edges that enhance its alignment with the observed data. The subsequent Backward Phase of the algorithm is depicted in Figure 3.

Algorithm 3
Delete Operation (Backward Phase)

Input *A* dataset *D* and a CPDAG G
Output *A* CPDAG *H*
Let *M* be a set of CPDAGs obtainable form *G* by *deleting* one edge
Let *K* be a CPDAG from *M* with the highest score *S(D,K)*

If *S(D,G) < S(D,K)* ***then***
H ← K
else
H ← G
else if
return **H**

Fig 3. Algorithm 3 of the GES method

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume xx, Number xx, month Year
DOI : https://doi.org/10.33395/sinkron.xxx.xxx
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Fig 3 presents the algorithmic process used to remove an edge from the Completed Partially Directed Acyclic Graph (CPDAG) in order to further refine its structural framework. During this stage, once an edge has been removed from the initial graph G, the algorithm generates a new set of candidate graphs M. Among these, the graph K with the highest score, denoted as S(D, K), is selected to assess the model's effectiveness based on the dataset D. If the removal of an edge results in a higher score compared to the original graph G, the current graph H is updated to K; otherwise, H remains equivalent to G. The elimination of edges that do not provide substantial contributions helps improve the overall causal structure of the model. Furthermore, once the Forward Phase is completed and no additional edges can enhance the score, the GES algorithm transitions to the Backward Phase. The complete procedural framework of the GES algorithm is illustrated in Fig 4.

Algorithm 4
Greedy Equivalence Search Algorithm

Input *A* dataset *D*
Output *A* CPDAG *G*

while *S(D, Add(D,G)) > S(D,G)* ***do***
    G ← *Add(D,G)*
end while
while *S(D, Del(D,G)) > S(D,G) do*
    G ← *Del(D,G)*
end while
return G

Fig 4. Algorithm 4 of the GES Method

Fig 4 presents the Greedy Equivalence Search (GES) Algorithm, which is used to learn the structure of a causal graph (CPDAG) from data. The algorithm operates in two main phases. In the first phase, it greedily adds edges to the graph as long as each addition improves the scoring function. Once no further beneficial additions can be made, the algorithm proceeds to the second phase, where it removes unnecessary edges if their removal increases the score. After both phases are completed, the algorithm returns the final graph, which represents the optimal structure according to the chosen scoring criterion.

Peter Clark Algorithm

The Peter Clark (PC) algorithm is a causal discovery method that identifies causal relationships by testing conditional independence (CI) among variables. The pseudocode of the PC algorithm is presented in Fig 5.

Algorithm 1 PC Algorithm

Input Dataset D with a set of variables V and significant level α
Output The undirected graph G with a set of edges E
Assume all nodes are connected initially
Let depth d = 0

1: **repeat**
2: for each orderes pair of adjacent vertices *X* and *Y* in *G* do
3: if |adj (*X,G*) \ {*Y*}|≥ d **then**
4: **for** each subset *Z* ⊂ adj (*X,G*) \ {*Y*} and |*Z*| = d **do**
5: Test *I* (*X,Y*|*Z*)
6: **if** *I* (*X,Y*|*Z*) then
7: Remove edge between *X* and *Y*
8: Save *Z* as the separating set of (*X, Y*)
9: Update *G* and **E**
10: break
11: **end if**
12: **end for**
13: **end if**
14: **end for**
15: Let *d* = d + 1
16: until |adj, (*X, G*) \ {*Y*} < for every pair of adjacent vertices in *G*

Fig 5. Algorithm from Peter Clark

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume xx, Number xx, month Year
DOI : https://doi.org/10.33395/sinkron.xxx.xxx
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The algorithm operates in two primary stages. In the first stage, the PC algorithm begins with a fully connected graph, assuming that all variables are initially dependent on one another. It then progressively removes edges that do not represent causal connections by performing conditional independence tests on every pair of variables. In the second stage, the algorithm orients the remaining edges according to established causal direction rules such as separation and orientation constraints to determine the directionality of relationships, ensuring that only causally relevant edges remain. The final output of the PC algorithm is a causal model represented as a Completed Partially Directed Acyclic Graph (CPDAG) [9].

Fig 5 illustrates that an edge between variables X and Y is removed when the two variables are found to be conditionally independent, without accounting for other variables in the model. Subsequently, distinct variable sets are retained separately in each iteration as the search depth (d) increases and additional variables are considered. This iterative process continues until no further conditional separations can be identified, resulting in a graphical structure that represents the causal relationships among variables. The final output of the Peter–Clark (PC) algorithm is a causal model comprising both directed and undirected edges, which is typically represented as a Completed Partially Directed Acyclic Graph (CPDAG) that satisfies the edge-completeness criterion. This rule is grounded in the concept of acyclicity and the notion that the selected variables yield a unique pattern of conditional independencies, thereby refining the overall v-structure of the causal graph.

Research Stages

This study was conducted through several sequential phases, including a literature review, data preprocessing, causal modeling analysis, evaluation, and dissemination. The detailed framework of the research process is illustrated in Fig 6.

Fig 6. Research Stages

The research process began with a literature review phase, which aimed to examine findings from previous studies relevant to this work [3, 8, 9, 10, 11], particularly those addressing the issue of stunting. The second phase involved data preprocessing, which focused on assessing dataset completeness, identifying missing values, and examining data distribution. Missing data were handled through data-cleaning procedures using the R command NewData <- Data[complete.cases(Data)].

The third phase involved applying the Greedy Equivalence Search (GES) and Peter–Clark (PC) algorithms to the prepared dataset. Model computation was performed using the R programming language, specifically employing the “pcalg” and “GaussL0penObsScore” packages. The output of this stage was a causal model representing the underlying determinants of stunting among children under five. The objective of this causal discovery process was to identify the causal structure, describing patterns of influence among predictor variables and their relationships to the target response variable. The process concluded with model evaluation and the dissemination of the resulting causal model.

RESULT

The findings of this study are presented to develop a causal model using the Peter Clark (PC) algorithm and the Greedy Equivalence Search (GES) algorithm. The demographic characteristics of the study subjects were described based on multiple variables, including sex, age, height (H), weight (W), height-for-age (TB/U), weight-for-age (BB/U), maternal education, maternal occupation, exclusive breastfeeding, frequency of Posyandu visits, immunization status, maternal age at marriage, source of drinking water, toilet ownership, presence of smoking family members in the household, and maternal nutritional status during pregnancy. The details of the dataset are illustrated in Fig 7.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume xx, Number xx, month Year
DOI : https://doi.org/10.33395/sinkron.xxx.xxx
e-ISSN : 2541-2019
p-ISSN : 2541-044X

```
> str(kausal)
'data.frame':   99 obs. of  13 variables:
 $ BB.U                          : int  11 11 11 11 12 8 7 13 14 10 ...
 $ TB.U                          : int  90 84 94 88 95 72 66 96 93 86 ...
 $ PBL                           : int  48 41 46 47 48 48 47 47 48 36 ...
 $ PendidikanIbu                 : int  0 1 0 1 0 0 3 1 2 1 ...
 $ PekerjaanIbu                  : int  1 1 1 1 1 1 1 1 1 1 ...
 $ ASIEksklusif                  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ KunjunganPosyandu             : int  1 1 1 1 1 1 1 1 1 1 ...
 $ Imunisasi                     : int  1 1 1 1 1 0 0 1 1 0 ...
 $ UsiaIbuMenikah                : int  19 19 18 21 19 19 19 17 18 17 ...
 $ SumberAirMinum                : int  1 2 2 1 2 1 1 1 1 1 ...
 $ KepemilikanJamban             : int  1 1 1 1 1 1 1 0 1 1 ...
 $ AnggotaKeluargaMerokokDalamRumah: int  1 1 1 1 1 1 1 1 1 1 ...
 $ StatusGiziIbuSaatHamil        : int  0 0 1 0 0 0 0 0 1 1 ...
```

Fig 7. Stunting Factor Data

Fig 7 presents the detailed variables included in the stunting dataset. Prior to analysis, the researchers conducted a data preprocessing phase to identify and handle missing values. Missing data were examined and removed using the R command DataBaru <- Data[complete.cases(Data),]. Subsequently, Variables selected for computation were identified based on correlation tests, using a significance level of 0.05 (p-value < 0.05) to determine statistical significance. The distribution of stunting cases among children under five is illustrated in Fig 8.

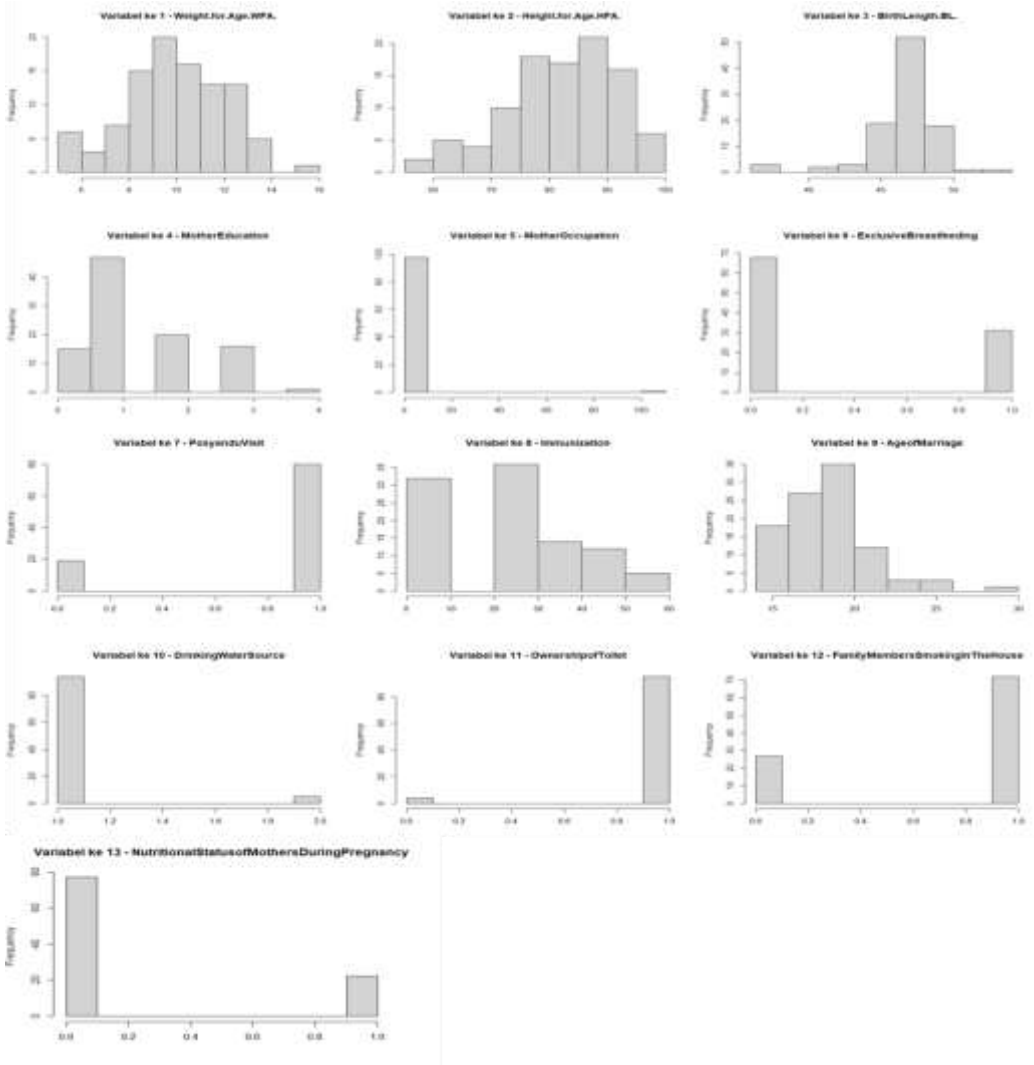

Fig 8. Distribution of Stunting Data

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume xx, Number xx, month Year
DOI : https://doi.org/10.33395/sinkron.xxx.xxx
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Fig 8 illustrates the data distribution of several variables employed in this study to analyze the determinants of stunting among children under five using the Peter Clark (PC) and Greedy Equivalence Search (GES) algorithms. Each histogram represents the frequency distribution of values for the respective research variables. These variables encompass both height for age (TB/U), weight for age (BB/U), maternal education, exclusive breastfeeding, frequency of visits to Posyandu, immunization status, and smoking habits of family members at home.

The visualization of these distributions is crucial, as it provides an initial overview of the data patterns that will be utilized in the causal analysis process. Through these distribution patterns, researchers can identify variables that potentially exert strong influences on the incidence of stunting. The resulting dataset was subsequently analyzed using the Peter Clark (PC) algorithm to construct a causal model representing the interrelationships among variables, thereby identifying the dominant factors contributing to stunting among children under five. Fig 9 dan Fig 10 present the outcomes of the causal modeling based on the stunting dataset.

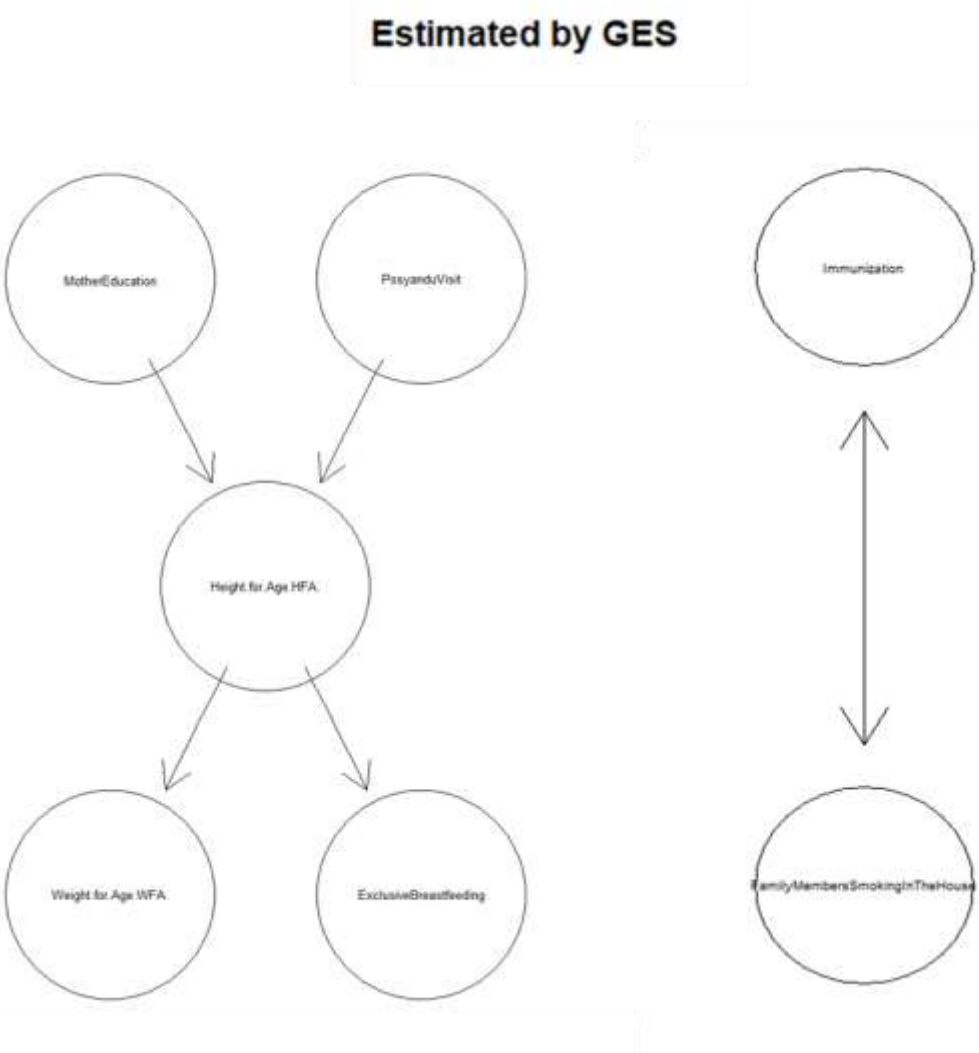

Fig 9. Causal Stunting Model using the GES Algorithm

Fig 9 illustrates the causal structure estimated from the stunting dataset using the GES algorithm. Several relationships appear consistent with established causal ordering, while others require careful interpretation due to potential violations of temporal precedence. Maternal education and Posyandu visit frequency show directed effects toward height-for-age (TB/U), suggesting that socioeconomic factors and access to health services may contribute to improved child growth. TB/U then appears connected to weight-for-age (BB/U), which is consistent with the biological relationship between linear growth and weight status.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume xx, Number xx, month Year
DOI : https://doi.org/10.33395/sinkron.xxx.xxx
e-ISSN : 2541-2019
p-ISSN : 2541-044X

However, the initial GES output also produced an edge from TB/U to exclusive breastfeeding. This direction is not causally plausible, as breastfeeding occurs early in life and precedes the child's growth outcomes. The presence of this edge is therefore interpreted as a statistical artefact likely arising from correlated household factors or unmeasured confounding rather than a true causal influence. In the interpretation of the model, only temporally coherent directions such as exclusive breastfeeding influencing TB/U are considered meaningful.

The model additionally identifies a link between immunization status and the presence of smoking family members at home. Although the algorithm suggests a bidirectional association, this is more reasonably interpreted as an indication of shared household behavioral and environmental conditions rather than mutual causation. This emphasizes the importance of considering domain knowledge, data limitations, and temporal ordering when interpreting causal discovery outputs in public health studies.

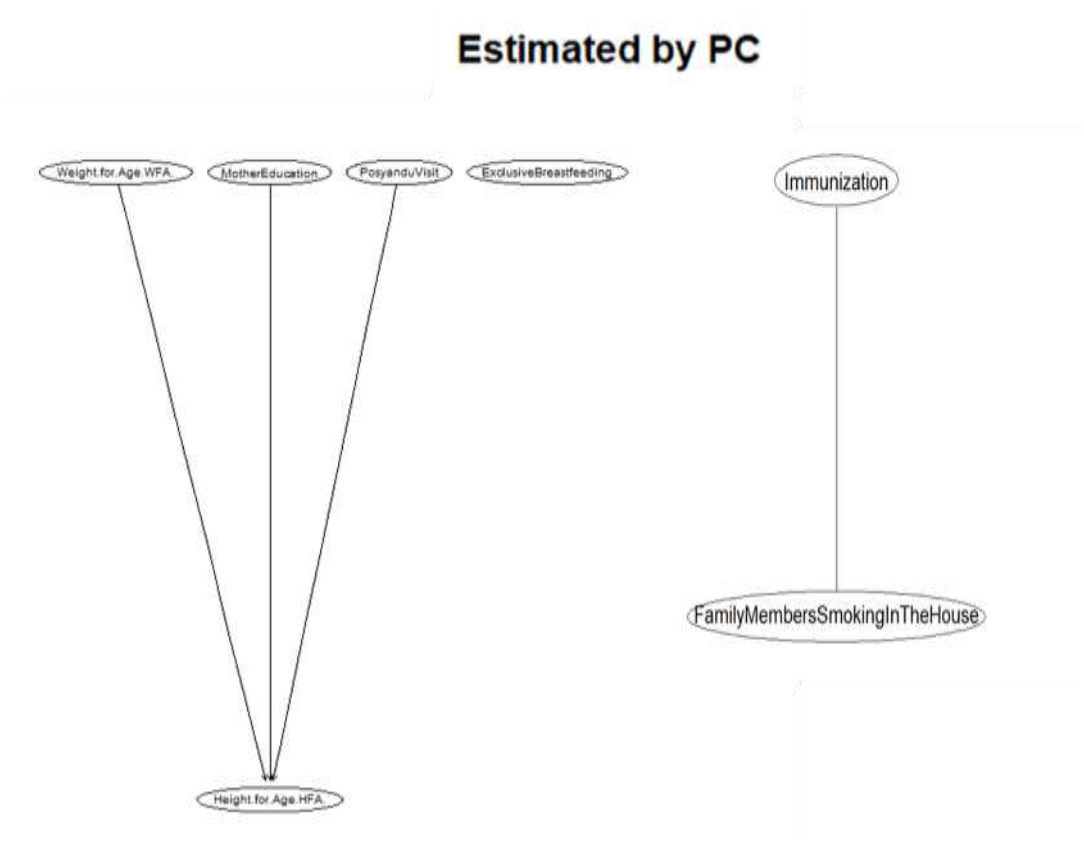

Fig 10. Causal Stunting Model using the Peter-Clark Algorithm

Fig 10 presents the causal structure derived from the Peter-Clark (PC) algorithm, which identifies the key determinants influencing the height-for-age (TB/U) indicator as a proxy for stunting among children under five. The model reveals that maternal education, Posyandu visit frequency, exclusive breastfeeding, and weight-for-age (BB/U) directly influence TB/U, indicating that both maternal and child health behaviors, as well as nutritional factors, play crucial roles in determining child growth outcomes.

Notably, the PC model also identifies a distinct independent relationship between immunization status and the presence of smoking family members within the household, suggesting that environmental risk factors may operate through separate pathways not directly linked to anthropometric indicators.

This structure highlights that maternal education and healthcare engagement (through Posyandu visits and breastfeeding practices) serve as dominant causal drivers of child growth, while environmental health behaviors influence complementary but independent dimensions of child well-being. The model supports the notion that effective stunting reduction strategies require integrated interventions addressing both maternal knowledge and household environment. The results of the causal modeling evaluation are presented in Table 1.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume xx, Number xx, month Year
DOI : https://doi.org/10.33395/sinkron.xxx.xxx
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 1. Directed Density (DD) and Causal Density (CD) Evaluation Metrics

Metric	GES Algorithm	PC Algorithm
DD	4/5 = 0.80	5/5 = 1.0
CD	4 / [7(7-1)] = 0.10	5 / [7(7-1)] = 0.12

Error! Reference source not found. shows the formal definition of Directed Density (DD) is the proportion of directed edges relative to the total number of edges in the CPDAG, expressed as DD = k / m, where k is the number of directed edges and m is the total edges (directed + undirected). Based on the CPDAG GES shown in Figure 9, there are 4 directed edges and 1 undirected edge, so that the total m = 5 and obtained DD = 4/5 = 0.80. The calculation of Causal Density (CD) follows the definition of CD = k / [n(n−1)], where n is the number of variables. With n = 7 and k = 4, the CD value for GES is 4/42 = 0.10. For the CPDAG PC, all identified edges are directed with a total of 5 directed edges, so DD = 1.00 and CD = 5/42 = 0.12. This calculation ensures consistency between the definitions of DD and CD according to standards used in the causal discovery literature.

DISCUSSIONS

Causal modeling results between the Greedy Equivalence Search (GES) and Peter-Clark (PC) algorithms show different causal structures in identifying the determinants of stunting in children. Although both methods aim to infer causal relationships, the underlying mechanisms differ significantly, resulting in variations in the estimated causal graphs.

The GES model (Figure 9) applies a score-based learning approach that optimizes the Bayesian Information Criterion (BIC), producing a structure that prioritizes the overall goodness of fit over strict conditional independence requirements. The model identifies maternal education and Posyandu visit frequency as direct causal determinants of height-for-age (TB/U). Mechanistically, this aligns with established that maternal education improves health literacy, feeding practices, and service utilization, while Posyandu attendance enhances early detection and guidance on child growth. Together, these factors support improved linear growth, which is further reflected in the model's pathway connecting TB/U to weight-for-age (BB/U). (Sutarno & Telkom, 2021)(Sumardiyono, 2020)

Although the GES algorithm initially produced an implausible edge from TB/U to exclusive breastfeeding, this is best interpreted as an artefact of correlated household behaviors rather than a true causal direction. The likely mechanism is that households practicing exclusive breastfeeding also demonstrate better overall child-care behaviors, leading to improved TB/U thus reversing the temporal interpretation seen in the raw GES output.

Importantly, the GES structure also highlights indirect pathways, such as the linkage between household smoking and immunization. This does not imply direct causation but rather indicates potential latent confounders, such as household socioeconomic status, parental health awareness, cultural norms, or exposure to local health programs. These unobserved factors may simultaneously influence smoking behavior and immunization uptake, generating the appearance of a bidirectional relationship in the GES graph (Sumardiyono, 2020)(Ayuni et al., 2024).

In contrast, the PC algorithm (Fig 10) employs a constraint-based approach that relies on conditional independence tests, resulting in a more conservative and sparser causal structure. For example, maternal education, Posyandu visits, and exclusive breastfeeding are all directly connected to TB/U, suggesting independent causal influences without mediation through other variables. This outcome reflects the PC algorithm's sensitivity to latent confounding: paths that cannot be conditionally distinguished due to unmeasured variables are often pruned, producing a model that emphasizes robust, contextually reasonable dependencies. The isolated bidirectional association between immunization and household smoking likewise suggests a shared unmeasured determinant likely health awareness or socio environmental conditions rather than a direct causal interaction (Rezaeizadeh et al., 2024).

From a mechanistic perspective, both algorithms offer complementary insights. The GES algorithm emphasizes broader propagation of socioeconomic and behavioral influences, capturing multi-level causal chains in which maternal knowledge, household environment, and service utilization interact to shape child growth outcomes. Meanwhile, the PC algorithm yields more tightly constrained pathways that align with temporally and biologically plausible mechanisms, such as maternal education and exclusive breastfeeding influencing TB/U (de Onis & Branca, 2016).

Overall, the consistency of both methods in identifying maternal education, exclusive breastfeeding, and Posyandu attendance as key determinants supports existing evidence on the importance of maternal and community health practices in stunting prevention. The presence of potential latent confounders such as socioeconomic status, cultural caregiving norms, household health awareness, or availability of community health

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

services highlights the need for cautious interpretation and suggests that future studies incorporate additional contextual variables to fully capture the complexity of stunting determinants.

Methodological Considerations and Limitations

Both the GES and PC algorithms operate under several assumptions such as causal sufficiency, acyclicity, and the faithfulness condition which, when unmet due to unmeasured contextual factors or bidirectional relationships, may lead to uncertain or incorrect edge orientation; in addition, sparse network connectivity and sensitivity to sample size can further limit the stability of causal direction estimates. These methodological constraints may also explain domain-specific inconsistencies, for example when the model produces an edge height-for-age (TB/U) → exclusive breastfeeding, despite literature showing that exclusive breastfeeding reduces infection and undernutrition risk; such a result is likely not a true causal effect, but rather a reflection of residual correlation, unmeasured confounding, or edge-orientation difficulties inherent to causal discovery algorithms.

CONCLUSION

This study using the Directed Density (DD) and Causal Density (CD) metrics show that the PC Algorithm produces a more complete and denser causal structure than GES. A PC DD value of 1.0 indicates that all edges in the CPDAG are successfully oriented, so that the direction of causal relationships can be interpreted without ambiguity. In contrast, GES only achieved a DD of 0.80, indicating that some edges are still unoriented. Meanwhile, the PC CD value of 0.12 is slightly higher than GES's 0.10, reflecting that PC identifies more causal relationships between variables. Overall, these findings indicate that PC tends to be more exploratory in mapping complex causal structures, while GES produces a more parsimonious and conservative structure. The choice of algorithm can be adjusted according to the analysis objective: whether prioritizing completeness of causal orientation (PC) or a simpler and more stable model (GES).

Methodologically, these findings demonstrate the utility of causal discovery approaches for health data, offering interpretability advantages over purely predictive or regression-based models. Substantively, the resulting causal pathways provide actionable insights for policy formulation. Strengthening maternal health literacy programs, improving the regularity and quality of Posyandu services, and promoting exclusive breastfeeding emerge as evidence-based intervention priorities. Collectively, the study provides a scientific foundation for designing integrated maternal–child health policies to reduce stunting prevalence in West Bangka Regency.

ACKNOWLEDGMENT

Thank you to LPPM Universitas Bangka Belitung for its support of this research program through the 2025 Young Researcher Funding Grant. Thank you also to the West Bangka Regency Health Office for their great assistance in ensuring the smooth running of this research.

REFERENCES

Ayuni, S. S., Rizqi, E. R., & Isnaeni, L. M. A. (2024). Faktor-Faktor Penyebab Kejadian Stunting pada Balita Usia 12-24 Bulan di Desa Karya Mulya, Provinsi Riau. *Jurnal Ilmu Gizi Dan Dietetik*, *3*(1), 48–55. https://doi.org/10.25182/jigd.2024.3.1.48-55

de Onis, M., & Branca, F. (2016). Childhood stunting: A global perspective. *Maternal and Child Nutrition*, *12*, 12–26. https://doi.org/10.1111/mcn.12231

Dewanti, C., Ratnasari, V., & Rumiati, T. (2019). Pemodelan Faktor-faktor yang Memengaruhi Status Balita Stunting di Provinsi Jawa Timur Menggunakan Regresi Probit Biner. *Jurnal Sains Dan Senit ITS*, *8*(2), 2337–3520.

Efrizal, W. (2020). Berdampakkah Pandemi Covid-19 Terhadap Stunting Di Bangka Belitung? *Kebijakan Kesehatan Indonesia*, *09*(03), 154–157.

Herliansyah, V., Latuconsina, R., & Dinimaharawati, A. (2021). Prediksi Stunting pada Balita dengan Menggunakan Algoritma Klasifikasi Naive Bayes. *E-Proceeding of Engineering*, *8*(5), 6642–6649.

Julianti, E., & Elni. (2021). Faktor-Faktor yang Berhubungan dengan Kejadian Stunting pada Balita. *JITK Bhamada*, *12*(2), 2–7.

Khairani, Mursyita, A., & Darmawa, S. (2020). *Situasi Stunting di Indonesia*. Pusat Data dan Informasi Kementrian Kesehatan RI.

Lonang, S., & Normawati, D. (2022). Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination. *Jurnal Media Informatika Budidarma*, *6*(1), 49–56. https://doi.org/10.30865/mib.v6i1.3312

Nur, Y. S. R., Rahmadi, R., & Effendy, C. (2021). Causal Modeling Between Factors on Quality of Life in Cancer Patients Using S3C-Latent Algorithm. *Journal of Information Systems Engineering and Business*

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Intelligence, *7*(1), 74. https://doi.org/10.20473/jisebi.7.1.74-83

Rahmadi, R. (2019). *Finding stable causal structures from clinical data*. 63–78.

Rahmadi, R., Groot, P., Heins, M., Knoop, H., Heskes, T., & Optimistic, T. (2017). Causality on Cross-Sectional Data : Stable Specification Search in Constrained Structural Equation Modeling. *Appl. Soft Comput.*, *52*, 687–698.

Rahmadi, R., Groot, P., & Heskes, T. (2018). The stablespec package for causal discovery on cross-sectional and longitudinal data in R. *Neurocomputing*, *275*, 2440–2443. https://doi.org/10.1016/j.neucom.2017.10.064

Rezaeizadeh, G., Mansournia, A., Keshtkar, A., Farahani, Z., Zarepour, F., & Sharafkhah, M. (2024). Articles Maternal education and its in fl uence on child growth and nutritional status during the fi rst two years of life : a systematic review and meta-analysis. *EClinicalMedicine*, *71*, 102574. https://doi.org/10.1016/j.eclinm.2024.102574

Sumardiyono, S. (2020). Pengaruh Usia, Tinggi Badan Dan Riwayat Pemberian Asi Eksklusif Terhadap Stunting Pada Balita. *Medika Respati : Jurnal Ilmiah Kesehatan*, *15*(1), 1. https://doi.org/10.35842/mr.v15i1.269

Sumarni, S., Oktavianisya, N., & Suprayitno, E. (2020). Pemberian Air Susu Ibu Eksklusif Berhubungan dengan Kejadian Stunting pada Balita di Pulau Mandangin Kabupaten Sumenep Provinsi Jawa Timur. *Jurnal Riset Hesti Medan Akper Kesdam I/BB Medan*, *5*(1), 39–43. https://doi.org/10.34008/jurhesti.v5i1.174

Sutarno, H. H., & Telkom, U. (2021). *PREDIKSI STUNTING PADA BALITA DENGAN MENGGUNAKAN ALGORITMA KLASIFIKASI K-NEAREST NEIGHBORS STUNTING PREDICTION IN CHILDREN USING K-NEAREST NEIGHBORS CLASSIFICATION ALGORITHM*. *8*(5), 6657–6661.

Tou, N., Effendy, C., & Rahmadi, R. (2021). Causal Relations of Factors Representing the Elderly Independence in Doing Activities of Daily Livings Using S3C-Latent Algorithm. *International Journal of Artificial Intelligence Research*, *5*(1), 65–77. https://doi.org/10.29099/ijair.v5i1.206

Widia Pebrianti, S., Astuti, R., & M Basysyar, F. (2024). Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Status Stunting Balita Di Desa Bojongemas. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(2), 2479–2488. https://doi.org/10.36040/jati.v8i2.8448

*name of corresponding author