# Triangulation for causal loop diagrams: constructing biopsychosocial models using group model building, literature review, and causal discovery

Check for updates

Jeroen F. Uleman [1] ✉, Maartje Luijten [2], Wilson F. Abdo[3], Jana Vyrastekova [4], Andreas Gerhardus[5], Jakob Runge[5,6], Naja Hulvej Rod[1,7] & Maaike Verhagen[2,7]

The complex nature of many health problems necessitates the use of systems thinking tools like causal loop diagrams (CLDs) to visualize the underlying causal network and facilitate computational simulations of potential interventions. However, the construction of CLDs is limited by the constraints and biases of specific sources of evidence. To address this, we propose a triangulation approach that integrates expert and theory-driven group model building, literature review, and data-driven causal discovery. We demonstrate the utility of this triangulation approach using a case example focused on the trajectory of depressive symptoms in response to a stressor in healthy adults. After triangulation with causal discovery, the CLD exhibited (1) greater *comprehensiveness*, encompassing multiple research fields; (2) a modified *feedback* structure; and (3) increased transparency regarding the *uncertainty* of evidence in the model structure. These findings suggest that triangulation can produce higher-quality CLDs, potentially advancing our understanding of complex diseases.

In epidemiology and related fields, the adoption of causal diagrams has marked a significant leap forward in drawing causal inferences from observational data. These visual tools synthesize theory-based assumptions about causal relations, enabling the estimation of causal effects for well-defined hypothetical interventions[1,2]. However, health problems are often complex and characterized by dynamic feedback, adaptation, and interactions across numerous variables at multiple scales of hierarchical organization[3–5], complicating the search for the underlying network of causality. Consequently, researchers increasingly recommend adopting a multitude of approaches to causality[3,4,6] and, in particular, complex systems thinking[5–8].

Among complex systems thinking methods, the causal loop diagram (CLD) is increasingly recognized as a valuable tool in health research[9,10]. CLDs are causal diagrams that describe hypothesized links between many system variables across relevant space and time scales. A hallmark of CLDs is the inclusion of feedback loops, which are integral to understanding nonlinear changes over time[11]. Crucially, CLDs provide a conceptual basis for computational system dynamics models that quantify causal relations and can simulate potential 'what-if' scenarios[10], such as personalized interventions on multiple modifiable risk factors[12].

When constructing CLDs, well-established participatory methods, such as group model building (GMB), are commonly used to exploit the collective insights of a diverse group of domain experts[13,14]. This method of eliciting and synthesizing expert knowledge is particularly beneficial for complex problems where data tend to be scarce and unified causal theories are lacking. However, overreliance on domain experts as a single source of evidence can compromise the reproducibility and structural validity of CLDs[15]. To mitigate this, researchers increasingly combine multiple approaches, such as GMB and literature review[9,10,15]. Nevertheless, such approaches may still be prone to subjectivity, relying on the researcher's or expert group's assumptions and limitations, which can introduce biases, especially when empirical data are inadequately considered[5]. Therefore, a mixed-method approach that integrates qualitative and quantitative data is

[1]Copenhagen Health Complexity Center, Department of Public Health, University of Copenhagen, Copenhagen, Denmark. [2]Behavioural Science Institute, Faculty of Social Sciences, Radboud University, Nijmegen, The Netherlands. [3]Department of Intensive Care Medicine, Radboud University Medical Center, Nijmegen, The Netherlands. [4]Institute for Management Research, Department of Economics, Radboud University, Nijmegen, The Netherlands. [5]German Aerospace Center (DLR), Institute of Data Science, Jena, Germany. [6]Technische Universität Berlin, Berlin, Germany. [7]These authors contributed equally: Naja Hulvej Rod, Maaike Verhagen. ✉e-mail: Jeroen.uleman@sund.ku.dk

warranted to develop CLDs that more accurately represent real-world systems[16].

A promising quantitative approach to identifying causal diagrams from observational data is causal discovery[17,18]. A famous and classic example is Granger causality[19], together with its various refinements (see, e.g., Camps-Valls et al.[20] for a brief overview). More recently, initiated by the influential works of Pearl[1] and Spirtes, Glymour, and Scheines[21], causal discovery has emerged as an established field of research in statistics and machine learning. While the majority of works in this field were primarily intended for cross-sectional data, which captures information at a single point in time, more recent algorithms can efficiently handle longitudinal data, which involves repeated observations of the same subjects over a period of time[18,22–24]. By assessing variables over multiple time points (provided this feedback is not faster than the temporal resolution), dynamic feedback loops can be identified. Although causal discovery does not require a priori domain knowledge of the causal links, its application to real-world data is challenging as causal discovery algorithms can be strongly affected by unmet assumptions regarding, for example, latent confounders, measurement error, selection bias, sample size, and missing data[18]. Consequently, the resulting causal diagrams might deviate from theory-driven models[25] and should be compared to – and validated by – domain knowledge[26].

To overcome the individual limitations of causal discovery and GMB, we propose triangulation[27] to combine domain expertise from GMB, literature review, and causal discovery (Fig. 1). Triangulation can improve causal inferences by integrating multiple approaches with different and unrelated sources of potential biases[4,28,29]. For example, in GMB, experts can propose theory-based links not found in the available data. However, the experts may overlook certain links, especially for large systems where considering all possibilities is infeasible[10,16] or discipline boundaries do not fully integrate the existing empirical evidence. In contrast, causal discovery provides a data-driven perspective where all possible links are assessed. Thus, it has the potential to identify causal links that experts have overlooked[25] or that are not yet established in the scientific literature.

By triangulating across data sources and methods, causal evidence can, therefore, become increasingly compelling[4,27,29]. Combining GMB, literature review, and causal discovery in a triangulation approach should enable the accumulation of evidence for specific links between variables, leading to more reproducible and valid causal diagrams while pointing out uncertainties (e.g., if GMB and causal discovery disagree) where further research is needed[27].

This paper provides an empirical case example of a mixed methods triangulation approach to CLD development. The case example seeks to map the biopsychosocial feedback mechanisms underlying the course of

depressive symptoms in response to a stressor in a healthy adult population. We combine theory-based insights from GMB and literature review with causal discovery using longitudinal data from the Healthy Brain Study[30]. In this way, we aim to assess how triangulation can impact the development of CLDs.

## Results

In this section, we consecutively discuss the case example results from the GMB process, literature review, causal discovery, and their combination through triangulation. Our group of four domain experts included 14 variables in the CLD. These variables came from four domains covered by two experts each: biological, psychological, behavioral, and social. Table 1 provides the baseline characteristics of the measures from the Healthy Brain Study utilized to operationalize the included CLD variables. Supporting evidence for the CLD is provided in the supplementary Excel table and an online interactive version of the CLDs is presented here: https://hbscld. kumu.io/triangulation-for-causal-loop-diagrams.

### Group model building
The experts mapped the links between the variables in the GMB sessions. A complete consensus was reached among the expert group regarding each of the 33 links they identified. Figure 2 shows the CLD resulting from these GMB sessions.

### Literature review
After each GMB session, a literature review was conducted for every newly added link. Based on the results of the literature review, the experts adjusted the CLD. The updated CLD is shown in Fig. 3, containing 31 links overall. No literature was found for four of the links. Out of these four, two were removed by the experts (*Daily hassles → Sleep disturbance* and *Proinflammatory cytokines → Perceived stress*) because upon further discussion they considered them less plausible. The other two links (*Depressive symptoms → Perceived stress* and *Perceived stress → Sleep disturbance*) were kept as the experts still found them plausible with reasonable potential mechanisms of action (dotted arrows, Fig. 3). Nevertheless, these links were considered 'hypothetical' due to the lack of literature evidence. Finally, the literature review also sparked a discussion among the experts about the direction of one of the links, leading to the change of *Proinflammatory processes → Loneliness* into *Loneliness → Proinflammatory processes*.

### Causal discovery
We then conducted the causal discovery analysis using the J-PCMCI+ algorithm[31]. The algorithm identified 12 links (Fig. 4 and Table 2) between the variables identified by the experts. We also conducted several sensitivity tests to assess these findings' robustness. Firstly, we assessed the impact of the significance level of the regression-based conditional independence test, resulting in the same graph with four links missing (Supplementary Information C). Secondly, we assessed the impact of using a different independence test and found the same graph (Supplementary Information C). Finally, to explore the possible impact of selective attrition, we repeated the analysis with only those participants ($N = 51$) who had complete data at all three assessments (i.e., time points) (Supplementary Information C), where only six out of 11 links corresponded to those found in Table 2. This suggests that causal discovery analysis alone, while informative, may best be complemented with theory-based methods to address data-based limitations.

### Triangulation
Finally, the experts participated in the final GMB session, combining the information from GMB, literature review, and causal discovery. Figure 5 shows the final CLD resulting from this triangulation process, which contains 36 causal links. An overview of feedback loops is provided in Supplementary Information D.

We distinguished between three possibilities for each link based on the GMB and causal discovery results. (1) Both GMB and causal discovery
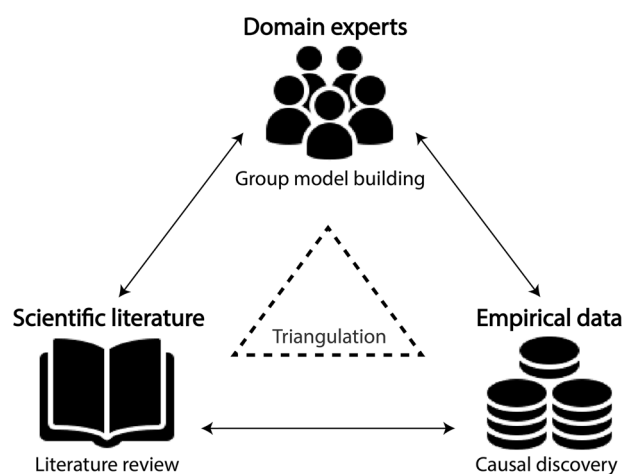


**Fig. 1 | Evidence triangulation for causal loop diagrams.** The triangulation of evidence for causal loop diagrams with sources of evidence (domain experts, scientific literature, empirical data) and methods (group model building, literature review, causal discovery).

**Table 1 | CLD variables with corresponding measures from the Healthy Brain Study (HBS) and their baseline characteristics (*N* = 403)**

| CLD variables | HBS measures | Mean (SD) | Available per assessment (1/2/3) |
|---|---|---|---|
| – | Sex [% Females] | 60% | 94% |
| – | Highest completed education level [% University or applied university] | 77% | 93% |
| – | Country of birth [% Dutch] | 91% | 93% |
| Depressive symptoms | Self-report inventory of depressive symptomatology (IDS-SR) [0–27] | 4.4 (3.3) | 80/60/62% |
| Perceived stress | Perceived stress scale (PSS) [0–40] | 12 (6.0) | 80/63/64% |
| Loneliness | UCLA loneliness scale (UCLS) [20–100] | 38 (11) | 77/60/62% |
| Perceived social support | Multi-dimensional scale of perceived social support (MSPSS) [1–7] | 6.0 (0.9) | 77/60/62% |
| Rumination | Cognitive emotion regulation questionnaire – rumination subscale (CERQ-R) [4–20] | 11 (3.8) | 78/60/62% |
| Sleep disturbance | Pittsburgh sleep quality index (PSQI) [0–21] | 4.7 (2.8) | 88/70/68% |
| Smoking | Substance Matrix Mate-Q (SMQ) [number of cigarettes in the past 30 days] | 6.4 (25) | 90/72/71% |
| Prosocial behavior | Public goods game (PGG) [unconditional contribution to a common project, 0–5 euro] | 3.2 (1.6) | 70/63/65% |
| Mindfulness | Five facet mindfulness questionnaire (FFMQ) [24–120] | 83 (10) | 93/72/70% |
| Body mass index | Body mass index (BMI) [kg/m²] | 25 (4.2) | 97/78/75% |
| Physical inactivity | Sedentary behavior questionnaire (SBQ) [Hours of sitting in a typical week] | 76 (34) | 93/72/70% |
| Daily hassles | Daily hassles scale (DH) [Sum of hassles and their severity in the past month, 0–354] | 32 (23) | 79/61/63% |
| Anti-inflammatory processes | Interleukin 10 (IL-10) [pg/ml] | 21 (18) | 84/54/35% |
| Proinflammatory processes | Tumor Necrosis Factor alpha (TNF-alpha), Interleukin 6 (IL-6) [pg/ml]* | 6.3 (4.3), 6.0 (10.5) | 84/54/35% |

Reported are percentages or means and standard deviations (SD) with the percentage of participants for whom the measure was available in assessments 1/2/3. Total scores of questionnaires were used unless otherwise specified, and scale ranges were provided between the square brackets unless otherwise specified (e.g., units). *TNF-alpha and IL-6 were turned into Z-scores based on their baseline means and standard deviations, after which the average of these Z-scores was used to operationalize Proinflammatory processes.

indicate a link. (2) Causal discovery indicates a link, while GMB does not. (3) GMB indicates a link, while causal discovery does not. The literature review was used to address inconsistencies between the two methods. We took a link as indicated by causal discovery if it had the same polarity as the CLD link and had either the same direction or was undirected (i.e., did not have the opposite direction). Although the 'o-o' notation (Table 2) does not indicate a feedback loop, as the method assumes acyclicity for the contemporaneous links, it does not exclude the possibility of a loop in case the assumption of acyclicity is violated. The 'o-o' notation could suggest a link in either direction. Therefore, we interpret 'o-o' as potential evidence for both links, though it should be noted that this evidence is less strong than that of a directed link.

1) Five links in Table 2 were indicated by both GMB and causal discovery. Each of these links was kept in the CLD by the experts. Four links could not be oriented by J-PCMCI+ (Table 2), so the opposite directions were also considered. This resulted in the addition of *Depressive symptoms → Sleep disturbance* to the CLD, after which supporting literature was found.

2) Seven links in Table 2 were indicated by causal discovery but not GMB. For these links, the experts either adopted or rejected a link, depending on the presence or absence of scientific literature. One of the links, *Anti-inflammatory processes → Proinflammatory processes*, was discussed in detail. The experts decided its direction was likely incorrectly identified by J-PCMCI+ and took the link as support for *Proinflammatory processes → Anti-inflammatory processes*, consistent with the balancing loop (B1) in Figs. 2 and 3. For four of the links, no supporting literature was found (Table 2). One of these links, *Loneliness → Perceived social support*, was not deemed plausible anymore by the experts and, therefore, not added. Although no literature was found for the other three links, the experts still considered them plausible and added them to the CLD.

3) Most of the links indicated by GMB were not indicated by causal discovery. This discrepancy could arise from data limitations, such as the links not being sufficiently strong within the observed time intervals to be detected. To avoid false negatives, the experts were, therefore, prompted to reevaluate these links only when they were also not supported by the literature review (hypothetical links in Fig. 3). Since the undirected link in Table 2 supported *Depressive symptoms → Perceived stress*, the discussion was limited to *Sleep disturbance → Perceived stress*. The experts took into account that, in the causal discovery analysis, *Sleep disturbance* and *Perceived stress* were only independent after conditioning on other variables. Out of these variables, *Depressive symptoms* most strongly diminished the effect. Therefore, the experts considered the indirect pathway *Sleep disturbance → Depressive symptoms → Perceived stress* more plausible than the direct link, which they removed. The opposite link, *Perceived stress → Sleep disturbance*, was, however, kept since scientific literature was found to support it.

## Discussion

In this study, we employed a novel mixed-methods triangulation approach to construct a biopsychosocial CLD, mapping the complex interplay of factors involved in the course of depressive symptoms in response to a stressor in healthy adults. Our CLD provides a preliminary causal mapping of relevant variables, spanning four research domains seldom considered in the same model: biological, psychological, behavioral, and social, relating to our expert group's fields of expertise. The CLD blends three layers of scientific evidence: consensus from a multi-disciplinary group of domain experts via GMB, knowledge extracted from a literature review, and data-driven detection of causal links through causal discovery using a unique multi-dimensional longitudinal cohort: the Healthy Brain Study[30].

While previous studies have compared causal discovery results to expert-based diagrams (e.g., Petersen et al.[25]), the present study is, to our knowledge, the first to combine causal discovery with GMB, a formal participatory method. Petersen et al.[25] found that causal discovery identified links missing from expert-based directed acyclic diagrams that were considered plausible by those same experts. Similarly, in our study, seven links indicated by J-PCMCI+ were not initially indicated by GMB (Table 2), but five of these were incorporated into the final CLD (Fig. 5). This suggests that
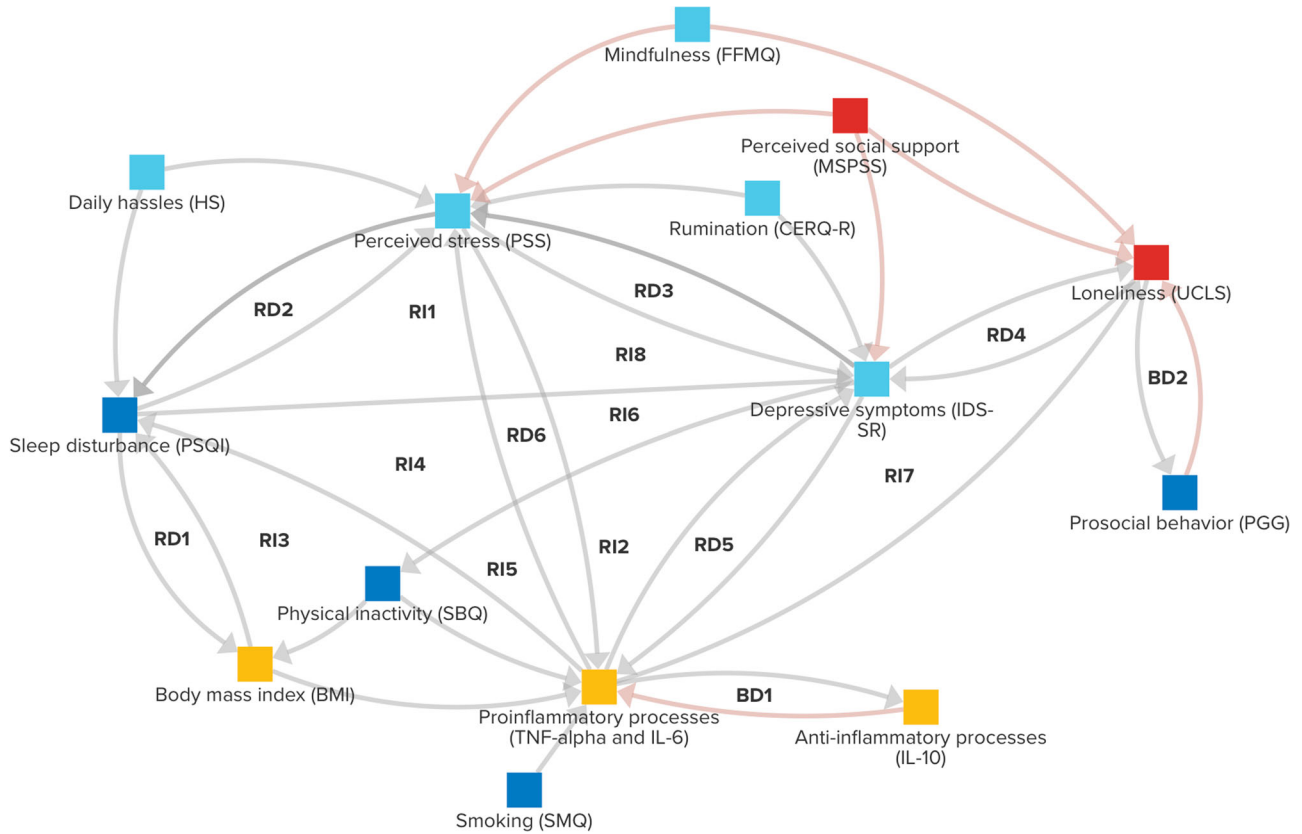
**Fig. 2 | Expert-driven causal loop diagram developed through group model building.** Reinforcing (R) and balancing (B) feedback loops consisting of two (RD, BD) or three variables (RI, BI) have been annotated (see Supplementary Information D for an overview). Grey connections have positive polarity, whereas red connections have negative polarity. The variables' colors correspond to different dimensions: Social = Red; Behavioral = Dark blue; Psychological = Light blue; Biological = Yellow.

causal discovery can be particularly beneficial for proposing links that were missed by the experts – who may not fully explore all possible interconnections – or indicate potential research gaps. Furthermore, Petersen et al.[25] found large differences between individual expert contributions. Participatory methods like GMB can facilitate greater agreement by actively building consensus through the alignment of mental models[32], plausibly leading to higher-quality diagrams. Determining the most effective method for developing causal diagrams, however, requires further empirical exploration through future triangulation studies[33].

Based on our case example, we identify three main areas where triangulation positively affects CLD development: comprehensiveness, feedback, and uncertainty. Firstly, by conducting a literature review and causal discovery analysis, the CLD became more *comprehensive* as several links were added and a few were changed or omitted. Secondly, these changes to the CLD led to modifications in the *feedback* structure. Feedback loops are critical drivers for the nonlinear behavior of complex systems[11], and we found that feedback loops changed between the different steps of the triangulation process. For instance, the reinforcing feedback loop between *Perceived stress* and *Proinflammatory processes* (RD6) was removed after the literature review (Fig. 3), and the reinforcing feedback loop between *Body mass index* and *Perceived social support* (RD9) was introduced after causal discovery (Fig. 5). Finally, the CLD became more transparent regarding *uncertainty*, elucidating where experts, literature, and data agreed (i.e., thick and solid lines, Fig. 5), leading to greater confidence, and where they disagreed (i.e., thin or dotted lines, Fig. 5). For example, the link *Body mass index → Perceived social support* might warrant further empirical research, while *Perceived stress → Depressive symptoms* is more firmly established. This uncertainty regarding causal links can also be incorporated into a computational version of the CLD: a system dynamics model[34]. Such a model could

be used for model comparison with and without hypothetical links, enabling hypothesis testing and a more precise uncertainty quantification[10].

Although the CLD was constructed as a proof-of-concept for explorative purposes, it already provides preliminary insights into potential drivers of the course of depressive symptoms over time (the problem behavior; see Fig. 7), such as reinforcing feedback loops[35]. The CLD reveals short reinforcing feedback loops between *Depressive symptoms* and *Proinflammatory processes* (RD5), *Perceived stress* (RD3), *Loneliness* (RD4), and *Sleep disturbance* (RD4) (Fig. 5), illustrating potential self-reinforcing patterns. For example, heightened *Perceived stress* due to *Daily hassles* can lead to *Depressive symptoms*, activating loop RD3. Consequently, *Perceived stress* and *Depressive symptoms* can both cause *Sleep disturbances*, triggering loop RD4, which, in turn, can worsen (the experience of) *Daily hassles*, forming a longer feedback loop with the shorter loops nested within it (for the sake of clarity, we did not annotate these longer loops in Figs. 2, 3 and 5, see Supplementary Information D). Furthermore, the CLD also describes 'cross-scale' loops that involve multiple research domains. For instance, *Proinflammatory processes* can lead to *Loneliness*, linking back to *Depressive symptoms* via loop RD4 and to *Proinflammatory processes* via loop RI7. We also find biological pathways linking *Body Mass Index* to *Perceived social support* (RD9) and *Sleep disturbance* (RD1), which both impact *Prosocial behavior*. Interestingly, prosocial behavior is increasingly identified as a major public health asset[36], which our CLD suggests can be strongly affected by multiple domains and might help identify intervention strategies to protect an individual's health.

Limitations of our approach arise from the limitations of the individual methods. Causal discovery is limited by the specific datasets used, which often contain biases arising from, e.g., a significant number of missing values (Table 1), possibly leading to selective loss to follow-up. In our case example,
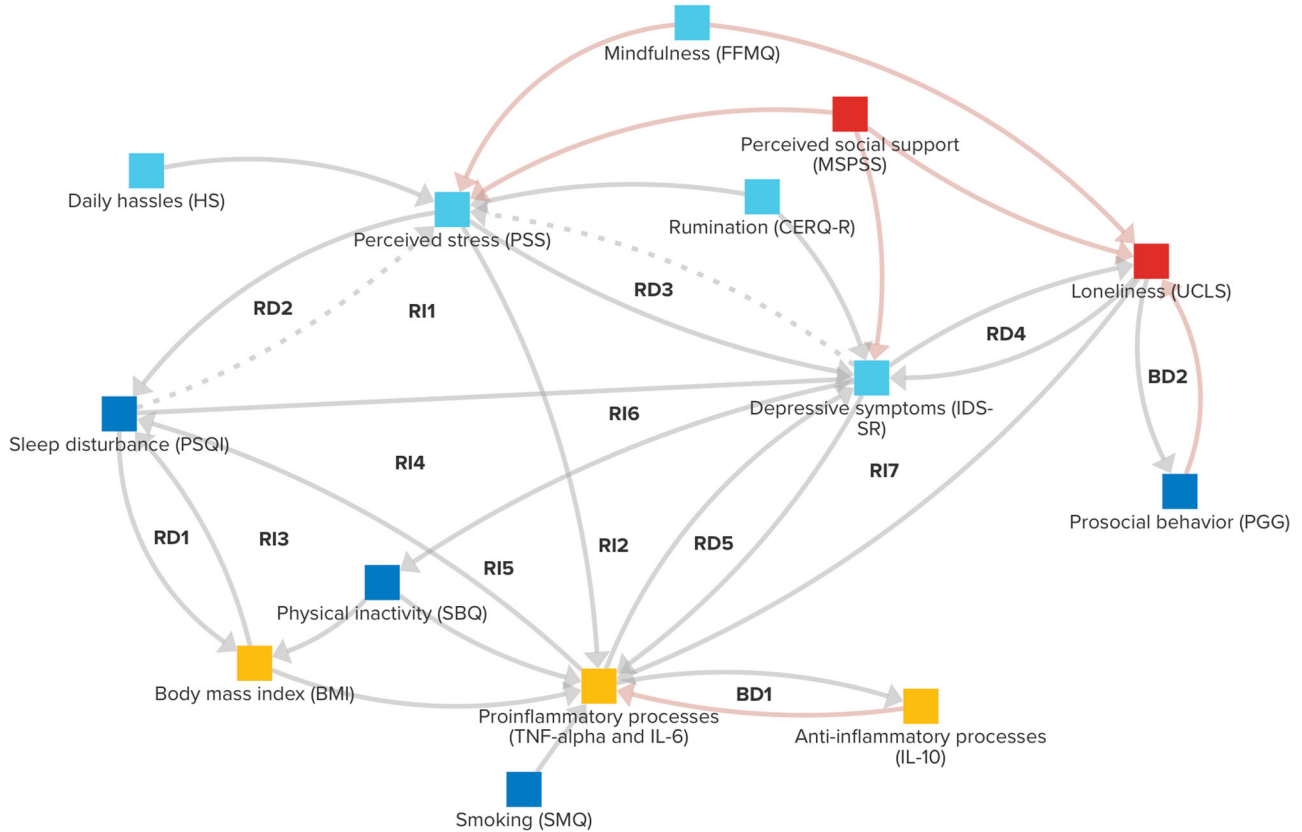
**Fig. 3 | Expert- and literature-driven causal loop diagram developed through group model building and literature review.** Reinforcing (R) and balancing (B) feedback loops consisting of two (RD, BD) or three variables (RI, BI) have been annotated (see Supplementary Information D for an overview). For dotted links, no supporting literature was found. Grey connections have positive polarity, whereas red connections have negative polarity. The variables' colors correspond to different dimensions: Social = Red; Behavioral = Dark blue; Psychological = Light blue; Biological = Yellow.
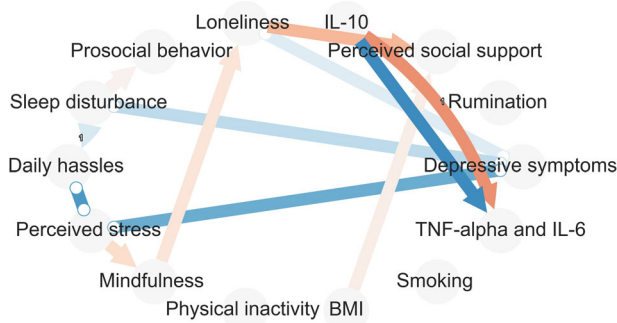


**Fig. 4 | The causal graph returned by our application of J-PCMCI+.** Straight links correspond to contemporaneous direct causal effects, and curved links correspond to lagged direct causal influences. In our analysis, all lagged links are at lag 1, corresponding to six months. The orange links correspond to a negative effect and the blue links to a positive effect. Lighter colored lines correspond to smaller absolute effects. The algorithm could not orient undirected links (o-o; for example, Daily hassles o-o Perceived stress).

the causal discovery results were indeed sensitive to removing participants with incomplete data from the analysis (Supplementary Information C). This dependency further highlights the need for triangulating across different methods, relying on not only data but also theory. J-PCMCI+ can, in principle, account for unmeasured confounders that are either constant at the subject level or constant over time[31]. However, we were unable to do so because the Healthy Brain Study includes only three measurements over the span of one year, limiting our ability to correct for context-related confounding beyond sex and education. Additionally, some variables may change at different rates, affecting which links can be found by the algorithm (Table 2). Future studies with longer or more frequent follow-ups may address these issues, though balancing the number of variables with the number of measurements remains challenging. Similarly, the Latent-PCMCI (LPCMCI) algorithm could be considered to account for unmeasured confounding[37], but this would also require larger sample sizes to obtain reliable results. On a positive note, our results *were* robust to changing the independence tests and the method's hyperparameter settings (Supplementary Information C), and all test statistics' polarities agreed with the GMB-based CLD link polarities, which adds credibility to the analysis.

Other limitations relate to the literature review. Although a systematic review of scientific literature for all possible links in the CLD would be optimal, this was beyond the scope of our case example and will be infeasible for most research projects (e.g., in our case example, this would have entailed $14 \times 13 = 182$ separate reviews). In future studies, artificial intelligence tools, e.g., based on natural language processing, might automate some of this work, facilitating a more comprehensive and systematic literature assessment. Considering the above, it should be noted that triangulation is a time-consuming process that introduces complexities in dealing with the different data and conflicting findings. Hence, achieving an optimal balance between efficiency and comprehensiveness is crucial[27], depending on the available resources.

Future work could involve alternative approaches to evidence triangulation. For instance, the process could be instantiated with causal discovery instead of GMB, with domain experts interpreting the results and adjusting the diagram according to their domain knowledge, which could then serve as constraints for further causal discovery. Data-driven methods could also be used for variable selection in such an approach. While we focused on triangulation in the mapping phase, triangulation could also

## Table 2 | Links identified by causal discovery (J-PCMCI+) using statistical significance level $pc_\alpha = 0.05$

| Origin | Type | Destination | Coeff | P-value | Lit | Add |
|---|---|---|---|---|---|---|
| **Indicated by group model building** | | | | | | |
| Anti-inflammatory processes | →t + 1 | Proinflammatory processes | −0.46 | 0.01 | Yes | – |
| Daily hassles | o-o | Perceived stress | +0.58 | $1 \times 10^{-5}$ | Yes | – |
| Sleep disturbance | o-o | Depressive symptoms | +0.26 | 0.01 | Yes | – |
| Depressive symptoms | o-o | Loneliness | +0.12 | 0.004 | Yes | – |
| Perceived stress | o-o | Depressive symptoms | +0.50 | 0.001 | Yes | – |
| **Not indicated by group model building** | | | | | | |
| Sleep disturbance | →t + 1 | Daily hassles | +0.16 | 0.02 | No | Yes |
| Anti-inflammatory processes | → | Proinflammatory processes | +0.63 | $6 \times 10^{-38}$ | Yes | * |
| Mindfulness | → | Loneliness | −0.13 | $3 \times 10^{-5}$ | Yes | Yes |
| Loneliness | → | Perceived social support | −0.33 | $2 \times 10^{-9}$ | No | No |
| Body mass index | → | Perceived social support | −0.46 | 0.008 | No | Yes |
| Perceived stress | → | Mindfulness | −0.18 | $2 \times 10^{-4}$ | No | Yes |
| Sleep disturbance | → | Prosocial behavior | −0.04 | 0.02 | yes | Yes |

The 'Type' refers to the direction of the links. The algorithm could not orient undirected links (o-o). The column 'coeff' provides standardized regression coefficients that measure the strength of the individual links (in units standard deviation of the 'Destination' variable divided by the standard deviation of the 'Origin' variable), with the '+' and '-' signs indicating their polarity. To calculate these coefficients, we performed standardized (linear) regressions of the respective effect variable (in the 'Destination' column) on all its parents according to the graph (including the cause variable in the 'Origin' column). For variables that formed feedback loops in the CLD, we assessed both directions and found the same polarities with Loneliness →+0.27 Depressive symptoms and Depressive symptoms →+0.41 Perceived stress (as compared to Depressive symptoms →+0.12 Loneliness and Perceived stress →+0.50 Depressive symptoms, as mentioned in the table). Links not indicated by GMB are annotated with 'yes' under 'Add' if they were added in the final group model building session. Links indicated by the literature review are annotated with 'yes' under 'Lit.' All links were contemporaneous except those indicated 't + 1,' which had a time lag of six months. *Given the positive polarity, the domain experts assumed J-PCMCI+ found the wrong direction for the link Proinflammatory processes → Anti-inflammatory processes.
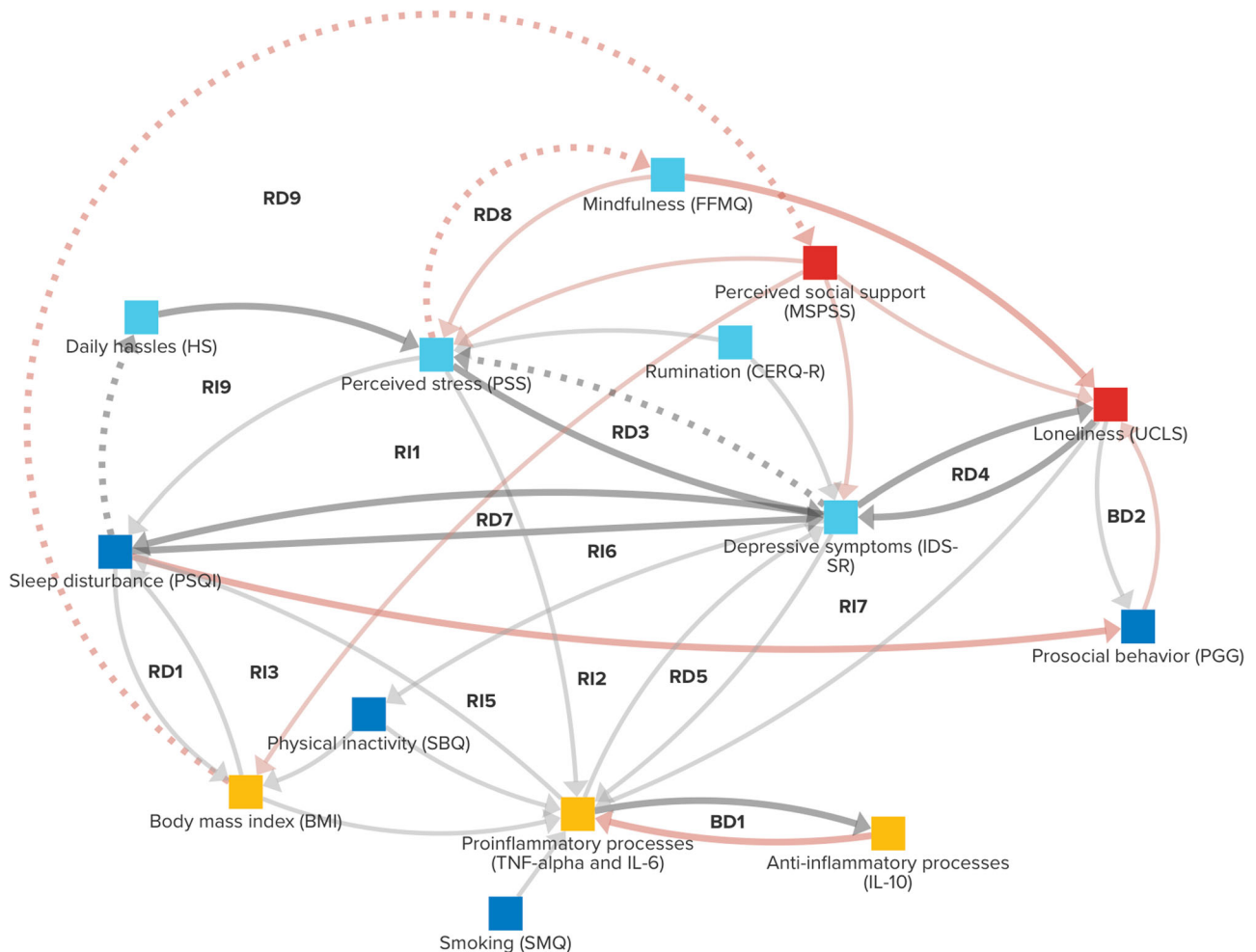


**Fig. 5 | Expert-, literature-, and data-driven causal loop diagram based on an evidence triangulation process using group model building, literature review, and causal discovery.** Reinforcing (R) and balancing (B) feedback loops consisting of two (RD, BD) or three variables (RI, BI) have been annotated (see Supplementary Information D for an overview). Bold links are supported by causal discovery. For dotted links, no supporting literature was found. Grey connections have positive polarity, whereas red connections have negative polarity. The variables' colors correspond to different dimensions: Social Red; Behavioral = Dark blue; Psychological = Light blue; Biological = Yellow.

Fig. 6 | Steps of our approach. Our approach to developing a causal loop diagram through evidence triangulation from group model building with domain experts, a review of scientific literature, and causal discovery based on empirical data.
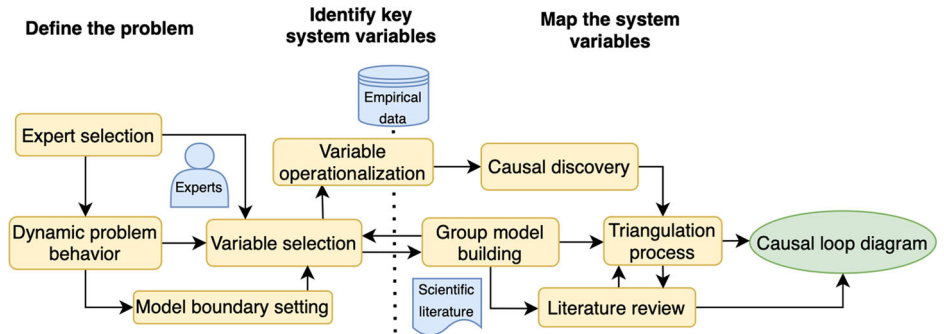


enhance the identification of key system variables[16]. For example, applying explainable machine learning to the Healthy Brain Study data could help identify variables with high predictive accuracy for depressive symptoms. The causal discovery could also be conducted on panel data from multiple studies, and alternative causal discovery methods could also be explored (e.g., TPC[23]). Lastly, the CLD we developed may be further extended. Future extensions of the CLD could broaden the scope of our case example to encompass more extended time frames and incorporate additional variables that were not yet considered relevant in the demographic of healthy 30-year-olds, such as hippocampal volume.

To conclude, the mixed-methods approach presented in this paper illustrates how evidence triangulation can contribute to constructing more comprehensive CLDs of complex problems with more robust feedback loops and transparency regarding uncertainties in the model's structure. Our case example combined insights from domain experts, literature, and empirical data, leading to several changes to the CLD during the process and offering a basis for future mental health research. The addition of the literature review and causal discovery helped the experts critically examine their understanding, making the triangulation process recommendable for future studies. As we move forward, the evolution of causal discovery, artificial intelligence, and platforms for interdisciplinary collaborations may help streamline and refine the evidence triangulation process. In turn, this promises to advance our ability to discern the causal networks underlying complex health problems, enabling the development of more valid computational models and intervention strategies with greater effectiveness.

## Methods

In CLDs, variables are presented with assumed causal links between them represented as directed arrows. These links can be positive (a change in X gives a change in Y in the same direction) or negative (a change in X gives a change in Y in the opposite direction). Together, the links may form feedback loops. When the number of links with negative polarity is even, these loops are reinforcing, leading to exponential growth or decline. When the number of links with negative polarity is odd, the loops are 'balancing,' leading to equilibrium.

Various modeling processes for developing CLDs are documented in the literature[11,38–40]. While these processes may vary in specifics, their fundamental steps are generally comparable. In this project, we employed a CLD development process specifically outlined for health research[40]. This approach consists of three developmental phases, as depicted in Fig. 6: 'Define the problem,' 'Identify key system variables,' and 'Map the system variables.' In the problem definition phase, we selected a group of domain experts and defined a relevant dynamic problem behavior to model and a model boundary. In the key system variables identification phase, we listed and operationalized an initial set of core variables necessary to explain the problem behavior. Finally, the system variables mapping phase involved connecting the variables in the system to explain the dynamic problem behavior of interest. Since multi-method approaches have already been introduced to address variable selection[16], we focus on the triangulation of

evidence in the system variables mapping phase. This study was not preregistered.

### Define the problem

**Experts selection.** Before starting a participatory group modeling process, a multi-disciplinary group of domain experts should first be selected to represent the relevant research domains associated with the research question. The experts should sufficiently cover these domains to promote beneficial debate in the group. In our case example, the expert group for the GMB process was involved in designing the Healthy Brain Study, ensuring the relevance of the measures used to operationalize CLD variables. The Healthy Brain Study is an ongoing project of the Radboud University, the Netherlands, rooted in team science[30]. The study collected data on a population-based sample comprising 1000 healthy participants aged between 30 and 39. By design, the study seeks to address common limitations in brain research among young adults[30] and is, therefore, a good starting point for addressing our research question. In particular, the study takes a holistic approach, providing a wide range of biopsychosocial measures, including cognitive, behavioral, and physiological testing, bio-sampling, questionnaires (e.g., targeting social support and loneliness), and neuroimaging[30]. Furthermore, the study consists of three assessments (four-month intervals) over one year to gain insight into the within-participant variation of these different measures[30], which allows us to address dynamic behavior.

Our specific expert group was formed with the explicit purpose of developing a biopsychosocial health model focused on the longitudinal interplay between social functioning and systemic inflammation on mental health outcomes. The experts, who are all co-authors of this paper, covered the relevant social (MV, JV), behavioral (JV, WFA), psychological (ML, MV), and biological (WFA, ML) research domains.

**Dynamic problem behavior.** To provide a clear purpose for the GMB process, a relevant "dynamic," i.e., time-dependent, problem should be defined[11,41]. Such dynamic problems are commonly visualized using one or multiple reference modes. Reference modes are plots that depict typical problem behavior over time used to illustrate trends, cycles, or patterns in the data, helping researchers identify underlying dynamics and develop hypotheses about the causes of the observed behaviors[11,13,39]. For instance, if the system moves towards or oscillates around an equilibrium, balancing loops should be included in the CLD[40] (e.g., to study concussion[42]). In our case example, we aimed to map out critical biopsychosocial mechanisms underlying the onset and persistence of depressive symptoms in response to a stressor (e.g., daily hassles). Before starting the GMB process, the experts were involved in two one-hour online brainstorming sessions between March and June 2022, during which the reference mode in Fig. 7 was developed. This reference mode served as a guide for the GMB process. The reference mode shows two hypothetical scenarios. In one scenario, a high frequency and intensity period of daily hassles triggers several reinforcing feedback loops, e.g., involving stress, inflammation, and lifestyle factors. This leads to a rapid

**Fig. 7 | Reference mode.** Reference mode involving two hypothetical scenarios where depressive symptoms change over time as a function of a stressor (red).
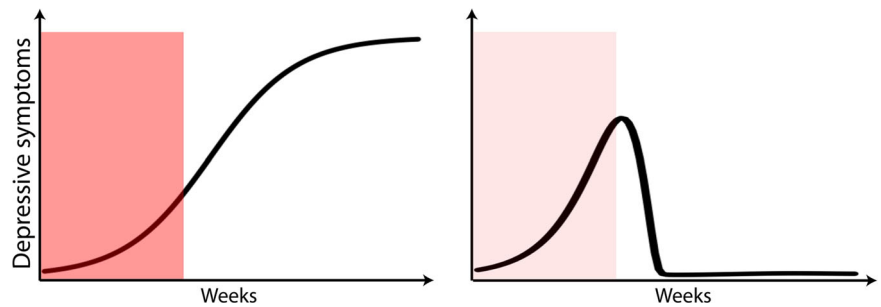


increase in depressive symptoms, which are self-limiting but are sustained over time after the daily hassles cease. In the other scenario, a less frequent and intense period of daily hassles also triggers these loops, but to a lesser degree. Consequently, once the daily hassles cease, balancing feedback loops become dominant and return the individual to a normal mood. Given that this reference mode exhibits balancing, equilibrium-seeking behavior, we focused on identifying both potentially balancing and reinforcing mechanisms.

**Model boundary setting.** Once the dynamic problem behavior is defined, a model boundary should be established around the problem. The model boundary includes a context of validity with a time frame of interest and the level of aggregation[15]. In our case example, the expert group also considered the model boundary in the two brainstorming sessions. The boundary was partly driven by the design of the Healthy Brain Study; namely, we assumed a context of healthy 30-year-olds and a total time frame of one year, matching the sample population and follow-up time. Additionally, set a boundary at the individual level, e.g., an individual's social functioning, but not, e.g., their partner's. Due to the expertise within our group, we primarily focused on the biological mechanisms related to inflammatory processes.

### Identify key system variables
**Variable selection.** Once the model boundary is established, key variables should be listed that are important for explaining the dynamic problem behavior. A common approach is to involve the GMB experts in selecting the variables[11,38–40]. A standard method for eliciting variables in a GMB process is the nominal group technique[43,44], in which the experts individually identify relevant variables and then alternate proposing these variables while arguing for the variable's relevance.

In our case example, the domain experts individually reviewed relevant scientific literature in their respective domains to identify critical variables. We then applied the nominal group technique during two online sessions of one hour each between June and July 2022. A variable was only included if there was complete consensus by the experts. Although this approach may obscure some uncertainty regarding the CLD, consensus would also be expected to improve the reliability and reproducibility of the CLD, which is beneficial for its integration with causal discovery. JFU was the facilitator during the nominal group technique and GMB sessions. Although the experts could select any variable relevant to the dynamic problem behavior, they primarily focused on variables measured in the Healthy Brain Study. While a CLD should never be limited by measured variables in a specific study, for the case example it made sense to focus on the variables in the Healthy Brain Study, where the data collection was designed for the biopsychosocial context of human brain functioning among young adults[30]. Nevertheless, for most studies, we recommend conducting GMB first and only then relating the variables to real-world data, as this process might limit the experts to available measures.

Ultimately, a core selection of 20 variables was identified, out of which 14 were ultimately used in the GMB procedure. The experts distinguished

between dynamic variables and variables considered static and not yet relevant in the considered age group and time frame (e.g., cognitive functioning, health status, and hippocampal volume). Although incorporating these variables into the CLD was beyond the scope of the present case example, they could be included at a future stage, e.g., when considering adults later in life. Since we focus on young and healthy adults, we assumed these variables do not yet play an important role.

**Variable operationalization.** A critical step in complementing GMB with causal discovery is operationalizing the CLD variables based on quantitative data. To do this, the experts should consider which measures in the available data best represent the CLD variables. Longitudinal data are ideally used, but cross-sectional data may also suffice, although that will prevent the causal discovery from identifying feedback loops. In our case example, we used panel data (i.e., consisting of multiple individuals over multiple time points) from the Healthy Brain Study resource under release number 2023-1, with cytokine data analyzed by the Radboud Biobank[45]. For instance, we used experimental behavioral measures to assess prosocial behavior (i.e., the public goods game[46]), biosamples to assess inflammatory processes, and a validated questionnaire for loneliness[47]. As the Healthy Brain Study has only recently completed data collection and most of the data is still being analyzed, a specific cohort was defined once 300 participants completed the third and final assessment. By that point, 410 participants had participated in the study, each of whom was incorporated into the cohort. Complete follow-up was available for 73% of individuals and another 4% for the second but not the third assessment. The primary reasons for loss to follow-up were (in order of frequency) the perceived study-related burden, a lack of reachability, and pregnancy or the receipt of a medical diagnosis or treatment.

Not all measures in for the CLD variables were obtained for every participant at every assessment. Various factors contributed to this, including incomplete questionnaires and challenges in obtaining blood samples (e.g., due to concerns about fainting or difficulties in drawing blood samples). Regarding the specific HBS measures of interest, data points were available in the first assessment for 403 participants, with complete data for 189 participants. During the third assessment, 302 participants had measures, of which 98 had complete data. Only 51 participants had complete data across all assessments. Histograms and spaghetti plots are provided for the measures in Supplementary Information B.

### Map the system variables
The next phase is to map the hypothesized causal links between the core variables identified during the 'Identify key system variables' phase, thereby identifying the dynamic network of causation underlying our problem behavior. In our approach, this phase involves four steps. First, a CLD can be developed with the expert group. Second, the experts review scientific literature for each causal link they identify. Third, empirical data (preferably longitudinal) are used to identify a causal model with causal discovery. Fourthly, these different sources of evidence should be combined into a single causal loop diagram in a triangulation process with the experts. These

steps are explained below in further detail and in relation to the case example.

**Group model building.** GMB is a participatory method for developing systems models, including CLDs. During GMB sessions, a group of expert participants engages in facilitated discussions to align their perspectives and develop a consensus-based model that is more comprehensive than each expert's mental model. Although CLDs can also be developed from individual expert interviews, GMB produces more comprehensive and coherent CLDs that better explain the system's dynamic behavior[48].

After selecting a core variable list in our case example using the nominal group technique, we conducted seven online GMB sessions with the domain experts between September 2022 and January 2023 to map proposed causal links between the variables. Each session lasted one hour, and at least three out of four experts were present during every session. We also had two additional one-hour discussion meetings where only two experts were present. Between sessions, summary reports were sent to all experts. After each session, the expert who proposed a new link performed a literature review to scrutinize it.

**Literature review.** Based on the proposed causal links in the GMB, each expert could be asked to conduct a scoping review[49] to scrutinize the available evidence for each link belonging to their expertise and provide at least one reference to supporting evidence, if available. During the review, inferring causality from a body of evidence is a nuanced process that lacks a single best prescription[4]. However, suggested indicators of causality include experimental evidence, temporality (i.e., the exposure preceding the outcome in time), robustness across studies, and plausible intermediary mechanisms consistent with the experts' knowledge of the system[10,50,51]. In our case example, the experts conducted reviews after each GMB session for the causal links proposed. Although these reviews were not fully systematic in this explorative project, the experts looked for longitudinal associations that were ideally robust across multiple studies. If no literature supported the link, the group could either remove it if they had lost confidence in it or keep it in the CLD but consider it 'hypothetical.'

**Causal discovery.** In parallel to formulating an initial CLD based on expert knowledge and scientific literature, we propose conducting causal discovery whenever sufficient data are available. Ideally, these data are longitudinal, allowing for the identification of time-dependent feedback loops, and comprehensive, like the Healthy Brain Study, where most variables are measured within a single data set so that they can be included in the analysis. Although different types of algorithms exist for causal discovery[18], constraint-based algorithms are particularly advantageous due to their flexibility. First, they utilize tests to identify conditional independencies in the data, which can suggest the absence of causal relationships. Moreover, constraint-based algorithms can be applied with a variety of conditional independence tests that are able to deal with different data dependencies (e.g., nonlinearity) as well as data types. In the present study, we had to account for mixed-type data containing both discrete and continuous variables. Second, our study involved multiple time series data sets, and the constraint-based algorithm we utilized (J-PCMCI+, see below) allowed us to handle such instances. We are unaware of a score-based method to deal with multiple time series datasets for mixed-type data. This makes constraint-based algorithms useful for our demands. Moreover, constraint-based causal discovery can be useful even with limited data, as they can exclude certain causal links by detecting these independencies. However, it is important to carefully consider the potential impact of omitted variables, as they can lead to spurious links and incorrect conclusions.

After formulating the initial CLD in our case example, we performed causal discovery on the longitudinal data from the Healthy Brain Study using a simplified version of the J-PCMCI+ algorithm[31], which is implemented in the Python package Tigramite version 5.2 (www.github.com/jakobrunge/tigramite). The J-PCMCI+ algorithm is an extension of PCMCI+[52], which, in turn, is a time-series adaptation of the widely applied PC algorithm[18]. Like PC and PCMCI+, J-PCMCI+ is a constraint-based causal discovery algorithm that performs a sequence of statistical independence tests to infer a undirected graph whose (yet undirected) links represent direct causal influences. Then, the algorithm uses the test results and several graphical orientation rules to establish the directionality of these links, that is, to tell apart cause and effect. However, some of the links might remain undirected. Despite its assumption of acyclicity, incorporating time-dependent variables permits the identification of feedback loops within the system, provided these feedback loops do not act on a time scale smaller than the temporal resolution. The algorithm can detect and orient both time-lagged and non-time-lagged (so-called "contemporaneous") causal links, where the contemporaneous links are assumed to define an acyclic graph and some of the contemporaneous links might remain undirected. Given that the data has only three assessments per individual and that J-PCMCI+ employs a sliding-window approach to generate samples for independence testing, statistically valid tests in JPCMCI+ are possible for, at most, a single time lag. For more details on the algorithm, see Supplementary Information A.

The specific feature of J-PCMCI+ is that it can utilize data from different so-called contexts (here, these are the individual participants) to remove the confounding effect of so-called context variables[31]. These are variables that are (a) exogenous to the system and (b) constant in time or constant across all contexts (in our case, constant across the participants). In our case example, we included education and sex as context variables. By using a dummy-variables approach, J-PCMCI+ can, in principle, also handle unobserved context variables[31]. However, in our case example, we did not use this dummy-variables approach for the following reasons: First, to remove the effect of unobserved context-related confounders, a sufficient number of time steps is needed. Since the Healthy Brain Study has only three assessments, this is not the case in our study. Second, time-confounding was deemed unlikely since all measurements were taken within one year. Notwithstanding, the inclusion of education and sex as context variables accounts for at least part of the context-related confounding.

Like all causal discovery algorithms, J-PCMCI+ operates under specific assumptions. For J-PCMCI+, these are the causal Markov and causal faithfulness conditions (see, e.g., Spirtes et al.[21] and Peters et al.[53]), acyclicity of the contemporaneous links (as already mentioned above), stability of the causal structure within the timeframe of the data and across individuals, and the absence of unobserved confounders (except confounders that are constant across time or contexts when dummy variables are used, see above). For independence testing, we used a regression-based independence test for datasets with categorical (i.e., education, sex) and continuous (others) variables with linear dependencies, as implemented in the Tigramite package (RegressionCI). This test is asymptotically equivalent to the test for mixed data suggested in Section 2 of Tsagris et al.[54], making the assumption that the respective requirements of the employed regression methods (linear or logistic regression) are met (e.g., independent and homoscedastic Gaussian residuals). However, the regression methods are, to a certain degree, robust against violations of these assumptions[54,55]. We used a significance level of $pc_\alpha = 0.05$ for the independence tests in the main analysis.

To quantify the strength of the identified links, we then fit linear models of each variable (except context variables) on their parents as estimated with J-PCMCI+ (more details in Runge et al.[56]). To do this, we assumed that the undirected links (i.e., o-o) had the same direction as in the CLD developed by the domain experts and, in case of a feedback loop, we assessed both directions. Finally, we used the sign of the estimated effects to infer the polarity of the links for the CLD (+ or –).

The J-PCMCI+ algorithm consistently handles missing values for each participant by using only those sliding-window samples with complete data. For instance, say a participant has data for all assessments for some variable 'A' but only the first assessment for another variable 'B.' Then, when testing for independence of $A_t$ and $B_t$, the algorithm will

only draw a single sample from that specific participant (namely, from the first assessment), as opposed to three samples from a second participant who has all assessments for both A and B. Moreover, when testing for independence of $A_{t+1}$ and $A_t$, the algorithm will draw two samples from both these participants.

**Triangulation process.** Once all sources of evidence have been collected, we suggest combining them through a triangulation process. In our case example, we adopted the approach that domain experts ultimately decide which links are included to develop a consensus-based diagram[10]. After obtaining the causal diagram through causal discovery, we thus started a triangulation process with the domain experts to merge it with the expert- and literature-based CLD. We conducted a final two-hour GMB session in September 2023, during which the causal discovery results were discussed with the experts. If an expert considered a link introduced by causal discovery potentially plausible, they were made responsible for the literature review and reported their findings back to the group later.

## Data availability
In the future, the full Healthy Brain Study dataset can be requested by researchers around the world. Please visit the Healthy Brain Study website for more information and any updates on the release. For questions and inquiries, please contact hbs-data@radboudumc.nl.

## Code availability
The J-PCMCI+ algorithm is implemented in the Python package 'Tigramite' (https://github.com/jakobrunge/tigramite). The specific code used to preprocess the data and run the causal discovery can be made available upon reasonable request from the corresponding author.

## References
1. Pearl, J. *Causality*. (Cambridge University Press, 2009).
2. Hernán, M. A. & Robins, J. M. *Causal Inference: What If*. (Chapman & Hall/CRC, 2020).
3. Vandenbroucke, J. P., Broadbent, A. & Pearce, N. Causality and causal inference in epidemiology: the need for a pluralistic approach. *Int. J. Epidemiol.* **45**, 1776–1786 (2016).
4. Krieger, N. & Davey Smith, G. The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. *Int. J. Epidemiol.* **45**, 1787–1808 (2016).
5. Galea, S., Riddle, M. & Kaplan, G. A. Causal thinking and complex system approaches in epidemiology. *Int. J. Epidemiol.* **39**, 97–106 (2010).
6. Rod, N. H. et al. Complexity in Epidemiology and Public Health: Studying and acting on complex health problems through a mix of epidemiological methods and data. *Epidemiology* **34**, 505–514 (2023).
7. Rutter, H. et al. The need for a complex systems model of evidence for public health. *Lancet* **390**, 2602–2604 (2017).
8. Marshall, B. D. L. & Galea, S. Formalizing the role of agent-based modeling in causal inference and epidemiology. *Am. J. Epidemiol.* **181**, 92–99 (2015).
9. Uleman, J. F. et al. Mapping the multicausality of Alzheimer's disease through group model building. *GeroScience* 1–15 https://doi.org/10.1007/s11357-020-00228-7 (2020).
10. Crielaard, L. et al. Refining the causal loop diagram: a tutorial for maximizing the contribution of domain expertise in computational system dynamics modeling. *Psychol. Methods* **29**, 169 (2022).
11. Sterman, J. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. (Irwin/McGraw-Hill, 2000).
12. Uleman, J. F., Melis, R. J., Hoekstra, A. G., Rikkert, M. G. O. & Quax, R. Exploring the potential impact of multi-factor precision interventions in Alzheimer's disease with system dynamics. *J. Biomed. Inform.* **145**, 104462 (2023).
13. Vennix, J. A. M. *Group Model Building: Facilitating Team Learning Using System Dynamics Field*. (Wiley, 1996).
14. Andersen, D. F., Vennix, J. A., Richardson, G. P. & Rouwette, E. A. Group model building: problem structuring, policy simulation and decision support. *J. Oper. Res. Soc.* **58**, 691–694 (2007).
15. Cassidy, R. et al. How to do (or not to do)…using causal loop diagrams for health system research in low and middle-income settings. *Health Policy Plan* **37**, 1328–1336 (2022).
16. Dhirasasna, N. & Sahin, O. A multi-methodology approach to creating a causal loop diagram. *Systems* **7**, 42 (2019).
17. Spirtes, P. & Zhang, K. Causal discovery and inference: concepts and recent methodological advances. in *Applied Informatics* **3** (Springer, 2016).
18. Glymour, C., Zhang, K. & Spirtes, P. Review of causal discovery methods based on graphical models. *Front. Genet.* **10**, 524 (2019).
19. Granger, C. W. J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438 (1969).
20. Camps-Valls, G. et al. Discovering causal relations and equations from data. *Phys. Rep.* **1044**, 1–68 (2023).
21. Spirtes, P., Glymour, C. N. & Scheines, R. *Causation, Prediction, and Search*. (MIT Press, 2000).
22. Runge, J., Nowack, P., Kretschmer, M., Flaxman, S. & Sejdinovic, D. Detecting and quantifying causal associations in large nonlinear time series datasets. *Sci. Adv.* **5**, eaau4996 (2019).
23. Petersen, A. H., Osler, M. & Ekstrøm, C. T. Data-driven model building for life-course epidemiology. *Am. J. Epidemiol.* **190**, 1898–1907 (2021).
24. Runge, J., Gerhardus, A., Varando, G., Eyring, V. & Camps-Valls, G. Causal inference for time series. *Nat. Rev. Earth Environ.* **4**, 487–505 (2023).
25. Petersen, A. H., Ekstrøm, C. T., Spirtes, P. & Osler, M. Constructing causal life course models: Comparative study of data-driven and theory-driven approaches. *Am. J. Epidemiol.* **192**, kwad144, https://doi.org/10.1093/aje/kwad144 (2023).
26. Mäkelä, J. et al. Technical note: Incorporating expert domain knowledge into causal structure discovery workflows. *Biogeosciences* **19**, 2095–2099 (2022).
27. Denzin, N. The fundamentals. *Introd. Triangulation Rugg Ed. UNAIDS Monit. Eval. Fundam. Geneva Switz.* **12** (2010). https://www.shortcutstv.com/wp-content/uploads/2021/01/Introducing-Triangulation.pdf.
28. Lawlor, D. A., Tilling, K. & Davey Smith, G. Triangulation in aetiological epidemiology. *Int. J. Epidemiol.* **45**, 1866–1886 (2016).
29. Hammerton, G. & Munafò, M. R. Causal inference with observational data: the need for triangulation of evidence. *Psychol. Med.* **51**, 563–578 (2021).
30. Consortium, H. B. S. et al. Protocol of the Healthy Brain Study: An accessible resource for understanding the human brain and how it dynamically and individually operates in its bio-social context. *PLOS ONE* **16**, e0260952 (2021).
31. Günther, W., Ninad, U. & Runge, J. Causal Discovery for time series from multiple datasets with latent contexts. in *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence* 766–776 (PMLR, 2023).
32. Andersen, D. F., Richardson, G. P. & Vennix, J. A. M. Group model building: adding more science to the craft. *Syst. Dyn. Rev.* **13**, 187–201 (1997).
33. Petersen, A. H., Ekstrøm, C. T., Spirtes, P. & Osler, M. Causal discovery and epidemiology: A potential for synergy. *Am. J. Epidemiol.* **101** https://doi.org/10.1093/aje/kwae101 (2024).
34. Uleman, J. F. et al. Simulating the multicausality of Alzheimer's disease with system dynamics. *Alzheimers Dement* **19**, 2633–2654, https://doi.org/10.1002/alz.12923 (2023).

35. Wittenborn, A. K., Rahmandad, H., Rick, J. & Hosseinichimeh, N. Depression as a systemic syndrome: mapping the feedback loops of major depressive disorder. *Psychol Med* **46**, 551–562 (2016).

36. Kubzansky, L. D., Epel, E. S. & Davidson, R. J. Prosociality should be a public health priority. *Nat. Hum. Behav.* 1–3 https://doi.org/10.1038/s41562-023-01717-3 (2023).

37. Gerhardus, A. & Runge, J. High-recall causal discovery for autocorrelated time series with latent confounders. in *Advances in Neural Information Processing Systems* **33** 12615–12625 (Curran Associates, Inc., 2020).

38. Haraldsson, H. V. *Introduction to System Thinking and Causal Loop Diagrams*. (2004).

39. Hovmand, P. S. *Community Based System Dynamics*. (Springer Science & Business Media, 2013).

40. Uleman, J. F., Stronks, K., Rutter, H., Arah, O. A. & Rod, N. H. Mapping complex public health problems with causal loop diagrams. *Int. J. Epidemiol.* **53**, dyae091 (2024).

41. Sterman, J. D. All models are wrong: reflections on becoming a systems scientist. *Syst. Dyn. Rev.* **18**, 501–531 (2002).

42. Kenzie, E. S. et al. The dynamics of concussion: mapping pathophysiology, persistence, and recovery with causal-loop diagramming. *Front. Neurol.* **9**, 203 (2018).

43. Gustafson, D. H., Delbecq, A. L. & Van de Ven, A. H. Group techniques for program planning-a guide to nominal group and Delphi processes. *Group Organ. Stud.* **1**, 256–256 (1986).

44. Hovmand, P. S. et al. Group model-building 'Scripts' as a collaborative planning tool. *Syst. Res. Behav. Sci.* **29**, 179–193 (2012).

45. Manders, P., Lutomski, J. E., Smit, C., Swinkels, D. W. & Zielhuis, G. A. The Radboud Biobank: A central facility for disease-based biobanks to optimise use and distribution of biomaterial for scientific research in the Radboud University Medical Center. *Nijmegen* **5**, 2 (2018).

46. Volk, S., Thöni, C. & Ruigrok, W. Temporal stability and psychological foundations of cooperation preferences. *J. Econ. Behav. Organ.* **81**, 664–676 (2012).

47. Russell, D., Peplau, L. A. & Cutrona, C. E. The revised UCLA Loneliness Scale: Concurrent and discriminant validity evidence. *J. Pers. Soc. Psychol.* **39**, 472–480 (1980).

48. Valcourt, N., Walters, J., Javernick-Will, A. & Linden, K. Assessing the efficacy of group model building workshops in an applied setting through purposive text analysis. *Syst. Dyn. Rev.* **36**, 135–157 (2020).

49. Munn, Z. et al. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med. Res. Methodol.* **18**, 143 (2018).

50. Rothman, K. J. & Greenland, S. Causation and causal inference in epidemiology. *Am J Public Health* **95**, S144–S150 (2005).

51. Hill, A. B. The environment and disease: association or causation? *Proc. R. Soc. Med.* **58**, 295–300 (1965).

52. Runge, J. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. in *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)* 1388–1397 (PMLR, 2020).

53. Peters, J., Janzing, D. & Schölkopf, B. *Elements of Causal Inference: Foundations and Learning Algorithms*. (The MIT Press, 2017).

54. Tsagris, M., Borboudakis, G., Lagani, V. & Tsamardinos, I. Constraint-based causal discovery with mixed data. *Int. J. Data Sci. Anal.* **6**, 19–30 (2018).

55. Lumley, T., Diehr, P., Emerson, S. & Chen, L. The importance of the normality assumption in large public health data sets. *Annu. Rev. Public Health* **23**, 151–169 (2002).

56. Runge, J. et al. Identifying causal gateways and mediators in complex spatio-temporal systems. *Nat. Commun.* **6**, 8502 (2015).

57. World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA* **310**, 2191–2194 (2013).

## Author contributions
J.F.U., N.H.R., J.V., M.L., M.V., and W.F.A. contributed to the study conception and design. J.V., M.L., M.V., and W.F.A. were domain experts in the group model-building process and conducted literature reviews. J.F.U. facilitated the group model-building sessions and performed the data analysis. A.G. and J.R. supported the causal discovery analysis and its interpretation from a methodological point of view. J.F.U. wrote the first draft of the manuscript, and all authors commented on several versions of the manuscript. All authors read and approved the final manuscript.

## Competing interests
The authors declare no competing interests.

## Ethics approval
The Institutional Review Board of Radboud University Medical Center approved the HBS on the 23rd of May, 2019 (reference number: 2018–4894) in accordance with the latest revision of the Declaration of Helsinki[57].

## Consent
Participants were contacted via phone and invited to a face-to-face information meeting on the Radboud campus. Participants provided written informed consent at this meeting before participation.

## Additional information
**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s44260-024-00017-9.

**Correspondence** and requests for materials should be addressed to Jeroen F. Uleman.

**Reprints and permissions information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.