**Investigating mechanisms of change in the GLA:D Back program**

Bernard Liew, University of Essex, UK   
+ GLAD Back Team (Jan, Per [WP2], Inge [WP2])

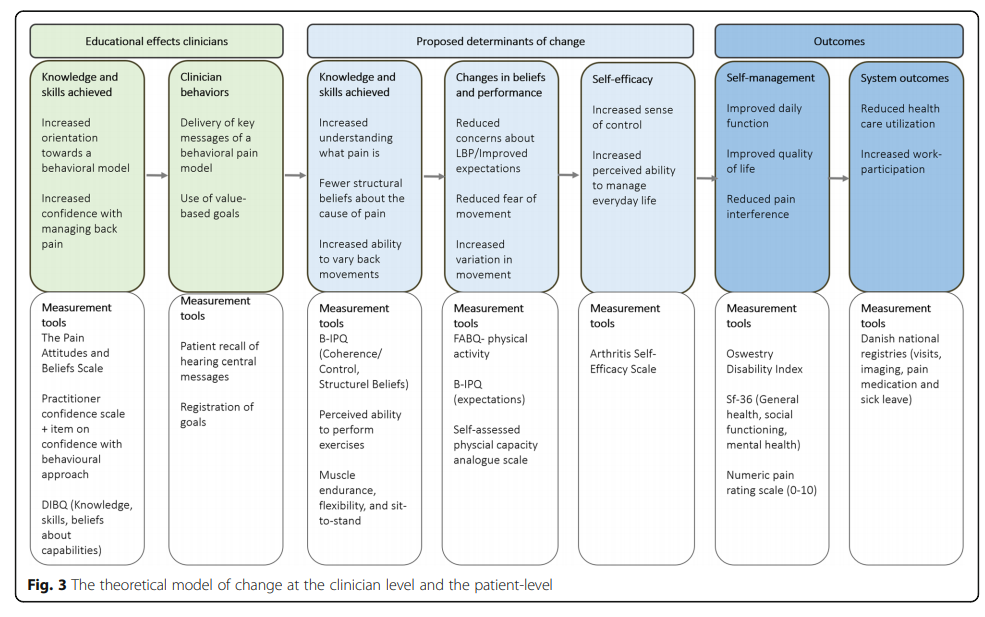


Figure 1. Theoretical model underpininig the GLAD program.

# Research question

1. To understand if clusters of homogenous individiauls with low back pain LBP can be derived from baseline demographic, pain, psychological, and physical characteristics (work package [WP] 1).
2. To verify if the mechanism of effect of the GLAD program follows the hypothesized pathway of change (see above) (WP1).
3. In the circumstance where the hypothesized variables did not mediate the outcome, the secondary aim is to propose alternative pathway of change effects (WP2).

# Methods

The study includes all patients enrolled in the registry at least 12 months prior to extracting the data set. If ongoing work about enrolment patterns suggests that there is an “upstart phase” in clinics with a different patient profile, we might exclude patients from that phase.

## WP1 – part 1

### Exploratory

To determine the optimal number of clusters of homogenous individiauls with low back pain LBP derived from baseline demographic, pain, psychological, and physical characteristics. The primary purpose of this analysis is to inform subsequent work packages of the potential for sub-group analysis. For example, if there exist different clinical sub-groups (e.g. high pain-low fear vs low pain-high fear), then it is predicted that different causal mechanisms may mediate the effect of the GLAD program on disability. In this instance, the effect of *clusters* could either be added into the analysis of work package 2 and 3 as a moderating variable (e.g. in work package two), or separate analyses could be performed on each cluster, sample size permitting (e.g. performing separate mediation analysis on participants of each cluster).

### Variables

1. Fear
2. Illness perception
3. Self-efficacy
4. Physical performance
5. Disability
6. Pain characteristics
7. Demographic characteristics
8. General health and comorbidities
9. Clinical location

All statistical analyses will be performed in R software. Hierachical clustering will be used to determine homogenous clinical subgroups using the previously defined baseline variables. The optimal number of clusters will be determined using a statistical threshold combined with expert clinical opinion (1). The statistical determination of the optimal number of clusters will be done using the elbow method – which looks at the total within-cluster sum of squares (WSS) as a function of the number of clusters. One should choose a number of clusters so that adding another cluster doesn’t improve much better the total WSS. In addition, an expert judgement of the meaningfulness of the resultant clusters will be made based on several factors e.g.: 1) if the clusters seems “clinically/substantially” different, and 2) if the clusters have a sufficiently large number of individuals.

## WP1 – part2

### Hypotheses

The variables - fear, illness perception, self-efficacy and physical capacity will mediate the relationship between the independent variables (patient knowledge and skills) and the dependent variables of (disability or pain).

### Independent variables ()

1. (baseline to 3m)
2. (baseline to 3m)

### Mediator variables ():

1. (3m to 6m)
2. (3m to 6m)
3. (3m to 6m)
4. performance (3m to 6m)

### Outcome variables ()

1. (6m to 12m)
2. (6m to 12m)

### Structural equation modelling (SEM)

Structural equations modelling (SEM) will be used to test a hypothesised causal chain (Figure 1) whereby participating in the GLAD program leads to changes in potential mediating variables causing change scores in the study outcomes. Multi-group SEM will be conducted with the group being the clusters identified in WP1-part 1 (Figure 2). First, a SEM model will be created where the path coefficients will be constrained to be equal across all participants. Second, another SEM model will be created whereby all path coefficients will be allowed to vary across each clinical subgroup. Given that both these models are nested, an Analysis of Variance will be conducted. A significant differences between models indicate that some paths in the SEM model vary between clinical subgroups.

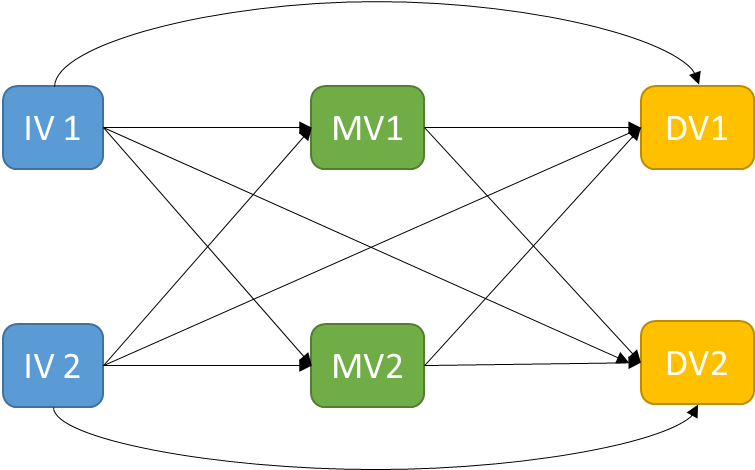


Figure 2. Basic structural model with only two independent, mediating, and dependent variables. The model will be expanded to use the number of mediating and dependent variables present in the study.

SEM will be performed using the *lavaan* package in R. Model fit will be assessed using 4 goodness of fit indices: p-value of chi-square, root mean square error of approximation (RMSEA) with 90% confidence interval, Adjusted Goodness of Fit Index (AGFI), and Bentler Comparative Fit Index (BCFI). An acceptable model fit is expected to reach p>0.05, RMSEA<0.05, AGFI≥0.90 and BCFI≥0.95.

## Work package 2

### Exploratory

If the hypothesized structural model in WP 1 does not provide an adequate model fit, a data-driven statistical approach to learn the multi-variate structural relationships will be undertaken.

### Variables included in the Bayesian Network

1. (baseline to 3m)
2. (baseline to 3m)
3. (3m to 6m)
4. (3m to 6m)
5. (3m to 6m)
6. performance (3m to 6m)
7. (6m to 12m)
8. (6m to 12m)

### Bayesian network analysis.

All analyses will be performed in R software(3) using the “bnlearn” package (4). BN is a graphical modelling technique (5) used increasingly in the health sciences to understand causal relationships. BN can handle some missing data (6), which makes them practical in longitudinal studies where data sets are often incomplete. BN quantifies the relationships among a set of variables X = {*X1*, …, *XN*}, where *N* is the number of different variables, using a directed acyclic graph (DAG). Each variable is associated with a node and directed arcs represent conditional dependencies between pairs of nodes. Building a BN model using a data-driven approach involves two stages: 1) structural learning - identifying which arcs are present in the DAG; and 2) parameter learning - estimating the parameters that regulate the strength and the sign of the corresponding relationships.

BN can easily include prior knowledge, sourced from the literature and experts, during the model building process. In the BN framework, prior knowledge can be included in the model as blacklist and whitelist arcs. Blacklist arcs are those which contravene known biological/physical mechanisms. We will only blacklist arcs that point backwards in time in the present study.

We will make use of model averaging to reduce the potential of including spurious relationships in the BN, using bootstrap resampling (*B = 200*) and performing structure learning on each of the resulting sample using Structural Expectation-Maximization (EM) (6). Structural EM is a technique which can build BN models in the presence of missing data (6). It does so by building an initial empty BN model using the original complete data, using it to impute missing data, rebuilding the BN model using the imputed complete data, and repeating this sequence until convergence. We will compute an “average” consensus DAG by selecting those arcs that have a frequency of > 50% in the bootstrapped samples, to create a sparse and interpretable network (3).

The structure of the BN models will be used for subsequent SEM analysis (). For each clinical subgroup, the theoretical model () will be compared against the to see if an alternative model better explains the data. For each subgroup, the following tests will be undertaken to determine which models best fit the data (Merkle et al., 2016):

1. Global equivalence of two models will be assessed using the *net* function (Bentler and Satorra, 2010) within the *semTools* package. Models are considered equivalent if the difference between their chi-square fit statistics is less than 0.0001. Two models are considered equivalent if they provide the same fits to the sample data as well as to the population data.
2. If models are not found to be equivalent, Vuong’s omega tests will be undertaken (Vuong, 1989) to test if two models are distinguishable from one another. Two models are distinguishable if they provide the same fit to a population of interest, but not necessarily to an observed sample.
3. If the omega test is significant (P < 0.05), indicating distinguishability, it is followed by Vuong’s closeness test (a z-test of the difference in model-predicted probabilities) to test for differences in the fit of the two models. The closeness test simultaneously quantifies if model 1 provides a better fit to the data than model 2, and vice-versa, indicative by a significance level of < 0.05.

# References

1. Kent P, Jensen RK, Kongsted A. A comparison of three clustering methods for finding subgroups in MRI, SMS or clinical data: SPSS TwoStep Cluster analysis, Latent Gold and SNOB. BMC Med Res Methodol. 2014;14:113.

2. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R Package for Causal Mediation Analysis. J Stat Softw. 2014;59(5):38.

3. Scutari M, Nagarajan R. Identifying significant edges in graphical models of molecular networks. Artif Intell Med. 2013;57(3):207-17.

4. Scutari M. Learning Bayesian Networks with the bnlearn R Package. J Stat Softw. 2010;35(3):22.

5. Nagarajan R, Scutari M, Lèbre S. Bayesian networks in R with applications in systems biology. Verlag, New York: Springer; 2013.

6. Friedman N. Learning Belief Networks in the Presence of Missing Values and Hidden Variables. Proceedings of the Fourteenth International Conference on Machine Learning. 657145: Morgan Kaufmann Publishers Inc.; 1997. p. 125-33.