
Supplementary Material to “Thinning Measurement Models and Questionnaire Design”

Ricardo Silva

Department of Statistical Science
University College London
Gower Street, London WC1E 6BT
ricardo@stats.ucl.ac.uk

Abstract

We provide a detailed description of the preprocessing done to the dataset *National Health Service National Staff Survey*, which we had to omit from the main paper due to lack of space.

1 Variable Selection and Binarization

The *National Health Service National Staff Survey* contains 206 questions divided into 45 sections. A PDF copy of the questionnaire, and the data itself, can be downloaded from the *Economic and Social Data Service* at <https://www.esds.ac.uk/findingData/snDescription.asp?sn=6570>. Restrictions might apply concerning permission to download the data. The documentation is freely accessible.

The aims of this survey are stated in the first page of the questionnaire:

This is an independent survey of your experience of working at your Trust. The overall aim is to gather information that will help to improve the working lives of NHS staff and so provide better care for patients.

Given this data, we design a measurement model by postulating one latent variable per section in the questionnaire. For instance, Section 24 has the header “Please indicate how frequently you feel this way about your job?”, followed by three items

- I look forward to going to work.
- I am enthusiastic about my job.
- Time passes quickly when I am working.

to which the answer should be in a Likert scale “Never”, “Rarely”, “Sometimes”, “Often” and “Always”.

In this case, a new latent variable X_{24} was created with these three responses as indicators. This is a common practice in structural equation models (Bollen, 1989). We did not attempt to optimize the structure to allow indicators to measure several latent variables.

Other questions have “Yes” \ “No” answers. A few are qualitative (e.g., “What is your ethnic background?”). Qualitative questions were excluded in our preprocessing. Likert variables were binarized by setting to 1 all values equal to or above the central value (that is, in example above we have {“Sometimes”, “Often”, “Always”} coded as 1), and the rest to zero. Some questions had a “Don’t know” as an option. In this case, “Don’t know” was given a value of zero.

Finally, three extra steps: we excluded all variables (after binarization) in which 95% of more of the responses were either 0 or 1. We excluded those with more than 50% of missing responses. After this exclusion, some latent variables end up with a single indicator as a child. We excluded these latent variables (and their indicators) on the grounds that it is not possible to identify the parameters associated with them (Bollen, 1989). After this preprocessing, we ended up with 126 observed variables and 28 latent variables.

2 Model Fitting and Further Postprocessing

Given the model structure specified in the previous section, we perform (Bayesian) model fitting by using the posterior expected value estimator of the parameters. Non-zero coefficients λ_{ij} linking indicator Y_i to latent variable X_j were given independent Gaussian $\mathcal{N}(0, 3)$ priors. The latent covariance matrix Σ was given an inverse Wishart with parameters $(28, 28\mathbf{I})$. Gibbs sampling was used, estimates being computed using a run of 20,000 steps after a burn-in period of 1,000. To simplify inference, we used a subsample of 50,000 respondents only.

For simplicity, we treated missing values as if they were missing at random, although this is not necessarily the case: there are questions which are answered depending only on the response to other questions. We ignored this, since low response variables ($< 50\%$) would have been excluded in the first step to being with. For identifiability purposes, we also do the following: for each latent variable X_j , we choose one of its indicators Y_i and fix the correspond coefficient λ_{ij} to 1 (Bollen, 1989). After calculating the estimates, we rescale the latent variable covariance matrix Σ so that all latent variables have variance of 1 (with the corresponding adjustment to the coefficients Λ_i): the motivation is to make the statistics $m_{\mathcal{F}}$ more meaningful, since we average over latent variables.

After calculating the estimates, we finalize the model building stage by excluding all indicators with a reliability index less than 0.2. Recall that we define the index of Y_i as $\Lambda_i^T \Sigma \Lambda_i^T / (\Lambda_i^T \Sigma \Lambda_i^T + 1)$ (the proportion of variance of \mathcal{Y}_i^* explained by the latent variables). The motivation is to make the problem harder for a fixed level K : otherwise, such variables are too obvious candidates to be removed.

3 Brief Explanation of Results

All 63 items we preselected are briefly listed in Appendix A. Below, we show the 13 items that were removed by the tree-structured method when we choose to keep 80% of the items.

- Group 4:
Health and safety training (e.g. fire training, manual handling)
How to handle confidential information about patients / service users
- Group 6:
My immediate manager can be counted on to help me with a difficult task at work
- Group 8:
In the last 12 months, as part of your KSF development review, appraisal, personal development review, did you agree a Personnel Development Plan?
- Group 9:
Does the team meet regularly and discuss its effectiveness and how it could be improved?
- Group 12:
My level of pay
- Group 13:
I get clear feedback about how well I am doing my job
- Group 17:
I know how my role contributes to what my Trust is trying to achieve
- Group 18:
I feel that my role makes a difference to patients / service users

- Group 19:
Healthcare professionals and managers in non-clinical roles work well together
- Group 23:
During the last 12 months have you been injured or felt unwell as a result of the following problems at work?
- Group 26:
Overall, how would you rate your health during the last four weeks?
- Group 27:
In general, my job is good for my health

One example of insight that follows this analysis: one of the questions that was chosen to be removed, “I get clear feedback about how well I am doing my job”, is tied to the same factor as the question “I do not have time to carry out all my work” (see Appendix), which was preserved. It raises the possibility that time management is perhaps the single most important aspect of how a NHS employee perceives his or her success – particularly in the light of other related questions that were preserved in Group 17, which concern the Trust more than the individual staff member.

Disclaimer

The original data creators, depositors or copyright holders, the funders of the Data Collections and the UK Data Archive bear no responsibility for our analysis or interpretation.

References

- K. Bollen. *Structural Equations with Latent Variables*. John Wiley & Sons, 1989.
- Care Quality Commission and Aston University. Aston Business School, National Health Service National Staff Survey, 2009 [computer file]. *Colchester, Essex: UK Data Archive [distributor], October 2010. Available at* [HTTPS://WWW.ESDS.AC.UK](https://www.esds.ac.uk), SN: 6570, 2010.

Appendix A: All 63 Items

This provides a sketch of the 63 questions provided as input at the beginning of our measurement thinning process. We separate them by target factor. Please refer to (Care Quality Commission and Aston University, 2010) for more details on each item.

- Group 1:
My Trust is committed to helping staff balance their work and home life
- Group 2:
Job share with someone else
- Group 3:
Having a mentor
e-learning / online training
- Group 4:
Health and safety training (e.g. fire training, manual handling)
Infection control (e.g. guidance on hand-washing, MRSA, waste management, disposal of sharps / needles)
How to handle confidential information about patients / service users
- Group 5:
Training, learning development... has helped me to do my job better
Training, learning development... has helped me stay up-to-date with my job
- Group 6:
My immediate manager encourages those who work for her/him to work as a team
My immediate manager can be counted on to help me with a difficult task at work
My immediate manager gives me clear feedback on my work

- Group 7:
In the last months, have you had an appraisal or KSF development review?
Did the appraisal/ review had helped them agree clear objectives for their work?
- Group 8:
In the last 12 months, as part of your KSF development review, appraisal, personal development review, did you agree a Personal Development Plan?
Have you received training, learning or development that was identified in that plan?
- Group 9:
Does the team meet regularly and discuss its effectiveness and how it could be improved?
How many core members are there in your team?
- Group 10:
I often have trouble working out whether I am doing well or poorly in this job
I am involved in deciding on changes introduced that affect my work area / team / department
I cannot meet all the conflicting demands on my time at work
There are enough staff at this Trust for me to do my job properly
- Group 11:
As soon as I can find another job, I will leave this Trust
- Group 12:
The support I get from my work colleagues
The amount of responsibility I am given
The extent to which my Trust values my work
My level of pay
- Group 13:
I do not have time to carry out all my work
I get clear feedback about how well I am doing my job
- Group 14:
The people I work with treat me with respect
The people I work with seek my opinions
- Group 15:
Senior managers here try to involve staff in important decisions
Communication between senior management and staff is effective
On the whole, the different parts of the Trust communicate effectively with each other
- Group 16:
There are opportunities for me to progress in my job
I am supported to keep up to date with developments in my field
I am encouraged to develop my own expertise
- Group 17:
I know how my role contributes to what my Trust is trying to achieve
I know how my Trust contributes to what the NHS is trying to achieve
I would recommend my Trust as a place to work
- Group 18:
I feel that my role makes a difference to patients / service users
I am able to make suggestions to improve the work of my team / department
- Group 19:
There are frequent opportunities for me to show initiative in my role
I am able to make improvements happen in my area of work
Healthcare professionals and managers in non-clinical roles work well together

Senior managers act on staff feedback
I look forward to going to work

- Group 20:
 - I am enthusiastic about my job
 - Time passes quickly when I am working
- Group 21:
 - My Trust encourages us to report errors, near misses or incidents
 - My Trust treats reports of errors, near misses or incidents confidentially
 - When errors, near misses or incidents are reported, my Trust takes action to ensure that they do not happen again
 - We are informed about errors, near misses and incidents that happen in the Trust
- Group 22:
 - My trusts takes effective action if staff are physically attacked by patients / service users, their relatives or other members of the public
- Group 23:
 - My trusts takes effective action if staff are physically attacked by other members of staff
 - My trusts takes effective action if staff are bullied, harassed or abused by patients / service users, their relatives or other members of the public
 - My trusts takes effective action if staff are bullied, harassed or abused by other members of staff
 - During the last 12 months have you been injured or felt unwell as a result of the following problems at work?
- Group 25:
 - Hot water, soap and paper towels, or alcohol rubs, are available when they are needed by: Staff
- Group 26:
 - Overall, how would you rate your health during the last four weeks?
- Group 27:
 - In general, my job is good for my health
 - My immediate manager takes a positive interest in my health and well-being
- Group 28:
 - In the last three months have you ever come to work despite not feeling well enough to perform your duties?