Hospital quality improvement in context: a multilevel analysis of staff job evaluations

U Krogstad, D Hofoss, M Veenstra, P Gulbrandsen, P Hjortdahl

Objective: To investigate how much of the variance in data on nurse evaluation of different aspects of hospital work can be attributed to individual, ward, department and hospital levels, and to discuss the implication of the findings on quality improvement strategies.

Design and method: National survey data of work experiences were collected from hospital nurses working at 124 hospital wards in 36 departments in 15 hospitals across Norway during the autumn of 1998. The multilevel structure of the variation of nine indices of job satisfaction was explored by fitting four-level random intercept models (nurse, ward, department and hospital).

Results: A total of 2606 nurses (66%) responded. The indices showed varying clustering to organizational units. Intraclass correlations (ICCs) varied from 0.05 to 0.38, representing considerable higher level variation. The ward level was the dominating level for the clustering of nurses' job aspect evaluations.

Conclusion: Multilevel modelling of staff work experiences may identify which improvement goals can be addressed at which organizational level. Improvement efforts should be directed specifically towards each aspect of work and at its most relevant organizational level. Strategies aimed at the micro-organizational level (ward management) rather than the individual level or the macro level (hospital top management) might prove worthwhile.

METHODS

Participants

Data were derived from a large 1998 survey program of the experiences of hospital staff in a selection of 15 Norwegian somatic hospitals representing all geographical regions and all types of hospitals (table 1). Eleven of the 14 hospitals stratified as a national sample accepted our invitation to participate and four other hospitals asked to be included in the study. All medical and surgical departments in these hospitals were included. The survey was approved by the National Board of Health and the Norwegian Data Inspectorate.

All doctors, registered nurses (RNs), and auxiliary nurses working in 124 wards in 36 departments (general internal medicine, heart, lung, general surgery, orthopaedic surgery, neurosurgery) in the 15 hospitals were included in the study. Staff names, addresses, and unit of work were drawn from the hospitals' administrative databases. Non-responders received one reminder after 2–3 weeks. After the reminder, all personal identifications were deleted. This paper analyses responses from RNs and auxiliary nurses only, as these are the groups that can be assigned to wards (doctors in this material can only be grouped by department).

The mean number of responders per ward was 21 (range 7–43), and the mean number of wards per departments was 3.4.
Hospital staff surveys. The questionnaire included questions by the Foundation for Health Services Research's Work Research and Quality Improvement Questionnaire (WORQUA) developed for health services research. Staff work experiences were measured by the Work Research Questionnaire. By principal axis factoring with promax rotation, we used factor analysis for data reduction; 29 items covering various domains of work were selected from the questionnaire. For the purpose of this article, we used factor analysis for data reduction; 29 items covering various domains of work were selected from the questionnaire. By principal axis factoring with promax rotation, we reduced the number of domains to nine multi-item indices. The indices of workload, competence, and physical layout represented the structural domain of hospital organisation. The indices of leadership, cooperation, and communication, clarity of tasks, patient continuity, and system continuity reflect processual and cultural aspects of work. Indices and items are listed in Appendix 1. Each index was transformed into a 0–100 score, higher scores indicating more positive evaluation. If at least 50% of the items in a specific subscale were answered, the mean of the answered items for the missing responses was substituted (the half rule). The nine indices reflect areas of central as well as local levels of responsibility and decision. We therefore expected them to cluster at different organizational levels.

### Statistical analysis

Differences between responders and non-responders were tested for statistical significance by \( \chi^2 \) tests. The internal consistency of the indices was assessed by Cronbach's coefficient alpha. \( \text{MLWin}^{18} \) was used to test if staff work experiences were hierarchically clustered. The partitioning by organizational level of the response variance was identified by fitting random intercept models with four levels: nurse, ward, department, and hospital. For each domain of work experience we first applied the full four level model. Levels at which variances appeared non-significant by the crude criterion of having a variance less than twice its standard error (usually the hospital and department levels) were then removed from the model and the change in the model's \(-2LL\) value was inspected for significance. The removed levels were then re-entered separately and accepted where their re-inclusion produced a significant reduction (\( p<0.05 \)) in \(-2LL\).

The degree of clustering was measured by the size of the intraclass correlation coefficient (ICC), calculated as the percentage of the variance of organizational levels divided by the total variance in each index. \( \text{Table 1} \)

<table>
<thead>
<tr>
<th>Property</th>
<th>Observed range</th>
<th>No of items in index</th>
<th>Cronbach's alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workload</td>
<td>0–100</td>
<td>2</td>
<td>0.70</td>
</tr>
<tr>
<td>Colleague competence</td>
<td>0–100</td>
<td>2</td>
<td>0.76</td>
</tr>
<tr>
<td>Physical layout</td>
<td>0–100</td>
<td>3</td>
<td>0.68</td>
</tr>
<tr>
<td>Quality of leadership</td>
<td>0–100</td>
<td>7</td>
<td>0.87</td>
</tr>
<tr>
<td>Quality of cooperation</td>
<td>0–100</td>
<td>5</td>
<td>0.73</td>
</tr>
<tr>
<td>Clarity of tasks</td>
<td>0–100</td>
<td>3</td>
<td>0.79</td>
</tr>
<tr>
<td>Quality of communication</td>
<td>0–100</td>
<td>2</td>
<td>0.68</td>
</tr>
<tr>
<td>System continuity</td>
<td>0–100</td>
<td>3</td>
<td>0.70</td>
</tr>
<tr>
<td>Continuity of care</td>
<td>0–100</td>
<td>2</td>
<td>0.57</td>
</tr>
</tbody>
</table>

The questionnaire

Staff work experiences were measured by the Work Research and Quality Improvement Questionnaire (WORQUA) developed by the Foundation for Health Services Research's hospital staff surveys. The questionnaire included questions on general job satisfaction and psychosocial working conditions, workload, quality of leadership, competence of colleagues, continuity of patient care, system continuity (familiarity with procedures and work patterns), clearness of tasks, quality of cooperation, and physical layout of workplace. Most questions had a 5-point Likert scale format, but some were measured on 10-point scales with unique anchoring statements. Exploratory factor analysis was used to investigate the empirical pattern of meaning in developing the WORQUA questionnaire. For the purpose of this article, we used factor analysis for data reduction; 29 items covering various domains of work were selected from the questionnaire. By principal axis factoring with promax rotation, we reduced the number of domains to nine multi-item indices. The indices of workload, competence, and physical layout represented the structural domain of hospital organisation. The indices of leadership, cooperation, and communication, clarity of tasks, patient continuity, and system continuity reflect processual and cultural aspects of work. Indices and items are listed in Appendix 1. Each index was transformed into a 0–100 score, higher scores indicating more positive evaluation. If at least 50% of the items in a specific subscale were answered, the mean of the answered items for the missing responses was substituted (the half rule). The nine indices reflect areas of central as well as local levels of responsibility and decision. We therefore expected them to cluster at different organizational levels.

### Statistical analysis

Differences between responders and non-responders were tested for statistical significance by \( \chi^2 \) tests. The internal consistency of the indices was assessed by Cronbach's coefficient alpha. \( \text{MLWin}^{18} \) was used to test if staff work experiences were hierarchically clustered. The partitioning by organizational level of the response variance was identified by fitting random intercept models with four levels: nurse, ward, department, and hospital. For each domain of work experience we first applied the full four level model. Levels at which variances appeared non-significant by the crude criterion of having a variance less than twice its standard error (usually the hospital and department levels) were then removed from the model and the change in the model's \(-2LL\) value was inspected for significance. The removed levels were then re-entered separately and accepted where their re-inclusion produced a significant reduction (\( p<0.05 \)) in \(-2LL\).

The degree of clustering was measured by the size of the intraclass correlation coefficient (ICC), calculated as the percentage of the variance of organizational levels divided by the total variance in each index. \( \text{Table 1} \)

<table>
<thead>
<tr>
<th>Index</th>
<th>Mean</th>
<th>Median</th>
<th>No of items in index</th>
<th>Cronbach's alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workload</td>
<td>76</td>
<td>77</td>
<td>0–100</td>
<td>2</td>
</tr>
<tr>
<td>Colleague competence</td>
<td>37</td>
<td>32</td>
<td>0–100</td>
<td>3</td>
</tr>
<tr>
<td>Physical layout</td>
<td>69</td>
<td>71</td>
<td>0–100</td>
<td>7</td>
</tr>
<tr>
<td>Quality of leadership</td>
<td>68</td>
<td>67</td>
<td>0–100</td>
<td>5</td>
</tr>
<tr>
<td>Quality of cooperation</td>
<td>62</td>
<td>67</td>
<td>0–100</td>
<td>3</td>
</tr>
<tr>
<td>Clarity of tasks</td>
<td>69</td>
<td>75</td>
<td>0–100</td>
<td>2</td>
</tr>
<tr>
<td>Quality of communication</td>
<td>71</td>
<td>75</td>
<td>0–100</td>
<td>3</td>
</tr>
<tr>
<td>System continuity</td>
<td>76</td>
<td>75</td>
<td>0–100</td>
<td>2</td>
</tr>
</tbody>
</table>

### The questionnaire

Staff work experiences were measured by the Work Research and Quality Improvement Questionnaire (WORQUA) developed by the Foundation for Health Services Research's hospital staff surveys. The questionnaire included questions on general job satisfaction and psychosocial working conditions, workload, quality of leadership, competence of colleagues, continuity of patient care, system continuity (familiarity with procedures and work patterns), clearness of tasks, quality of cooperation, and physical layout of workplace. Most questions had a 5-point Likert scale format, but some were measured on 10-point scales with unique anchoring statements. Exploratory factor analysis was used to investigate the empirical pattern of meaning in developing the WORQUA questionnaire. For the purpose of this article, we used factor analysis for data reduction; 29 items covering various domains of work were selected from the questionnaire. By principal axis factoring with promax rotation, we reduced the number of domains to nine multi-item indices. The indices of workload, competence, and physical layout represented the structural domain of hospital organisation. The indices of leadership, cooperation, and communication, clarity of tasks, patient continuity, and system continuity reflect processual and cultural aspects of work. Indices and items are listed in Appendix 1. Each index was transformed into a 0–100 score, higher scores indicating more positive evaluation. If at least 50% of the items in a specific subscale were answered, the mean of the answered items for the missing responses was substituted (the half rule). The nine indices reflect areas of central as well as local levels of responsibility and decision. We therefore expected them to cluster at different organizational levels.

### Statistical analysis

Differences between responders and non-responders were tested for statistical significance by \( \chi^2 \) tests. The internal consistency of the indices was assessed by Cronbach's coefficient alpha. \( \text{MLWin}^{18} \) was used to test if staff work experiences were hierarchically clustered. The partitioning by organizational level of the response variance was identified by fitting random intercept models with four levels: nurse, ward, department, and hospital. For each domain of work experience we first applied the full four level model. Levels at which variances appeared non-significant by the crude criterion of having a variance less than twice its standard error (usually the hospital and department levels) were then removed from the model and the change in the model's \(-2LL\) value was inspected for significance. The removed levels were then re-entered separately and accepted where their re-inclusion produced a significant reduction (\( p<0.05 \)) in \(-2LL\).

The degree of clustering was measured by the size of the intraclass correlation coefficient (ICC), calculated as the percentage of the variance of organizational levels divided by the total variance in each index. \( \text{Table 1} \)

<table>
<thead>
<tr>
<th>Index</th>
<th>Mean</th>
<th>Median</th>
<th>Observed range</th>
<th>No of items in index</th>
<th>Cronbach's alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workload</td>
<td>40</td>
<td>39</td>
<td>0–100</td>
<td>2</td>
<td>0.70</td>
</tr>
<tr>
<td>Colleague competence</td>
<td>76</td>
<td>78</td>
<td>0–100</td>
<td>2</td>
<td>0.76</td>
</tr>
<tr>
<td>Physical layout</td>
<td>37</td>
<td>32</td>
<td>0–100</td>
<td>3</td>
<td>0.68</td>
</tr>
<tr>
<td>Quality of leadership</td>
<td>69</td>
<td>71</td>
<td>0–100</td>
<td>7</td>
<td>0.87</td>
</tr>
<tr>
<td>Quality of cooperation</td>
<td>68</td>
<td>67</td>
<td>0–100</td>
<td>5</td>
<td>0.73</td>
</tr>
<tr>
<td>Clarity of tasks</td>
<td>62</td>
<td>67</td>
<td>0–100</td>
<td>3</td>
<td>0.79</td>
</tr>
<tr>
<td>Quality of communication</td>
<td>69</td>
<td>75</td>
<td>0–100</td>
<td>2</td>
<td>0.68</td>
</tr>
<tr>
<td>System continuity</td>
<td>71</td>
<td>75</td>
<td>0–100</td>
<td>3</td>
<td>0.70</td>
</tr>
<tr>
<td>Continuity of care</td>
<td>76</td>
<td>75</td>
<td>0–100</td>
<td>2</td>
<td>0.57</td>
</tr>
</tbody>
</table>

### Results

A total of 2606 questionnaires from nurses were returned (response rate 66%). The number of responders was about 10% of the occupationally active nurses and auxiliary nurses in Norway in the year of study. The distribution of respondents across hospital types, fraction of part time employed, and percentage of nurses and auxiliaries corresponded very well with the national corps of nursing staff.

The response rate was lower for nurses working in university hospitals (62%) than in county or local hospitals (67%; \( p<0.001 \)). The characteristics of responders are presented in table 1. There were no significant differences between responders and non-responders with regard to type of hospital or department. As shown in table 2, the index means (observed ranges 0–100 for all indices) differed considerably. Two domains of work stood out as areas of
less satisfaction: physical layout of workplace and workload. Cronbach’s alpha was lowest for continuity of care (0.57) and communication (0.68) and highest for leadership (0.87) and clarity of tasks (0.79).

The proportions of variance at each organizational level and the ICCs for each index are shown in Table 3. ICCs ranged from 0.05 for quality of communication and for system continuity to 0.35 for physical layout of workplace and 0.28 for quality of leadership.

One domain of job evaluation—physical layout of workplace—had much of its variation at hospital level. All other domains of job evaluation had most of its organization level variance at ward level. This was particularly true for quality of leadership, workload, and clarity of tasks.

### Discussion

In the four-level random intercept analyses, staff job experiences varied significantly across organizational units. Of the nine indices, four had ICCs above 0.20 and none were below 0.05. ICCs between 0.05 and 0.20 are considered to indicate strong clustering of scores, and ICCs above 0.20 are interpreted as high.19 27

Trying to catch the flow of local work, we asked about their experiences at “their” ward. The ward is the organizational, professional, social, and cultural frame of work for nursing staff.30 19 Even if directing their attention to this level, the variance of ward experiences might vary between hospitals and departments. The ward was, however, the dominating organizational level for the clustering of the variation in nurses’ job assessment. Other studies also confirm that, in the smallest hospitals, wards specialize in diagnoses and in patient groups and run a selected repertoire of procedures. Nurses working in different wards therefore experience nursing jobs under different work conditions. This differentiation also contributes to the building of shared experiences, values, and attitudes, and shapes ward-specific local cultures.

The relatively strong clustering of our data probably reflects the fact that our responders were a relatively stable workforce: more than 40% of them had worked “at this ward” for more than 5 years and two thirds had worked “at this hospital” for more than 5 years.

Ward level variation was particularly strong for nurse evaluation of leadership, workload, clarity of tasks, and physical layout of workplace. For these indices, the grouping of responses by ward identified important similarities between the evaluations of nurses in the same wards. Yet one should not jump to the conclusion that efforts to improve the quality of these four aspects of work should be made at ward level only. Discussion of the level at which an intervention should be directed points directly to a lack of information that links quality problems to the right level of decision.25 16 19

### Key messages

- National strategies for improving quality of hospital care are strikingly similar across national borders; they have mainly been directed towards the organizational macro level.
- Improvement strategies need to be directed to specific organizational levels of decision making.
- Multilevel modelling of staff work experiences may identify which improvement goals can be addressed at which organizational level.
- The work experiences of nurses provide substantial information about improvement opportunities at the ward level.

![Table 3 Mean scores and intraclass correlation coefficients (ICCs) by job aspect and level of variation](http://qualitysafety.bmj.com/content/11/3/36.F1)

<table>
<thead>
<tr>
<th>Job aspect</th>
<th>Mean satisfaction score</th>
<th>Hospital level variation (% of total variance)</th>
<th>Department level variation (% of total variance)</th>
<th>Ward level variation (% of total variance)</th>
<th>Responder level variation (% of total variance)</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workload</td>
<td>40.3</td>
<td>18.6 (11.1)</td>
<td>84.6 (18.6)</td>
<td>351.5 (77.3)</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>Competence of colleagues</td>
<td>76.2</td>
<td>7.8 (2.2)</td>
<td>16.6 (4.7)</td>
<td>324.3 (93.1)</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>Physical layout of workplace</td>
<td>37.1</td>
<td>80.8 (17.6)</td>
<td>77.8 (16.9)</td>
<td>460.3 (65.5)</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>Quality of leadership</td>
<td>68.5</td>
<td>16.6 (4.0)</td>
<td>95.9 (24.1)</td>
<td>285.8 (71.9)</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>Quality of cooperation</td>
<td>68.0</td>
<td>7.0 (3.3)</td>
<td>20.4 (9.7)</td>
<td>182.1 (86.9)</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Clarity of tasks</td>
<td>61.9</td>
<td>84.9 (17.5)</td>
<td>29.4 (5.0)</td>
<td>564.2 (95.0)</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Quality of communication</td>
<td>68.8</td>
<td>5.9 (1.9)</td>
<td>30.2 (95.1)</td>
<td>376.3 (91.6)</td>
<td>0.08</td>
<td></td>
</tr>
</tbody>
</table>

Empty cells represent levels where the re-inclusion of variance did not produce a significantly improved model.

functional and better suited to their tasks than others. Closer inspection of our data (not shown) confirmed that the highest rated hospitals were the newest and most modern (and vice versa).

Our results reflect evaluations by nurses and auxiliary nurses. Their viewpoint is the ward. They are excellently located to assess patient care at the sharp end of the system, but may have a less precise view of the advantages and problems at other organizational levels.

Finally, one should take care not to read low ICCs and/or low variance at any particular organizational level as an indication that no quality improvement effort should be taken at that level. Little variation in a bad score may signal that the situation is equally problematic in all units at that level, and one cannot rule out the possibility that the roots of the problem and the key to its solution may be found at just that level.

Major restructurings of hospital care at the macro level have taken place all over Europe. Few of the reforms, however, have been based on evidence—or produced evidence for—quality improvement. It has even been suggested that the concentration on the macro level is more a part of the problem than its solution. Our study demonstrates substantial variation in how nurses evaluate different domains of work. It also shows that a significant part of the variance was clustered at the ward level, indicating that this level is an important focus for quality improvement efforts. We suggest that quality improvement leverage points should be sought by multilevel analyses and ICCs.

Organising for improvement of hospital services has been a major topic in health services all over the world for a long time. Many of the problems documented are strikingly similar across national borders. We suggest that national governments should not only instigate macro reforms of the hospital system (such as changes in ownership and financing systems), but also stimulate internal hospital improvement work at the department and ward levels. After all, that’s where patients are.

Authors’ affiliations
U Krogstad, Norwegian Health Services Research Centre, Oslo, Norway
D Hofoss, Institute of Community Medicine, University of Tromsø, Norway
D Hofoss, M Veenstra, Rikshospitalet University Hospital, Oslo, Norway
P Gulbrandsen, Akershus University Hospital, University of Oslo, Norway
P Hjortdahl, Institute of Public Health and Community Medicine, University of Oslo, Norway

Competing interests: none.

APPENDIX 1: INDICES OF WORK EXPERIENCES BASED ON ITEMS FROM THE WORQUA QUESTIONNAIRE

INDEX 1: WORKLOAD
- How do you rate your physical workload this autumn?
- How do you rate your mental workload this autumn?

Original item scale: 1 = very light; 10 = intolerably heavy

INDEX 2: PHYSICAL LAYOUT OF WORKPLACE
- How well do these statements fit?
  - “This hospital has a very functional layout”
  - “At this department we experience a major lack of space”
  - “This ward is built exactly to suit its tasks”

Original item scale: 1 = fits completely; 5 = does not fit at all

INDEX 3: COMPETENCE OF COLLEAGUES
- How would you rate the competence of the nurses at this ward?
- How would you rate the competence of the doctors at this ward?

Original item scale: 1 = very insufficient; 10 = fully sufficient

INDEX 4: LEADERSHIP
- How well do these statements fit?
  - “My immediate superior always speaks clearly”
  - “My immediate superior knows my job situation well”
  - “My immediate superior has clear goals for the future development of this unit”
  - “My immediate superior provides feedback so that I know whether I’m doing a good job”
  - “My immediate superior provides continuous information about goals and results”
  - “The department management knows the job situation at the wards”
  - “The head nurse knows the daily capacity and professional challenges at this ward”

Original item scale: 1 = fits completely; 5 = does not fit at all

INDEX 5: COOPERATION
- How well do these statements fit?
  - “At this hospital, interdepartmental cooperation is very good”
  - “Information about patients gets to the right place at the right time”
  - “At this ward, interprofessional cooperation is very good”
  - “At this ward, all professions have common goals for the patient’s stay”
  - “At this ward, members of the other professions ‘know’ their patients”

Original item scale: 1 = fits completely; 5 = does not fit at all

INDEX 6: CLEAR TASKS AND RESPONSIBILITIES
- How well do these statements fit?
  - “Expectations from my superiors are clear”
  - “Tasks are clearly defined”
  - “New employees are being supervised and followed up”

Original item scale: 1 = very good; 10 = quite unsystematic

INDEX 7: COMMUNICATION
- How well do these statements fit?
  - “I discuss patient information with other professions many times a day”
  - “Patient information from other professions is vital to my work”

Original item scale: 1 = fits completely; 5 = does not fit at all

INDEX 8: SYSTEM CONTINUITY
- How well do these statements fit?
  - “I know this hospital inside out”
  - “I am thoroughly familiar with the job routines of this ward”
  - “I have firm knowledge of the most typical patient groups at this ward”
INDEX 9: CONTINUITY OF PATIENT CARE

- Item scale: 1 = fits completely; 5 = does not fit at all

How well do these statements fit?

- “By and large I observe the patients that I am responsible for every day during their stay at this ward”
- “Our work organization imply that we actually follow up the same patients”
- Original item scale: 1 = fits completely; 5 = does not fit at all

REFERENCES


27. Van Yperen NW, Snijders TA. A multilevel analysis of the demands-control model: is stress at work determined by factors at the group level or the individual level? J Occup Health Psychol 2000;5:182–90.

www.qph.com