

Assessment of inter-rater consistency in classification of adverse events

The national reporting and learning system for serious adverse events in specialized health care

Ylva Helland PhD, Eli Saastad PhD, Øystein Flesland MD/PhD. The National reporting and learning Unit, The Norwegian Directorate of Health

INTRODUCTION

In Norway, all serious adverse events in specialized care health institutions are to be reported to the national reporting and learning system (RLS). All patient safety reports are manually screened and classified. Classification facilitates analysis on an aggregated level with the overall goal of learning. Consistent use of the classification system across case handlers is therefore essential in order to enhance data quality. There have been indications of low inter-rater consistency, but this has not previously been assessed.

OBJECTIVE

The aim was to assess the degree of consistency in classification across case handlers.

METHODS

19 case handlers independently classified 41 adverse events case reports. The case reports were taken from a previous study, representing a range of incident types (1). The classification system is a modified version of the WHO's International Classification for Patient Safety. The category *incident type* has three levels: main categories (level 1) and subcategories on two levels (level 2 and 3). However, three out of the nine main categories only have levels 1 and 2. It is mandatory to classify on the lowest subcategory level available. Inter-rater consistency was analyzed by distributions of categories and subcategories by case handlers. Classification were assessed on each category level with a stepwise aggregation of results on the higher levels.

RESULTS

Classification of *incident type* showed low degree of consistency across case handlers on the subcategory levels. However, inter-rater consistency in classification increased on the higher category levels (Table 1). On subcategory level 2 classification was consistent across raters in three out of 41 case reports. On the main category level (level 1) inter-rater consistency increased to 12 out of 41 case reports. Only one case report had total agreement among all case handlers independent of category level (Figure 1).

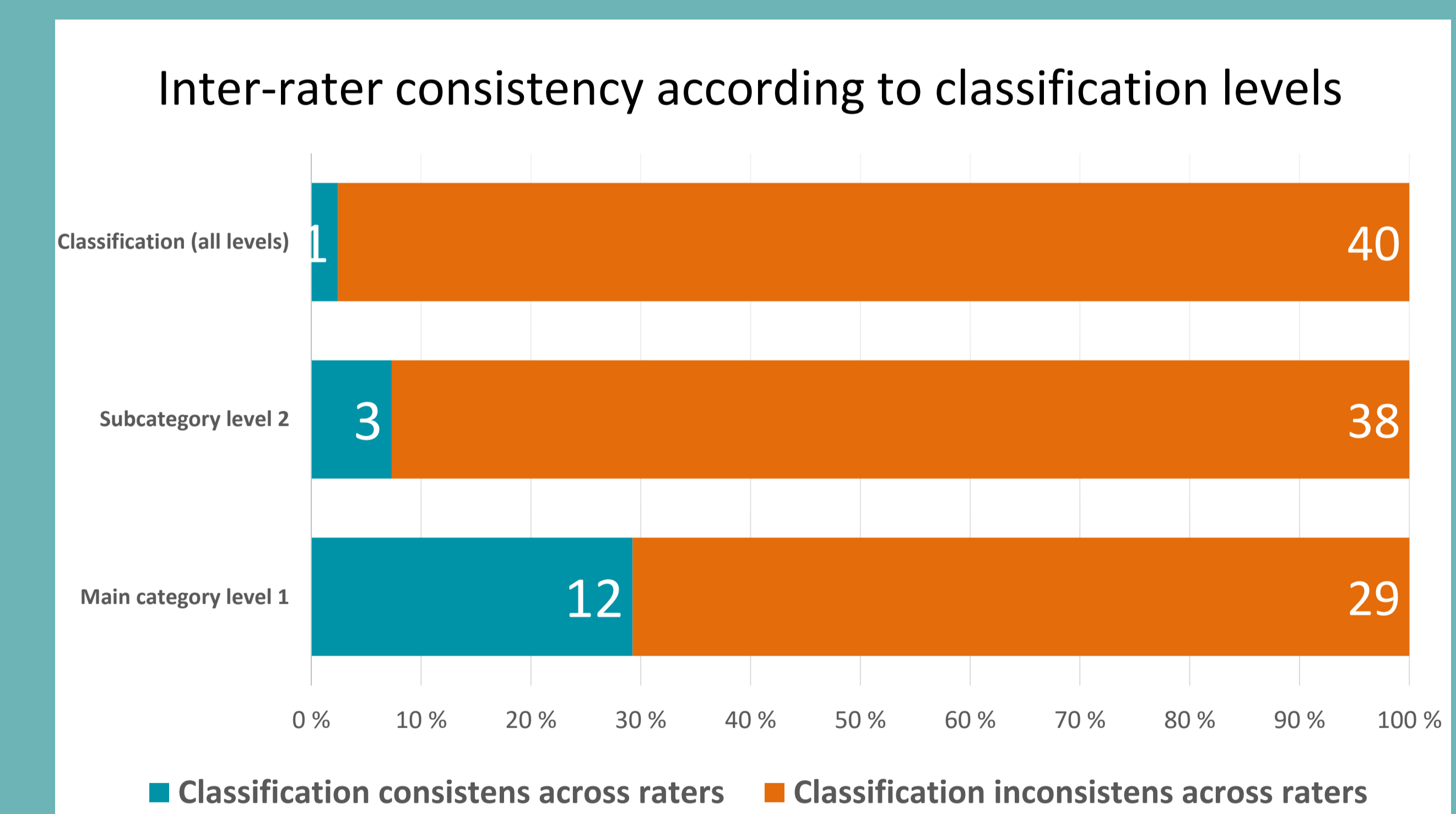
There was no clear pattern between case reports with low and high inter-rater consistency in terms of number of words in case descriptions, and category characteristics (general, specific, broad).

Table 1

Consistency across case handlers	Classification (all levels)	Subcategory level 2	Main category level 1
Minimum categories pr case report (n)	1	1	1
Maximum categories pr case report (n)	10	7	4
Median categories pr case report (n)	5	4	2

Number of case handlers n=19
Number of adverse events n=41

Figure 1



CONCLUSIONS

Accuracy in classification and consistency across case handlers was low. However, inter-rater consistency increased for each increasing level of the classification categories. Investigations of the reasons for the limited accuracy in classification across case handlers is warranted.

Referanse: Mikkelsen KL, Thommesen J, Andersen HB. Validation of Danish adaption of the World Health Organization's International Classification for Patient Safety classification of patient safety incident types. *Int J of Qual Health Care* 2013;25:132-140