**Sample weighting**

Weighting procedures were used to achieve a) the representativeness of the DS sample for the reference population (N=9617), and the comparability of b) the NS with the DS sample and c) the civilian sample with the DS sample. In a) post-stratification weights were used to account for differences that occurred due to the joint effects of sampling, eligibility, and response. Ten strata were defined, eight among males (all possible combinations of combat vs non-combat units, rank high vs. low and deployment location (Kunduz vs other)) and two among females (Kunduz vs other location). These ten strata were defined as a trade-off between all information we had on the source population (24 strata would have been possible) and ensuring a sufficient number of cases per stratum for stable estimates. The weights in the DS sample were then calculated on the basis of the inverse sample versus reference population frequency count per stratum. For b) the following military and demographic characteristics were considered: rank, unit, number of years in service, operational area (army, air force, basic fighting, forces, marine, medical), marital status, gender, age (linear, squared and cubic terms vs. four categories), educational level and having children. We used the propensity score method based on logistic regression (Kurth et al. 2005) with the outcome deployment status (DS versus NS) to adjust the NS sample to the weighted DS sample in a). After testing for interactions, backward selection of main effects and interaction terms, fitting different alternative models (e.g. age dimensional vs. age categorical) and inspecting the balance of the weighted marginal distributions of the variables between DS and NS, we chose the model that included the following main effects: rank (2 dummy variables), years of deployment, operational area (4 dummy variables), combat unit (2 dummy variables), gender, age (3 dummy variables representing 4 categories), educational level (2 dummy variables), having children and family status (3 dummy variables) (Wittchen et al. 2012a). With the associated model-based probabilities for DS (propensity score), a categorization into 5 blocks turned out to yield the best balance of all variables considered. The weights in the NS sample were then based on the NS to DS ratios of the relative frequencies of these blocks (DS weighted as in a)), and the PSCORE command in Stata (Stata Corp. 2012) was used. For this publication, we omitted 154 females (110 in DS and 54 in NS). Note that the weights are still valid in the subsample of male soldiers since a) was done conditional on gender and in b) interactions with gender in predicting DS versus civilian sample were investigated but not found, suggesting that the predictive values of variables other than gender did not depend on gender. For c) the same propensity score method was used, but we did not have the military characteristics for adjustment, nor had all demographic variables been collected in the same way in the civilian sample. Age and marital status could be used for this comparison. The propensity score was based on the model predicting DS versus civilians with the linear age term and marital status main effects (squared age and interaction terms were not significant), and was categorized into 6 blocks for the best balance.