ONLINE SUPPORTING MATERIAL

Supplemental Methods 1: Exploratory factor analysis (EFA)

METHODS

The scheme of analysis for EFA was comparable to the PCA analysis and is presented in Figure 1. Five factor solutions ranging from 2 to 6 patterns were extracted in the derivation sample. All solutions were subsequently confirmed in the replication sample using CFA (a). Internal reproducibility was checked by performing EFA in the replication sample and the entire study population (b). Finally, the Kaiser criterion, the scree test optimal coordinate and a visual inspection of the scree plot were applied to identify the number of patterns to retain (c). Data analysis was performed with R.2.15.2 (1).

EFA with principal axis factoring, based on the correlation matrix and with varimax rotation, was used to extract solutions ranging from 2 to 6 factors in the derivation sample (factor.pa() function in R-package psych) (2). Each of the obtained solutions was then separately confirmed in the replication sample using CFA (cfa(); lavaan) (3, 4). Food groups with factor loadings $\geq |0.25|$ in EFA ($FL_{EFA}$) were included in the models for CFA. We retained food groups with factor loadings $\geq |0.20|$ in CFA ($FL_{CFA}$). Patterns that contained only two dominant food groups were excluded for CFA, e.g. pattern D of the 4-factor solution (Supplemental Table 3), as it is recommended to confirm only patterns with at least 3 dominant food group loadings (5). In order to summarize the results, the ratio of food groups not confirmed ($n$ food groups with $FL_{CFA}\leq |0.20|$) to the total number of food groups ($n$ food groups with $FL_{EFA}\geq |0.25|$) and the deviation in factor loadings between EFA and CFA, defined as $Deviation(EFA - CFA) = \frac{1}{k} \cdot \sum_{i=1}^{k} 100 \cdot |(FL_{EFA} - FL_{CFA})/FL_{EFA}|$, where $k = n$ food
ONLINE SUPPORTING MATERIAL

groups, were computed. For example, a value of 20% means that $F_L^{CFA}$ of a particular food group differs on average 20% from $F_L^{EFA}$.

RESULTS

Confirmation success of all factor solutions obtained by EFA is presented in Supplemental Table 3. The two measures for the reliability of the patterns were well in line with each other, i.e. high ratios of food groups not confirmed to the total number of food groups (NC) were mostly accompanied by high deviations in food group loadings between EFA and CFA. Confirmation success differed considerably between patterns within the same solution. Compared to component solutions (PCA) factor solutions (EFA) were generally better confirmed, with mean deviation ranging from 13.3 to 20.3% (compared to 24.5 to 33.0% in PCA). This most likely reflects that CFA is statistically more related to EFA than PCA. Confirmation success differed considerably between different solutions. The 3- and 4-factor solutions were best confirmed, although the latter contained a pattern consisting of only two food groups (4D) (Supplemental Table 4).

After visual inspection of the scree plot, one might select 3 patterns, while the scree test optimal coordinates as well as the Kaiser criterion suggested extracting 8 patterns. EFA and PCA extracted similar underlying patterns, but for PCA component loadings were higher and component solutions contained considerably more food groups. For example, the 3 factor solution (EFA) contained resp. 10, 5 and 8 food groups, while the 3 component solution (PCA) contained resp. 13, 11 and 11 food groups.

EFA in the whole study population led to comparable food group loadings for all patterns as for the split-half samples (data not shown).
ONLINE SUPPORTING MATERIAL

As for PCA, reliability of dietary patterns differed considerably between different EFA solutions and the scree plot was the only criterion indicating the same solution as the reliability results. EFA and PCA extracted similar underlying patterns, but component loadings (PCA) were higher than factor loadings (EFA) leading to the inclusion of more food groups in components compared to factors. Theoretically this can be explained by conceptual differences between PCA and EFA (5-8). Since dietary intake estimation contains measurement error (9) and correlations between food group variables are mostly low to moderate, it could be that dietary patterns derived from PCA are inflated.

Literature cited


