THE LANCET **Global Health**

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Sheikh M, Shakeri R, Poustchi H, et al. Opium use and subsequent incidence of cancer: results from the Golestan Cohort Study. *Lancet Glob Health* 2020; **8:** e649–60.

Supplementary File 1

Details of the interviews ¹

Each subject was interviewed by a trained general physician and a trained nutritionist, either in the local language (Turkmen) or in the national formal language (Persian), depending on the participant's preference. Two structured questionnaires were administered: a life-style questionnaire and a food frequency questionnaire (FFQ). Following the questionnaires and a limited physical examination, samples of blood, urine, hair, and nails were collected by a trained technician.

Validity and reliability of the questionnaires 2,3

In the pilot phase of the Golestan Cohort Study 1,057 were interviewed, and a repeat interview was performed on 131 subjects 2 months after the first interview. The kappa statistics for agreement were above 0.7 for most variables, including tobacco, nass, opium, and alcohol consumption. The validity of the questionnaire data about opium use was assessed in 150 subjects by comparing their questionnaire responses with the presence of codeine or morphine in their urine; the questionnaire responses had a sensitivity of 0.93 and a specificity of 0.89 for identifying subjects with these urinary opium metabolites. There was also a good agreement between selfreported current tobacco smoking or nass use and positive urinary cotinine. To validate the study FFQ, twelve 24-h recall questionnaires (one every month) and four FFQs (one in each season) were administered to 131 participants during 1 year. There was good correlation between FFQ and recall data on food group and nutrient intakes, and there was acceptable correlation between FFQ data and biomarker measurements.

To examine the repeatability of the data collected in the actual cohort, we repeated the entire enrollment process, including interviews and sample collections, in 698 cohort participants from rural areas. The mean interval between the first and second enrollments was 45 months. The results showed very good agreement between data collected at the two interviews.

Methods of creating the wealth score ⁴

We have previously published the details of the methods we used for creating the wealth score for GCS participants⁴. Briefly, to build a composite score for wealth based on appliances and other variables, we utilized multiple correspondence analysis (MCA) on personal car, motorbike, B/W TV, colour TV, refrigerator, freezer, vacuum and washing machine ownership variables, as well as house ownership, house structure, house size (tertiles), having a bath in the residence and occupation. MCA may be used as an exploratory tool and is appropriate for qualitative variables.

Brief description of multiple correspondence analysis used for creating the wealth score 4

MCA is a technique that may be used to gain insight into dependencies within data cross-classified by discrete factors. It has many similarities to principal components analysis (PCA), though the latter is designed for continuous variables. Since PCA is easier to understand, we briefly describe its use first. Then, we discuss correspondence analysis (CA), which is appropriate for data cross classified by two factors, and MCA. PCA constructs a set of linear combinations of variables with the following properties. The first principal component is the linear combination of the variables that has the greatest variance amongst all the linear combinations in that model. The second principal component is the combination that has the second greatest proportion, and is orthogonal to the first, and so on for the third, fourth, and more principal components. The output is a set of weights which define the set of linear combinations. The aim is often to find a small number of linear combinations that explain a large proportion of the variability in the data. Technically, the weights and the proportion of variation accounted for are obtained by carrying out a singular value decomposition of the variance-covariance matrix of the data, to produce the eigenvalues and eigenvectors.

The basic approach in CA is similar, though the discreteness of the variables suggests that the variancecovariance matrix is not an appropriate quantity to decompose. Instead, the inertia, which is a scaling of the Pearson's chi-squared statistic, is examined. The inertia is a linear combination of elements that measure the distance of each contingency cell entry from independence, and the matrix of such entries are decomposed. In MCA, a two-way table may be constructed by cross classifying the data by the levels of each of the variables. Again, in a similar fashion to PCA, one may plot the weights of the first combinations in order to see the contributions of each variable to the scores. In our study, we plot the first two combinations. The weights given to each level of each of the variables in each linear combination may be used to construct an index.. We used weights in the first linear combination (illustrated as the first dimension in Figure 1) to produce our wealth score, since it explained the majority of the chi-square variation of the wealth-related variables. For example, the weights for owning or not owning a refrigerator were 0.038 and -0.733 , and for owning or not owning a freezer were 0.542 and – 0.105, respectively. If a subject owned a refrigerator but did not have a freezer, the corresponding weights $(0.038 \text{ and } -0.105)$ were summed up; this procedure was continued until the weights of all MCA variables were included in this calculation. More information about MCA is available in Encyclopedia of Statistical Sciences.

Methods for creating the Healthy Eating Index (HEI) dietary score ⁵

We have previously published the details of the methods we used for creating the HEI dietary score for GCS participants⁵. To create components of the HEI, we converted the daily intakes from grams to cup and ounce equivalents using the Food Patterns Equivalents Database (FPED) 2013–2014⁶. The FPED units are ounce and cup equivalents and can be converted to standard units as follows: $1 \text{ oz} = 28.35 \text{ g}$ and $1 \text{ cup} = 225 \text{ ml}$. For fruits and vegetables, we used an extensive list of one cup equivalent weights for fruits and vegetables in the FPED⁶. For example, for canned fruit in light syrup, 65% fruit was assumed. For grain products such as bread, dough and cake, made with four, each 16 g of four present in a food was used as the basis for defning a 1-oz grain equivalent, the rationale being that one standard slice of bread has been defned as equal to 1-oz grain equivalent, which will contain 16 g of four ⁶. For intact grains such as rice and pasta, cooked grains were converted to the uncooked forms with conversion factors 0.36 and 0.37, respectively $\bar{7}$, and 1-oz equivalent of grains was defned as 28.35 g⁶. In the FFQ, multi-ingredient foods such as pizza were not asked, so we did not have to disaggregate the foods. However, protein foods were further disaggregated to lean fraction and fat as follows: meat and poultry were disaggregated to lean meat and solid fat fractions; and seafood and nuts were disaggregated to lean protein and oil fractions. Similarly, dairy foods were further disaggregated to a low fat dairy fraction, similar to skim milk, and a solid fat fraction⁶.

References

- 1 Pourshams A, Khademi H, Malekshah AF, *et al.* Cohort Profile: The Golestan Cohort Study--a prospective study of oesophageal cancer in northern Iran. *Int J Epidemiol* 2010; **39**: 52–9.
- 2 Pourshams A, Saadatian-Elahi M, Nouraie M, *et al.* Golestan cohort study of oesophageal cancer: feasibility and first results. *Br J Cancer* 2005; **92**: 176–81.
- 3 Malekshah AF, Kimiagar M, Saadatian-Elahi M, *et al.* Validity and reliability of a new food frequency questionnaire compared to 24 h recalls and biochemical measurements: pilot phase of Golestan cohort study of esophageal cancer. *Eur J Clin Nutr* 2006; **60**: 971–7.
- 4 Islami F, Kamangar F, Nasrollahzadeh D, *et al.* Socio-economic status and oesophageal cancer: results from a population-based case–control study in a high-risk area. *Int J Epidemiol* 2009; **38**: 978–88.
- 5 Hashemian M, Farvid MS, Poustchi H, *et al.* The application of six dietary scores to a Middle Eastern population: a comparative analysis of mortality in a prospective study. *Eur J Epidemiol* 2019; **34**: 371–82.
- 6 Bowman SA, C.J., Friday JE, Lynch KL, and Moshfegh AJ. Food patterns equivalents database 2013–14: Methodology and User Guide. Food Surveys Research Group, Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Maryland. http://www.ars.usda. gov/nea/bhnrc/fsrg (accessed Sept 20, 2019).
- 7 Bowman SA, M.C., Carlson JL, Clemens JC, Lin B-H, and Moshfegh AJ. Food intakes converted to retail commodities databases: 2003–08: methodology and user guide, A.R.S. U.S. Department of Agriculture, Beltsville, MD, and U.S. Department of Agriculture, Economic Research Service, Editor. Washington, D.C. p. 48., 2013 https://data.nal.usda.gov/dataset/food-intakes-converted-retail-commodities-databases-ficrcd (accessed Sept 20, 2019).

Supplementary File 2

Dose-response associations, stratified by tobacco use status, between duration of opium use and risk of all cancers combined (Figures A & D), gastrointestinal cancers combined (Figures B & E), and respiratory cancers combined (Figures C & F).

The vertical axis shows the hazard ratios in a model that used age as the time-scale and was adjusted for sex, ethnicity (Turkman/non-Turkman), residence (urban/rural), wealth score quartiles, smoking cigarettes (ever/never), cumulative pack-years of smoked cigarettes (continuous variable), and regular alcohol drinking (never/ever).

1. Never Tobacco Users

A. All cancers combined (n=1,361)

B. Gastrointestinal cancers combined (n=685)

C. Respiratory cancers combined (n=59)

2. Ever Tobacco Users

D. All cancers combined (n=472)

E. Gastrointestinal cancers combined (n=229)

F. Respiratory cancers combined (n=95)

Supplementary Tables

Table footnotes:

HR: Hazards Ratio, CI: Confidence Interval, N: Number

‡: Adjusted Model 1 used age as the time-scale and was adjusted for sex, ethnicity (Turkman/non-Turkman), residence (urban/rural), wealth score quartiles, smoking cigarettes (ever/never), cumulative pack-years of smoked cigarettes (continuous variable), and regular alcohol drinking (never/ever)

¶: Adjusted Model 2 included all variables in model 1, and further included chewing nass (never/ two halves of nass-years), regular consumption of water pipe (never/ever), predominant household fuel (natural gas/kerosene/biomass/mixed), and tertiles of the healthy eating index

Supplementary Table2. Individual and combined effects of opium and tobacco use on the risk of developing

HR: Hazards Ratio, CI: Confidence Interval, N: Number

‡: this model used age as the time-scale and was adjusted for sex, ethnicity (Turkman/non-Turkman), residence (urban/rural), wealth score quartiles, and regular alcohol drinking (never/ever)

HR: Hazards Ratio, CI: Confidence Interval, N: Number

HR: Hazards Ratio, CI: Confidence Interval, N: Number

HR: Hazards Ratio, CI: Confidence Interval, N: Number

Supplementary Table 6. The association between different types of opium and risk of cancers in the Golestan Cohort

٦

HR: Hazards Ratio, CI: Confidence Interval, N: Number

Supplementary Table 7. Using opium and cancer risk analysis, stratified by socioeconomic status, in the Golestan Cohort Study.

SES: Socioeconomic status

HR: Hazards Ratio, **CI:** Confidence Interval, **N:** Number

HR: Hazards Ratio, CI: Confidence Interval, N: Number

HR: Hazards Ratio, CI: Confidence Interval, N: Number

HR: Hazards Ratio, CI: Confidence Interval, N: Number