Aalborg Universitet

AALBORG UNIVERSITY

DENMARK

Corrigendum to "A validated clinical risk prediction model for lung cancer in smokers of all ages and exposure types

A HUNT study" [EBioMedicine 31 (2018) 36-46]

Markaki, Maria; Tsamardinos, Ioannis; Langhammer, Arnulf; Lagani, Vincenzo; Hveem, Kristian; Røe, Oluf Dimitri

Link to publication from Aalborg University

Corrigendum to "A validated clinical risk prediction model for lung cancer in smokers of all ages and exposure types: A HUNT study" [EBioMedicine 31 (2018) 36−46]

Maria Markaki,[a] Ioannis Tsamardinos,[b] Arnulf Langhammer,[c] Vincenzo Lagani,[b,1] Kristian Hveem,[d] and Oluf Dimitri Røe[e]*

[a]University of Crete, Department of Computer Science, Voutes Campus, Heraklion, GR 70013, Greece
[b]University of Crete, Department of Computer Science, Voutes Campus, Heraklion, GR 70013, Greece; Gnosis Data Analysis PC, Palaiokapa 64, Heraklion, GR 71305, Greece
[c]HUNT Research Centre, Department of Public Health and Nursing, Norwegian University of Science and Technology, Forskningsvegen 2, Levanger, NO 7600, Norway
[d]HUNT Research Centre, Department of Public Health and Nursing, Norwegian University of Science and Technology, Forskningsvegen 2, Levanger, NO 7600, Norway; K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health an Nursing, Norwegian University of Science and Technology, NO 7491 Trondheim, Norway
[e]Norwegian University of Science and Technology, Department of Clinical Research and Molecular Medicine, Prinsesse Kristinsgt. 1, Trondheim, NO 7491, Norway; Levanger Hospital, Nord-Trøndelag Hospital Trust, Cancer Clinic, Kirkegata 2, Levanger, NO 7600, Norway; Clinical Cancer Research Center, Department of Clinical Medicine, Hobrovej 18-22, Aalborg, DK 9000, Denmark

The authors recently became aware of some minor errors in the algorithms provided in the Supplementary, pages 4 and 5.

In the correct algorithm, to avoid logarithm of zero, the argument of logarithm is incremented by 1 in the risk model equations. Those variables are the following: pack years, quit years, BMI and smoke exposure hours. This increment was mistakenly not written in the Supplementary text.

We also noticed that some of the coefficients reported in the Supplementary were slightly rounded values, probably generated automatically.

In the text, the *ln* and *log* were both used. The correct is that all are natural logarithm, and therefore we changed the *log* to *ln* and added one line for clarification: "We indicate with ln the natural logarithm."

Here we provide the full algorithm with non-rounded coefficients and the +1 increment.

Absolute risk estimation in the Cox model for events within 16 years

To estimate the absolute risks per patient, the baseline survival function $S_o(t)$ was estimated according to van Houwelingen[17]: the general Cox model $S(t|X) = S_o(t)^{\exp(X\beta)}$ can be rewritten as $\ln(-\ln(S(t|X)) = ln(-\ln(S_o(t))) + PI(X)$, with $PI(X) = X\beta$ the prognostic index of the model. Using the survfit function, we calculated the baseline survival probability $S_o(t)$ at corresponding max follow-up time $t = 16.4$ years for the smokers' Cox model. A simplification based on the Weibull model specifies that

$$ln(-\ln(S_o(t))) = \beta_o + \beta_1 \ln(t).$$

Baseline survival used in the final model is:

$$S_o(t) = \exp(-\exp(-0.929 + 0.708 \ln(t))) = 0.06$$

We can calculate an individual's estimated risk within max follow-up of $t = 16$ years given data on the risk factors using the risk model. We indicate with ln the natural logarithm.

DOI of original article: http://dx.doi.org/10.1016/j.ebiom.2018.03.027

*Corresponding author at:
E-mail address: oluf.roe@ntnu.no (O.D. Røe).

[1] Current affiliation: Biological and Environmental Sciences and Engineering Division (BESE), King Abdullah University of Science and Technology KAUST, Thuwal, Saudi Arabia.

Correction

The prognostic index $X\beta$ (based on Cox model coefficients in Table 2) is:

$$X\beta = 0.12058192\,(male) - 2.00205575\left[\left(\frac{age}{100}\right)^{-1}\right] + 1.16301808\,[\ln(pack\,yrs + 1)]$$

$$-0.02954059(cigarettes\,per\,day) - 0.24079981[\ln(quit\,yrs + 1)] - 1.24626558[\ln(BMI + 1)]$$

$$+0.16632008\,[\ln(smoke\,exposure\,hrs + 1)] + 0.4059355(cough\,daily)$$

and the calculation of the predicted risk within 16 yrs proceeds as follows:

Risk of Lung cancer $= 1 - 0.06^{\exp(X\beta)}$.

Logistic regression prediction model of lung cancer risk for events within 6 years

We also developed a logistic regression model where risk was the probability of a lung cancer diagnosis within a 6 years period. We indicate with ln the natural logarithm. The prognostic index $X\beta_{6yrs}$ (based on logistic regression model coefficients) is:

$$X\beta_{6yrs} = 1.18203062 + 0.31573217(male) - 1.98496138\left[\left(\frac{age}{100}\right)^{-1}\right] + 1.11994217\,[\ln(pack\,yrs + 1)]$$

$$-0.04002877(cigarettes\,per\,day) - 0.24019955[\ln(quit\,yrs + 1)] - 1.70238304[\ln(BMI + 1)]$$

$$+0.0807242\,[\ln(smoke\,exposure\,hrs + 1)] + 0.49212668(cough\,daily)$$

and the calculation of the predicted risk within 6 years proceeds as:

6 yrs Risk of lung cancer $= \dfrac{1}{1 + exp(-X\beta\_\{6yrs\})}$

To be totally transparent with the research community, and to allow fellow researchers to easily perform their own calculations and validations, we provide the Excel files for calculating the 6 and 16 years risk.

Another unintentional mistake was noticed in the main manuscript.

In section 3.5. two examples of use of the calculator are shown. Unfortunately, the base 10 logarithm instead of the natural logarithm was used by error. The corrected results for this section are:

> «By applying this threshold, using either the HUNT Lung Cancer Model nomogram or the online calculator, a 40-year old person with 15 pack-years and full-score contribution with low BMI (e.g. 22), low smoke intensity (e.g. 10 cigarettes per day), periodical or daily cough, and many hours of indoor smoke exposure (e.g. 10 h, total risk score > 15, LC risk = 1.77% at 16 years and 0.44% in 6 years would be assessed as a medium- or high-risk individual. A 56-year-old with 15 pack-years, high BMI (e.g. 33), high smoke intensity (e.g. 40 cigarettes per day), no periodical or daily cough, and no indoor smoke exposure would be assigned a lower than cut-off risk and would not be eligible for screening (risk score 12.5, LC risk = 0.84% at 16 years and 0.14% at 6 years (Figure 2a, b).»

After correction, the numbers are different but the message is unchanged: compared to an older individual (56 years), a younger person (40 years of age) with the same number of pack-years apparently has a higher risk score within 6 and 16 years if he scores high on the other negative predictive factors.

The correct algorithms were used for all the other results presented in the manuscript. These results, their interpretation and overall conclusions of the manuscript have thus not changed. The authors apologize for any inconvenience caused.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2022.104187.