Risk factors for COPD spirometrically defined from the lower limit of normal in the BOLD project

R Hooper[1], P Burney[1], WM Vollmer[2], MA McBurnie[2], Þ Gíslason[3], WC Tan[4], A Jithoo[1], A Kocabaş[5], T Welte[6], AS Buist[7]

[1]National Heart and Lung Institute, Imperial College, London UK; [2]Kaiser Permanente Center for Health Research, Portland, OR, USA; [3]Medical Faculty, Landspitali and University of Iceland, Reykjavik, Iceland; [4]iCapture Center for Cardiovascular and Pulmonary Research, University of British Columbia, Vancouver, Canada; [5]School of Medicine, Cukurova University, Adana, Turkey; [6]Respiratory Medicine Department, Medical Faculty of the University of Hannover, Hannover, Germany; [7]Oregon Health & Sciences University, Portland, OR, USA

Online Supplement

Methods: Model fitting and coding of risk factors

Risk factors investigated were: age, sex, body mass index (BMI, categorised as normal/underweight (<25), overweight (25-30), obese (>30)), education (categorised as none, primary or middle school, secondary school, technical/vocational college or university), father's education (categorised as for subject), hospitalisation for breathing problems before age 10 years, pack-years of smoking (number of cigarettes smoked per day divided by 20 and multiplied by years of smoking), current smoking, passive smoking (somebody else smoking in the subject's home in the last two weeks), doctor diagnosed tuberculosis (ever), family history of COPD (doctor ever diagnosed mother, father, sister or brother with emphysema, chronic bronchitis or COPD), years working in dusty jobs, regular exposure to dust in present job, regular exposure to fumes in current job, years of heating home with biomass fuel (coal, coke, peat, wood, crop residue or dung), and equivalent years of continuous exposure to cooking fires using biomass fuel (calculated by multiplying number of years over which subject was exposed by number of hours exposure per day divided by 24).

For both heating and cooking, use of coal, coke or peat was reported separately on the questionnaire to use of wood, crop residue or dung. However, heating (or cooking) with one type of fuel was closely associated in most centres with heating (or cooking) with the other – in fact participants frequently reported identical lengths of time exposed to each, and the distribution of this time exposed was similar to that among people using just one type of fire. Thus we did not consider it appropriate to add the time exposed to coal, coke or peat fires to the time exposed to wood, crop residue or dung fires – instead we calculated the maximum of the two reported durations of exposure for each subject.

Effects of subject's education and father's education were modelled as linear trends across the categories. Education levels differed widely between centres and this was the simplest way to guarantee an estimable effect of education that could be compared across all centres. Father's education was reported as unknown for around a quarter of participants: this was coded as no education, and an additional covariate was included in the regression to compare those with known and unknown father's education. The other risk factors which could be reported on the questionnaire as unknown were subject's own education, and hospitalisation for breathing problems before age 10 years: in the few cases where this occurred these values were treated as missing.

We attempted to include all risk factors in the regression analyses of every centre, but in the analysis of COPD there were occasions where the exposed subgroup at a centre was small and

did not include any cases of COPD. This meant that the subgroup had to be excluded from the analysis, and the effect of the factor was not estimated for this centre.

Additional decisions about the modelling of risk factors were made by looking at the results of the logistic regression analyses of COPD. The same decisions were then carried over to the analyses of standardised FEV1/FVC ratio. This attention to model fitting was to help ensure that any heterogeneity observed in the effects of a risk factor was not due to an incorrectly specified model. Thus, age, pack-years of smoking, years working in dusty jobs, years exposed to biomass heating, and years exposed to biomass cooking were modelled as continuous variables, but to allow for possible non-linear effects we first considered a model in each centre in which these variables had quadratic effects, and looked at the results of meta-analysing each quadratic term. In the case of age, years working in dusty jobs, and years exposed to biomass cooking, non-linear effects were homogeneous and non-significant: we therefore used linear effects of these variables in the final regression model. In the case of pack-years of smoking, non-linear effects were heterogeneous and significant overall: we therefore retained a quadratic effect of this variable in the final regression model. The non-linear effect of years exposed to biomass heating was heterogeneous, but non-significant overall. Models which included quadratic effects of years exposed to biomass heating, and models which divided years exposed to biomass heating into three categories, did not fit the data noticeably better at any centre than models which used a linear effect of years exposed to biomass heating, according to Hosmer-Lemeshow goodness-of-fit tests.* In each case the only centre with a significantly poor fit ($P<.05$) was Guangzhou, China. We decided, for simplicity, to model years exposed to biomass heating as a linear effect in the final regression models. We also tested for interactions between sex and all other risk factors in each centre. Of those interactions that could be estimated, none were statistically significant using a Simes procedure to allow for the multiple testing.† We did not therefore include any interactions in the final regression models.

Modelling pack-years of smoking as a categorical variable would have simplified presentation of the effects of smoking in some respects (though they would still have been heterogeneous between centres), but would have left the possibility of residual confounding. Because smoking was known *a priori* to be an important risk factor for COPD, and in order to adjust for its effects as accurately as possible, we preferred to model pack-years of smoking as a continuous variable with a non-linear effect.

Results: forest plots

The graphs included with this supplement are forest plots showing effects of risk factors on stage 1 or higher COPD,‡ adjusted for all risk factors in the regression model. The effect of pack-years of smoking, which was modelled as a non-linear effect, is not shown here because the separate coefficients of pack-years and pack-years squared are difficult to interpret (Table 2 of the main paper shows how pack-years of smoking affected prevalence of COPD at different centres). As well as an overall pooled estimate, pooled estimates are calculated separately for centres with and without data on biomass exposure. ES is effect size, *i.e.* odds ratio.

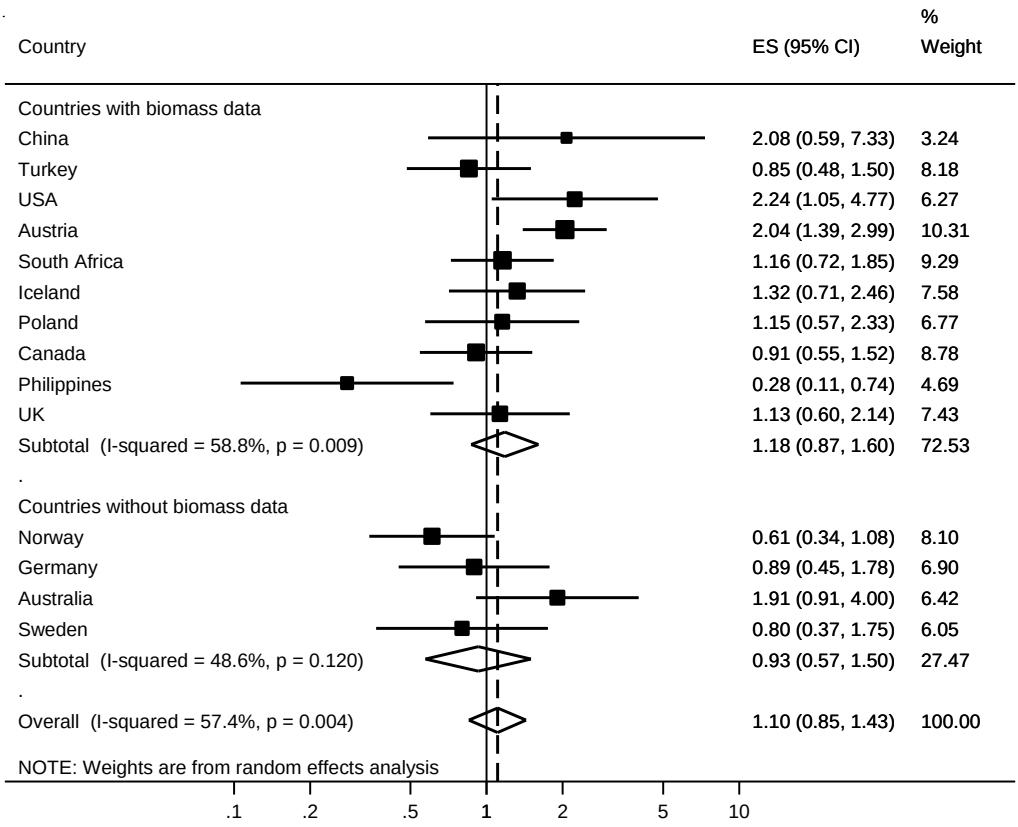
---

* Hosmer DW, Lemeshow S. *Applied logistic regression (2nd ed)*. Wiley (New York) 2000

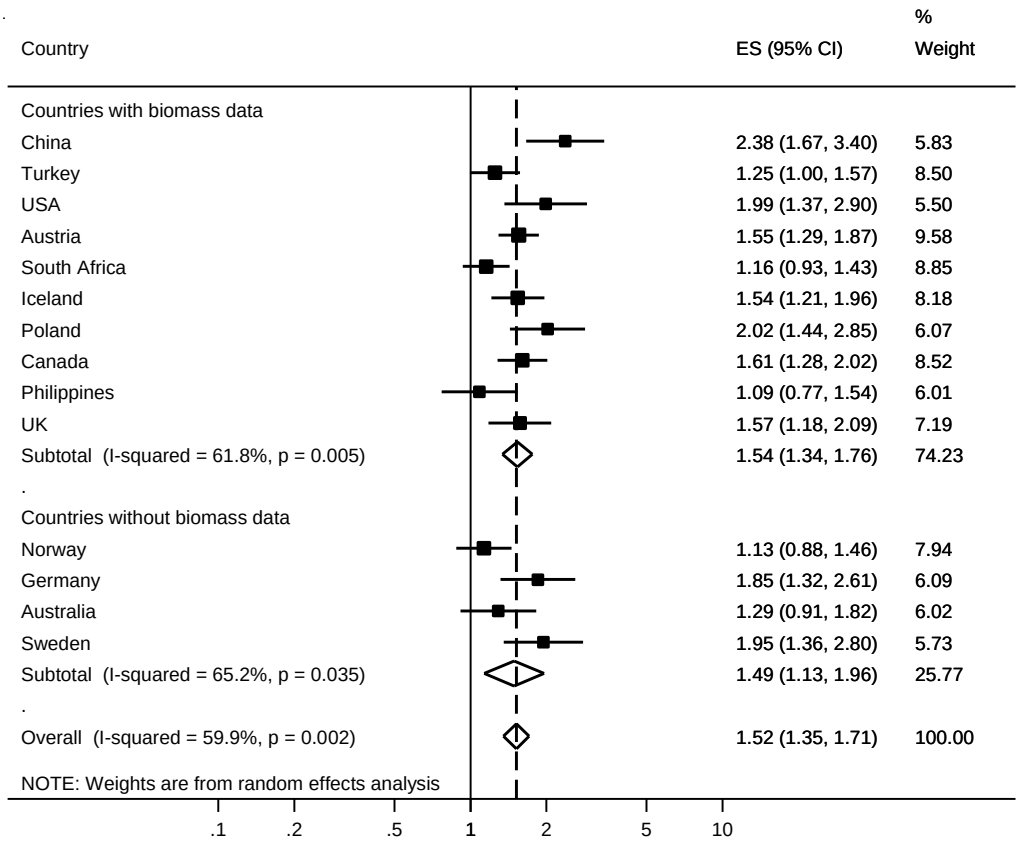
† Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995;57:289-300

‡ Lewis S, Clarke M. Forest plots: trying to see the wood for the trees. *BMJ* 2001;322:1479-1480

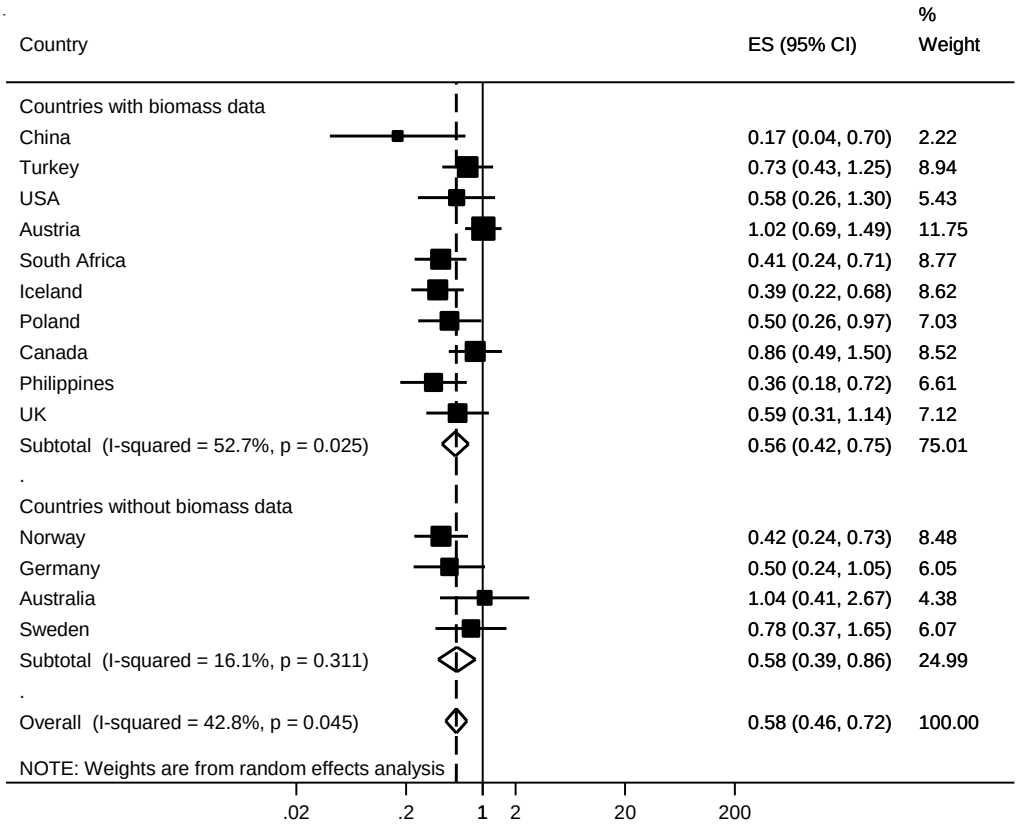
(a) Female sex


Country
ES (95% CI)
% Weight
Countries with biomass data
China 2.08 (0.59, 7.33) 3.24
Turkey 0.85 (0.48, 1.50) 8.18
USA 2.24 (1.05, 4.77) 6.27
Austria 2.04 (1.39, 2.99) 10.31
South Africa 1.16 (0.72, 1.85) 9.29
Iceland 1.32 (0.71, 2.46) 7.58
Poland 1.15 (0.57, 2.33) 6.77
Canada 0.91 (0.55, 1.52) 8.78
Philippines 0.28 (0.11, 0.74) 4.69
UK 1.13 (0.60, 2.14) 7.43
Subtotal (I-squared = 58.8%, p = 0.009) 1.18 (0.87, 1.60) 72.53
Countries without biomass data
Norway 0.61 (0.34, 1.08) 8.10
Germany 0.89 (0.45, 1.78) 6.90
Australia 1.91 (0.91, 4.00) 6.42
Sweden 0.80 (0.37, 1.75) 6.05
Subtotal (I-squared = 48.6%, p = 0.120) 0.93 (0.57, 1.50) 27.47
Overall (I-squared = 57.4%, p = 0.004) 1.10 (0.85, 1.43) 100.00
NOTE: Weights are from random effects analysis
.1
.2
.5
1
2
5
10

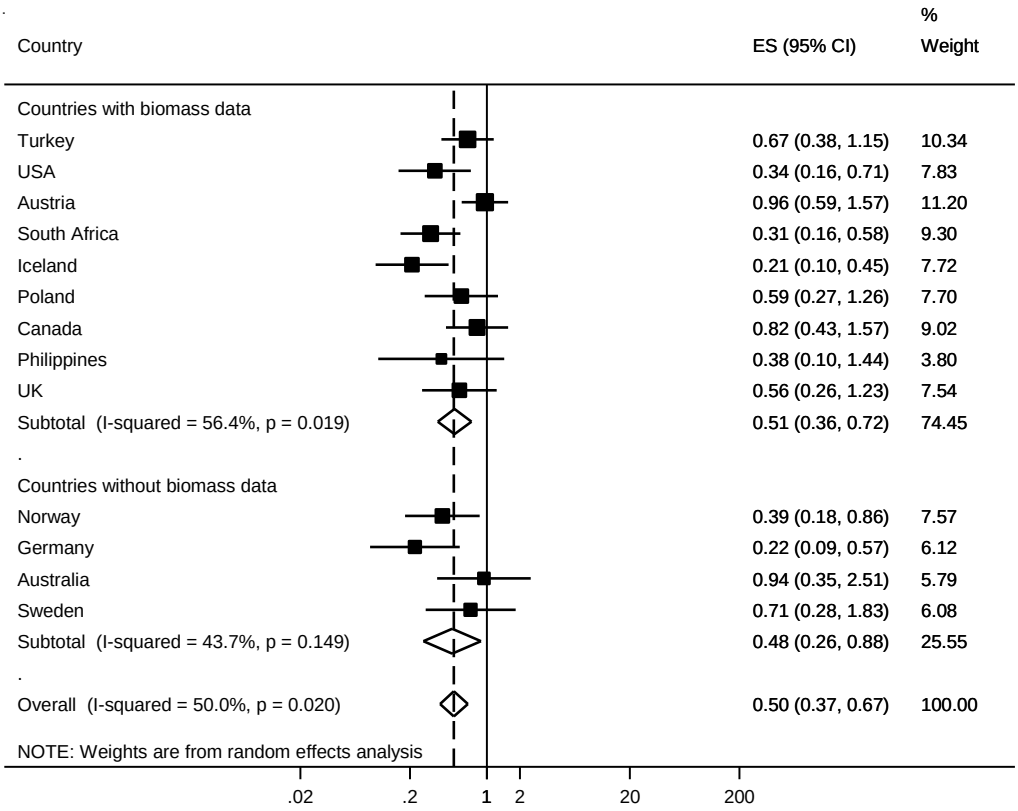
(b) Age (per 10 years, assuming a linear effect)



(c) BMI – overweight compared with normal weight/underweight

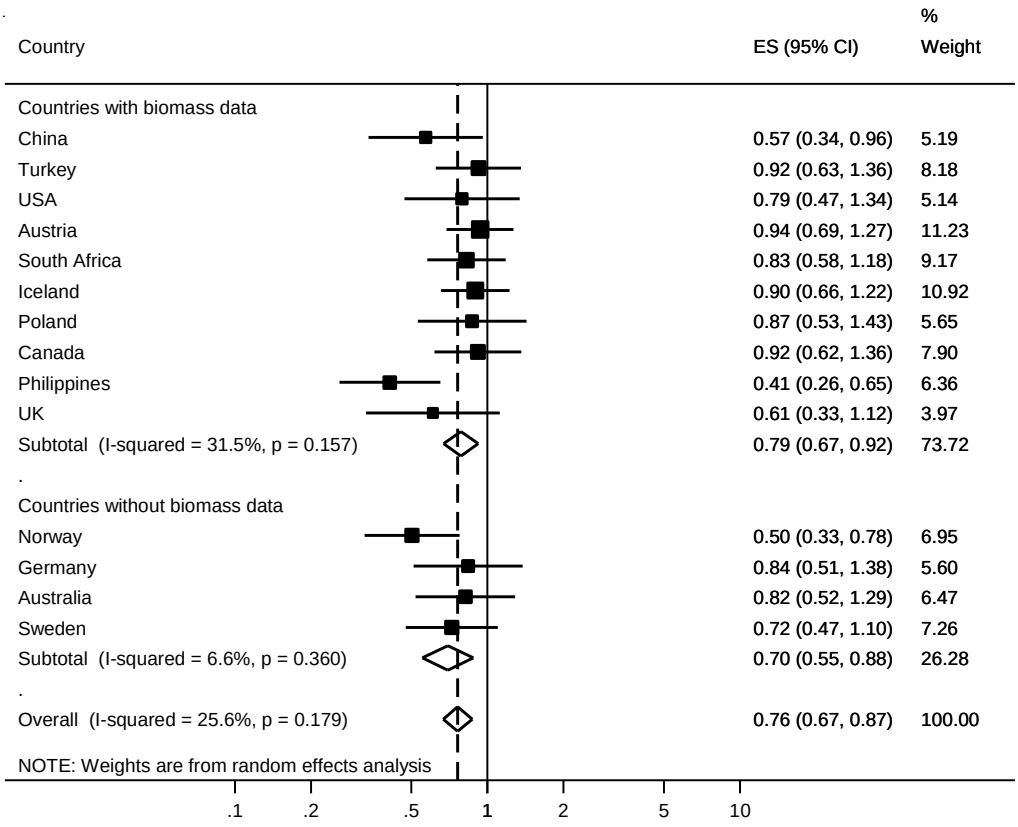


(d) BMI – obese compared with normal weight/underweight

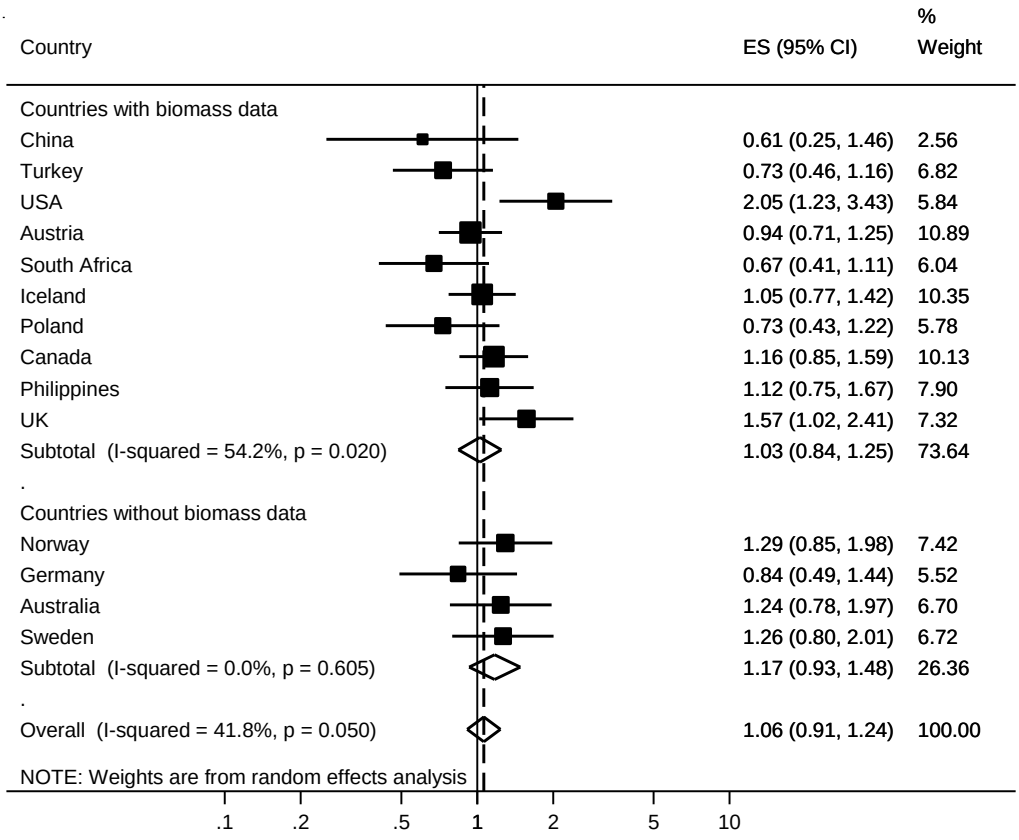


China excluded because none of the 16 obese participants had COPD

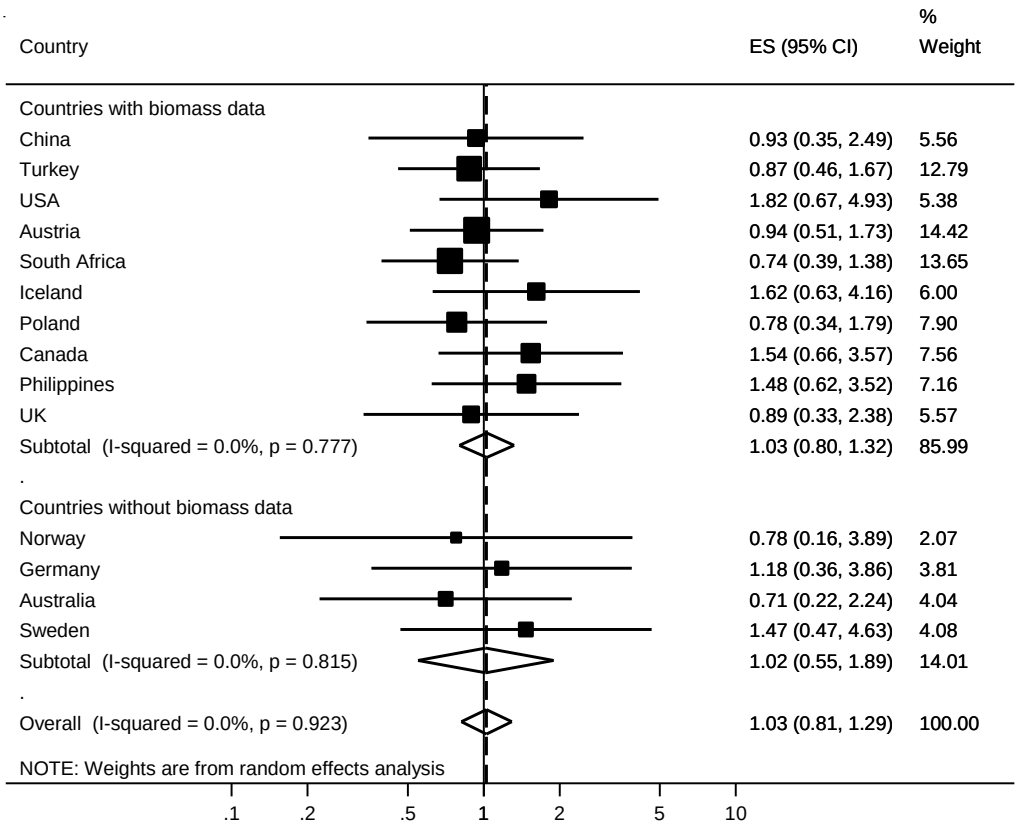
(e) Education (per group, assuming a linear trend across the four groups)



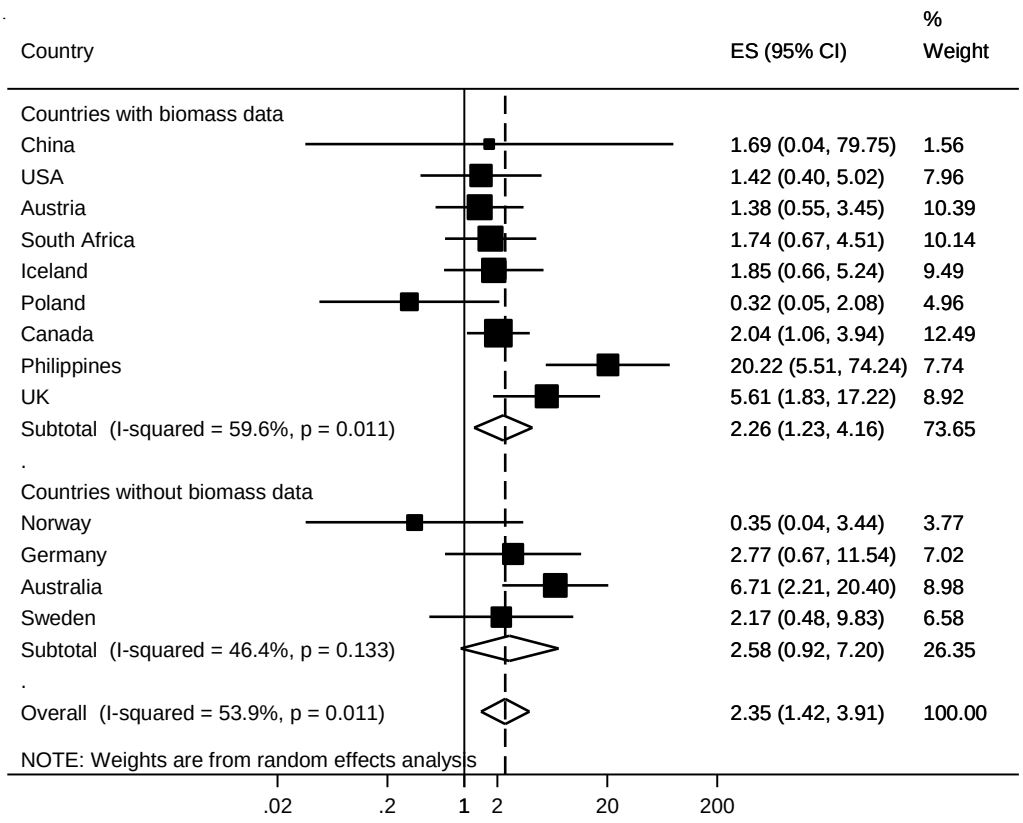
(f) Father's education (per group, assuming a linear trend across the four groups)



(g) Father's education unknown

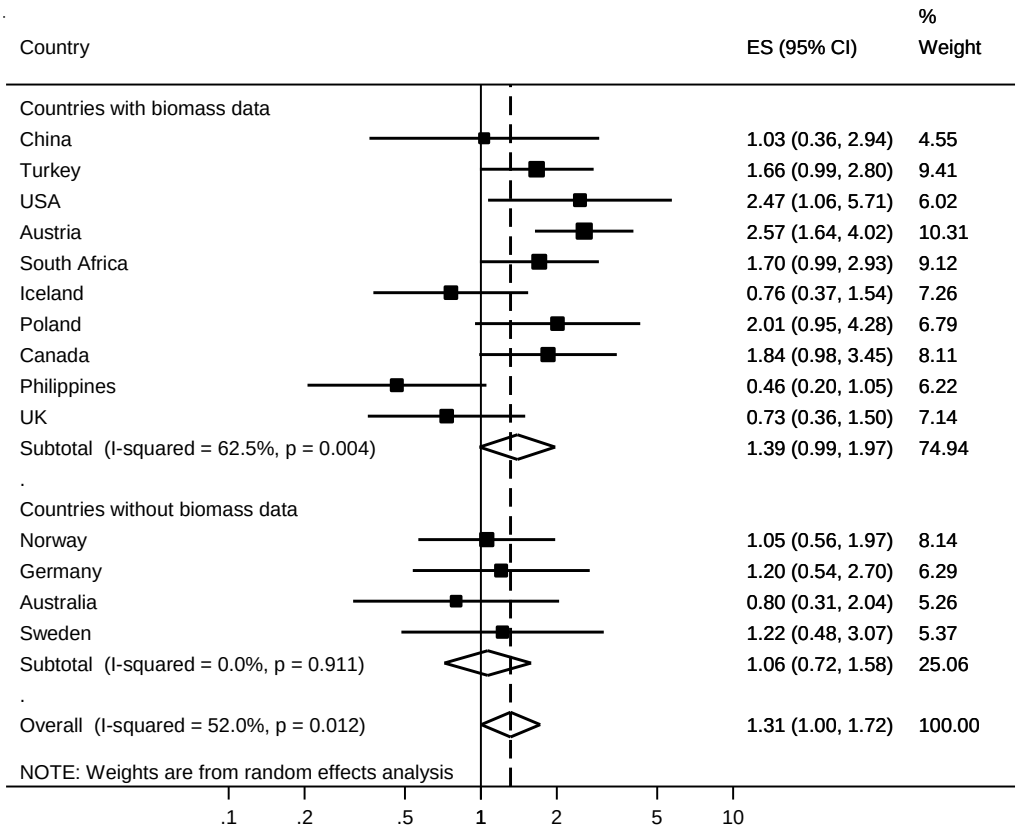


(h) Hospitalisation for breathing problems before age 10 years

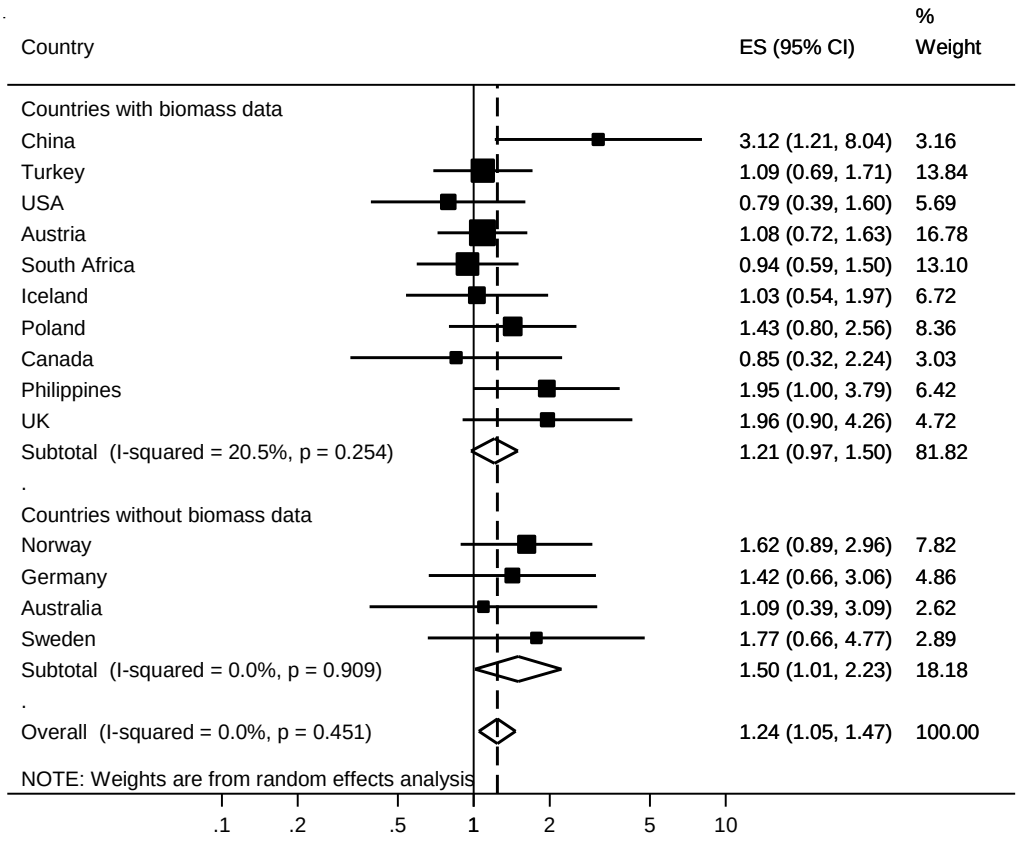


Turkey excluded because none of the 3 participants who were hospitalised had COPD

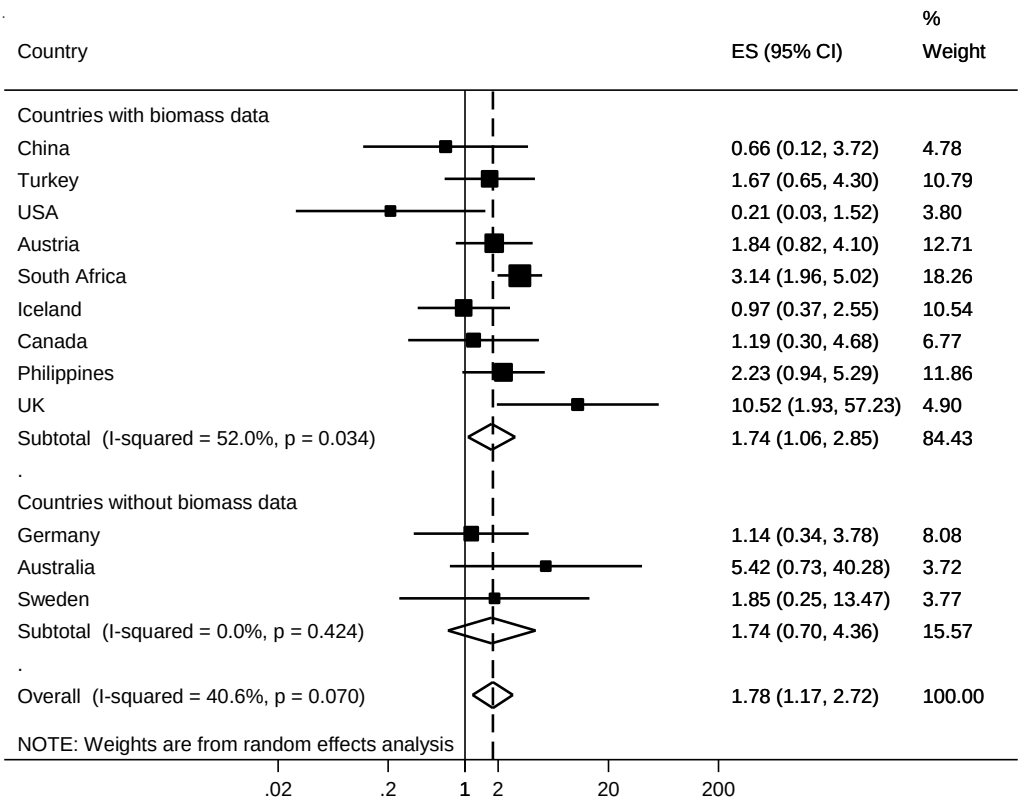
(i) Current smoking



(j) Passive smoking

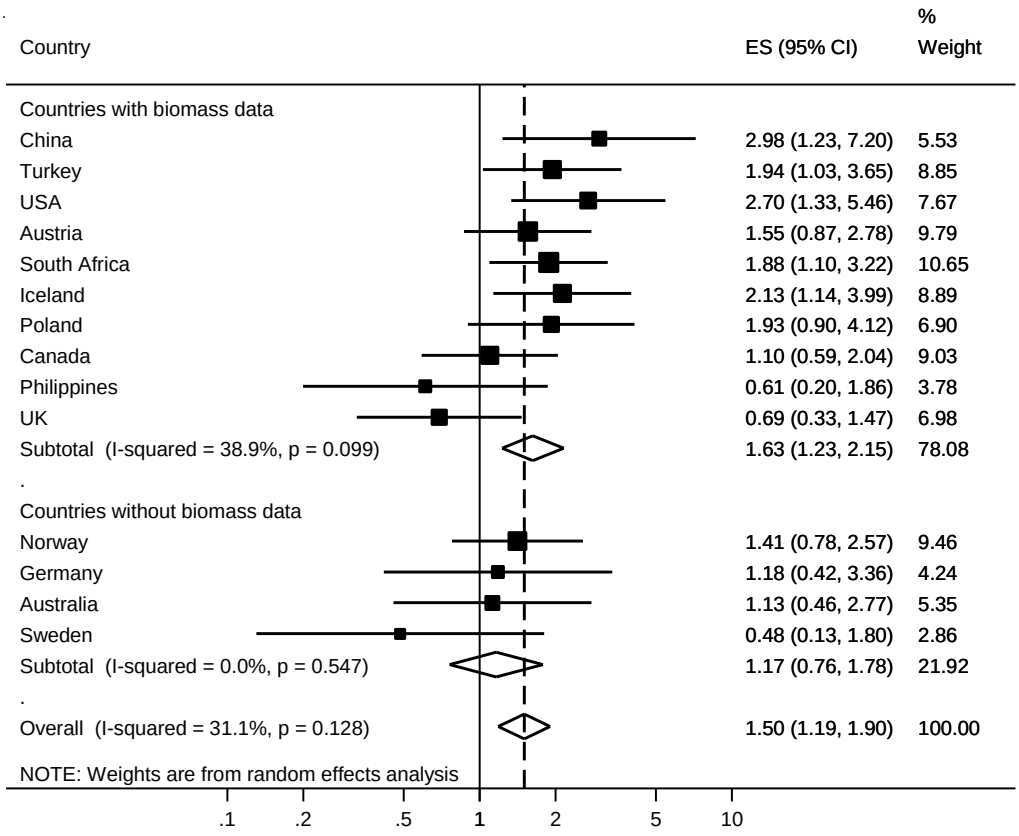


(k) Doctor ever diagnosed tuberculosis

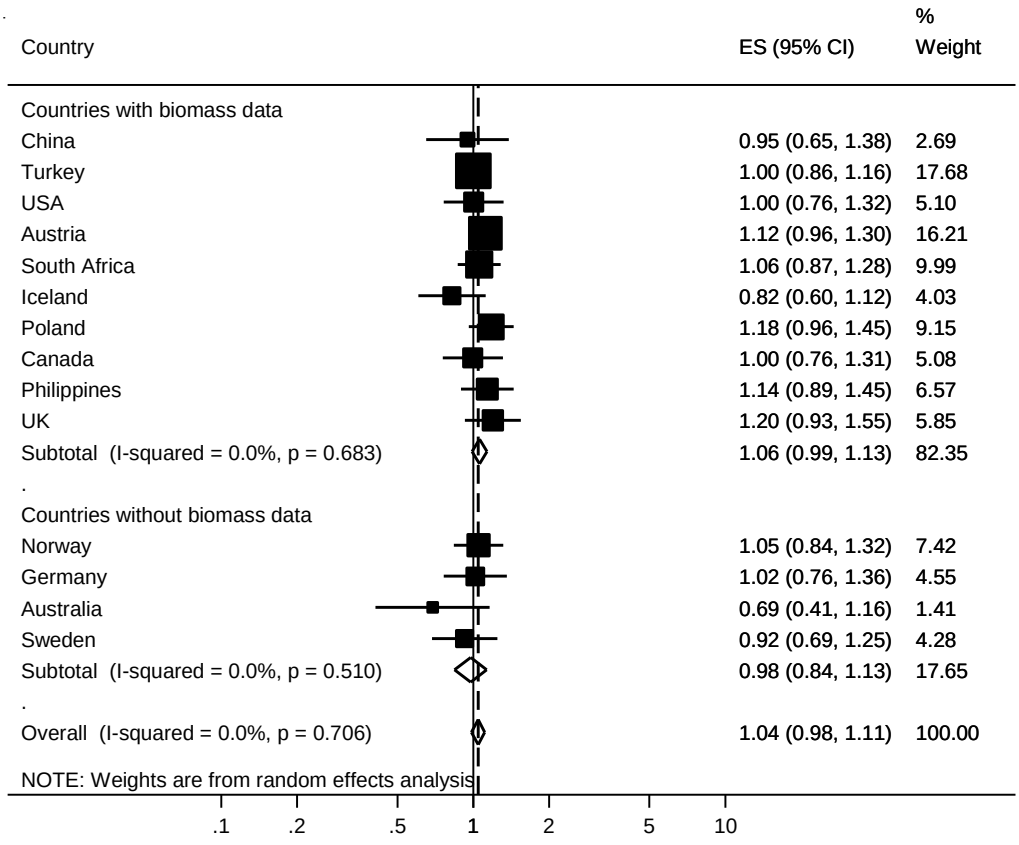


Poland excluded because none of the 14 participants with tuberculosis had COPD
Norway excluded because neither of the 2 participants with tuberculosis had COPD

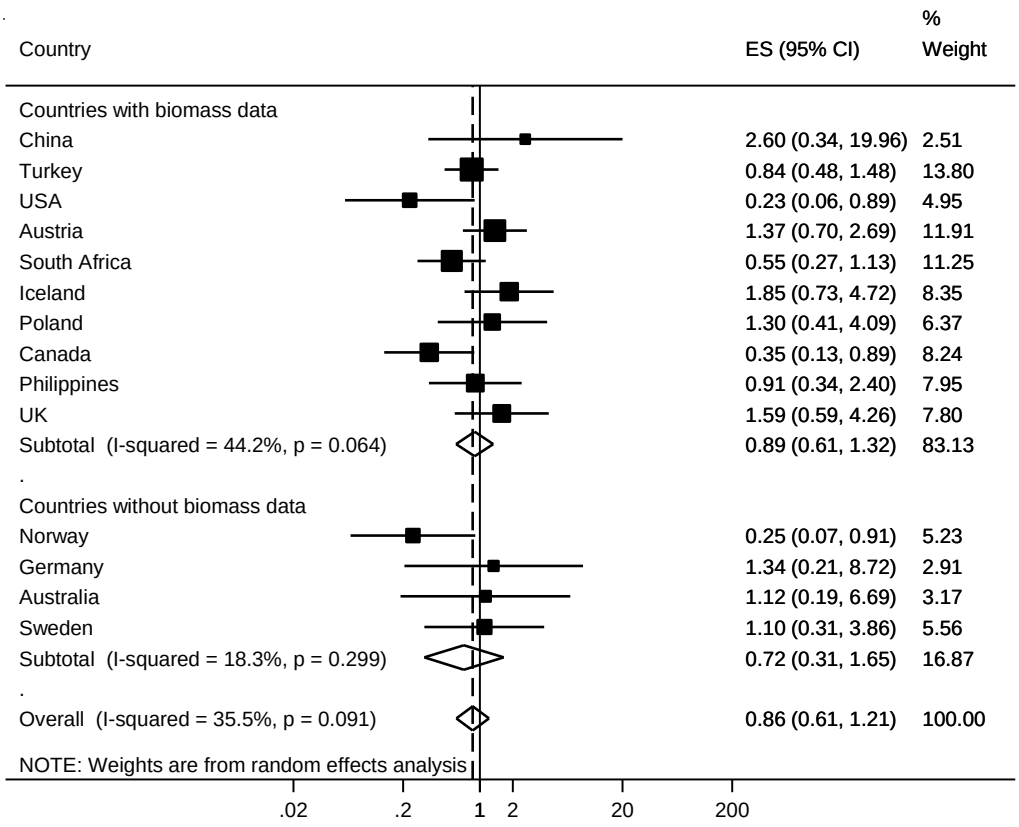
(l) Family history of COPD



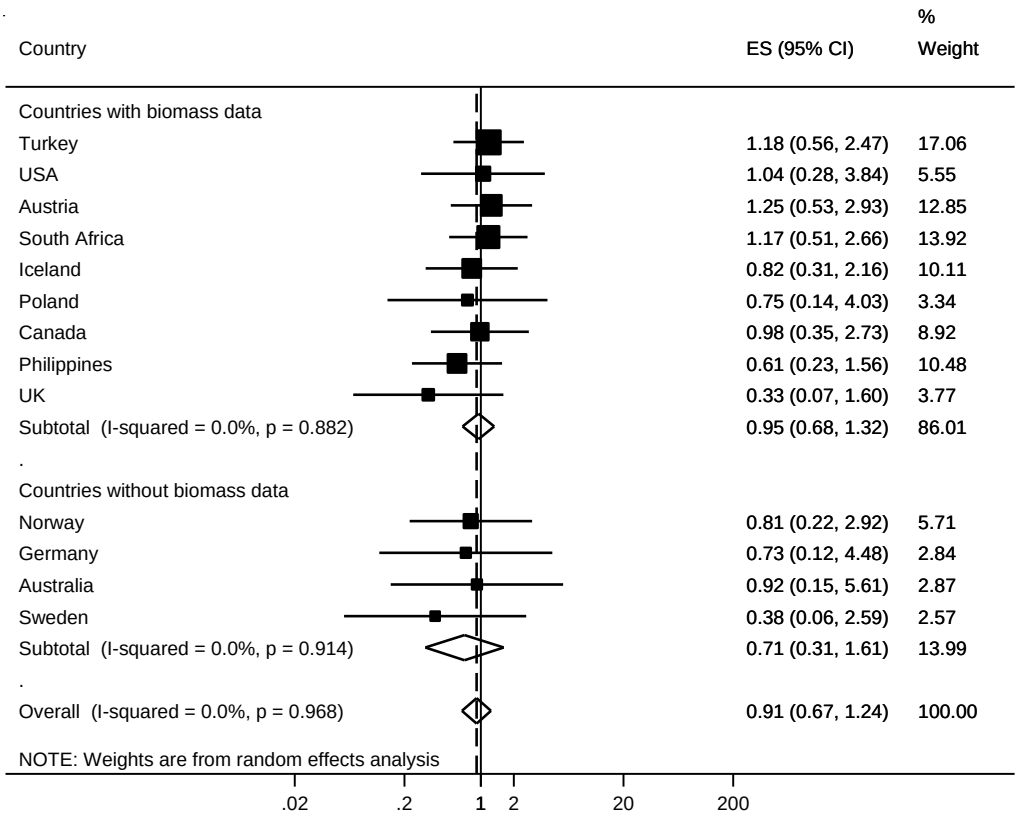
(m) Years working in dusty jobs (per 10 years, assuming a linear effect)



(n) Regular exposure to dust in present job

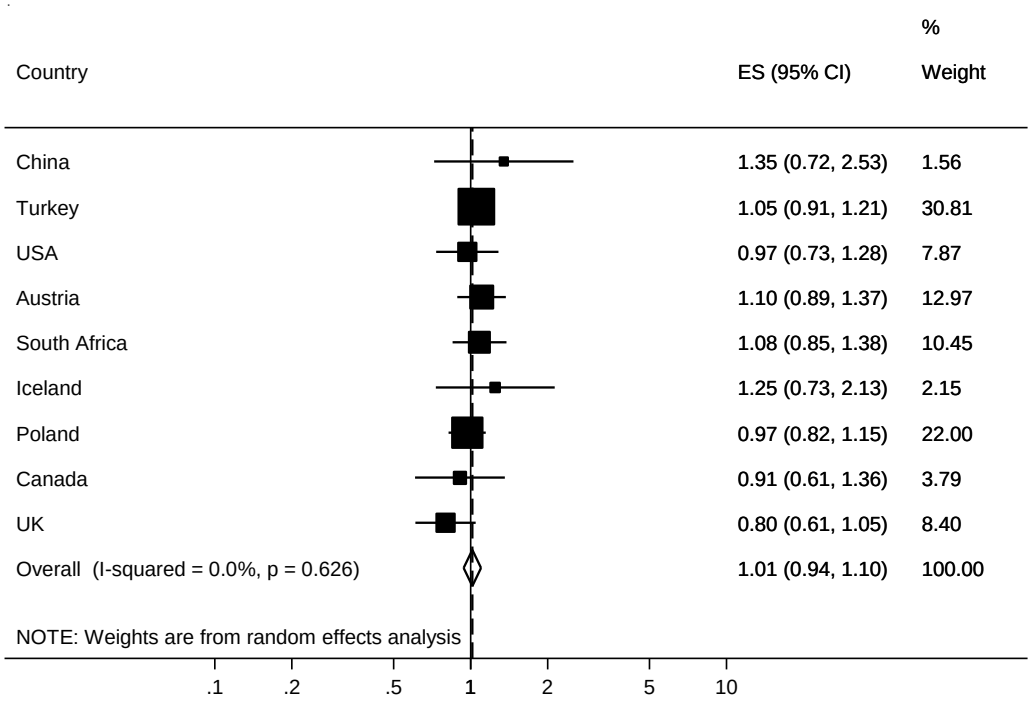


(o) Regular exposure to fumes in present job



China excluded because none of the 7 participants exposed to fumes had COPD

(p) Exposure to biomass fires for heating (per 10 years, assuming a linear effect)



Philippines excluded because none of the 7 participants exposed to biomass fires for heating had COPD

(q) Exposure to biomass fires for cooking (per 10 years of equivalent continuous exposure, assuming a linear effect)