Supplemental Material

Figure S1. Article Identification flow chart following the PRISMA guidelines.[137]

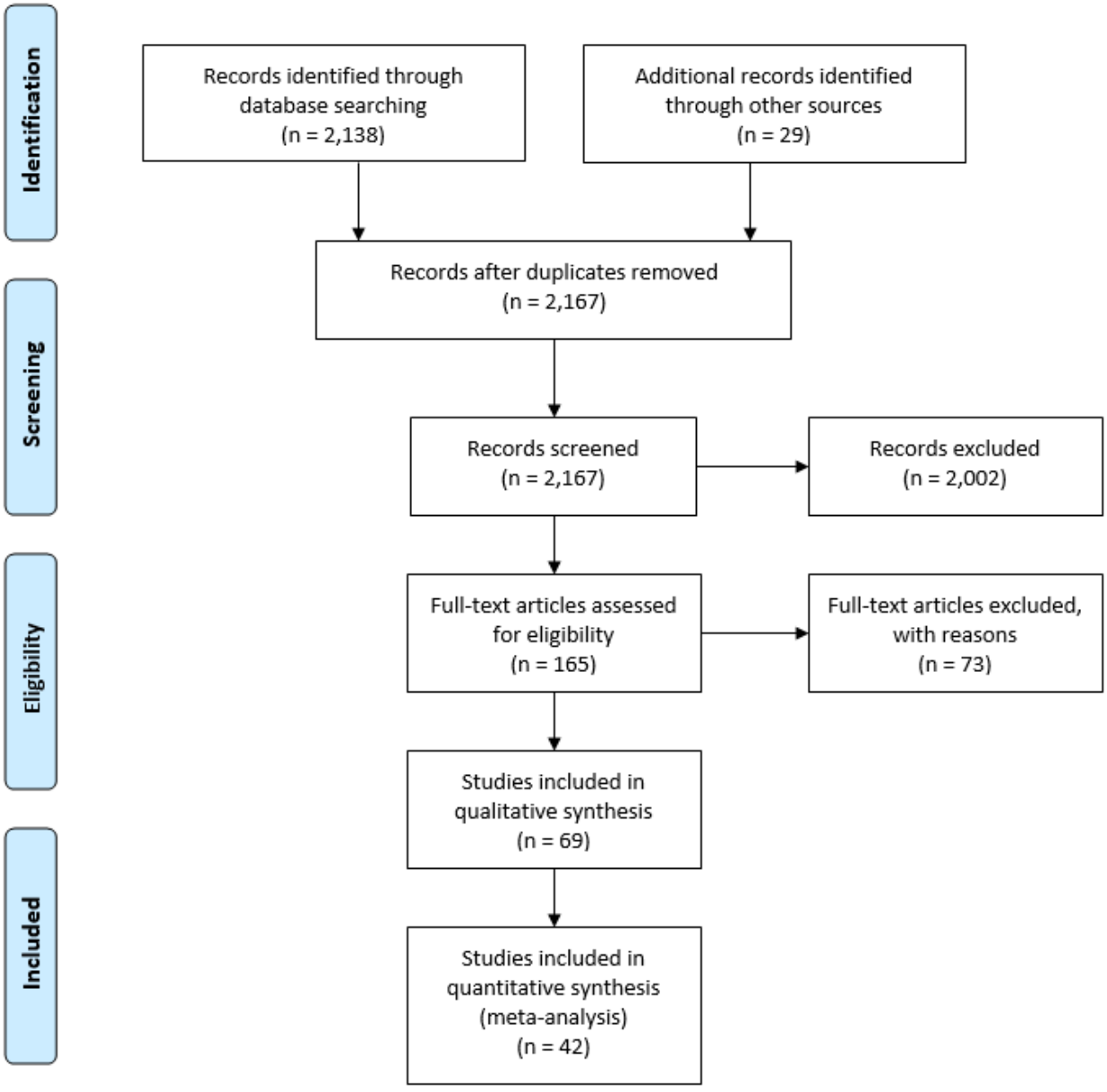

Figure S2. Number of publications by year reporting the association of long-term $PM_{2.5}$ exposure and CVD events.

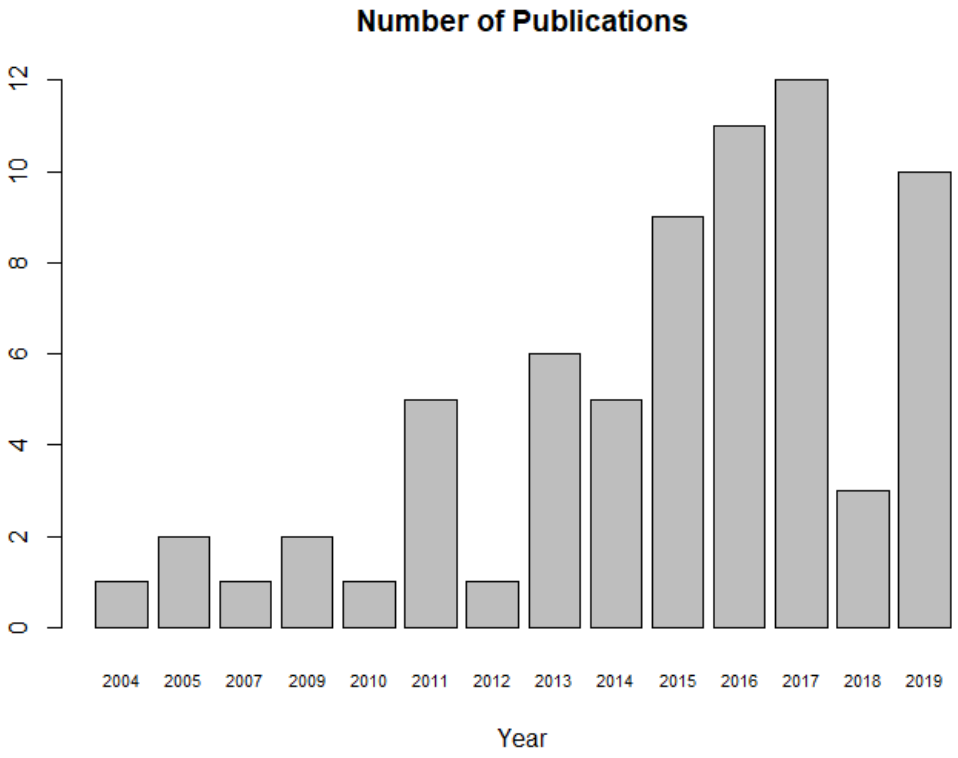

Figure S3. Funnel plots indicating extreme outlier studies in meta-analyses.

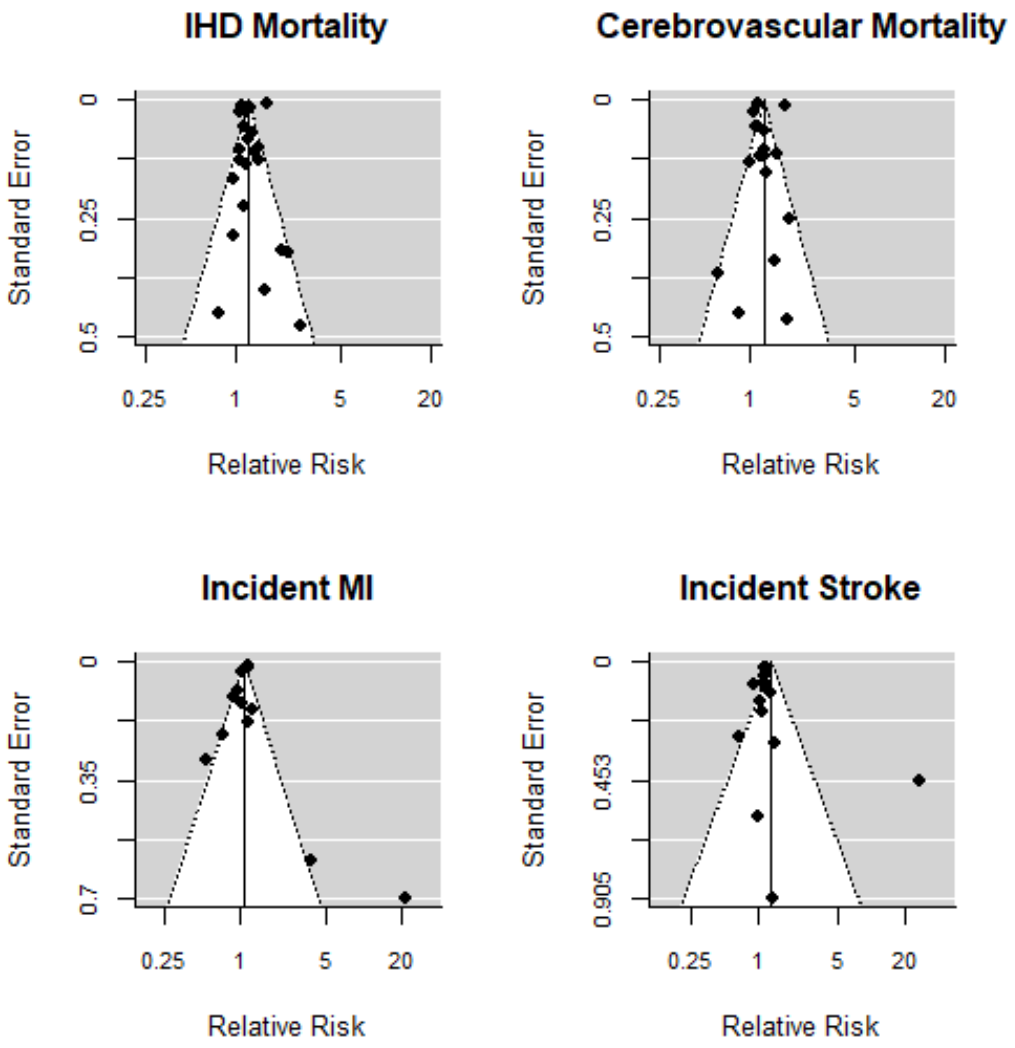

Figure S4. Meta-analysis of the relative risk of cardiovascular mortality per 10 µg/m^3 increase in long-term $PM_{2.5}$ exposure, combining effects of 28 studies.

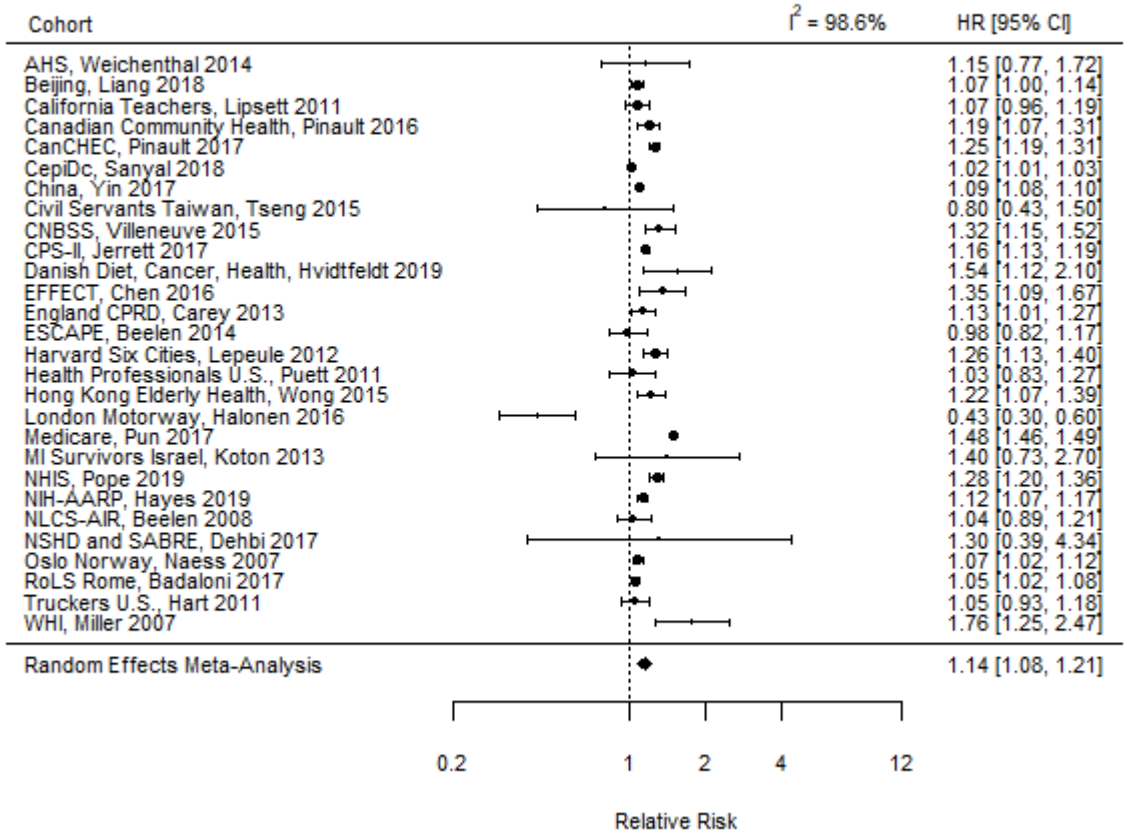

Figure S4 shows the results of our meta-analysis for long-term $PM_{2.5}$ exposure and CVD mortality, combining associations of 28 studies. We found that the combined relative risk of CVD mortality was 1.14 (95% CI 1.08 to 1.21) per 10 µg/m^3 increase in long-term $PM_{2.5}$ exposure. The meta-analysis excluded one extreme outlier study of residents of Seoul, Korea that reported a relative risk of 21.65 (95% CI: 2.95 to 158.88) per 10 µg/m^3 increase in long-term $PM_{2.5}$ exposure (Kim et al. 2017). As a sensitivity analysis, we recomputed the meta-analysis with this outlier study included, which resulted in a combined relative risk of 1.14 (95% CI 1.08 to 1.22). Thus, results of the meta-analysis were not sensitive to the inclusion or exclusion of this extreme outlier study.

Table S1. Newcastle-Ottawa Scale[*] for assessing the quality of nonrandomized studies.

Study, Cohort	Selection[†]	Comparability	Outcome	Total Stars
Atkinson 2013, England CPRD	***	**	***	8
Badaloni 2017, RoLS Rome	***	**	***	8
Bai 2019, ONPHEC	****	**	***	9
Beelen 2014, ESCAPE	***	**	***	8
Cakmak 2018, CanCHEC	***	**	***	8
Carey 2013, England CPRD	***	**	***	8
Cesaroni 2013, RoLS Rome	***	**	***	8
Chen 2005, AHSMOG	***	**	***	8
Chen 2016, EFFECT	**	**	***	7
Crichton 2016, London Stroke Register	****	**	***	9
Dirgawati 2019, HIMS	***	**	***	8
Gan 2011, Vancouver	****	**	***	9
Gandini 2018, Italian Longitudinal Study	****	**	***	9
Hart 2011, Truckers U.S.	**	**	***	7
Hart 2015, Nurses Health Study	***	**	***	8
Hartiala 2016, Cleveland Clinic	***	**	***	8
Hayes 2019, NIH-AARP	**	**	***	7
Huang 2019, China-PAR project	****	**	***	9
Jerrett 2017, CPS-II	**	**	***	7
Lipsett 2011, California Teachers	***	**	***	8
Ljungman 2019, Sweden cohorts	****	**	***	9

Loop 2018, REGARDS	****	**	***	9
Miller 2007, WHI	***	**	***	8
Parker 2018, NHIS	***	**	***	8
Pinault 2016, Canadian Community Health	***	**	***	8
Pinault 2017, CanCHEC	***	**	***	8
Pope 2019, NHIS	***	**	***	8
Puett 2009, Nurses Health Study	***	**	***	8
Puett 2011, Health Professionals U.S.	***	**	***	8
Pun 2017, Medicare	***	**	***	8
Qiu 2017, Hong Kong Elderly Health	***	**	***	8
Shin 2019, ONPHEC	****	**	***	9
Stafoggia 2014, ESCAPE	****	**	***	9
Tseng 2015, Civil Servants Taiwan	**	**	***	7
Turner 2016, CPS-II	**	**	***	7
Villeneuve 2015, CNBSS	**	**	***	7
Weichenthal 2014, AHS	**	**	***	7
Wolf 2015, ESCAPE	****	**	***	9
Wong 2015, Hong Kong Elderly Health	**	**	***	7
Yin 2017, China	**	**	***	7

Table S1 displays the Newcastle-Ottawa Scale for cohort studies. Selection included 4 possible stars (1. Representativeness of the exposed cohort, 2. Selection of the non-exposed cohort, 3. Ascertainment of exposure, and 4. Demonstration that outcome of interest was not present at start of study). Rankings for selection biases varied across studies, with differences mostly driven by representativeness of the exposed cohort (selected groups were not given a star) and demonstration that the outcome of interest was not present at the start of the study (incident events). Comparability included two possible stars for the degree of comparability of cohorts on the basis of the design or analysis. All studies received two stars for comparability as they appropriately controlled for important covariates (specific details are provided in Table 1).

Outcome included three possible stars (1. Assessment of outcome, 2. Length of follow-up, 3. Adequacy of follow up). Regarding the outcome section, all studies received a star for assessment of outcome (reference to secure records and/or use of ICD codes), long enough follow-up time, and adequacy of follow-up.

* The scale includes three categories (Selection, Comparability, and Outcome), each with numbered items ranking the possibility for bias. Stars were given to each numbered item according to study design and quality, with a maximum of 9 stars per study. One study (Madrigano et al. 2013) used a case-control study design and therefore was not compared to the other cohort studies using NOS.

† We gave every study a star for selection of non-exposed cohort ("drawn from the same community as the exposed cohort") and a star for ascertainment of exposure ("secure record"). For these air pollution studies, there are no "unexposed" subjects but instead "less exposed" subjects drawn from the same cohort. All studies also used sufficient sources of air pollution data.