World Health Organization
REGIONAL OFFICE FOR Europe

METHODOLOGY FOR SYSTEMATIC EVIDENCE REVIEWS FOR THE WHO ENVIRONMENTAL NOISE GUIDELINES FOR THE EUROPEAN REGION

By:
Marie-Ève Héroux, WHO Regional Office for Europe;
Jos Verbeek, Finnish Institute of Occupational Health

ABSTRACT

Exposure to environmental noise has been demonstrated to have adverse effects on health. WHO has developed new environmental noise guidelines for the European Region, based on the latest scientific evidence retrieved and assessed using predefined systematic review methodology.

This paper includes a description of the methodology used to conduct these systematic evidence reviews. It includes two protocols: one for the systematic review of health effects resulting from environmental noise and one for the systematic review of noise interventions.

Keywords

EVIDENCE-BASED MEDICINE - METHODS
RESEARCH DESIGN
META-ANALYSIS AS TOPIC
REVIEW LITERATURE AS TOPIC
NOISE - ADVERSE EFFECTS, PREVENTION AND CONTROL
ENVIRONMENTAL EXPOSURE - ADVERSE EFFECTS, PREVENTION AND CONTROL
GUIDELINES AS TOPIC

Address requests about publications of the WHO Regional Office for Europe to:
Publications
WHO Regional Office for Europe
UN City, Marmorvej 51
DK-2100 Copenhagen Ø, Denmark
Alternatively, complete an online request form for documentation, health information, or for permission to quote or translate, on the Regional Office web site (http://www.euro.who.int/pubrequest).

Contents

Acknowledgements

This document has been prepared in the context of the development of the WHO Environmental Noise Guidelines for the European Region, coordinated by the WHO European Centre for Environment and Health.

The authors would like to thank the members of the Guideline Development Group and all the authors of the systematic evidence reviews, performed as part of the development of the WHO environmental noise guidelines for the European Region, for their input and comments during the development of the protocols. We also thank Stephanie Sangalang, PhD candidate, University of Cologne, Germany, and Lydia Wanstall, copy-editor, for assisting with language editing and formatting of this document.

WHO Regional Office for Europe acknowledges funding and in-kind contribution from the Swiss Federal Office for the Environment and the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety, and the German Environment Agency for the work on the WHO Environmental Noise Guidelines for the European Region.

Abbreviations

AMSTAR	a measurement tool to assess systematic reviews
dB	decibel
GATE	graphical appraisal tool for epidemiological studies
GRADE	grading of recommendations assessment, development and evaluation
ICBEN	International Commission on Biological Effects of Noise
NICE	National Institute for Health and Clinical Excellence (United Kingdom)
TNO	Organization for Applied Scientific Research (the Netherlands)

1 Background

Exposure to environmental noise has been demonstrated to have adverse effects on health. WHO has developed new environmental noise guidelines for the European Region, based on the latest scientific evidence retrieved and assessed using predefined systematic review methodology. Systematic reviews commissioned to inform the guidelines assessed both the relationships between environmental noise and health outcomes and the effects on health outcomes of interventions to change exposure to environmental noise.

Exposure to noise can lead to auditory and nonauditory effects on health. Through direct injury to the auditory system, noise leads to effects such as hearing loss and tinnitus. In addition, noise is a nonspecific stressor that has been shown to affect human health adversely, especially following long-term exposure. Nonauditory effects include cardiovascular disease, metabolic diseases, effects on sleep, annoyance, cognitive impairment, quality of life, mental health and well-being, and adverse birth outcomes. They are caused by psychological and physiological distress as well as a disturbance of the organism's homeostasis and increasing allostatic load.

Several interventions to change exposure to noise were often carried out over an extended period of time and might have comprised multiple components. They might have involved multiple governmental (or nongovernmental) sectors including environment, transport, energy and health. Further, such interventions might not have led to immediate changes in noise exposure or health outcomes: significant lag times might exist between the implementation of the intervention and the detection of its effect. This complexity, as well as the multiple, interacting environmental and biological pathways leading to a health response, greatly complicate the assessment of these effects (Burns et al., 2014).

This paper includes a description of the methodology used to conduct the systematic evidence reviews. It includes two protocols: one for the systematic review of health effects resulting from environmental noise and one for the systematic review of noise interventions. Specific results from the evidence reviews are available as open access papers in the Special Issue "WHO Noise and Health Evidence Reviews" of the International Journal of Environmental Research and Public Health (IJERPH) at http://www.mdpi.com/journal/ijerph/special_issues/WHO_reviews.

2 Methods

2.1 Protocol for the systematic review of health effects resulting from environmental noise

2.1.1 Description of health conditions

The evidence was reviewed for the following health outcomes:

- effects on sleep
- annoyance
- cognitive impairment
- quality of life, mental health and well-being
- cardiovascular and metabolic diseases
- hearing impairment and tinnitus
- adverse birth outcomes.

2.1.2 Description of exposure to environmental noise

Environmental noise was broadly defined as noise emitted from all sources except sources of occupational noise exposure in workplaces. The following keywords (see Table 1) were used to describe environmental noise.

Table 1. Environmental noise keywords indicating suitable types of noise exposure

Aircraft noise Airport noise Classroom noise Combined exposure to noise and air pollution Combined exposure to noise and vibration Combined noise exposure Community noise Entertainment noise Environmental noise High-volume music High-volume noise Hospital noise	Household noise Leisure noise Leisure-time noise Low-frequency noise Motor vehicle noise Neighbourhood noise Noise exposure Noise from children's toys Noise from mobile phones Noise from mp3 players Noise from personal audio devices Noise from personal electronic devices	Noise from personal music players Noise load Noise nuisance Railway noise Road traffic noise School noise Traffic noise Train noise Transportation noise Truck noise Wind farm noise Wind farm sound Wind turbine noise Wind turbine sound

2.1.3 Objectives

The key objectives of the evidence review were to assess the strength of association between exposure to environmental noise and incidence or prevalence of adverse health effects and, where possible, to quantify the risk of these health effects with an incremental increase in noise exposure. The main research question was:

In the general population exposed to environmental noise, what is the exposure-response relationship between exposure to environmental noise (reported as various noise indicators) and

the proportion of people with a validated measure of health outcome (see section 2.1.4 for types of outcome measure), when adjusted for main confounders?

2.1.4 Criteria for considering studies for this review

To be considered for inclusion in the review, studies had to meet the following criteria:

Types of study considered were prospective and retrospective cohort studies, case-control studies and observational or experimental cross-sectional studies of people exposed to environmental noise. Where relevant – for example, for the health outcome "cardiovascular and metabolic diseases" – ecological studies were also included.

Types of study participant considered were members of the general population, as well as specific segments of the population particularly at risk, such as children or vulnerable groups. Studies including participants exposed to noise in occupational settings were included only if relevant – for example, if they considered combined occupational and environmental noise exposures.

Types of exposure measurement considered were noise exposure levels either measured or calculated and expressed in decibel (dB) values that aimed to be representative of the individual exposure of the study participants (for most observational studies, this would be the dwelling location or home). Calculated levels for transportation noise (road, rail, air) had to be based on traffic data reflecting the use of roads, railway lines and in- and outbound flight routes at airports. Studies that used hearing loss or defective hearing as a proxy for (previous) noise exposure were excluded. Surveys that assessed noise exposure on the basis of subjective ratings, such as those given by subjects in a questionnaire, were excluded.

Types of confounder (other risk factors that may confound the relationship between exposure to noise and a health outcome): no inclusion or exclusion criteria were applied; however, for every study, the possible confounders taken into account were assessed.

Types of outcome measure considered were assessment of the following seven primary outcomes:

- effects on sleep, such as insomnia (trouble in the initiation or maintenance of sleep for at least 15 days in a month), sleep medication use, subcortical and autonomic arousals (e.g. increases in blood pressure or heart rate) during sleep, cortical arousals during sleep, probability/number of awakenings, self-reported sleep disturbance, sleep duration and quality, changes in waking or daytime cognitive performance following exposure to nocturnal noise, morning or daytime tiredness/fatigue, perceived well-being, mood changes and injury;
- annoyance, such as proportion of self-reported annoyed or highly annoyed people, average self-reported annoyance assessed on a continuous (if possible, standardized) scale, activity disturbance (communication (including speech interference), recreation, rest, work at home) and all annoyance other than that relating to sleep;
- cognitive impairment, such as reading and oral comprehension in children, short-term and long-term memory in children, measures of attention in children, impairment assessed through standardized assessments such as standard assessment tasks, cognitive impairment in the elderly and working age population (reduced concentration, speech intelligibility, etc.), executive function deficit (working memory capacity, reasoning, task flexibility, problem solving) and hyperactivity;
- quality of life, mental health and well-being, such as self-reported quality of life (well-being, health status, vitality) using assessments such as the short-form health survey, general health questionnaire,

WHO full and abbreviated quality of life and health-related quality of life assessments, medication intake for treatment of anxiety and depression, self-reported depression, anxiety and psychological symptoms (scale), interview measures of depressive and anxiety disorders, hospital admission data for psychiatric disorders, emotional and conduct disorders in children (e.g. assessed by instruments such as the strengths and difficulties questionnaire and a revised questionnaire to assess health-related quality of life in children and adolescents), helplessness and behavioural issues;

- cardiovascular and metabolic diseases, such as ischaemic or coronary heart disease including myocardial infarction and angina pectoris, stroke, cardiovascular mortality, heart failure, hypertension (self-reported, doctor-diagnosed, medication use, blood pressure readings in accordance with WHO criteria), mean heart rate variability, mean blood pressure (children), metabolic syndrome, waist circumference, obesity, diabetes and hormonal response (cortisol or adrenaline or nor-adrenaline or epinephrine or nor-epinephrine) in blood, urine and other media (e.g. saliva);
- hearing impairment, such as hearing loss, hearing impairment and tinnitus;
- adverse birth outcomes, such as gestational age/prematurity/ preterm birth, low birth weight (term low birth weight, small for gestational age) and developmental indices (growth retardation, cognitive/behavioural development).

2.1.5 Search for available systematic reviews

Initially, an information specialist performed a search of all available systematic reviews and meta-analyses on environmental noise. The databases searched included Medline/PubMed, Scopus (including Embase), PsycINFO, Web of Science, the database of the Organization for Applied Scientific Research of the Netherlands (TNO) and ScienceDirect.

The systematic review team included any other systematic reviews of which they were aware, and consulted the publications databases of recognized institutions to identify systematic reviews in technical reports that had not been published in scientific journals. They also hand-searched reference lists of relevant literature reviews and articles.

Online conference proceedings – such as for the International Commission on Biological Effects of Noise (ICBEN) and Inter-Noise – were not consulted as they were unlikely to include full systematic reviews containing enough detail (such as about comprehensive literature searches performed) to assess suitably for quality. ICBEN summaries by the organization's international noise team chairs were used, as relevant, to identify potential systematic reviews.

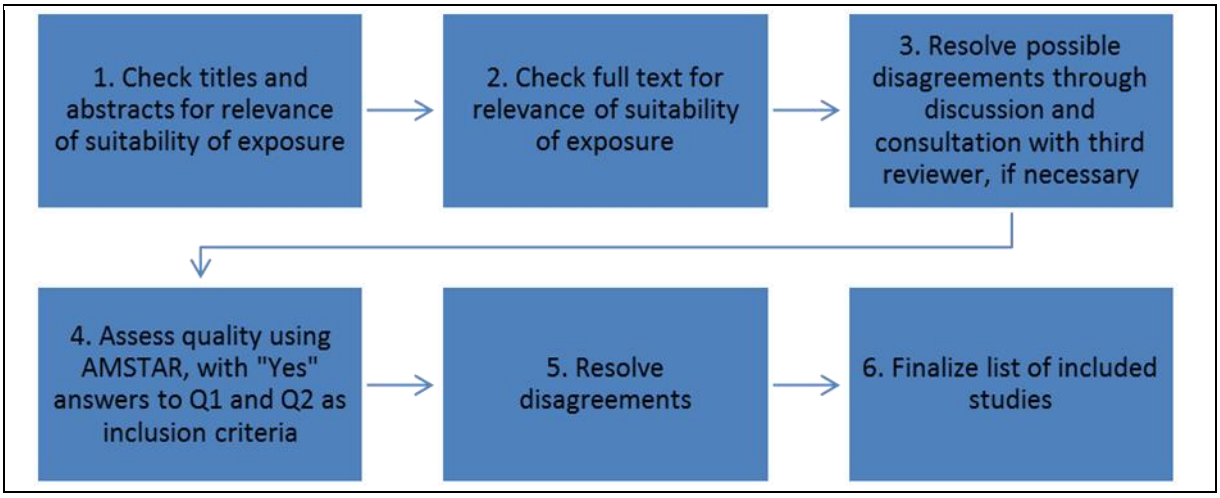
The search included systematic reviews and meta-analyses published in or after 2000 and aimed to include papers in all languages. In the unlikely case that the full text of a paper was not available, the paper's corresponding author was contacted and asked to provide it. If this was not successful, the paper was excluded. When a paper was available only in a language with which the systematic review team was not familiar, they consulted other noise experts outside the group to assess the quality of the review and the suitability of the references.

2.1.6 Selection and quality assessment of available systematic reviews

The systematic review team assessed and documented the suitability and quality of the systematic reviews retrieved to decide whether they were relevant for the purposes of the guidelines. The assessment (see Fig. 1) was performed independently, in duplicate, by two authors.

Each author first checked whether the titles and abstracts of systematic reviews that came up in the search were related to environmental noise (i.e. the relevance of suitability of exposure; see section 2.1.4) – if not, they were excluded. In addition, multiple articles that described the same study were excluded to avoid duplication, as each study would be reviewed only once. As a second step, the full texts of the articles resulting from this selection were assessed to see which studies should be included. Any disagreement on inclusion was resolved by discussion. If no consensus was reached, a third reviewer was consulted. This resulted in a list of systematic reviews to assess further for quality.

Fig. 1. Process of selecting and assessing the quality of available systematic reviews



For the quality assessment, the following questions from AMSTAR (a measurement tool to assess systematic reviews; see Annex 1) were used to establish inclusion or further consideration.

Q1. Was an “a priori” design provided? (The research question and inclusion criteria should be established before conducting the review.)

Q2. Was a comprehensive literature search performed? (At least two electronic sources should be searched. The report must include years and databases used (e.g. Medline/PubMed, Scopus (including Embase). Keywords and/or medical subject heading terms must be stated and, where feasible, the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers or experts in the particular field of study, and by reviewing the references in the studies found.)

The minimum criterion for inclusion was a positive answer to both questions. Again, any disagreement on inclusion was resolved by discussion and if no consensus was reached, a third author was consulted. Once the final list of systematic reviews to be included was prepared, the systematic review team needed to decide how to best to use them for this review, whether by:

- using the results as provided in the original;
- updating the results; or
- conducting a new or modified systematic review.

Expert judgement was used to decide how the results of this search affected the search strategy for individual studies, based on the quality of the systematic reviews (as per the answers in the AMSTAR template), as well as the coherence between the main research question of the evidence reviews (see section 2.1.3).

2.1.7 Search for individual studies

The results of the search for available systematic reviews were used to design the search strategy for each systematic evidence review, which was adapted to retrieve relevant individual papers accordingly. The systematic review team checked the references from existing systematic reviews and explored whether additional papers had been mentioned in the many available expert commentaries published in response. A search strategy appropriate for the various databases was developed using the criteria provided in section 2.1.4, especially focusing on the types of study, exposure and outcome measures.

Online conference proceedings for ICBEN and Inter-Noise, supranational and national reports, identified through a search of publication databases of recognized institutions – such as TNO; the Netherlands National Institute for Public Health and the Environment; United Kingdom Department for Environment, Food and Rural Affairs; German Federal Environment Agency; National Aeronautics and Space Administration; and Imperial College London – WHO publications and European Union reports about specific projects were initially considered for inclusion but generally not consulted due to time constraints and limited availability of resources.

The systematic review team searched for individual papers published since the last suitable systematic review available for a particular outcome measure. Where no systematic review had ever been conducted, the search was conducted with no restriction on the date of publication. The aim was to include papers in all languages.

2.1.8 Selection of individual studies

For each systematic review of a health outcome, two reviewers independently checked whether the titles and abstracts that came up in the search failed to fulfil one or more of the inclusion criteria (see section 2.1.4), and could thus be excluded. The full texts of the remaining articles were assessed by the same two reviewers to see which fulfilled all inclusion criteria (see Annex 2 for a list of inclusion and exclusion criteria). Any disagreement on inclusion was resolved by discussion. If no consensus was reached, a third reviewer was consulted. Multiple articles that described the same study were excluded to avoid duplication, as each study would be reviewed only once. This resulted in a list of individual studies to be included, presented separately for each systematic review.

2.1.9 Data extraction and management

Two reviewers then independently extracted data from the articles on the inclusion list, including study characteristics (such as design, country of origin, year of study), participants (such as number, response rate), noise exposure characterization, health outcome and confounding factors.

As both the data extraction process and the resulting data were different for all the reviews, they are described separately in the various evidence reviews. All data extractions were done in duplicate by the two reviewers independently. Any disagreement was resolved by discussion. If no consensus was reached, a third reviewer was consulted. The studies and their main characteristics were listed in table format to be able to derive the most appropriate comparisons. Please refer to the separate evidence reviews for a description of the individual studies and their main characteristics.

2.1.10 Assessment of risk of bias in included studies

A checklist (see Annex 3) to assess the quality of observational studies was adapted from Shamliyan et al. (2010a; 2010b) and was used as a starting point to develop a risk of bias assessment tool to meet the specific needs of the reviews. The most important characteristics related to the quality of the studies were reported separately. A predefined set of criteria was used to identify studies with a high risk of bias.

2.1.11 Measures of health effects

If the health outcome was dichotomous, relative risks or odds ratios (as relevant) were used as estimates of the effect of exposure. In exceptional cases, another measure of health effects was used (such as awakenings, where the measure was the probability between 0 and 1). If odds ratios were reported and the outcome prevalence was higher than 10%, they were recalculated as relative risks. Standardized mortality ratios from register-based studies were also included. If the health effects were measured on a continuous scale, mean differences were used.

2.1.12 Dealing with missing data and data transformation

If data necessary for analysis were missing from the articles, the systematic review team asked the article authors for additional information. If they could not be reached or if, for example, standard errors were needed but only p-values were available, the systematic review team tried to calculate the missing data from the available statistics. If these were not available or could not be obtained, the original results were reported.

The generalized least squares for trend estimation method (a Stata procedure described by Orsini, Bellocco & Greenland (2006)), or an acceptable equivalent, was used to transform category-specific risk estimates into an incremental risk estimate. First, a linear relationship between the natural logarithm of relative risk and increasing exposure to noise was assumed. The data were then tested to establish whether a reason existed to assume another relationship, such as logarithmic or cubic. The systematic review team assigned a single exposure value to each noise exposure category: for closed categories we assigned the midpoint and for open categories the median value, assuming a normal distribution for noise exposure. We used either a 5 or 10 A-weighted dB noise exposure as one incremental step of increased exposure.

2.1.13 Assessment of heterogeneity

The systematic review team first assessed studies for similarity of participants, exposure and outcome measurement. Children up to 18 and adults over 18 years of age were considered separately. Noise was categorized according to the type of noise exposure described in section 2.1.4. All outcome measurements that fulfilled the inclusion criteria were deemed similar.

Statistical heterogeneity is useful in estimating consistency between studies included in a meta-analysis (Ryan & Cochrane Consumers and Communication Review Group, 2014). Consistency is important because it indicates that differences in reported results are genuine, rather than caused by chance. The systematic review team assessed statistical heterogeneity by means of the I^2 statistic. The I^2 values of 25%, 50% and 75% were taken as low, moderate and high degree of heterogeneity, respectively. When enough data were available, funnel plots were used to assess heterogeneity visually (without statistical testing, for example, as can be done with the Egger test), as relevant.

2.1.14 Assessment of reporting biases

The reviews tried to avoid language bias by including studies in any language. A wide range of studies were initially considered, including supranational and national studies and project reports that had not been published in scientific journals. When there were enough data available, publication bias was assessed using a funnel plot and applying the Egger test (Egger et al., 1997) to the included studies.

2.1.15 Data synthesis

Where possible, effect sizes were combined per 5 or 10 A-weighted dB increase. For continuous outcomes, other feasible estimates of exposure – such as mean differences – were considered. The natural logarithms of the relative risk or odds ratio most adjusted by the authors were used as input for a random effects meta-analysis. Studies with different study designs were generally analysed separately. The systematic review team used software appropriate for data analysis, such as Stata. Where statistical pooling was not possible, a narrative synthesis of the data was produced.

2.1.16 Subgroup analysis and investigation of heterogeneity

The authors evaluated whether the outcomes varied according to types of participant, including vulnerable subgroups. They also checked whether results differed between older and more recent studies, and between participants from different countries.

2.1.17 Sensitivity analysis

The authors evaluated whether the results were sensitive to the inclusion of low-quality studies with a high risk of bias by undertaking separate analyses for studies with low and high risk of bias. They also evaluated how sensitive the results were to assumptions made about the level of exposure. They used random effects meta-analysis to check how sensitive the results were to the model assumptions.

2.1.18 Grading/strength of quality of the evidence

The GRADE (grading of recommendations assessment, development and evaluation) approach was adapted to assess the overall quality of evidence (Guyatt et al., 2008). GRADE allowed the quality of the body of evidence to be systematically and transparently assessed for each outcome grouping, based on specific factors (see Table 2).

Table 2. Factors determining the quality of evidence

Factors decreasing quality of evidence	Factors increasing quality of evidence
Study limitations	Large magnitude of effect
Inconsistency of results	Plausible confounding, which would reduce a demonstrated effect
Indirectness of evidence	Presence of dose-response gradient
Imprecision	
Publication bias	

Based on these criteria, each outcome grouping was graded (see Table 3).

Table 3. Grading of outcome groupings

Quality grading	Description

High quality	Further research is very unlikely to change our confidence in the estimate of effect
Moderate quality	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low quality	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very low quality	Any estimate of effect is uncertain

The main adaptations made to the GRADE approach, which properly reflect the understanding of the Guideline Development Group (GDG) on the quality of the evidence on environmental noise, are the following:

1. Start level for grading of studies: The best available study designs to assess a particular outcome would start as high quality:

In most cases, this is a longitudinal observational study design (cohort or case-control study), evaluating an exposure-response relation with sufficient follow-up time for the outcome of interest to develop. The remaining epidemiological studies would start as low (cross-sectional studies) or very low (ecological studies) quality. However, for annoyance, the best study design is considered to be cross-sectional, as it is the only study design available to assess this particular outcome.

Once the decision has been made on the starting point of the level of evidence, the quality would be upgraded or downgraded according to the various GRADE criteria described earlier.

Upgrading for magnitude of effect: According to the original GRADE, upgrading for magnitude of effect is possible when a RR is >2. When it comes to that specific step in the GRADE framework, a relative risk (RR) >1.3, applied to the whole range of noise exposure, was used as a criteria for considering upgrading the quality of evidence. Using 45dB and 75dB as generic lowest and highest noise exposure categories in studies, this is equivalent to a relative increase of 3 times 10 dB over this range. Taking a RR of 2 as a big magnitude over this whole range (as is advised by GRADE), this would translate back to an incremental RR of (third root 2) RR = 1.26. Therefore we decided to use a RR of 1.3 as a round figure. This upgrading should only take place if there are no major concerns of bias, and not when there is only one single study informing the relationship between noise and a particular health outcome. This RR of 1.3 is considered large for fields of epidemiological or occupational health, and in the context of environmental noise can be justified by the large population health impact of exposure to noise and the fact that, potentially, small RRs translate into a large burden of disease.

GRADE rating for single studies: In some cases, only one study informs the rating for a particular noise source/health outcome pair. In those particular cases, the GDG decided that the overall GRADE rating should not be higher than moderate quality, to reflect the fact that the evidence base is limited. As a result, in some cases downgrading for inconsistency was done as it is not possible to assess consistency with other studies.

The main GRADE adaptations for the assessment of the quality of the evidence for health effects resulting from environmental noise are summarized in Table 4.

Table 4. Summary of main GRADE adaptations

Study design	Initial quality of a body of evidence	Lower if	Higher if	Final quality of a body of evidence
Best study design(s) to assess a particular relationship: in most cases: longitudinal (cohort, case control) for annoyance: cross-sectional Other study designs: Cross-sectional Ecological (for cardiovascular and metabolic diseases only)	High → Low Very low	Risk of bias Inconsistency (including for single studies) Indirectness Imprecision Publication bias	Large effect (RR>1.3) Dose response All plausible residual confounding	High Moderate Low Very low

Summary of findings tables were created to summarize this information for the seven evidence reviews on the outcomes listed in section 2.1.1.

2.2 Protocol for the systematic review of noise interventions

2.2.1 Description of health conditions

The evidence was reviewed for the following health outcomes associated with environmental noise:

- effects on sleep
- annoyance
- cognitive impairment
- quality of life, mental health and well-being
- cardiovascular and metabolic diseases
- hearing impairment and tinnitus
- adverse birth outcomes.

2.2.2 Description of the intervention

An intervention was defined as one of the following:

- a measure that aimed to change noise exposure and associated health effects;
- a measure that aimed to change noise exposure, with no particular evaluation of the impact on health;
- a measure designed to reduce health effects of noise exposure, but that may not have included a reduction in noise exposure directly.

2.2.3 Objective

The objective of the evidence review was to assess the effect on health outcomes of interventions that aimed to change exposure to environmental noise. The main research question was:

> In the general population exposed to environmental noise, are interventions effective in reducing exposure to and/or adverse health outcomes from environmental noise?

2.2.4 Criteria for considering studies for this review

To be considered for inclusion in the review, studies had to meet the following criteria.

Types of study considered were both experimental and observational study designs, including randomized controlled trials, non-randomized controlled trials, interrupted time series, repeated cross-sectional studies and both controlled and uncontrolled before-and-after studies (also called cohort studies). Modelling studies were also considered, where relevant.

Types of study participant considered were all members of the general population: interventions that aimed to change environmental noise levels were usually intended for the general population, and may have been of global, regional or local relevance. Exposure levels that have been shown to affect human health can be experienced by both children and adults, in rural and urban settings, and in both developed and developing countries. For this reason, no exclusions were made with regard to age group or other participant or setting characteristics.

Types of intervention considered included measures that aimed to change noise exposure and associated health effects; measures that aimed to change noise exposure, with no particular evaluation of the impact on health; and measures that aimed to reduce the health effects of noise but that did not have direct effects on noise exposure. Interventions were categorized according to the target noise source (rail, road, aircraft, wind turbines, personal electronic devices, other) and specific settings (residential, school, hospital, public venues, other). Interventions aimed at changing noise exposure that originated from multiple sources were also included. Each may have been comprised of multiple components, including technological or infrastructural, educational, policy and regulatory components (Burns et al., 2014). Some interventions had a geographical focus (such as a particular roadway, residential area or airport); others did not (such as a policy to reduce source noise levels of new motor vehicles). Interventions in occupational settings were excluded. In cases where there was a control group, the comparison was no intervention or an alternative intervention.

Types of outcome measure considered were assessment of the following outcomes:

- environmental noise exposure – a study was only included if the noise exposure had been measured or calculated and expressed in dB. Calculated levels for transportation noise (road, rail, air) must have been based on traffic data reflecting the use of roads, railway lines and in- and outbound flight routes at airports.
- human health – see the criteria listed in the protocol for the systematic review of health effects resulting from environmental noise (section 2.1.4).

2.2.5 Search for available systematic reviews

A single search was conducted to retrieve all relevant existing systematic reviews on environmental noise, health effects and interventions. This search was detailed in the protocol for the systematic review of health effects resulting from environmental noise (section 2.1.5).

2.2.6 Selection and assessment of quality of available systematic reviews

See the protocol for the systematic review of health effects resulting from environmental noise (section 2.1.6).

2.2.7 Search for individual studies

See the protocol for the systematic review of health effects resulting from environmental noise (section 2.1.7).

2.2.8 Selection of individual studies

See the protocol for the systematic review of health effects resulting from environmental noise (section 2.1.8).

2.2.9 Data extraction and management

Two reviewers independently extracted data from the articles on the list of included individual studies, provided separately in the systematic review. As considerable differences in intervention type were expected, the focus was on the relevant data to describe the intervention thoroughly. The following specific details about the interventions were extracted: study design, population, noise source and settings, intervention duration, level of implementation (local, regional, national, international), exposure and health outcomes of significance and other intervention characteristics. The systematic review team documented information and effect estimates for all primary outcomes (environmental noise exposure and human health) reported by the studies. See the protocol for the systematic review of health effects resulting from environmental noise (section 2.1.9) for additional information.

2.2.10 Assessment of risk of bias in included studies

See the protocol for the systematic review of health effects resulting from environmental noise (section 2.1.10).

2.2.11 Measures of health effects

If the health outcome was dichotomous, risk ratios were used as estimates of the effect of the intervention. If the outcome was measured on a continuous scale, mean differences or standardized mean differences were used in accordance with the recommendations given in the Cochrane handbook (Higgins & Green, 2011).

2.2.12 Dealing with missing data

See the protocol for the systematic review of health effects resulting from environmental noise (section 2.1.12).

2.2.13 Assessment of heterogeneity

Interventions were categorized according to their type, target noise source, setting and geographical scale. All the health outcomes (listed in section 2.1.1) were deemed to be separate categories, but within each category as many outcomes as possible were combined. See the protocol for the systematic review of health effects resulting from environmental noise (section 2.1.13) for additional information.

2.2.14 Assessment of reporting biases

See the protocol for the systematic review of health effects resulting from environmental noise (section 2.1.14).

2.2.15 Data synthesis

We originally intended to combine studies with similar participants, interventions and outcomes and to pool results using a statistical programme that included meta-analysis. However, as statistical pooling was not possible the findings were presented in narrative form, including tables, figures and harvest plots to aid data presentation, where appropriate.

2.2.16 Subgroup analysis and investigation of heterogeneity

The systematic review team evaluated whether the outcomes varied according to types of participant and checked whether results differed between studies carried out with participants from high-income countries versus participants from low- or middle-income countries.

2.2.17 Sensitivity analysis

See the protocol for the systematic review of health effects resulting from environmental noise (section 2.1.17).

2.2.18 Grading/strength of the quality of the evidence

The GRADE approach was adapted, as described in the protocol for the systematic review of health effects resulting from environmental noise (section 2.1.18).

The main specific adaptations made to the GRADE approach for interventions, which properly reflect the understanding of the GDG on the quality of the evidence on environmental noise interventions, are the following:

1. Start level for grading of studies: The best available study designs to assess a particular outcome would start as high quality. For noise interventions, the ideal study design is one which would provide longitudinal assessment of human response, measure potential confounders longitudinally, and include steady-state controls in the study design. Given the wide diversity in studies and study designs for interventions, the start level for the GRADE rating for noise interventions/health outcome

pairs was based on how well most of the studies in a particular pair approached the ideal study design. The start level was assigned as follows:

a. High quality was assigned where the bulk of the studies in the pair were:
    i. Before-and-after longitudinal studies using prospective cohort with controls
    ii. Before-and-after (uncontrolled) longitudinal studies
    iii. Controlled before-and-after studies
    iv. Interrupted time series studies.
b. Moderate quality was assigned where there were two or less studies from the list above, and the other(s) were cross-sectional; or there was a single controlled before-and-after study, or a single study with patients recruited to different conditions.
c. Low quality was assigned where the majority of studies were cross sectional studies.
d. Very low quality was assigned where the before-and-after studies used retrospective assessment.

Once the decision has been made on the starting point of the level of evidence, the quality would be upgraded or downgraded according to the various GRADE criteria described earlier.

2. Upgrading for dose-response: for noise interventions, given the wide diversity in studies, it was not possible to combine the estimates to provide a quantitative assessment of the dose-response. However, upgrading was done in cases where there was a significant trend in the observed magnitude of the change in health outcome as predicted by the relevant exposure-response function, or when a significant excess response was demonstrated.
3. GRADE rating for single studies: In some cases, only one study informs the rating for a particular noise source/health outcome pair. In those particular cases, the GDG decided that the overall GRADE rating should not be higher than moderate quality, to reflect the fact that the evidence base is limited. As a result, in some cases downgrading for inconsistency was done as it is not possible to assess consistency with other studies.

The main GRADE adaptations for the assessment of the quality of the evidence for environmental noise interventions are summarized in Table 5.

Table 5. Summary of main GRADE adaptations

Study design	Initial quality of a body of evidence	Lower if	Higher if	Final quality of a body of evidence
Best study design(s) to assess a particular relationship: longitudinal assessment of human response, measures potential confounders longitudinally, and if possible, includes steady-state controls in the study design Other study designs: Cross-sectional Studies with retrospective assessment	High Moderate → Low Very low	Risk of bias Inconsistency (including for single studies) Indirectness Imprecision Publication bias	Large effect Dose response (observed magnitude of change) All plausible residual confounding	High Moderate Low Very low

References

Burns J, Boogaard H, Turley R, Pfadenhauer LM, van Erp AM, Rohwer AC et al. (2014). Interventions to reduce ambient particulate matter air pollution and their effect on health (Protocol). Cochrane Database of Systematic Reviews. doi:10.1002/14651858.CD010919.

Egger M, Smith G, Schneider M, Minder C (1997). Bias in meta-analysis detected by a simple, graphical test. BMJ.315(7109):629–34.

Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P et al. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ.336(7650):924–6.

Higgins JPT, Green S (editors) (2011). Cochrane handbook for systematic reviews of interventions, version 5.1.0 [updated March 2011]. London: Cochrane Collaboration (http://community.cochrane.org/handbook, accessed 5 August 2015).

Orsini N, Bellocco R, Greenland S (2006). Generalized least squares for trend estimation of summarized dose-response data. Stata J.6(1):40–57.

Ryan R, Cochrane Consumers and Communication Review Group (2014). Heterogeneity and subgroup analyses in Cochrane Consumers and Communication Review Group reviews: planning the analysis at protocol stage. London: Cochrane Collaboration (http://cccrg.cochrane.org/sites/cccrg.cochrane.org/files/uploads/Heterogeneity_subgroup_analyses.pdf, accessed 18 July 2015).

Shamliyan T, Kane RL, Dickinson S (2010a). A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. J Clin Epidemiol.63(10):1061–70.

Shamliyan T, Kane RL, Jansen S (2010b). Quality of systematic reviews of observational nontherapeutic studies, Prev Chronic Dis.7(6):A133 (http://www.cdc.gov/pcd/issues/2010/nov/pdf/09_0195.pdf, accessed 5 August 2015).

Annex 1

AMSTAR TEMPLATE FOR ASSESSMENT OF QUALITY OF SYSTEMATIC REVIEWS ON ENVIRONMENTAL NOISE

Reviewer initials:	Date of review:
Author:	Year of study:
Journal:	Unique number:

Main topic (circle all that apply)	▪ Health outcomes: ○ effects on sleep ○ annoyance ○ cognitive impairment, mental health and well-being ○ cardiovascular disease, diabetes and metabolic diseases ○ hearing impairment and tinnitus ○ adverse birth outcomes ○ other, please specify: ▪ Interventions/risk management: ○ change in noise exposure with assessment of associated health impacts ○ change in noise exposure with no assessment of associated health impacts ○ measure to reduce health effects of noise, but that may not include a change in noise exposure ○ other, please specify: ▪ Other, please specify:

1. Was an "a priori" design provided? The research question and inclusion criteria should be established before conducting the review.

- Yes
- No
- Can't answer
- Not applicable

Does the design meet the criteria for inclusion? ☐Yes ☐No *(EXCLUDE)* ☐Unclear

2. Was a comprehensive literature search performed? At least two electronic sources should be searched. The report must include years and databases used (e.g. Medline/PubMed, Scopus (including Embase)). Keywords and/or medical subject heading terms must be stated and, where feasible, the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers or experts in the particular field of study, and by reviewing the references in the studies found.

- Yes
- No

- Can't answer
- Not applicable

Does the search meet the criteria for inclusion? ☐Yes ☐No *(EXCLUDE)* ☐Unclear

Summary of assessment for inclusion	
Include in review ☐	Exclude from review ☐
Independently assessed and then compared? Yes ☐ No ☐	Any initial differences resolved between reviewers' assessments? Yes ☐ No ☐
Request further details? Yes ☐ No ☐	Contact details of authors:
Notes:	

DO NOT PROCEED IF PAPER IS EXCLUDED FROM REVIEW

3. Was there duplicate study selection and data extraction? At least two independent data extractors are required and a consensus procedure for disagreements should be in place.

- Yes
- No
- Can't answer
- Not applicable

4. Was the status of publication (i.e. grey literature) used as an inclusion criterion? The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.

- Yes
- No
- Can't answer
- Not applicable

5. Was a list of studies (included and excluded) provided? A list of all studies should be provided.

- Yes
- No
- Can't answer
- Not applicable

6. Were the characteristics of the included studies provided? In an aggregated form, such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The range of characteristics in all the studies analysed (e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity or other diseases) should be reported.

- Yes
- No
- Can't answer
- Not applicable

7. Was the scientific quality of the included studies assessed and documented? "A priori" methods of assessment should be provided (e.g. for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo-controlled studies, or allocation concealment as inclusion criteria); for other types of studies, alternative items will be relevant.

- Yes
- No
- Can't answer
- Not applicable

8. Was the scientific quality of the included studies used appropriately in formulating conclusions? The methodological rigour and scientific quality should be considered in the analysis and the conclusions of the review and explicitly stated in formulating recommendations.

- Yes
- No
- Can't answer
- Not applicable

9. Were the methods used to combine the findings of studies appropriate? For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. a Chi-squared test for homogeneity, I^2 index). If heterogeneity exists, a random effects model should be used and/or the clinical appropriateness of combining taken into consideration (i.e. is it sensible to combine?).

- Yes
- No
- Can't answer
- Not applicable

10. Was the likelihood of publication bias assessed? An assessment of publication bias should include a combination of graphical aids (e.g. funnel plot, other available tests) and/or statistical tests (e.g. Egger regression test).

- Yes

- No
- Can't answer
- Not applicable

11. Was the conflict of interest included? Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

- Yes
- No
- Can't answer
- Not applicable

Conclusions

Authors' conclusions:
Reviewer's conclusions:

Note: "Can't answer" is chosen when the item is relevant but not described by the authors; "not applicable" is used when the item is not relevant, such as when a meta-analysis has not been possible or was not attempted by the authors.

Annex 2

INCLUSION AND EXCLUSION CRITERIA FOR INDIVIDUAL STUDIES

Individual studies should meet the following inclusion criteria to be included in the evidence reviews. The criteria can be adjusted if needed for each review; detailed justification should be given.

Category	Inclusion criteria	Exclusion criteria
Population: general population in settings (hospitals, residences, public venues, educational facilities)	▪ Studies including members of the general population ▪ Studies including specific segments of the population particularly at risk, such as children or vulnerable groups ▪ Studies including participants exposed to noise in occupational settings only if relevant with combined exposure to environmental noise	▪ Does not meet inclusion criteria ▪ Studies including participants exposed to noise in occupational settings not relevant with combined exposure to environmental noise
Exposure: exposure to high levels of environmental noise from various noise sources	▪ Noise exposure levels either measured or calculated and expressed in dB values, which should aim to be representative of the individual exposure of the study participants (for most observational studies, this would be the dwelling location or home) ▪ Calculated levels for transportation noise (road, rail, air) based on traffic data reflecting the use of roads, railway lines and in- and outbound flight routes at airports	▪ Does not meet inclusion criteria ▪ Studies using hearing loss or defective hearing as a proxy for (previous) noise exposure ▪ Surveys assessing noise exposure on the basis of subjective ratings, as given by the subjects in a questionnaire
Comparator: no noise exposure or lower levels of noise exposure	▪ Should have comparator group (corresponding to no exposure or lower-level exposure)	▪ Does not meet inclusion criteria
Confounding: adjusted for confounding	▪ No inclusion criteria applied; however, for every study, the systematic review team will assess which possible confounders have been taken into account	▪ No exclusion criteria applied
Outcome: assessment of outcome	▪ Data about outcomes taken from medical records or interview using a known scale or validated assessment method ▪ Self-reported data about outcome taken from questionnaire	▪ Does not meet inclusion criteria

Annex 3

TEMPLATE FOR ASSESSMENT OF QUALITY AND RISK OF BIAS OF INDIVIDUAL STUDIES

Study name:	Assessor name:	Date assessed:	
Study design	▪ Cohort ▪ Case-control ▪ Cross-sectional		
Domain	**Description of criteria for judgement**	**Quotation from article on which the judgement is based**	**Judgement of risk of bias**
1. Noise exposure assessment leading to information bias	The noise level (in dB) is expressed in L_{den} and L_{night} or its components (L_{day}, $L_{evening}$, L_{night} and the duration in hours of L_{night})[1] **AND** ▪ **for long-term average noise level:** a. is based on a noise map using as input the actual traffic volume, composition and speed per 24 hours per road/railway/airport, or the type and sound power of an industrial installation and the size in terms of either production volume or people employed **OR** b. is based on measurements for a minimum of one week by qualified staff, and adjusted for data under point a, as well as meteorological conditions when necessary **OR** c. is based on a noise map reported in a separate publication but which fulfils conditions a or b ▪ **for short-term noise level:** d. is based on measurements for a sufficient time by qualified staff		Low

[1] Lnight: night-time noise indicator; Lday: day-time noise indicator; Levening: evening-time noise indicator; Lden: day-evening-night level indicator. Definitions taken from Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 relating to the assessment and management of environmental noise.

	The noise level is not expressed in dB **OR** is not expressed in Lden and Lnight or its components **OR** ▪ **for long-term average noise level:** a. is based on a map that does not use as input the actual traffic volume, composition and speed per 24 hrs per road/railway/airport, or the type and sound power of an industrial installation and the size in terms of either production volume or persons employed **OR** b. is based on measurements of less than one week **OR** not adjusted for data under point a or meteorological conditions when necessary **OR** by unqualified staff **OR** c**.** is based on a noise map reported in a separate publication but which does not fulfil conditions a or b ▪ **for short-term noise level:** **d**. is based on measurements for an insufficient time **OR** by unqualified staff		High
	Insufficient information is reported to decide on one of the above		Unclear
2. Bias due to confounding (At least the following confounders should be incorporated for a valid assessment for the relation between noise and XXX outcome: 1….2….3….4….[to be completed by reviewer])	All-important confounders are taken into account either through matching or restriction or in the analysis		Low
	Only one or no confounders are taken into account **OR** subjects in exposed and unexposed groups differ for one or more important confounders but no adjustment is made in the analysis		High
	Less than all but more than one important confounder(s) are taken into account **OR** insufficient information is reported to decide on one of the above		Unclear
3. Bias due to selection of participants	Participants are randomly sampled from a known population **AND** the response rate is higher than 60% **AND** the attrition rate is less than 20% in follow-up studies		Low
	No random sampling is done **OR** the response rate is less than 60%		High
	Insufficient information is reported to decide on one of the above		Unclear

4. Health outcome assessment leading to information bias (i)	The health outcome of interest is objectively measured **OR** taken from medical records **OR** taken from questionnaires or interviews using a known scale or validated assessment method		Low
	The health outcome of interest is self-reported and not assessed using a known scale or validated assessment method		High
	Insufficient information is reported to decide on one of the above		Unclear
5. Health outcome assessment leading to information bias (ii)	The health outcome of interest is assessed blind for exposure information in cohort and cross-sectional studies or exposure is assessed blind for being a case in case-control studies		Low
	The health outcome and/or exposure assessment is not blinded		High
	Insufficient information is reported to decide on one of the above		Unclear
Total risk of bias in study	At least 4/5 judgements of low risk of bias, including for domains 1, 2 and 3		Low
	Any other		High

The WHO Regional Office for Europe

The World Health Organization (WHO) is a specialized agency of the United Nations created in 1948 with the primary responsibility for international health matters and public health. The WHO Regional Office for Europe is one of six regional offices throughout the world, each with its own programme geared to the particular health conditions of the countries it serves.

Member States

Albania
Andorra
Armenia
Austria
Azerbaijan
Belarus
Belgium
Bosnia and Herzegovina
Bulgaria
Croatia
Cyprus
Czechia
Denmark
Estonia
Finland
France
Georgia
Germany
Greece
Hungary
Iceland
Ireland
Israel
Italy
Kazakhstan
Kyrgyzstan
Latvia
Lithuania
Luxembourg
Malta
Monaco
Montenegro
Netherlands
Norway
Poland
Portugal
Republic of Moldova
Romania
Russian Federation
San Marino
Serbia
Slovakia
Slovenia
Spain
Sweden
Switzerland
Tajikistan
The former Yugoslav Republic of Macedonia
Turkey
Turkmenistan
Ukraine
United Kingdom
Uzbekistan

Original:English

World Health Organization
Regional Office for Europe

UN City, Marmorvej 51, DK-2100 Copenhagen Ø, Denmark
Tel.: +45 45 33 70 00 Fax: +45 45 33 70 01
Email: euwhocontact@who.int
Website: www.euro.who.int