Aviation noise and public health

Rapid evidence assessment

NatCen
Social Research that works for society

Authors: Chris Grollman, Imogen Martin, Joy Mhonda
Date: July 2020
Prepared for: The Independent Commission on Civil Aviation Noise

At **NatCen Social Research** we believe that social research has the power to make life better. By really understanding the complexity of people's lives and what they think about the issues that affect them, we give the public a powerful and influential role in shaping decisions and services that can make a difference to everyone. And as an independent, not for profit organisation we're able to put all our time and energy into delivering social research that works for society.

NatCen Social Research
35 Northampton Square
London EC1V 0AX
T 020 7250 1866
www.natcen.ac.uk

A Company Limited by Guarantee
Registered in England No.4392418.
A Charity registered in England and Wales (1091768) and Scotland (SC038454)
This project was carried out in compliance with ISO20252

Contents

Executive summary 5

1 Introduction 7

1.1 Background 7

1.1.1 Aviation noise in the UK 7

1.1.2 Motivation for the review 8

1.1.3 Existing evidence reviews 8

1.2 Research questions 8

2 Methodology 10

2.1 Data sources 10

2.2 Inclusion criteria 11

2.3 Search strategy 12

2.3.1 Databases 12

2.3.2 Websites 12

2.3.3 Conference proceedings 13

2.4 Data extraction and synthesis 13

2.4.1 From existing reviews 13

2.4.2 From search results 13

2.5 Results 13

3 Findings on links between aviation noise and health 15

3.1 Evidence from this update 15

3.1.1 Sleep 15

3.1.2 Cardiovascular and metabolic disorders 17

3.1.3 Mental health and wellbeing 19

3.1.4 Mediation through annoyance 20

3.2 Quality of the evidence, considering the WHO and Defra reviews and the current findings 21

4 Evidence gaps and potential for new research 22

4.1 Evidence gaps 26

4.1.1 Gaps in the evidence 26

4.1.2 How to prioritise filling evidence gaps 28

4.2 Research options and considerations 31

4.2.1 Potential study designs and their strengths and weaknesses 31

4.2.2 Measuring noise 34

4.2.3 Measuring health 37

4.2.4 Potential ways of filling evidence gaps 37

5 Discussion 41
6 References 42
Appendix A. The GRADE approach................................48
Appendix B. Strategies for searches..............................50
Appendix C. Data extraction tables................................53
Appendix D. Reason for exclusion at full text screening 60
Appendix E. Noise metrics used in included studies......62
Appendix F. GRADE for present REA outcomes only ...63
Appendix G. GRADE for WHO/Defra and present outcomes combined..73
Appendix H. GRADE for WHO and Defra review findings combined 79
Appendix I. GRADE for WHO or Defra reviews alone ..80

Acknowledgements
We thank Dr Charlotte Clark and Dr Ben Fenech for their valuable input.

Executive summary

Background and motivation

Aviation noise affects the quality of life and health of many people living close to airports and/or under flightpaths. In January 2019 the UK Government established the Independent Commission on Civil Aviation Noise (ICCAN), a new non-statutory advisory arm's length body, to act as a credible and impartial voice on civil aviation noise. The key objectives of this review are to collate and summarise the scientific evidence on the links between aviation noise and health, to identify evidence gaps and to suggest ways that further research could fill these gaps. This will support ICCAN to develop their expertise, authority and credibility in guiding aviation noise policy in the UK.

Methods

The review took the form of a rapid evidence assessment (REA) – a tool for systematically finding and synthesising available research as comprehensively as possible within a reduced timeframe. This REA was designed to build on existing systematic reviews conducted for the World Health Organisation (WHO) and the UK Department for Environment, Food and Rural Affairs (Defra). We searched academic databases and conference proceedings for findings published in the year since those reviews were conducted, in addition to the websites of relevant organisations. The findings of 12 new studies were combined with those of the WHO and Defra reviews, and the quality of evidence summarised across 58 health outcomes using a systematic approach.

Key findings

- The new evidence primarily focuses on health outcomes for sleep, quality of life, mental health and wellbeing, and cardiovascular and metabolic disorders. Several recent studies had small sample sizes – some were feasibility studies – and therefore can only give indicative findings.
- We made and collated ratings of the quality of evidence as 'high', 'moderate', 'low' or 'very low' for given health outcomes, using the GRADE approach (described in Appendix A). For a small number of outcomes, primarily in the areas of sleep and cognition, there is moderate quality evidence on the links between aviation noise and public health. Typically, it is difficult to achieve high quality evidence in environmental studies, and moderate quality evidence is therefore considered sufficiently robust to support strong policy recommendations.
- For most health outcomes, the evidence on the effects of aviation noise is low or very low quality. This low quality is primarily driven by the fact that most studies use a cross-sectional design and many have small sample sizes which limits their power.
- For some areas of health, including dementia and other neurodegenerative outcomes, cancer, and birth and reproductive outcomes, there is little or no evidence at all relating to aviation noise.
- There are therefore evidence gaps for the areas with limited or no evidence and those with low or very low-quality evidence. These areas present ICCAN and other stakeholders with opportunities for further research.

- Where evidence is of moderate quality, there is a need to quantify how interventions or operational changes impact health outcomes.
- ICCAN has a range of stakeholders, who are likely to have different priorities regarding areas for future aviation noise research. There has been relatively little data from the UK, despite having a large noise-exposed population including the busiest airport in Europe. It is welcome that two large research projects (ANCO and RISTANCO) are currently ongoing.
- In weighing up the areas for further research, ICCAN may take into account current priority areas in wider public health, including air pollution, mental health, and reducing health inequalities, as well as longer term ambitions.
- There are also opportunities for ICCAN to investigate the potential of retrospective cohorts combining noise maps with the wealth of data available in various UK cohort studies, as a means of obtaining high quality evidence without the costs and delay inherent in prospective longitudinal research.
- Collaboration among academic and other interested parties could support wider use of consistent research methods, such that even studies of lower individual quality could be combined robustly in support of stronger evidence.
- High quality evidence requires investment in longitudinal research. Whilst this is expensive, it would be an opportunity to gain insight into exposures beyond aviation noise, such as air pollution, that will be of interest to a broad range of public bodies.

1 Introduction

1.1 Background

1.1.1 Aviation noise in the UK

Aviation noise affects the quality of life and health of a substantial number of people in the UK. The impact of this noise includes health effects, such as an increased risk of hypertension, and effects of annoyance, cognitive impairment for children and lost productivity [1]. The number of people exposed to aviation noise in the UK varies according to how noise exposure is measured. For 2017, almost one million people (1.5% of the UK population) were exposed to aviation noise above 55 dB using the widely applied L_{den} 55 dBA indicator [2]. Around 65% of exposure at that level is caused by flights to/from Heathrow [3]. (L_{den} measures the average level of noise in a 24-hours period, with a penalty applied for noise in the evening and night time. Noise metrics are described in more detail in the recent ICCAN report [4].) The L_{den} indicator has also been used in guidelines published by WHO which recommend reducing noise levels produced by aircraft to below 45 dB L_{den}. Aircraft noise above this level is associated with adverse health effects [5].

The systematic reviews that informed the WHO Environmental Noise Guidelines for the European Region 2018 (WHO ENG2018) [6] assessed quality of evidence using the GRADE approach (Grading of Recommendations Assessment, Development, and Evaluation). This approach rates the quality of bodies of evidence as “high”, “moderate”, “low” or “very low”, with implications for the need for further research. This rating is based on the study designs, consistency and other features of the data on a given question. It was developed for clinical medicine [7] and has been adapted for use with environmental health exposures [8]. GRADE encourages transparency and consistency but its strict methods mean it is typically difficult to obtain high quality evidence for environmental health risks. Moderate quality evidence is therefore considered adequate to support making strong recommendations [5]. (There is more detail on the GRADE approach in Appendix A.)

The WHO reviews concluded that there is moderate quality evidence that aviation noise has a harmful effect on annoyance [9], some cognitive outcomes in children [10], some aspects of sleep disturbance [11] and change in waist circumference [12]. The reviews also show moderate quality evidence of no effect on stroke mortality [12]. There is low and very low quality evidence relating to a wide range of other health outcomes – including mental health outcomes [13], quality of life outcomes [13] and cardiovascular and metabolic outcomes [12]. This evidence generally indicates harmful effects. Due to the strict methods used to assess quality of evidence for environmental exposures such as noise via the GRADE approach (explained below in section 3.2), high quality evidence is limited.

In areas under flight paths, aviation noise is a salient issue for residents. Surveys conducted by the Civil Aviation Authority (CAA) in 2017–2018 [14] and the Department for Transport (DfT), the Department for Environment, Food and Rural Affairs (Defra) and the CAA in 2014–2015 [15] show substantial grievances about aircraft noise among residents. ICCAN published a review of the 2014–2015 Survey of Noise Attitudes findings given concerns that aspects of its methodology led to an underestimate of the impact of noise on annoyance [16].

1.1.2 Motivation for the review

In its role as an independent and impartial voice on civil aviation noise and how it impacts communities, ICCAN commissioned this rapid evidence assessment (REA) to update their knowledge on the links between aviation noise and health and bring all the evidence into one place. This work builds on the reviews conducted by the WHO and Defra. ICCAN wish to use the evidence from this REA to achieve the following:

- Identify new evidence that links aviation noise to health outcomes
- Identify evidence gaps in research that links aviation noise to health
- Put forward research methodologies that might be feasible to fill identified evidence gaps

This REA summarises the quality of the evidence relating to a wide range of health outcomes, from the WHO and Defra reviews and from the evidence published since those reviews. It also summarises the measurement metrics and research methodologies that might be used to fill identified evidence gaps.

1.1.3 Existing evidence reviews

There is a substantial body of recent evidence from many countries on health impacts of environmental noise, including aviation noise. Defra commissioned two systematic reviews, published in 2019 [17] and 2020 [18], on various types of environmental noise and a range of health outcomes. The first, prepared by the Dutch Institute for Public Health and the Environment (RIVM), covers noise effects on annoyance, sleep disturbance, cardiovascular and metabolic health outcomes. The second, prepared by Arup, covers mental health, wellbeing, quality of life, cancer, dementia, other neurodegenerative outcomes, birth and reproductive health, and cognitive health outcomes. For some outcomes, these systematic reviews identified evidence relating to other sources of noise but did not identify any evidence relating to aviation noise.

These reports followed the methodology of the eight systematic reviews that underpin the 2018 guidelines on environmental noise published by the World Health Organization (WHO) [6]. The WHO reviews covered evidence published from 2000 to 2014 or 2015 and the Defra reviews cover evidence published from the cut-off of the WHO reviews until March 2019.

1.2 Research questions

The aim of this REA is to identify and summarise evidence linking aviation noise to public health. It combines evidence from existing reviews and evidence published subsequent to those reviews (since March 2019) to identify gaps in research. The research questions for this REA are:

1. **What evidence exists about the links between aviation noise and health?**
    a. Based on this REA, what are the links between aviation noise and public health?
    b. What research approaches and methods have been used to research these links?
2. **Based on the REA, what are the key evidence gaps for research regarding links between aviation noise and health**?
    a. Where is evidence weak?
    b. What health conditions need further evidence?

c. What are the priority evidence gaps?
d. What research approaches and methods can be best used to fill these evidence gaps?

2 Methodology

This review followed the methodology and structure of a Rapid Evidence Assessment: *"A Rapid Evidence Assessment (REA) is a tool for getting on top of the available research evidence on a policy issue, as comprehensively as possible, within the constraints of a given timetable"* [19].This evidence assessment collates evidence on aviation noise from the existing WHO and Defra reviews on environmental noise, and updates that with evidence published since the cut-off of those reviews in March 2019.

This section provides a summary of our criteria and processes for searching for relevant evidence, determining the inclusion of studies, data extraction and the synthesis of findings.

2.1 Data sources

The starting point for our evidence search was the conclusions of the systematic reviews commissioned by WHO and Defra in recent years. These included:

- Systematic reviews for WHO on environmental noise and:
    - Adverse Birth Outcomes [20]
    - Cognition [10]
    - Cardiovascular and Metabolic Effects [12]
    - Sleep [11]
    - Quality of life, wellbeing and mental health [13]

The series commissioned by WHO further included a review on annoyance (which is outside the scope of this REA) and a review on permanent hearing loss and tinnitus (not considered as the sound levels causing these outcomes are higher than those caused by aviation noise for the general population).

- Systematic reviews for Defra on environmental noise and:
    - Mental health, wellbeing, quality of life, cancer, dementia, other neurodegenerative outcomes and birth, reproductive and cognitive health outcomes ("Defra-Arup") [17]
    - Annoyance, sleep disturbance, cardiovascular and metabolic health outcomes ("Defra-RIVM") [18]

We sourced new evidence on links between aviation noise and health from searches covering the period since the cut-off of the WHO and Defra systematic reviews in March 2019. In the present report, the search for evidence published subsequent to the existing systematic reviews is called the "update review", we also refer to it as the "ICCAN review". Searches included:

- Databases (Medline, Embase, Scopus and Epistemonikos)
- Online websites and repositories for relevant evidence published from 2015 onwards, as recent grey literature may not have been captured by the systematic reviews
- Proceedings of 2019 conferences

Details of the websites and conferences searched are given in Appendix B.

2.2 Inclusion criteria

Our criteria determining eligibility for inclusion are set out below. These are similar to the inclusion criteria used in the previous WHO/Defra reviews.

1. **Population:** Studies had to include evidence relating to aviation noise and health in a general human population. We excluded occupational exposure (e.g. of pilots).

2. **Exposure:** We included evidence where the exposure was aviation noise (either civil or military), measured or modelled, and expressed in decibels with no restriction as to the metric used. Noise levels had to be measured/calculated at an appropriate location for the exposure of the study participants (for observational studies, this would usually be the external noise level at the relevant location such as the home). Studies had to include people exposed across at least two sound levels, so that outcomes could be compared according to level of sound exposure (allowing, for example, conclusions about the effect of a 10 dB increase, or the effect of living in an area with average noise above 55 dB compared to an area with average noise below 45 dB). We excluded evidence where noise exposure was characterised by proxy (for example, distance or number of events) or subjectively (for example, self-reported noise exposure).

3. **Outcome:** We included evidence relating to any health condition including sleep disturbance, hypertension, strokes, heart attacks, coronary heart disease, dementia, cancer, diabetes and other metabolic conditions, cognition, birth and other reproductive outcomes, mental health, wellbeing, quality of life, and any other health conditions identified.

    a. We excluded studies where the outcome was annoyance but we included evidence where annoyance is treated as a factor that modifies the effect of noise on another health outcome. We made this exclusion because ICCAN is more confident in the evidence base regarding annoyance than regarding other health outcomes. ICCAN is already funding separate work to fill evidence gaps relating to annoyance and aviation noise.

    b. We excluded economic studies, burden of disease studies and health impact assessments as these do not report health outcomes *per se* and as such were outside scope.

4. **Study design:** We excluded review papers but included papers that presented new summary estimates derived from meta-analysis. We excluded experimental studies, such as laboratory studies or home-based studies with artificial playback of noise, due to concerns about their validity, in line with the approach taken in the WHO reviews. We did not restrict our search by any other study design and considered any primary or secondary research studies that used methodologies which appropriately addressed the research questions. This was largely quantitative evidence, but we also considered high quality qualitative evidence that linked aviation noise to quality of life, mental health or wellbeing.

5. **Publication characteristics**

    a. **Date of publication**: We included original studies published after the cut-off date of the WHO and Defra systematic reviews (April 2019) as we considered those reviews methodologically sound and comprehensive for the period they covered. We conducted our searches on 28th March 2020.

    b. **Language**: We applied no restriction based on language. Our search terms were in English only.

    c. **Type of publication:** We excluded editorials, discussion pieces, comments, errata, letters to the editor, encyclopaedia entries, results with only a title and no abstract (unless the title indicates very likely relevance) and studies for which full texts were not accessible.

    d. **Publication status**: We included all evidence coming through the database searches and conference proceedings, including published (journal) and unpublished (grey) literature. We considered grey literature identified through website searches of airport authorities and the CAA, or provided by ICCAN.

2.3 Search strategy

Studies were screened in two stages, at title and abstract, and at full text. Title and abstract screening were conducted manually. All studies meeting our inclusion criteria were screened at full text for their relevance to address all research questions and sub-questions. The full search strategy is described in Appendix B.

2.3.1 Databases

The database searches returned 552 unique results. Titles and abstracts were screened in Abstrackr [21] which is an online database screening tool which allows selections to be made by researchers. The first 30 records were checked with a second reviewer to ensure consistency. After title and abstract screening, 31 papers were included for full text review and eight were included for the update. Reasons for exclusion at full text screening are given in Appendix D. We treated references to “traffic noise” to mean road traffic rather than air traffic and excluded articles whose title or abstract did not suggest aviation noise.

2.3.2 Websites

A number of governmental, industry and aviation research websites were manually searched using a simplified version of our search strategy (see Appendix B). These searches provided a total number of 819 results which were screened at title and abstract level. Three evidence reviews from the CAA website were included for full text review [22] [23] [24], and included three papers also identified through the search of conference proceedings.

2.3.3 Conference proceedings

In total there were 1309 papers from the ICA Aachen conference (2019) and 893 from Internoise Madrid (2019). We completed a two-stage screening process. First, we screened session titles for potentially relevant sessions. Second, we screened the titles and abstracts of all proceedings within those sessions (N=123). In total we identified 10 potentially relevant papers for full text screening, of which we included four in our update review. These included three that had also been cited in the recent CAA reviews.

2.4 Data extraction and synthesis

2.4.1 From existing reviews

We summarised from the WHO and Defra reviews the GRADE assessments for the quality of the evidence relating aviation noise to individual health outcomes. In the Defra-RIVM review, which did not conduct a GRADE assessment, we summarised the findings without assessing quality of evidence. We describe the GRADE process in Appendix A.

2.4.2 From search results

After screening for final inclusion, core information about each paper was placed in an extraction table (see Appendix C) for analysis and subsequent report development. The extraction sheet was refined in consultation with ICCAN and included:

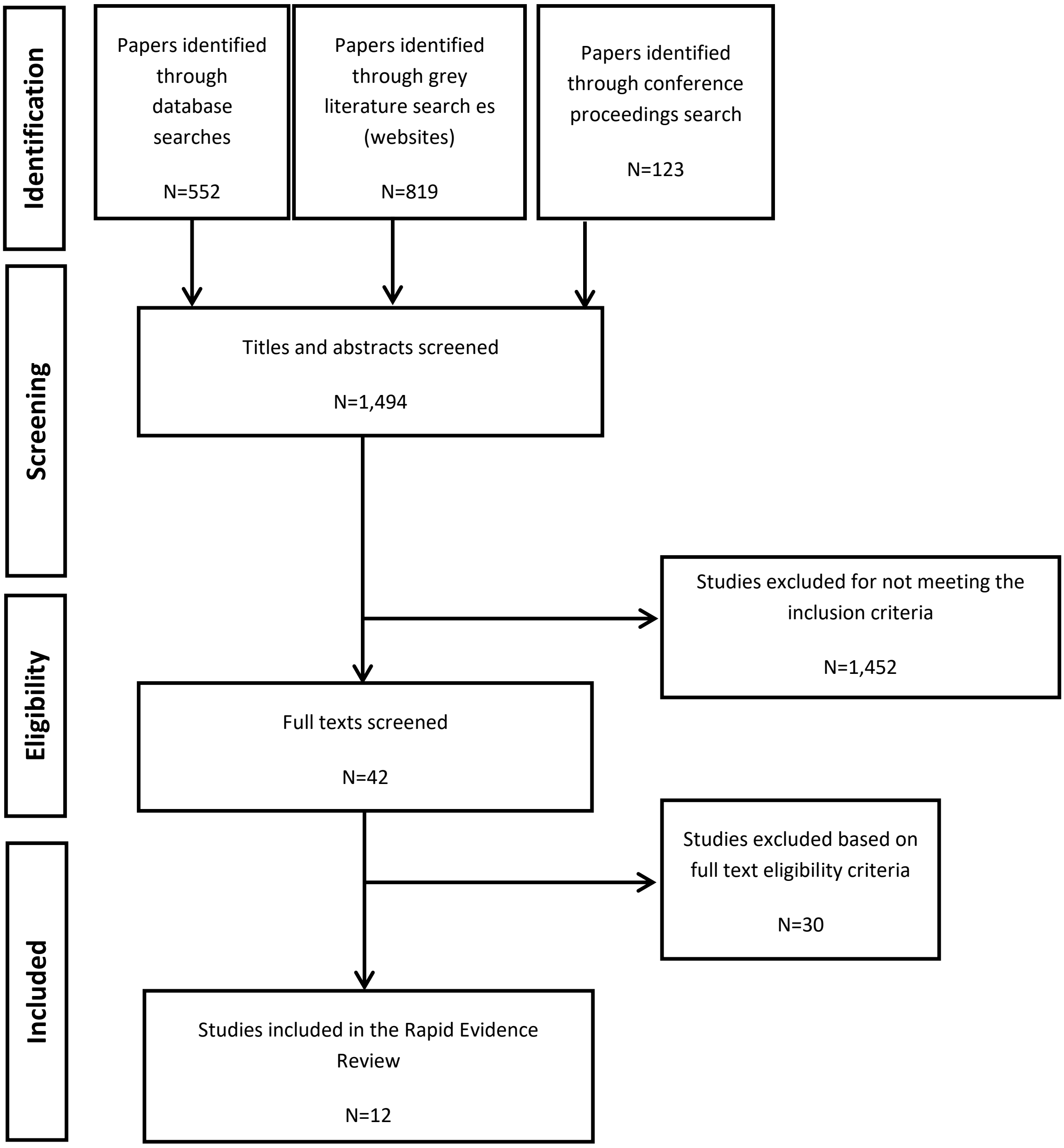
a. Sample size and study design
b. Setting/population of the research
c. Adjustments for confounders
d. Health conditions included in the paper
e. Measurement of health conditions
f. Noise assessment and noise metrics used
g. Effect size (metric and direction of association or effect)
h. Risk of bias assessment

2.5 Results

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) is an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses. The PRISMA flowchart below (Figure 1) summarises the REA's screening and inclusion processes.

There were 1,494 results returned from the systematic searches across the chosen databases, websites and conference proceedings. The search terms were designed to be highly sensitive, meaning that in order to make sure we identified all relevant evidence, we expected to have a large number of "false positive" results. In screening titles and abstracts, we excluded the vast majority of results (1,452). We screened the full text of the remaining 42 results, of which 12 met the inclusion criteria and were included for data extraction.

Figure 1 PRISMA Flowchart: REA screening and inclusion



3 Findings on links between aviation noise and health

This chapter presents the state of the evidence on the links between aviation noise and health. Section 3.1 presents the new evidence added in this update. Section 3.2 summarises the overall quality of evidence and direction of effect (whether aviation noise is harmful or has no effect), drawing together the conclusions on the quality of evidence from the existing reviews and integrating new evidence where possible. Together these sections address research questions 1a on the links between aviation noise and health and 2a and 2b on where evidence is weak and where further evidence is needed.

The reviews by WHO and Defra reported the evidence of the effects of transportation noise on specific health outcomes across seven broad health areas:

- Birth and reproductive outcomes
- Cognition
- Sleep
- Cardiovascular and metabolic outcomes
- Quality of life, mental health and wellbeing
- Cancer
- Dementia and other neurodegenerative outcomes

They included evidence on the effects of aviation noise in all areas except for *Dementia and other neurodegenerative outcomes*, for which there is no evidence relating to aviation noise. These health areas are largely exhaustive although there appears to be no evidence on auto-immune diseases.

Our update identified 12 papers (eight peer-reviewed journal papers from the database search and four papers from the conference proceedings) that presented new evidence across one or more of these three areas:

- Sleep (four papers)
- Quality of life, mental health and wellbeing (two papers)
- Cardiovascular and metabolic disorders (eight papers)

3.1 Evidence from this update

3.1.1 Sleep

Brink et al. (2019) [25] reported on sleep disturbance as part of the SiRENE study, which sampled 5592 people from the population of Switzerland and calculated aviation noise levels at the outer façade of the participant's home. The survey specified the source of noise in the questions asking about sleep disturbance. They found that the odds of the participant reporting being highly sleep disturbed (%HSD) increased significantly with increasing L_{night}: for every 1 dB increase the odds of being HSD increased by 13%.

Brink et al. (2019) also reported that for a given noise level, the effect of L_{night} on %HSD varied according to other factors (known as “effect modification”). First, they used the intermittency ratio (IR) to measure the intermittency or “eventfulness” of noise, that is how much loud events stand out from the background noise levels. A high IR means the loud event interrupts otherwise quieter background noise, while a low IR means the background noise is higher. The study found that with levels of L_{night} up to around 50 dB, participants with low IR (higher background noise) reported significantly lower levels of %HSD. They also found an effect modification in degree of urbanisation, whereby for a given level of L_{night}, %HSD is highest in rural areas, lower in towns/suburbs, and lowest in cities. These two effect modifications, by intermittency and by urbanisation, are clearly consistent with one another and closely related. The authors did not discuss reasons for the effect modifications. As noise was estimated at the external façade, one possible explanation may be that residents facing greater noise exposure (as with lower IR or greater urbanisation) take more steps to insulate their homes. Another possible explanation may be that the ongoing background noise makes noise events less noticeable.

Rocha et al. (2019) [26] conducted a pilot study around Atlanta international airport to test the feasibility of using postal surveys to recruit people to a national study about sleep. It is worth noting that as a pilot study of 268 people, this was not powered to elucidate precise associations or effects, and the results are only indicative. The question about sleep disturbance did not mention noise, although the title of the survey did. The study found a significant association between L_{night} and sleep disturbance, with 15% higher odds for being highly sleep disturbed for each 1 dB increase in L_{night}. A similar result was found for annoyance (17% higher for each 1 dB increase). Although the odds were lower than for annoyance or %HSD, they also found significant links between L_{night} and other sleep-related outcomes including overall sleep quality, trouble falling asleep, trouble sleeping at night, and trouble staying awake in the day. L_{night} was also associated with greater odds of using certain coping aids against noise when trying to sleep, including: alcohol (10% higher odds per 1 dB increase); TV (5% higher); music (7% higher); and closing windows (5% higher). After adjustment for covariates, L_{night} was not significantly associated with self-reported general health, use of sleep medication, or use of earplugs, medication, sound machines or fans to cope with noise.

Rocha et al. (2019) also asked participants whether they had any previous diagnosis of sleep disorder, hypertension, migraines, arrhythmia, heart disease, stomach ulcer or diabetes. There were no significant associations between L_{night} and the odds of reporting diagnoses of any of these conditions. The authors noted that “we were underpowered to detect the small effect sizes expected for these [chronic] health outcomes”.

Studies by Smith et al. (2020) and Basner et al. (2019) reported findings on sleep outcomes. **Basner et al.** (2019) [27] collected indoor noise measurements and ECG, movement and blood pressure data from 39 people living near Philadelphia airport and 40 controls not living close to an airport. This was also a pilot study and was not powered to elucidate precise associations or effects, and the results are only indicative. The median average noise (L_{Aeq}) during sleep was 43.2 dBA in the airport region and 31.8 dBA in the control region. There was no significant relationship between sleep fragmentation (awakenings per hour of sleep) and L_{Aeq} during sleep. The authors also investigated the effects of the maximum indoor sound level (L_{Amax}). They reported a significant relationship, with a 3% increase in the odds of awakening for every 1 dB increase in L_{Amax}. The percent awakened increased from around 3.5% at L_{Amax} 50 dB to 6% at 60 dB, 9% at 70 dB and 12% at 80 dB. This impact of maximum sound level rather than average sound level is consistent with the findings on intermittency reported by Brink et al. (2019).

Basner et al. (2019) also asked participants questions on sleep quality and on general health. There were significant differences in the expected direction between populations (airport vs. control) for several sleep aspects: people in the airport region were less likely to report that “My sleep was refreshing” or “I got enough sleep” and more likely to report “I had difficulty falling asleep”. Compared to the control population, they were also more likely to agree that “I expect my health to get worse” and less likely to agree that “My health is excellent”. There were no significant differences for the sleep items “My sleep was restless”, “I had trouble staying asleep”, “I had trouble sleeping” or “I was satisfied with my sleep”, or the general health items “I seem to get sick a little easier than other people” or “I am as healthy as anybody I know”. It is important to note that this was a small and underpowered pilot study whose primary aim was to test the feasibility of the data collection methods, and to keep that in mind when reading its findings (especially its findings of no effect).

Smith et al. (2020) [28] collected data on tiredness and awakenings related to noise during sleep from 34 people living around Atlanta international airport. This was also a pilot study and was not powered to elucidate precise associations or effects, and the results are only indicative. The maximum indoor noise level was marginally associated with the probability of awakening measured by ECG. This is consistent with the finding in Basner et al. (2019) above, and the authors suggest the marginal significance is likely to be due to the small sample size.

Smith et al. (2020) also reported effects of noise during sleep on questionnaire items reported by participants in the morning. They investigated both overall average sleep-time noise ($L_{Aeq,sleep}$) and maximum sleep-time noise ($L_{AS,max,sleep}$) during the time that each individual participant was measured to be asleep. In adjusted analysis, the only significant associations were for $L_{AS,max,sleep}$ with self-reported awakenings and with self-reported tiredness. They did not find any significant associations for either noise metric with sleep latency, sleepiness, difficulty falling asleep, sleep restlessness, sleep quality or disturbance by aircraft noise.

In 2018, **Trieu et al.** (2019) [29] surveyed residents living around Hanoi Noi Bai airport in two rounds, before (623 participants) and after (132 participants) an increase in night flights. The collected data was on annoyance, insomnia and a range of health problems and indicators including blood pressure and heart rate (reported below). There were associations between L_{den} and annoyance, and $L_{Aeq,night}$ and insomnia. After the increase in night flights there was greater insomnia at lower decibel exposure levels (up to around 60 dB $L_{Aeq,night}$) whereas at higher decibel levels (over 60 dB $L_{Aeq,night}$) insomnia was high before and remained high after.

3.1.2 Cardiovascular and metabolic disorders

Rojek et al. (2019) [30] investigated cardiovascular outcomes in a cross-sectional study of 201 residents of suburban Krakow, split evenly between areas exposed to high aircraft noise (>60 dB L_{den}) and low aircraft noise (<55 dB L_{den}). The outcomes investigated were a range of blood pressure measurements (measured in a study clinic and through 24-hour ambulatory monitoring), and arterial stiffness and a range of echocardiographic indicators selected for association with asymptomatic organ damage (measured in a study clinic). The study was designed to detect a difference in pulse wave velocity (PWV), a measure of arterial stiffness, in people living in the two areas. Greater arterial stiffness, indicating organ damage, means a higher PWV.

Significant differences were found between exposure groups for several outcomes. The results were stratified by hypertension status, as roughly half of participants in each area had hypertension. Among people who did not have hypertension (“normotensive” people), those in the exposed high noise area had higher PWV than those in the unexposed low noise area, and one measure of cardiac function was slower (the early diastolic mitral annulus mean velocity). This group also had higher diastolic blood pressure in both the clinic setting, and the ambulatory measure at night. Among hypertensive participants, the 24-hour heart rate and the central systolic blood pressure were higher in the exposed group.

Rojek et al. (2019) reported that in the unexposed group there is a trend for increasing PWV by age. In the noise-exposed group the increased PWV at younger ages means this trend by age is lessened and made non-significant, as though noise exposure causes premature aging related to this outcome. All differences between exposed and unexposed groups were in the expected direction, consistent with noise exposure causing worse cardiovascular health. No differences were found on a range of other parameters of blood pressure and cardiac health, including hypertension.

The authors also investigated the relationship between aircraft noise annoyance and PWV among exposed participants (only one unexposed participant reported annoyance), and found a significant trend for normotensive participants. This suggests a mediating effect of annoyance in the relationship between noise and increased PWV, which is consistent with other findings on the role of annoyance as an effect modifier of the relationship between noise and hypertension (as PWV and hypertension are strongly associated). It is important to note that the study was relatively small and the authors only stated that it is powered to detect the PWV outcome.

Basner et al. (2019) also reported that neither systolic nor diastolic morning blood pressure were associated with the average indoor noise level (L_{Aeq}) at night.

Baudin et al. (2019) [31] combined data from studies around several major European airports: seven airports in the HYENA (Hypertension and Exposure to Noise near Airports) study (London Heathrow, UK; Berlin Tegel, Germany; Amsterdam Schiphol, the Netherlands; Stockholm Arlanda and Bromma, Sweden; Milan Malpensa, Italy; and Athens International Airport Eleftherios Venizelos, Greece) and three French airports from the DEBATS (Discussion sur les effets du bruit des aéronefs touchant la santé – Discussion of the health effects of aircraft noise) study (Lyon Saint Exupéry, Toulouse-Blagnac, and Paris-Charles de Gaulle). They investigated the association between aviation noise and levels of cortisol, a stress hormone, found in saliva, for 1300 people. There is a natural daily cycle in which the production of cortisol varies. If the variation is reduced and there is less of a cycle and more of a constant level, this may indicate a less responsive hormonal system (specifically, disruption of the hypothalamus–pituitary–adrenal axis). Long-term exposure to stress, in this case noise, may disturb the stress response, with impacts on a range of biological outcomes. This study combined two existing cross-sectional studies, enabling analysis of cortisol outcomes stratified by sex.

The authors reported several significant associations: evening cortisol levels in women increased with increasing aircraft noise exposure measured by $L_{Aeq,16\,h}$, L_{den} and L_{night}. They also found significant reductions in cortisol variation per hour for women. This is an indicator of a poorly functioning stress response. Absolute variation per hour fell with increasing L_{night}, and relative variation per hour in women fell with increases in both L_{night} and L_{den}. Morning cortisol levels were unchanged with all noise exposure indicators. There were no statistically significant associations between aircraft noise exposure and cortisol levels for men.

Baudin et al. (2019) also found that the effects of noise exposure on cortisol were not modified by annoyance or noise sensitivity.

Nassur et al. (2019) [32] investigated associations between sleeping heart rate and several indicators of sound levels for people living near airports in Paris and Toulouse. This was a small study with 92 participants, self-selected from the larger DEBATS cross-sectional study and therefore with a moderate risk of selection bias and potentially underpowered. Looking at average sound levels across 15-second intervals, they found an increase in the heart rate associated with the sound from all sources. They found no association for the equivalent measurement for aviation noise alone and a smaller increase in heart rate looking across levels of sound from all sources exceeded for 90% of the measurement period.

Looking at maximum 1-second indoor sound levels during aviation noise events ($L_{Amax,1s}$) the authors found no difference in heart rate following events, but found an increase in heart rate amplitude during the event. Heart rate amplitude was the difference between the maximum and minimum heart rate during an event, and increased as the maximum 1-second sound level increased. The study recorded relatively low levels of $L_{Amax,1s}$, with a mean of 31 dB compared to 45+ dB in similar studies. The authors suggest this may be why there was no significant heart-rate elevation following an aircraft noise event.

In their survey of residents living around Hanoi Noi Bai airport, **Trieu et al.** (2019) [29] collected data on blood pressure. In the first round, all data were self-reported, while in the second round, blood pressure was measured. The prevalence of high blood pressure was 47% in round 1 and 62% in round 2, but the measurement differences mean direct comparison is not possible. The data showed a high prevalence of high blood pressure across the study population, but there was no significant association between high blood pressure and L_{den} (odds ratio 1.02, 95% CI: 0.97 to 1.08).

Vienneau et al. (2019) [33] published a meta-analysis of the impact of aviation noise on incidence of ischaemic heart disease (IHD) and diabetes. They found five new studies relating to aviation noise and IHD, giving a non-significant risk ratio of 1.03 (95% CI: 0.98 to 1.09) for every 10 dB increase in L_{den}. The authors found evidence of an increased risk of IHD from road traffic noise, but a similarly sized effect for aviation noise was non-significant and judged to be at a high risk of bias–2. They judged this estimate to be at high risk of bias as three of the five studies had high risk of bias, including one of the two large studies.

For diabetes incidence, the authors found three new studies resulting in a pooled risk ratio of 1.20 (95% CI: 0.88 to1.63) per 10 dB increase in L_{den}. This is a relatively large risk ratio but the wide confidence interval means the estimate is consistent with there not being a true effect. Estimates from the three contributing studies varied widely.

Weihofen et al. (2019) [34] published a meta-analysis of the impact of aviation noise on incidence of stroke. They included seven studies in the meta-analysis and found a pooled risk ratio of 1.013 (95% CI: 0.998 to 1.028), meaning a 1.3% increase in the incidence of stroke per 10 dB increase in L_{den}. The authors wrote that "the result is so close to the significance threshold that an actual effect seems likely". They also noted that noise is a marginal risk factor compared to other risk factors for stroke, and that even if people were universally exposed to high levels of aviation noise the effect on overall stroke incidence would still be minimal.

3.1.3 Mental health and wellbeing

Benz et al. (2019) [35] conducted a secondary analysis of the NORAH (Noise-Related Annoyance, Cognition, and Health) panel study around Frankfurt Airport. They investigated the relationship between noise and diagnosis of depression in the 12 months following operational changes comprising a new runway and a ban on night flights for 3319 participants. The authors also investigated the role of annoyance in mediating the relationship between noise and depression.

Benz et al. (2019) found that there was no direct association between L_{den} in the period after the new runway and night flight ban (t_1) and depression a year later (t_2). By contrast annoyance at t_1 was strongly associated with depression at t_2, and the authors showed that even though noise exposure had no direct effect on depression, there was a significant indirect effect from noise exposure to depression via annoyance. This suggests an important role of annoyance in mediating the relationship between noise and mental health outcomes. The authors also reported that the relationship between annoyance and depression may work in both directions, in that depression may also predict annoyance.

Spilski et al. (2019) [36] presented a secondary analysis of data from the NORAH panel study, looking at 8-year-old children's wellbeing and health as reported by children and their parents. 1200 children were included in the analysis. The authors hypothesised that increased aircraft noise exposure leads to increased stress responses in children and subsequently affects their well-being and health, mediated through aircraft annoyance. They also tested for effect modification by urbanisation and by imperviousness (that is, the level of sealed spaces such as buildings in the surrounding area: high imperviousness = many buildings, low imperviousness = many open spaces).

Physical wellbeing was estimated by two child-reported outcomes, "Last week I had a headache and stomach ache" and "Last week I felt sluggish and tired". Mental wellbeing was estimated by "Last week I laughed a lot and had a lot of fun" and "Last week I was bored". These outcomes were not commonly reported standardised measures. The parent-reported health outcomes were a set of diseases including asthma, migraine and speech and language disorders, and the intake of medically prescribed drugs.

The study found no significant direct effect of aviation noise on physical wellbeing. There were, however, significant indirect effects of noise on both indicators of physical wellbeing, mediated through annoyance. That is, where aviation noise increases annoyance, this in turn negatively affects physical wellbeing. The authors reported a similar finding for mental wellbeing on the boredom outcome but not on the outcome "Last week I laughed a lot and had a lot of fun".

There were neither direct nor indirect effects of aviation noise on children's health measured by parental report. However, after the inclusion of urbanisation and imperviousness in two extended models the direct relationship of noise with children's increased prescription drug use became significant in areas with medium levels of urbanisation and areas with low levels of imperviousness. The authors interpret this as suggesting that the impact of aircraft noise is greatest where "other stressors are less pronounced and therefore aircraft noise is more prominent".

3.1.4 Mediation through annoyance

Two studies found outcomes where there was no direct effect of noise exposure, but there was an indirect effect via annoyance. That is, for people who experienced annoyance due to aviation noise, there was an effect on the health outcome. These outcomes were prevalence of depression (Benz et al. 2019) [35] and general physical health of children (Spilski et al. 2019) [36]. There was also a role of annoyance in mediating the relationship between noise exposure and arterial stiffness (Rojek et al. 2019) [30]. There was no role of annoyance in mediating the relationship between aviation noise and cortisol levels (Baudin et al. 2019) [31].

3.2 Quality of the evidence, considering the WHO and Defra reviews and the current findings

Previous reviews assessed the quality of evidence relating aviation noise to given health outcomes using the GRADE approach, and we have taken the same approach for the new studies included in this REA. GRADE is a method of assessing quality of evidence in a structured and consistent manner. It was developed for assessing quality of evidence in clinical medicine, and has been adapted for use with environmental health risks. In this approach, quality is rated as 'High', 'Moderate', 'Low' or 'Very low'. These ratings have implications for the need for further research:

- High quality evidence means further research is *very unlikely* to change the certainty of the effect estimate
- Moderate quality evidence means further research is *likely* to have an important impact on the certainty of the effect estimate and *may* change the estimate
- Low quality evidence means further research is *very likely* to have an important impact on the certainty of the effect estimate and is *likely* to change the estimate
- Very low quality evidence means any effect estimate is *uncertain*

The GRADE process is described in Appendix A.

For some outcomes, evidence was only available in the update review (also referred to as the ICCAN review). We have made GRADE ratings for these outcomes (Appendix E). For some outcomes there was evidence from both the update review and the WHO/Defra reviews. For these outcomes, we took the conclusion of the WHO/Defra reviews as the starting level for the quality of evidence, applied the GRADE process to the additional evidence from the update review and decided whether to revise the GRADE rating (Appendix F). In the quality of evidence tables below (tables 1-6), this is referred to as the ICCAN synthesis. For some outcomes we combined the findings of the WHO and Defra reviews with one another (Appendix G). For outcomes only reported in either the WHO or Defra reviews, we report the GRADE ratings from those reviews (Appendix H). Tables Table 1 to

Table 6 summarise the quality of evidence across all these health outcomes, grouped by the health areas covered in the WHO/Defra reviews, with an additional “General health” category. In these tables we have indicated where the GRADE assessment of quality of evidence comes from. Where there are quality of evidence assessments from multiple sources, we have indicated these separately, along with a synthesis GRADE assessment conducted as part of the current REA. The above-named appendices present the detail of those synthesis assessments. The Defra–RIVM review did not conduct GRADE assessments but we include the conclusions of that review regarding the direction of effect. In the quality of evidence tables this is indicated with the label GRADE not conducted”.

Moderate or high quality ratings require a body of evidence based on multiple high quality studies with low risk of bias in their methods and consistent findings. This is a demanding threshold and consequently the quality of evidence for most outcomes is very low or low. This primarily reflects features of the studies that have contributed the evidence. First, most studies are cross-sectional rather than longitudinal, which means the evidence they provide is inherently of lower quality. Second, many studies have relatively small samples. This makes it hard to obtain high certainty that observed associations are not due to chance. Small samples also mean that it is harder to detect a real association if there is one. We discuss study designs in the next chapter.

Table 1 Summary of the quality of evidence for birth and reproductive health outcomes

Outcome	Quality of evidence – Direction of effect	Source of GRADE assessment
Congenital malformations	Very low quality – No overall effect stated in GRADE assessment but harmful effects reported in narrative review	WHO review
Low birth weight	Very low quality – No overall effect stated in GRADE assessment but harmful effects reported in narrative review	WHO review
Preterm birth	Very low quality – No overall effect stated in GRADE assessment but harmful effects reported in narrative review	WHO review

Table 2 Summary of the quality of evidence for cognition outcomes

Outcome	Quality of evidence – Direction of effect	Source of GRADE assessment
Assessments of student distraction	Very low quality – Harmful effect	Defra-Arup review
Attention	Low quality – No effect	WHO review
Executive function deficit (working memory capacity)	Very low quality – No effect	WHO review
Impairment assessed through SATs	Moderate quality – Harmful effect	WHO review

Outcome	Quality of evidence – Direction of effect	Source of GRADE assessment
Reading and oral comprehension	Moderate quality – Harmful effect Very low quality – Harmful effect Moderate quality – Harmful effect	WHO review Defra-Arup review ICCAN synthesis
Short-term and long-term (episodic) memory	Moderate quality – Harmful effect	WHO review

Table 3 Summary of the quality of evidence for sleep outcomes

Outcome	Quality of evidence – Direction of effect	Source of GRADE assessment
Physiologically measured awakenings in adults	Moderate quality – Harmful effect Low quality – Harmful effect Moderate quality – Harmful effect	WHO review ICCAN review ICCAN synthesis
Self-reported sleep quality	Very low quality – Harmful effect	ICCAN review
Self-reported sleep coping behaviours	Very low quality – Harmful effect	ICCAN review
Self-reported awakenings	Low quality – Harmful effect	ICCAN review
Self-reported sleep disorder	Very low quality – No effect	ICCAN review
Self-reported sleep disturbance in adults (source not specified)	Very low quality – Harmful effect	WHO review
Self-reported sleep disturbance in adults (source specified)	Moderate quality – Harmful effect GRADE not conducted – Harmful effect Low quality – Harmful effect Moderate quality – Harmful effect	WHO review Defra-RIVM review ICCAN review ICCAN synthesis

Table 4 Summary of the quality of evidence for cardiovascular and metabolic outcomes

Outcome	Quality of evidence – Direction of effect	Source of GRADE assessment
Arterial stiffness	Low quality – Harmful effect	ICCAN review
Blood pressure	Very low quality – No effect	ICCAN review
Blood pressure in children	Very low quality – No effect	WHO review
Cortisol levels	Very low quality – Harmful effect	ICCAN review
Diabetes incidence	Low quality – No effect GRADE not conducted – Harmful effect Low quality – Harmful effect Low quality – Harmful effect	WHO review Defra-RIVM review ICCAN review ICCAN synthesis
Diabetes prevalence	Very low quality – No effect	WHO review
Heart rate	Very low quality – Harmful effect	ICCAN review
Hypertension incidence	Low quality – No effect GRADE not conducted – Harmful effect Low quality – Harmful effect	WHO review Defra-RIVM review ICCAN synthesis

Outcome	Quality of evidence – Direction of effect	Source of GRADE assessment
Hypertension prevalence	Low quality – No effect	WHO review
Incidence of central obesity	GRADE not conducted – Harmful effect	Defra-RIVM review
Ischaemic heart disease incidence	Very low quality – Harmful effect GRADE not conducted – Harmful effect Low quality – Harmful effect Low quality – Harmful effect	WHO review Defra-RIVM review ICCAN review ICCAN synthesis
Ischaemic heart disease mortality	Low quality – No effect	WHO review
Ischaemic heart disease prevalence	Very low quality – No effect	WHO review
Asymptomatic heart damage	Very low quality – Harmful effect	ICCAN review
Obesity (change in BMI)	Low quality – No effect	WHO review
Obesity (change in waist circumference)	Moderate quality – Harmful effect	WHO review
Obesity (incidence of overweight)	GRADE not conducted – Harmful effect	Defra-RIVM review
Obesity (weight gain)	GRADE not conducted – Harmful effect	Defra-RIVM review
Self-reported diagnosis of arrhythmia	Very low quality – No effect	ICCAN review
Self-reported diagnosis of diabetes	Very low quality – No effect	ICCAN review
Self-reported diagnosis of heart disease	Very low quality – No effect	ICCAN review
Self-reported diagnosis of hypertension	Very low quality – No effect	ICCAN review
Stroke incidence	Very low quality – Harmful effect Moderate quality – Harmful effect Moderate quality – Harmful effect	WHO review ICCAN review ICCAN synthesis
Stroke mortality	Moderate quality – No effect GRADE not conducted – Harmful effect Moderate quality – No effect	WHO review Defra-RIVM review ICCAN synthesis
Stroke prevalence	Very low quality – No effect	WHO review

Table 5 Summary of the quality of evidence for quality of life, mental health and wellbeing outcomes

Outcome	Quality of evidence – Direction of effect	Source of GRADE assessment
Wellbeing of children	Very low quality – No effect	ICCAN review
Depression prevalence	Low quality – No effect	ICCAN review
Depression prevalence mediated by annoyance	Low quality – Harmful effect	ICCAN review

Outcome	Quality of evidence – Direction of effect	Source of GRADE assessment
Emotional and conduct disorders in children	Low quality – No effect	WHO review
Hyperactivity	Low quality – Harmful effect	WHO review
Interview measures of depression and anxiety	Very low quality – Harmful effect Low quality – Harmful effect Low quality – Harmful effect	WHO review Defra-Arup review ICCAN synthesis
Medication intake to treat anxiety and depression	Very low quality – Harmful effect	WHO review
Self-reported QOL or health	Very low quality – No effect Very low quality – No effect Very low quality – No effect	WHO review Defra-Arup review ICCAN synthesis
Wellbeing	Very low quality – Harmful effect	Defra-Arup review
Self-reported diagnosis of chronic headaches/migraine	Very low quality – No effect	ICCAN review
Children's medication intake	Very low quality – No effect	ICCAN review
Children's physical diseases	Very low quality – No effect	ICCAN review

Table 6 Summary of the quality of evidence for cancer and general health outcomes

Outcome	**Quality of evidence – Direction of effect**	**Source of GRADE assessment**
Incidence of breast cancer	Low quality – Harmful effect	Defra-Arup review
Self-reported general health	Very low quality – No effect	ICCAN review
Self-reported diagnosis of stomach ulcer	Very low quality – No effect	ICCAN review
General physical health of children	Low quality – No effect	ICCAN review
General physical health of children mediated by annoyance	Low quality – Harmful effect	ICCAN review

4 Evidence gaps and potential for new research

This chapter will outline ways of thinking about gaps in the evidence (section 4.1), and principles that could guide decision-making on what to prioritise in future research (4.2). It will address research questions 2b on where further evidence is needed, 2c on where the priority evidence gaps are, and 2d on research approaches and methods which could fill the evidence gaps.

4.1 Evidence gaps

4.1.1 Gaps in the evidence

There are several ways in which evidence on the effects of aviation noise is lacking. Health outcomes for which evidence is lacking include all of those relating to dementia and neurodegenerative outcomes, as well as many birth and reproductive outcomes; quality of life, mental health and wellbeing; and many outcomes relating to cardiovascular and metabolic health. Although there is a good representation of moderate quality evidence for sleep-related and cognitive outcomes, there are still many outcomes in these areas for which the quality is low or very low.

Across all outcomes where there is evidence, the large majority is of low or very low quality (of the 58 outcomes shown in Tables Table 1 to

Table 6, evidence for 16 is of a low quality and for 30 of a very low quality). Considering low quality evidence as a form of gap, the evidence base consists primarily of gaps. It is however important to distinguish between a lack of evidence and a lack of evidence *of an effect*. Quality of evidence relates primarily to study design and execution. It is possible to have high quality evidence of no effect.

The smaller the effect, the more difficult it is to gain evidence that allows us to be certain of the effect. We discuss study “power” and the difficulty of detecting small effects below. It is worth noting that even if the effects of noise across various outcomes are small, these may add up to a substantial health burden at a population level if there is a large number of people exposed. This may, however, be difficult to detect with certainty.

All the health outcomes considered have causes beyond aviation noise. The likely role of aviation noise in overall morbidity, compared to other environmental, social and genetic factors, will vary between outcomes. As noted by Weihofen et al. (2019) in relation to stroke [34], the low relative importance of aviation noise as a cause of most chronic disease outcomes means that even with universal exposure to high levels of aviation noise, the effect on overall morbidity and mortality would be small. We would therefore expect larger effects for outcomes where aviation noise was a more important exposure.

Given the difficulty in achieving “high” quality evidence on the GRADE scale for environmental exposures, WHO in its 2018 recommendations [6] used evidence of moderate quality as the basis for setting “strong” recommendations, which “can be adopted as policy in most situations”. It is reasonable to consider outcomes for which there is already moderate quality evidence, such as those in Table 7, as not a priority for further research. However, even for these outcomes, there would be value both in quantifying the adverse effect with more precision, and in assessing the potential for interventions and operational changes to reduce the harmful effect.

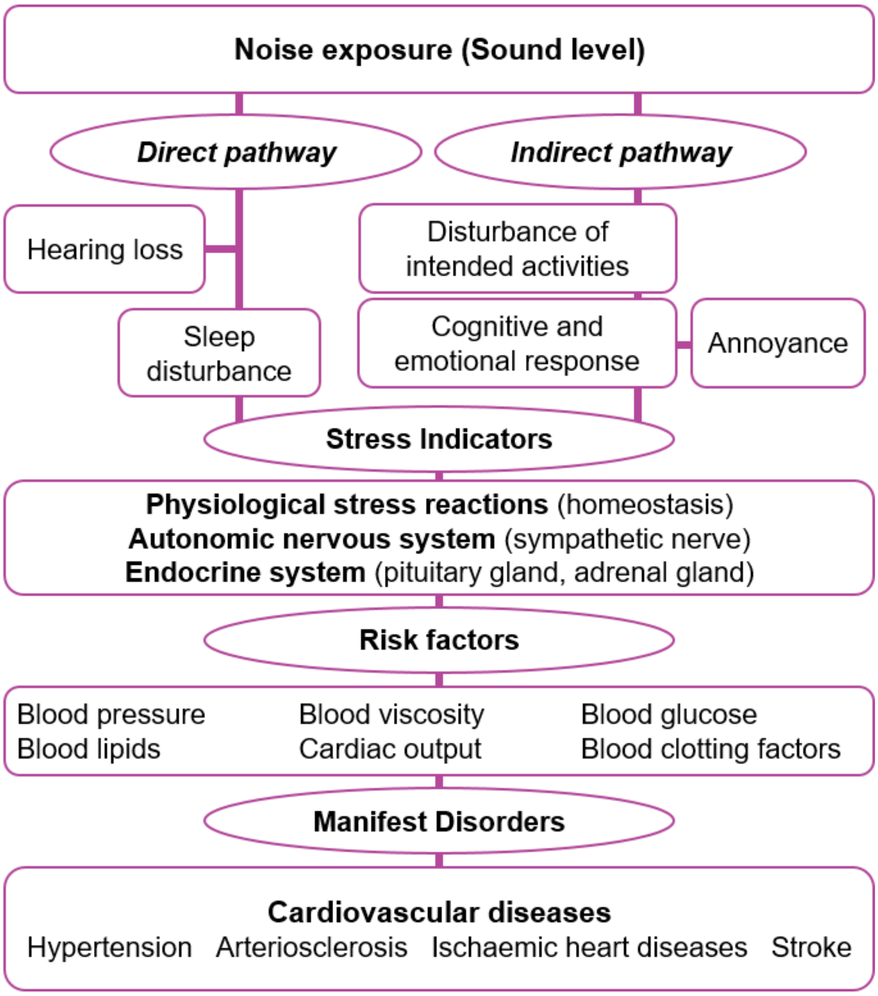
Table 7 Outcomes for which there is moderate quality evidence from WHO, Defra and ICCAN reviews

Outcome	**Direction of effect**
Stroke mortality	No effect
Stroke incidence	Harmful effect
Self-reported sleep disturbance in adults (source specified)	Harmful effect
Physiologically measured awakenings in adults	Harmful effect
Change in waist circumference	Harmful effect
Reading comprehension	Harmful effect
Impairment assessed through SATs	Harmful effect
Short-term and long-term (episodic) memory	Harmful effect

As well as chronic and acute health outcomes it is also possible to study the intermediate mechanisms by which aviation noise causes ill health. Figure 2 shows how noise might manifest in cardiovascular ill-health, and the mechanisms via stress indicators are plausible also for metabolic disorders. Measurable biological indicators (“biomarkers”) of stress responses include cortisol, which is regulated by the endocrine system. Our update includes a paper [31] which investigates the role of aviation noise in cortisol level and rhythms. Although there has long been an interest in the role of stress indicators in the relationship between noise and health outcomes (e.g. [37]), there is relatively little recent literature. The WHO and Defra reviews did not include evidence relating to these intermediate mechanisms, and how they are affected by aviation noise has not recently been systematically reviewed. This is an important area for further research.

Higher quality evidence on the relationship between aviation noise and risk factors (including stress hormones) may be easier to obtain than higher quality evidence on downstream disease outcomes, as risk factors are more prevalent. It is worth noting that evidence related to sleep disturbance meets this description, since sleep disturbance has a role in physiological stress reactions, as well as being a quality-of-life issue in its own right.

Figure 2 Noise effects pathway for cardiovascular diseases (from Babisch 2014 [38])



4.1.2 How to prioritise filling evidence gaps

There is no single answer as to which of these outcomes is “most important” and for which to seek higher quality evidence of the effect of aviation noise. This section outlines several possible approaches to choosing what to prioritise when seeking stronger evidence on the health impacts of aviation noise, and concludes with some suggestions for priority areas for new research.

What is adequate evidence to support action?

From the perspective of potential policy responses, it may be worth taking a step back from the specific outcome-related evidence to consider what policy options are available, and what evidence would be required to decide to undertake or to rule out those options. ICCAN engages with a wide range of stakeholders, whose different priorities may require different new evidence. These include residents affected by aviation noise, airlines, airports, local authorities and regulatory bodies. ICCAN aims to understand aviation noise issues from various perspectives and this approach may also be important when prioritising health outcomes. The responses to ICCAN's corporate strategy reflect the various health priorities for different stakeholders [39]. This ranges from campaign groups emphasising the mental health effects of aviation noise on communities, to local authorities calling for evidence to support their local public health and wellbeing commitments.

To take a specific example, there is moderate quality evidence of harmful effects on several cognitive outcomes related to children's learning and low/very low quality evidence on other outcomes (Table 2). Different actors may respond differently to this mixed evidence base. For example, the implications for action are different for those whose work relates to the exposure, namely those making aviation policy, compared to those whose work relates to the outcome, in this case those educating children in an area of high aviation noise.

For those educating children, it may be valuable to have better quality evidence across the whole range of cognitive outcomes, including those for which evidence is currently low quality, to support targeted remedial responses. From a policy perspective however, including for those setting aviation policy, one may conclude that the evidence of a harmful effect on reading comprehension is adequate to support efforts to reduce the impact of aviation noise on the school environment. Evidence across multiple measures of cognitive ability may not be needed if reading comprehension can be treated as a good proxy for subsequent attainment and life chances. Aviation policymakers, or industry actors seeking to minimise their impact on local communities, might focus attention and further research on how to achieve those reductions. Such mitigation efforts, including understanding the effects of different airspace organisation measures, fall beyond the scope of this REA but there is a further systematic review in the WHO series considering the topic of interventions [40].

Disease endpoints versus intermediate mechanisms

Some of the outcomes studied are disease endpoints, particularly chronic diseases such as diabetes, heart disease or depression. Other outcomes are intermediate mechanisms that increase individual risk for disease, such as cortisol levels, increased waist circumference or arterial stiffness. Outcomes related to sleep disturbance both have a short-term negative effect on people's lives and are risk factors for longer term morbidity. Cognitive outcomes may not cause physical or psychological morbidity, but affect people's lives through educational attainment and social development, in part determining future opportunities and challenges. Both exposures and outcomes are likely to reflect broader social and health inequalities characterised by factors such as social class and ethnicity.

Since effects of noise on sleep and on metabolic disruption potentially affect multiple chronic disease outcomes, a robust understanding of these effects on upstream factors would be beneficial for certainty about the extent of the role noise may play in these outcomes. There is already moderate quality evidence relating to some sleep effects, although more could be done. For indicators of metabolic disruption, the evidence base is lower quality, and this is a potential area of focus.

What is the disease burden (attributable to aviation noise)?

Burden-of-disease or economic studies can quantify the population impact of environmental health risks, which could be aggregated across health outcomes to give an overall burden attributable to aviation noise. This may be desirable, particularly to support cost–benefit analyses, but it is worth noting that these generally rely in turn on estimates of association for which the evidence may be of low quality. Where evidence comes from underpowered studies (see section 4.2.1 for a discussion of study power) the benefits may be underestimated as true effects may not have been detected.

Priority areas for new research?

The above considerations will help to guide decision-making on what new research areas to prioritise. Some possible starting points for prioritisation include the following.

Under-researched areas
There is currently no evidence on the effects of aviation noise on dementia and other neurodegenerative outcomes. Combined with the high prevalence of such disorders among the older population, this lack presents a rationale for seeking evidence on these outcomes.

Similarly, diabetes and hypertension are also sources of substantial morbidity at the population level, for which the evidence is currently only of low or very low quality. There is only evidence (low quality) of the impact of aviation noise on a single cancer outcome.

There is very low quality evidence on birth and reproductive outcomes. The potential contribution of aviation noise exposure, via maternal stress responses, to outcomes such as low birth weight or prematurity may be minimal compared to other exposures (as is true for many cardiovascular and metabolic outcomes). The importance of studying birth-related outcomes is increased by the long-term morbidity that they can cause and they could be considered as an area for further research.

Finally, neither the update review nor the systematic reviews for WHO and Defra include any evidence relating to auto-immune disorders.

Areas of high salience
ICCAN will know from its stakeholder engagement what topics are most salient for affected communities. In terms of areas of contemporary policy focus in the recovery from the coronavirus pandemic, it may be salient to emphasise outcomes related to the disadvantage caused by the pandemic and measures taken in response. These include educational attainment, which related to aviation noise through the effect on cognition, as well as potentially childhood obesity (on which there is currently no evidence related to aviation noise) and mental health outcomes at all ages.

In the short term, research relating to the impact of the massive reduction in flights during the pandemic response, and their subsequent resumption, may be of great public interest. However, the potential is limited as the most plausible short-term outcomes are those related to the quality of life, mental health or wellbeing, all of which will have been simultaneously impacted by the broader pandemic beyond its effect on aviation noise.

There are several priority areas for Public Health England in 2020 to 2025 [41] for which aviation noise may be a relevant exposure or co-exposure. Air pollution is an important exposure that commonly occurs alongside aviation noise. For cardiovascular disease especially, it would be beneficial to understand how these exposures interact. Mental health is a priority area of health policy and there are evidence gaps relating to the effect of aviation noise on adult mental health per se, and effects on people with pre-existing poor mental health. Regarding children’s mental health, there are systematic reviews currently underway on the effects of aviation noise. Research using standard mental health interview measures should be encouraged for comparability with other work on mental health. Childhood obesity is another priority outcome for which aviation noise may be a relevant exposure but for which there is no evidence. Aviation noise may also act in concert with other pollutants and social stressors, including air pollution and poverty, to exacerbate health inequalities, reducing which is another health priority.

Intermediate mechanisms
It would be valuable to improve our understanding of the role of aviation noise in causing stress responses that contribute to multiple chronic cardiovascular and metabolic disorders. Attention should be paid within such research to the potential for mediation by annoyance: the present update found that annoyance mediated effects on several outcomes but not on cortisol disruption. There is potential to study this relationship further to determine whether stress responses constitute a separate pathway to ill health from those mediated by annoyance. Similarly, it would be desirable to understand the role of sleep disruption as an intermediate mechanism for longer-term ill health.

4.2 Research options and considerations

This section will outline metrics and approaches that have been used in studies of aviation noise and health, and available study designs and their strengths and weaknesses. It addresses research questions 1b, on the research approaches and methods used to link aviation noise and health, and 2d on the research approaches and methods which could fill the evidence gaps.

4.2.1 Potential study designs and their strengths and weaknesses

There are a range of trade-offs that characterise the choice of study design, involving statistical certainty, public health relevance, cost, duration, timeliness and feasibility. Feasibility, in turn, relates to factors including the invasiveness/intrusiveness of data collection, and the likelihood of individuals adhering to the research protocol.

Study designs

Longitudinal studies
Longitudinal studies (which include cohort studies and panel studies) involve recruiting people who are followed over time, with repeated data collection on both exposures and outcomes. From the point of view of certainty in the findings, longitudinal studies are generally ideal as they allow the greatest confidence that the exposure preceded the outcome. This is particularly important for chronic disease outcomes that take a long time to manifest. It is also important when considering upstream risk factors as there is natural variation in, for example, the cortisol cycle. Knowing how such factors change over time alongside known noise exposure makes for greater certainty in any observed association. The starting level for longitudinal studies in the WHO review of cardiovascular and metabolic outcomes [12] was “high” quality.

The length of time required to generate the evidence is the main downside to longitudinal studies. The duration of cohort studies has cost implications, and from a policy-making point of view, may miss a window of policy relevance or opportunity. Cohort studies also have to be large and lengthy to detect differences in relatively rare outcomes with high certainty. One of the largest cohort studies to have contributed important evidence on the relationship of aviation noise to health is NORAH (Noise-Related Annoyance, Cognition, and Health, Germany [42]), which conducted three waves of data collection over three years, one wave before and two waves after a new runway opened at Frankfurt airport and night flights were banned.

Cross-sectional studies

Cross-sectional studies involve finding out the population health and noise exposure status at a single point in time. These findings are analysed to determine whether there are associations between health outcomes and noise exposures. For example, you might ask people exposed to different noise levels whether they had been diagnosed with heart disease in the last year. An association between heart disease and noise would mean that, for example, there were more diagnoses of heart disease among those who were exposed to higher noise levels. A cross-sectional study doesn't enable you to say with certainty that the noise caused the heart disease, but as long as other plausible explanations have been taken into consideration (through adjusting your results for confounders), an association helps to generate hypotheses, and may be sufficient to inform policy.

Although evidence gathered over time is ultimately stronger, some outcomes can reasonably be studied with cross-sectional approaches. This is particularly the case for those that occur on short timescales such as sleep-related outcomes or cognitive disruption.

Cross-sectional studies offer a way to generate evidence relatively quickly and at a lower cost than longitudinal studies. Most studies in the field of aviation noise and health are cross-sectional. The starting level for cross-sectional studies in the WHO review of cardiovascular and metabolic outcomes [12] was "low" quality, which contributes to the widespread low quality evidence for many health outcomes. One of the better known cross-sectional studies that have contributed evidence to the understanding of aviation noise effects on health is HYENA, a multicounty study in Europe.

Case–control studies

An alternative to longitudinal or cross-sectional methods, particularly suited to studying rare outcomes, is the case–control methodology. In a case–control study, you begin with a group of people called "cases" who have an outcome, and you seek to compare them to a group of people called "controls" who do not have the outcome. The controls are generally chosen to be similar to the cases in some ways, such as in their age or the neighbourhood they live in. Comparing those who developed an outcome with those who were similar but did not develop the outcome, can help to understand what the cases might have been exposed to that the controls were not. The starting level for case–control studies in the WHO review of cardiovascular and metabolic outcomes [12] was "high" quality. There were no case–control studies among the new studies included in this REA, but an example of this type of study is that by Zeeb et al. (2017) [43]. In that study, cases were all new diagnoses of hypertension in a large health insurance database, and controls were all those in the database without hypertension.

When there are small numbers of cases in the general population, you need a very large cross-sectional or longitudinal study to have adequate statistical power to detect real differences. Case–control studies avoid this problem by starting out with a group of cases. This generally makes them a cheaper study design for rare outcomes. Important disadvantages of the case–control method are the difficulty in choosing the controls so as to avoid selection bias, and the high potential for recall bias regarding what people were exposed to. Recall bias occurs when someone's outcome status (for example, having or not having a disease) affects their likelihood to recall what they were exposed to. Objective measures of exposure (such as address-based noise mapping, as mentioned below in the section Measuring noise) can reduce this risk.

Ecological studies

In contrast to longitudinal, cross-sectional or case–control studies, ecological studies do not assess health outcomes or noise exposure at the individual level. Instead, they assess outcomes and exposure at the population level. For example, health outcomes recorded at the level of electoral wards, publicly available in aggregated, anonymised datasets such as the Local Health dataset published by Public Health England [44], can be combined with noise maps to investigate broad, population-level associations. As they do not involve collecting data from individual participants, ecological studies are relatively cheap and subject to minimal selection biases. They lack precision in that there is no way to tell whether any relationship between exposure and outcome is true for individual people. It is also likely that within the area covered by, for example, an electoral ward there will be variation in sound levels, so the exposure assessment is necessarily crude.

Ecological studies can generally only investigate outcomes that are recorded in administrative datasets, and their data on confounders may be limited. They have the advantage that for those outcomes, their population coverage will be very high, potentially meaning fairly small differences or relatively rare outcomes can be studied, or high precision achieved. However, their lack of individual assessment of exposure and outcome mean they only provide low quality evidence. The starting level for ecological studies in the WHO review of cardiovascular and metabolic outcomes [12] was “very low” quality. There were no ecological studies among the new studies included in this REA, but an example of this type of study is that by Hansell et al. (2013) [45] who assessed hospital admissions for and mortality from cardiovascular conditions in areas exposed to different levels of noise around Heathrow airport.

Meta-analysis
Meta-analysis is a research method that combines the results of multiple studies to give a summary result across all those studies. This effectively increases the sample size, which increases study power and the precision of estimates. Studies included in a meta-analysis need to measure the same outcome in the same way, and the same exposure in the same way. For example, if studies use a cut-off to categorise noise exposure into “high” and “low” categories, this cut-off needs to be the same or very similar for the meta-analysis to be valid. The study populations should be similar, so that combining them is valid. For example, combining multiple studies of adults of similar age ranges from different settings may be fine, but combining studies with very different age eligibility for participants may not. This consistency of methods is the greatest challenge to meta-analysis, but where it can be achieved this is a powerful tool to make the most of existing studies. Meta-analysis doesn’t have a starting level for quality in GRADE. For our GRADE synthesis we have treated meta-analysis as having a starting level of “high” quality.

Table 8 provides a summary of the strengths and weaknesses of each of the potential study designs.

Table 8 Study design strengths and weaknesses

Study design (GRADE starting level [12])	**Strengths**	**Weaknesses**
Longitudinal (high)	High quality of evidence due to prospective assessment of exposures and outcomes Less potential for recall bias	Relatively long time to generate evidence High cost Potential differential attrition
Cross-sectional (low)	Relatively quick and low cost	Cannot assess causation
Case–control (high)	Efficient for rare outcomes	Subject to recall bias

		Require careful attention to confounding
Ecological (very low)	Low cost High population coverage	Descriptive only, individual-level analysis not possible
Meta-analysis (no starting level)	Can increase power and precision of estimates	Require multiple similar studies to be done robustly

Considerations when designing studies

Selection biases
Selection biases occur when people who take part in a study differ in a non-random way from the populations they are supposed to represent. These biases can affect all studies. This is especially true where individuals consent to take part (as opposed to studies using anonymised area-based medical records, for example). Selection bias can affect:

- who is considered for participation (if, for example, a sampling frame doesn't include all people living in an area)
- who is approached for participation (if, for example, recruitment is conducted via a channel that is not accessible to all participants, or at a time when some types of people are not at home)
- who consents to participate (if, for example, understanding or willingness to participate or motivation to participate differ by population group)
- who manages to participate (if, for example, ability or willingness to ultimately take part in the research differs for different types of people)

Longitudinal studies can additionally be subject to differential attrition: in addition to differences between who does and does not initially agree to take part, the people who remain in the cohort may be systematically different to those who drop out (or are "lost to follow-up"). People stop participating in studies for many reasons, including reasons related to the exposure or outcome. For example, people may stop participating because they are sick, or may move home due to aviation noise.

Statistical power
When you want to find something out about a whole population by looking only at a sample of the population, you might miss something that is true about the whole population because your sample was too small. Statistical power tells you how likely you are to detect that true finding in a sample of a given size. Studies ought generally to be designed to have statistical power to detect a given difference in a given outcome. Inadequate power can lead to findings of no effect when a larger study may have found a true effect. Uncertainty over power can therefore make it difficult to know how strongly to interpret the many findings of little or no significant difference.

In practice, it is unusual that authors report explicitly what their studies were powered to detect, but it can be particularly important where a study reports multiple outcomes. An example is the study by Rojek et al. (2019) on indicators of cardiac health, which reported over 40 combinations of outcome and population stratum. The authors reported that the study was powered to detect a difference in pulse wave velocity (PWV), and reported indicators related to asymptomatic heart damage alongside PWV. Among those indicators of asymptomatic heart damage, some had significant associations with the noise level and some did not. It is possible that the study lacked power to detect meaningful differences in some or all of those indicators. A study may in practice be powered for secondary outcomes, but it is good practice to specify a primary outcome and calculate the necessary sample size with regard to that outcome.

The smaller the absolute effect you wish to detect, the greater power is needed: to detect a difference of 30% vs 33% prevalence of an indicator requires more power than to detect a difference of 30% vs 40% prevalence. Likewise, smaller relative effects require greater power: to detect a difference of 3 percentage points between 30% and 33% requires greater power than to detect a difference of 3 percentage points between 10% and 13%. Power is directly related to sample size, and to get more precise estimates or detect smaller effects, larger sample sizes are needed.

4.2.2 Measuring noise

Variation in how noise exposure is assessed relates to choice of metric, measurement versus modelling and indoor versus outdoor measurement.

Choice of noise metrics

There is a thorough consideration of the range of noise metrics in the recent ICCAN review of aviation noise metrics and measurement [4]. Appropriate metrics depend on the health outcome of interest and the mechanism by which noise is thought to cause harm. Where the harmful noise exposure is thought to be the overall level, causing cumulative chronic stress, studies may choose to use average sound-level metrics based on L_{eq} such as the L_{den} metric. The weightings in L_{den} emphasise evening and night-time noise and thus incorporate the adverse consequences of noise into the metric itself. Where the harm primarily comes through short-term disturbance rather than overall level, as may be the case for sleep-related outcomes and cognitive outcomes, studies may use maximum sound level metrics (such as L_{Amax}), number above metrics (for example N65, the number of sound events exceeding 65 dB) or measures of intermittency.

The European Noise Directive [46] mandates strategic noise mapping using of L_{den} to assess annoyance and L_{night} to assess sleep disturbance. This requirement appears in English law in the Environmental Noise (England) Regulations 2006, and the legislation has also been transposed into law in Scotland, Wales and Northern Ireland. L_{den} and L_{night} are also the main indicators used in the WHO ENG2018 recommendations for the European region [6], a recommendation which in turn reflects their widespread use. To enable comparability between studies, it is important to include these metrics if average sound pressure level metrics are appropriate, although researchers report a wider range of metrics (and the aviation industry also uses a range of other metrics). It is valuable to be able to combine the findings of different studies in meta-analysis, which requires the use of comparable metrics. Particularly when using thresholds of “high” and “low” noise, researchers should consider in their study design and reporting how to ensure their findings will be comparable with others on the same topic.

Authors do not always describe why they choose specific metrics over others, and sometimes report similar metrics within the same paper. For example, in the study of salivary cortisol [31] the authors reported both L_{den} and $L_{Aeq,24hr}$, which are both average continuous sound pressure level metrics measured for the whole day, with L_{den} having a penalty added for evening and night-time noise. The authors did not state why they included both or how their interpretation of the presence or absence of an association with the outcome would vary according to which of the two metrics was associated.

The L_{eq} based metrics measure time-averaged sound pressure, whereas other metrics measure the degree to which sound is “eventful”. According to Brink et al. (2019), the intermittency ratio “expresses the energetic contribution of individual noise events from a specific noise source relative to the total sound energy (produced by all noise sources together) in a given time period” [25]. Another type of event-related metric is the “number above” metric which indicates the number of events within a specified time period exceeding a given decibel level. This metric is less frequently used than sound pressure metrics (either continuous or event-related) and did not feature in any of the studies included in our update. Nonetheless, similar to the intermittency ratio, it has the potential to quantify how much ongoing disturbance is caused by aircraft noise events, which a sound-pressure event measure such as L_{Amax} cannot.

Analytical approaches to noise exposure include categorisation by high and low noise areas (as in Rojek et al. (2019) [30]) and analysis by decibel level, using noise level as a continuous variable or bands of exposure (such as 5 dB or 10 dB bands).

For any given level of sound pressure (that is, physical energy), the human ear experiences the sound as more or less loud depending on the pitch. Up to very high frequencies, low sounds are experienced as quieter than high sounds, for any given level of sound pressure. In order to accurately assess noise as people experience it, noise studies use a long-established method called “A-weighting” which takes account of pitch. Most sound metrics (including L_{den} and L_{night}) are A-weighted, even if this is not explicitly stated in study reports.

Measurement and modelling

Noise modelling is an established and efficient method of determining external noise levels at a geographical location. A commonly used programme in the UK for modelling noise contours is ANCON (Aircraft Noise CONtour model) which is owned and operated by the CAA. Another modelling programme is AEDT (Aviation Environmental Design Tool), which is commercially available and developed by the FAA in the USA. Noise modelling uses multiple factors such as flight patterns and aircraft type to estimate how noise from aviation is experienced at ground level. Noise maps generated through modelling are routinely produced by airport authorities and regulators and provide the large-scale estimates of the numbers of people affected by given levels of aviation noise.

In contrast to noise modelling, monitoring noise involves using microphones to record the actual sound levels in a given setting. Modelling is used for estimating outdoor noise by the aviation industry, but measurements via monitoring are catered. This is an important part of ratifying the modelled outputs to real-world values. Monitoring is also useful for gathering specific local information.

Noise monitoring can be conducted by researchers both indoors and outdoors. In social and health research it is particularly important to have accurate levels of noise at the participant’s location indoors, most notably for studies of sleep-related and cognitive outcomes, to determine exposure more accurately than with outdoor measurements alone. Ideally it is possible to separate aviation noise from ambient noise, which is also recorded by measurement equipment, and some studies have attempted to do this (for example, Nassur et al. 2019 [32]).

Indoor and outdoor estimation

Whether noise levels are modelled or determined via monitoring outdoors, there is the inherent problem that outdoor noise does not necessarily determine indoor noise.

Residents may take different steps to mitigate their indoor noise level exposure depending on the levels of outdoor noise and personally perceived annoyance. Such steps may include installing double glazing or roof noise insulation. In terms of how this may influence effect estimates, if mitigation efforts are more likely with higher indoor noise levels, this would probably dampen any apparent effect comparing exposures classified by outdoor noise levels as the indoor level reduction will be greater for those who have installed insulation.

There are methods for estimating indoor noise from outdoor noise: for example, Brink et al. (2019) [25] describe accounting for the position of the bedroom within the dwelling, and also reducing the indoors exposure based on the position at which the participant keeps their window (open/half-open/closed).

Residential or other address-based measures all suffer from the limitation that people's noise exposure may not be the same as the noise level at their address, particularly for daytime levels. People may therefore experience noise exposures away from their homes and not experience noise exposure at their homes. The amount of time spent at home during the day will differ according to factors related to work (daytime work away from the home versus working from home, or shift work including daytime sleeping) and caring responsibilities (home-based carers for children, disabled people or elderly people).

4.2.3 Measuring health

Methods used to measure health outcomes include diagnoses of physical or mental health conditions, and short-term measurements by self-report, interview or monitoring. Harmonised, standardised methods are increasingly used, which is positive for encouraging comparability and the potential for meta-analyses.

Chronic diseases such as diabetes and cardiovascular outcomes tend to be measured by reported medical diagnosis. Mental health outcomes such as depression and anxiety are usually assessed by interview during data collection but can also be assessed by participant-reported diagnosis. Similarly, high blood pressure can be ascertained either by reported diagnosis, by reported medication use or by measurement during the study (either measurement by study staff or by self-administered equipment). Diagnoses can also be measured without involving participants, for outcomes recorded in administrative health databases (see below on ecological studies).

Some outcomes are mainly subjective, including many measures of quality of life or wellbeing (or annoyance), and are based primarily on self-reports although validated questionnaire instruments are available. That said, the WHO quality of life, mental health and wellbeing review included evidence, for non-aviation noise sources, related to a range of diagnosed conditions (such as children's hyperactivity or emotional and conduct disorders) or physical outcomes (such as measured cortisol levels).

Some outcomes are measured by both self-report and objective measures. For example, sleep quality and sleep disturbance can be measured by a range of self-reported measures (tiredness, trouble falling asleep, remembered awakenings etc.) and can also be defined by measures such as awakenings or movement derived from physiological monitoring equipment. Polysomnography is the gold standard approach to measuring sleep disturbance physiologically, and involves monitoring individuals' brain activity, eye movements, muscle tone, breathing, movement and other signals. Polysomnography is expensive and intrusive, however Basner et al. (2019) report that a simpler combination of monitoring heart activity and movement only performs almost as well [27].

For self-reported items, there are often standard questions, which enables comparability between studies. Among the studies summarised in this update, for example, sleep quality was measured using the Pittsburgh sleep quality index, children's wellbeing was assessed using the KINDL-R questionnaire, and sleep disturbance was measured using an adapted ICBEN scale. There are standardised tests for assessing children's cognitive abilities, and children's mental health (such as the Strengths and Difficulties Questionnaire).

Some intermediate risk factors might not commonly be ascertained outside the study setting, and require collection during the study. These include biomarkers which may be collected by the participant themselves (for less invasive procedures, such as a saliva sample for cortisol) or by study nurses (for procedures such as a blood sample for C-reactive protein).

4.2.4 Potential ways of filling evidence gaps

Here we outline some options for studies that could help improve the quality of evidence relating to many health outcomes.

Retrospective cohort methods

The UK has a series of high quality population cohorts that collect data on birth and other cohorts over many years. Birth cohorts include those of people born in 1958, 1970 and 2000, with around 18,000 members each and data collection every few years. The UK Household Longitudinal Survey has followed 40,000 households (100,000 individuals) with annual data collection since 2009 and is integrated with the British Household Panel Survey following 10,000 people back to 1991. The Avon Longitudinal Study of Parents and Children (ALSPAC) has intensively followed the families of 14,000 pregnant women recruited to the study in 1991 and 1992. Biobank recruited 500,000 participants aged between 40 and 69 in 2006 to 2010, and has followed them since, collecting biological and genetic samples and other health-related information. All these studies have rich data on a wide range of characteristics adequate to adjust for confounding, and some include biomarkers. Biobank also includes noise data modelled for participants' home addresses for some years.

Most cohorts do not contain noise data, so the feasibility of retrospective cohort methods to study aviation noise exposure would depend on the ability to map noise levels back on to study participants' addresses. The first step in considering such a study would be to seek expert opinion on such backward mapping including asking whether there are enough cohort members who experience high levels of aviation noise to have sufficient power to detect health effects of interest.

Perhaps the most promising cohort for retrospective noise mapping is the Southall And Brent REvisited Study (SABRE) which has followed the health of around 5000 people recruited in 1988 to 1991. Due to its West London location and proximity to Heathrow, this cohort has a greater chance of adequate numbers of participants exposed to aviation noise to be able to be powered for studying health outcomes. There is ongoing work analysing SABRE and Biobank data in the Aircraft Noise and Cardiovascular Outcomes (ANCO) study [48], which there may be potential to build upon with further funding. There is also an ongoing retrospective study to investigate short-term variation in cardiovascular outcomes associated with short-term changes in aviation noise exposure: the Reduced noise Impacts of Short-Term Aircraft Noise and Cardiovascular Outcomes (RISTANCO) study is using historical data on flight movements to generate address-based noise estimates linked to data on hospital admissions and mortality [49].

There are examples of similar work being done with non-aviation noise. For example, Smith et al. (2017) modelled road traffic noise onto addresses of a retrospective cohort of births to estimate the impact of road traffic noise and air pollution on birth weight in London [50], and Zeeb et al. (2017) used retrospective exposure mapping in their case–control study [43]. The latter study also describes the steps taken to ensure protection of sensitive data.

Baseline data for operational or infrastructure changes

Any infrastructure added or removed, or any operational change, presents an important opportunity to assess impacts of those changes, especially on shorter-term outcomes. Airspace change is strictly regulated and all potential changes ought to be notified to an appropriate agency. The lead time of operational or infrastructure changes varies and for longer term or larger project, there is potential to build research activities into the change process. In order to be able to generate evidence from shorter term changes, it may be worthwhile to pre-emptively collect baseline data on outcomes of interest from airports.

Further meta-analyses

Meta-analysis involves combining the results of existing studies on the same topic to get a single pooled estimate of the effect. This has the advantage of generally increasing the statistical power of the estimate and therefore the certainty of the effect. However, it requires studies to be similar in their definition of exposure and outcome and their study design, otherwise combining studies may be impossible or give spurious results. The WHO review on cognition [10] made a similar observation and added that "many studies group exposure into high and low, using different thresholds for high and low, which again makes combining study data challenging as the range of noise exposure within the high and low categories is often unknown and cannot be estimated reliably from the data provided. The potential to be able to conduct meta-analyses within this field will be greatly enhanced if future studies report effect estimates for a 1 dB and 5 dB increment in noise exposure" (p19).

In studies of quality of life, mental health and cognition, use of standard outcome measures should be encouraged so as to eventually make possible further meta-analyses. Researchers should also consider applying minimum quality cut-offs for inclusion in meta-analyses, to avoid undermining the certainty that might be derived from considering higher quality studies only. This rationale is why, for example, the Defra–RIVM review only included evidence from cohort studies and case–control studies to inform its findings on cardiovascular and metabolic outcomes.

Ideal study design

Here we outline an ideal study design to achieve the highest quality evidence, against which reasonable compromises relating to time, cost and priority can be assessed. The ideal study design to investigate effects of aviation noise on health in the UK would be a large cohort study of populations living around UK airports. Including sufficient airports would mean it could assess differences in exposures related to how airports operate, including night flights and flight path rotation. The study would want to have the following features:

- collecting data every year or two for a decade or more
- strong measures to minimise loss to follow-up including following cohort members who move
- data collection on multiple health outcomes, including:
    - recent and lifetime diagnoses of physical and mental ill health
    - sleep quality via ECG and actigraphy
    - self-completed quality of life and mental health measures
    - interview measures of psychiatric morbidity
- repeated collection of biomarkers from saliva and blood

- data on confounders and effect modifiers, including:
    - air pollution
    - access to green space
    - other noise
    - occupational noise exposure
    - annoyance
    - noise sensitivity
    - dwelling attributes
    - actions taken to mitigate noise
    - age, sex, ethnicity, household income, alcohol and tobacco use, diet, physical activity
- noise measurement and/or modelling to allow calculation of a range of metrics including equivalised overall noise measures of varied durations, maximum noise levels, intermittency ratios and number-above metrics
- noise measured in the bedroom during sleep and in the classroom for cognitive outcomes

Suggestions of specific studies

- A cohort study such as that outlined above would be expensive to set up and run, and would be best undertaken with a view to gathering evidence on a wider range of exposures than just aviation noise. It would be an important investment in generating evidence that could improve certainty of the relationship between aviation noise and a range of outcomes and should be explored as a priority. There may be scope to partner with other agencies to establish a longitudinal research programme that could also add value to evidence bases for exposures beyond aviation noise.
- It is highly advisable to investigate the potential of retrospective cohort methods using UK cohort data, to generate evidence in a relatively timely and cost-effective way. This would include evidence relating to chronic disease outcomes as well as birth and reproductive outcomes. This should begin with investigating the potential of building on the ANCO work to investigate further outcomes, and could also involve assessing the potential for aviation noise analyses with one or more of the large UK general population cohorts. Where cohorts cannot be used for such purposes, retrospective ecological studies using routine health datasets could be considered as an alternative. Such studies could build on the work of the RISTANCO study, which is due to end in late 2020.
- A longitudinal study including data on stress biomarkers, annoyance and disease outcomes would be valuable for being able to distinguish effects mediated through annoyance from those attributable directly to physiological stress responses.
- Despite the potential confounding effects of the pandemic and its response, short term surveys of outcomes including sleep, quality of life, mental health and wellbeing should be considered. These would be followed by further surveys of the same people in future waves over the following year or two to assess the impacts of “return to normal” after the present reduction. This is of course time sensitive and would require rapid action to achieve a baseline during the period of reduced flight activity.

    - This could be combined with a baselining study of operations at major airports to provide a comparator for when operations and therefore exposures change.

- The Defra-RIVM study suggested four new meta-analyses in the area of cardiovascular and metabolic health. Three of these (on IHD, stroke and diabetes) have been conducted and are reported above. The fourth, on hypertension, could be considered although this may already be in process.

5 Discussion

This REA has brought together the evidence available on the health effects of aviation noise. Between the existing WHO and Defra systematic reviews and evidence published subsequently, there is a wealth of data available on a wide range of health outcomes. However, the systematic assessment of quality of evidence using GRADE has found that the large majority of the evidence is of low or very low quality. There is therefore great potential for further research.

Our review found relatively little data from the UK, despite having a large noise-exposed population including the busiest airport in Europe. We have suggested some potential areas for further exploration, including under-researched health outcomes such as dementia and other neurodegenerative outcomes, and birth outcomes alongside health outcomes with low or very low quality evidence. There is no single way to determine what should be studied. Instead, decisions on the research for ICCAN to take forward should be informed by these gaps, combined with the priorities of its stakeholders and current priority areas in wider public health.

In focusing ideas for further research, study design should be a key consideration. This evidence update and the reviews conducted by the WHO and Defra identify design limitations, such as low sample sizes and cross-sectional studies, which tend to result in inconclusive results and therefore low or very low quality evidence.

Longitudinal studies are generally viewed as gold standard and tend to provide high quality evidence. However, they require a substantial budget and time investment. In section 4.2.4 we discuss how such a study could be approached. We have also put forward other types of studies including retrospective cohorts using the rich cohort data available in the UK which would be lower cost but benefit from some of the strengths of longitudinal methods. It remains to be seen what the coming months will bring with regard to the effects of the coronavirus pandemic on aviation noise, but there may be immediate opportunities to exploit the (presumably temporary) reduced exposure.

Generating an evidence base generally involves more than one study, however well designed. A further useful step toward improving the evidence base would be to support collaborative multi-study and international efforts to generate evidence using consistent methods. At present, particularly for sleep-related outcomes but also in other areas, there is some inconsistency in what specific measures are reported for particular outcomes (for example, what question(s) should be used to elicit self-reports of sleep disturbance). Multiple high-quality longitudinal studies may be difficult to achieve in practice, which makes a more consistent approach even more vital: meta-analysis of comparable cross-sectional studies, even if these studies individually offer low quality evidence, would make it possible to generate more precise estimates, which would strengthen the evidence base.

There have been such efforts in the past, including the European Network on Noise and Health (ENNAH) which reported in 2013 and in which UK universities participated [51]. The ENNAH project made a number of recommendations for further research, some of which have been acted upon but many of which remain. Those recommendations are oriented toward a specialist research community and continued engagement with such a community will be vital for ICCAN to make the most of the findings of this REA.

6 References

[1] Airports Commission, “Discussion Paper 05: Aviation Noise,” London, 2013.

[2] European Environment Agency, “The European environment — state and outlook 2020,” European Environment Agency, Luxembourg, 2019.

[3] European Environment Agency, “Reported data on noise exposure covered by Directive 2002/49/EC,” 2019. [Online]. Available: https://www.eea.europa.eu/data-and-maps/data/data-on-noise-exposure-7. [Accessed 22 June 2020].

[4] Independent Commission on Civil Aviation Noise, “Review of aviation noise metrics and measurement,” ICCAN, 2020.

[5] World Health Organisation, “Environmental Noise Guidelines for the European Region,” 2018. [Online]. Available: http://www.euro.who.int/__data/assets/pdf_file/0008/383921/noise-guidelines-eng.pdf?ua=1.

[6] World Health Organization Regional Office for Europe, “Environmental Noise Guidelines for the European Region,” WHO, Copenhagen, 2018.

[7] G. H. Guyatt, A. D. Oxman, G. E. Vist, R. Kunz, Y. Falck-Ytter, P. Alonso-Coello and H. J. Schünemann, “GRADE: an emerging consensus on rating quality of evidence and strength of recommendations,” *BMJ,* vol. 336, pp. 924-925, 2008.

[8] R. L. Morgan, K. A. Thayer, L. Bero, N. Bruce, Y. Falck-Ytter, D. Ghersi, G. Guyatt, C. Hoojimans, M. Landgendam and D. Mandrioli, “GRADE: Assessing the quality of evidence in environmental and occupational health,” *Environment International,* Vols. 92-93, pp. 611-616, 2016.

[9] R. Guski, D. Schreckenburg and R. Schuemer, “WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Annoyance,” *Int. J. Environ. Res. Public Health,* vol. 14, p. 1539, 2017.

[10] C. Clark and K. Paunovic, “WHO environmental noise guidelines for the european region: A systematic review on environmental noise and

cognition," *International Journal of Environmental Research and Public Health,* vol. 15, no. 2, 2018.

[11] M. Basner and S. McGuire, "WHO environmental noise guidelines for the european region: a systematic review on environmental noise and effects on sleep," *International Journal of Environmental Research and Public Health,* vol. 15, no. 3, 2018.

[12] E. van Kempen, M. Casas, G. Pershagen and M. Foraster, "WHO environmental noise guidelines for the European region: A systematic review on environmental noise and cardiovascular and metabolic effects: A summary," *International Journal of Environmental Research and Public Health,* vol. 15, no. 2, 2018.

[13] C. Clark and K. Paunovic, "WHO environmental noise guidelines for the European region: A systematic review on environmental noise and quality of life, wellbeing and mental health," *International Journal of Environmental Research and Public Health,* vol. 15, no. 11, 2018.

[14] Civil Aviation Authority, "Noise Impact Survey Summary of Responses," 2019. [Online]. Available: https://publicapps.caa.co.uk/docs/33/CAP1748%20-%20CAA%20Noise%20Impact%20Survey.pdf. [Accessed 22 June 2020].

[15] Civil Aviation Authority, "Survey of noise attitudes 2014: Aircraft," 2017.

[16] Independent Comission on Civil Aviation Noise, "Review of the Survey of Noise Attitudes 2014," 2019.

[17] Institute for Public Health and the Environment (RIVM), "Review of the evidence relating to environmental noise exposure and annoyance, sleep disturbance, cardio-vascular and metabolic health outcomes in the context of ICGB (N)," RIVM Report 2019-0088, London, 2019.

[18] Defra, "Review of evidence relating to environmental noise exposure and specific health outcomes in the context of the interdepartmental group on costs and benefits (ICGB(N): WP4," London, 2020.

[19] Government Social Research Service, "Rapid Evidence Assessment Toolkit index," 2008. [Online]. Available: https://webarchive.nationalarchives.gov.uk/20140402164155/http://www.civilservice.gov.uk/networks/gsr/resources-and-guidance/rapid-evidence-assessment.

[20] M. J. Nieuwenhuijsen, G. Ristovska and P. Dadvand, “WHO environmental noise guidelines for the european region: a systematic review on environmental noise and adverse birth outcomes,” *International Journal of Environmental Research and Public Health,* vol. 14, no. 10, 2017.

[21] B. C. Wallace, K. Small, C. E. Brodley, J. Lau and T. A. Trikalinos, “Deploying an interactive machine learning system in an evidence-based practice center: abstrackr,” *Proc. of the ACM International Health Informatics Symposium (IHI),* pp. 819-824, 2012.

[22] Civil Aviation Authority, “Aircraft Noise and Health Effects: A yearly update. CAP 1713,” Crawley, West Sussex, 2019.

[23] Civil Aviation Authority, “Aircraft Noise and Health Effects: A six-month update (April 2019- September 2019). CAP 1841,” Crawley, West Sussex, 2019.

[24] Civil Aviation Authority, “Aircraft Noise and Health Effects- a six monthly update. CAP 1993,” Crawley, West Sussex, 2020.

[25] M. Brink, B. Schäffer, D. Vienneau, R. Pieren, M. Foraster, I. C. Eze, F. Rudzik, L. Thiesse, C. Cajochen, N. Probst-Hensch, M. Röösli and J. M. Wunderli, “Self-reported sleep disturbance from road, rail and aircraft noise: exposure-response relationships and effect modifiers in the SiRENE study,” *Internation Journal of Environmental Research and Public Health,* vol. 16, no. 21, p. 4183, 2019.

[26] S. Rocha, M. G. Smith, M. Witte and M. Basner, “Survey results of a pilot sleep study near Atlanta international airport,” *International Journal of Environmental Research and Public Health,* vol. 16, p. 4321, 2019.

[27] M. Basner, M. Witte and S. McGuire, “Aircraft noise effects on sleep: results of a pilot study near Philadelphia International Airport,” *International Journal of Environmental Research and Public Health,* vol. 16, no. 17, p. 3178, 2019.

[28] M. G. Smith, S. Rocha, M. Witte and M. Basner, “On the feasibility of measuring physiologic and self-reported sleep disturbance by aircraft noise on a national scale: a pilot study around atlanta airport,” *The Science of the Total Environment,* vol. 718, p. 137368, 2020.

[29] B. L. Trieu, T. L. Nguyen and T. L. Bui, “Assessment of health effects of aircraft noise on residents living around Noi Bai International

Airport," in *Madrid internoise conference 2019: noise control for a better environment*, Madrid, Spain, 2019.

[30] M. Rojek, M. W. Rajzer, W. Wojciechowska, T. Drożdż, P. Skalski, T. Pizoń, A. Januszewicz and D. Czarnecka, "Relationship among long-term aircraft noise exposure, blood pressure profile, and arterial stiffness," *Journal of Hypertension,* vol. 37, no. 7, pp. 1350-1358, 2019.

[31] C. Baudin, M. Lefèvre, J. Selander, W. Babisch, E. Cadum, M.-C. Carlier, P. Champelovier, K. Dimakopoulou, D. Huithuijs, J. Lambert, B. Laumon, G. Pershagen, T. Theorell, V. Velonaki, A. Hansell and A.-S. Evrard, "Saliva cortisol in relation to aircraft noise exposure: pooled-analysis results from seven European countries," *Environmental Health,* vol. 18, no. 102, 2019.

[32] A. M. Nassur, D. Léger, M. Lefèvre, M. Elbaz, F. Mietlicki, P. Nguyen, C. Ribeiro, M. Sineau, B. Laumon and A.-S. Evrard, "Effects of aircraft noise exposure on heart rate during sleep in the population living near airports," *International Journal of Environmental Research and Public Health,* vol. 16, no. 2, p. 269, 2019.

[33] D. Vienneau, I. C. Eze, N. Probst-Hensch and M. Röösli, "Association between transportation noise and cardio-metabolic," in *Proceedings of the 23rd International Congress on Acoustics*, Aachen, Germany, 2019.

[34] V. M. Weihofen, J. Hegewald, E. Ulrike, P. Schlattmann, H. Zeeb and A. Seidler, "Aircraft Noise and the Risk of Stroke," *Deutsches Arzteblatt International,* vol. 116, no. 14, pp. 237-244, 2019.

[35] S. Benz and D. Screnckenberg, "Examination of the causal relationship between aircraft noise exposure, noise annoyance and diagnoses of depression using structural equation modelling," in *Proceedings of the 23rd International Congress on Acoustics*, Aachen, Germany, 2019.

[36] J. Spilski, M. Rumberg, M. Berchthold, K. Bergström, U. Möhler, T. Lachmann and M. Klatte, "Effects of aircraft noise and living environment on children's well-being and health," in *Proceedings of the 23rd International Congress on Acoustics*, Aachen, Germany, 2019.

[37] W. Babisch, "Stress hormones in the research on cardiovascular effects of noise," *Noise and Health,* vol. 5, no. 18, pp. 1-11, 2003.

[38] W. Babisch, "Updated exposure-response relationship between road traffic noise and coronary heart diseases: A meta-analysis," *Noise and Health,* vol. 16, no. 68, pp. 1-9, 2014.

[39] Independent Comission on Civil Aviation Noise, “ICCAN Corporate Strategy: Summary of responses,” 2019.

[40] A. Brown and I. Van Kamp, “WHO Environmental Noise Guidelines for the European Region: A Systematic Review of Transport Noise Interventions and Their Impacts on Health.,” *Int. J. Environ. Res. Public Health ,* vol. 14, no. 8, p. 873, 2017.

[41] Public Health England, “PHE Strategy 2020 to 2025,” Public Health England, 2019.

[42] Gemeinnützige Umwelthaus GmbH, “Overview of the NORAH Study,” [Online]. Available: http://www.laermstudie.de/en/norah-study/overview/. [Accessed 24 June 2020].

[43] H. Zeeb, J. Hegewald, M. Schubert, M. Wagner, P. Dröge, E. Swart and A. Seidler, “Traffic noise and hypertension - results from a large case-control study,” *Environ. Research,* vol. 157, p. 110–117, 2017.

[44] Public Health England, “Local Health,” 2020. [Online]. Available: https://www.localhealth.org.uk/.

[45] A. Hansell, M. Blangiardo, L. Fortunato, S. Floud, K. de Hoogh, D. Fecht, R. Ghosh and H. Laszlo, “Aircraft Noise and Cardiovascular Disease Near Heathrow Airport in London: Small Area Study,” *British Medical Journal,* vol. 347, p. f5432, 2013.

[46] EUR-Lex, “European Noise Directive,” [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32002L0049.

[47] Hales Swift, “A Review of the Literature Related to Potential Health Effects of Aircraft Noise,” The Partnership for AiR Transportation Noise and Emissions Reduction, Cambridge, MA, 2010.

[48] UK Research and Innovation, “Aircraft Noise and Cardiovascular Outcomes (ANCO),” [Online]. Available: https://gtr.ukri.org/projects?ref=MR/P023673/1.

[49] National Institute for Health Research, “Funding and Awards: Reduced noise Impacts of Short-Term Aircraft Noise and Cardiovascular Outcomes (RISTANCO),” 2016. [Online]. Available: https://fundingawards.nihr.ac.uk/award/15/192/13.

[50] R. Smith, D. Fecht, J. Gulliver, S. Beevers, D. Dajnak, M. Blangiardo, R. Ghosh, A. Hansell, F. Kelly, H. Anderson and M. Toledano, “Impact of London's road traffic air and noise pollution on birth weight:

Retrospective population based cohort study," *BMJ,* vol. 359, p. j5299, 2017.

[51] J. Lekaviciute, S. Kephalopoulos, S. Stansfeld, C. Clark and (eds), "European Network on Noise and Health: Final Report," European Commission Joint Research Centre, Luxembourg, 2013.

Appendix A The GRADE approach

The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) process is an approach to supporting the development of recommendations, the key aspect of which is assessing the quality of evidence. GRADE was initially developed for clinical medicine and has been adapted to other areas including environmental health. This approach encourages transparency consistency in assessing the quality of evidence for a relationship between an exposure and a health outcome [8].

GRADE assesses bodies of evidence against structured criteria to determine the overall quality of evidence for the presence or absence of a causal relationship. The process results in an assessment of “high”, “moderate”, “low” or “very low” evidence, with implications for the need for further research:

- High quality evidence means further research is *very unlikely* to change the certainty of the effect estimate;
- Moderate quality evidence means further research is *likely* to have an important impact on the certainty of the effect estimate and *may* change the estimate;
- Low quality evidence means further research is *very likely* to have an important impact on the certainty of the effect estimate and is *likely* to change the estimate;
- Very low quality evidence means any effect estimate is *uncertain*.

The WHO review of cardiovascular and metabolic outcomes [12] gives some commentary on how to use GRADE for environmental exposures. Study design is a key feature of quality and determines the “starting level” for the assessment. Where the bulk of evidence is from longitudinal or case–control studies, the starting level is “high”. Where it is largely from cross-sectional studies the starting level is “low”, and where it is from ecological studies, the starting level is “very low”. The authors of that review also downgraded the quality of evidence if based on only one study, regardless of the quality of that study.

From the starting level, quality of evidence can be downgraded across the following five domains. It is not always possible to assess each domain (for example, it was beyond the scope of the present review to assess publication bias).

- Study design (no downgrade if most studies have low risk of bias);
- Inconsistency (no downgrade if results across studies are consistent);
- Indirectness (on downgrade if studies are comparing like with like and have comparable populations and assessment of exposures and outcomes);
- Precision (no downgrade if the confidence intervals around the effect estimates are narrow); and
- Publication bias (no downgrade if no publication bias).

Although it is structured, GRADE is not a deterministic approach that gives an automatic outcome. GRADE is applied to bodies of evidence, taking into account all eligible data. As such, studies of differing quality are considered together and reviewers must ultimately judge the balance of that evidence. To avoid the results of inherently lower quality studies affecting the certainty derived from higher quality studies, some authors have treated only studies with a high quality starting level as eligible. While there is scope for reviewers to arrive at different conclusions, the structure and set criteria encourage consistency.

Appendix B Strategies for searches

Databases

Medline

1 ((aviation or aircraft or airport* or air-traffic* or "air traffic" or flight* or airfield* or "air base*" or airbase* or airline* or flight or flights or runway* or aerodrome* or airspace or "air space") adj5 (noise or sound or sounds or decibel* or respite)).ti,ab,kw.
2 Noise, Transportation/ or Environmental Exposure/ or Environmental Monitoring/
3 Aircraft/ae, lj [Adverse Effects, Legislation & Jurisprudence]
4 Aviation/ae, in, lj, pa, px [Adverse Effects, Injuries, Legislation & Jurisprudence, Pathology, Psychology]
5 Airports/
6 2 and (3 or 4 or 5)
7 1 or 6
8 limit 7 to yr="2019 -Current"

Embase

1 ((aviation or aircraft or airport* or air-traffic* or "air traffic" or flight* or airfield* or "air base*" or airbase* or airline* or flight or flights or runway* or aerodrome* or airspace or "air space") adj5 (noise or sound or sounds or decibel* or respite)).ti,ab,kw.
2 ((noise injury/ or noise pollution/ or noise/ or environmental monitoring/ or environmental exposure/) and (aviation/ or aircraft/ or airport/ or helicopter/)) or aircraft noise/
3 1 or 2
4 limit 3 to yr="2019 -Current"
5 limit 4 to exclude medline journals
6 limit 4 to embase
7 5 or 6

Scopus

((aviation OR aircraft OR airport* OR air-traffic* OR "air traffic" OR flight* OR airfield* OR "air base*" OR airbase* OR airline* OR flight OR flights OR runway* OR aerodrome* OR airspace OR "air space") W/5 (noise OR sound OR sounds OR decibel* OR nuisance)) AND ((health OR disease* OR disorder* OR mortality) OR (sleep* OR well-being OR wellbeing OR hypertension OR blood-pressure OR "blood pressure" OR "heart disease*" OR ihd OR angina-pectoris OR "angina pectoris" OR myocard*-infarct* OR "myocardial infarct*" OR cardio* OR *vascular OR stroke OR cva OR diabetes OR diabetic OR obes* OR overweight OR bmi OR body-mass-index OR "body mass" OR dementia OR cancer OR immun* OR endocrine* OR birth OR pregnan* OR fetus OR foetus OR preterm OR pre-term OR gestation OR infert* OR steril* OR malformation* OR labor OR labour OR *natal OR teratogen* OR depress* OR anxiety OR quality-of-life OR stress OR cortisol) OR (cogniti* OR memory OR hyperactiv* OR attention OR comprehen* OR read OR learn)) AND (LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019)) AND (LIMIT-TO (SUBJAREA , "ENVI") OR LIMIT-TO (SUBJAREA , "MEDI") OR LIMIT-TO (SUBJAREA , "SOCI") OR LIMIT-TO (SUBJAREA , "PSYC") OR LIMIT-TO (SUBJAREA , "DECI") OR LIMIT-TO (SUBJAREA , "MULT"))

Epistemonikos

(title:(((aviation OR aircraft OR airport* OR air-traffic* OR "air traffic" OR flight* OR airfield* OR "air base*" OR airbase* OR airline* OR flight OR flights OR runway* OR aerodrome* OR airspace OR "air space") AND (noise OR sound OR sounds OR decibel* or respite))) OR abstract:(((aviation OR aircraft OR airport* OR air-traffic* OR "air traffic" OR flight* OR airfield* OR "air base*" OR airbase* OR airline* OR flight OR flights OR runway* OR aerodrome* OR airspace OR "air space") AND (noise OR sound OR sounds OR decibel*)))) – **(April 2019-Dec 2020)**

Websites

We searched the following websites:

Appendix table 1 Online websites and repositories

Authority	Website address
UK Government	www.gov.uk
Civil Aviation Authority	www.caa.co.uk
Five busiest UK airports (Heathrow, Gatwick, Manchester, Stansted, Luton)	www.heathrow.com; www.gatwickairport.com; www.manchesterairport.co.uk; www.stanstedairport.com; www.london-luton.co.uk
Chartered Institute of Environmental Health	www.cieh.org/
International Transport Forum	www.itf-oecd.org/
Strategic Aviation Special Interest Group	www.sasig.org.uk/

UK Government
We searched www.gov.uk with the string “noise health” restricted to items published after 31/12/2014. We included results of the type “Research and statistics” or “Policy papers and consultations” under the following topics (number of results in brackets):

- Environment > Pollution and environmental quality (21)
- Business and industry > Business and the environment (5)
- Corporate information (11)
- Health and social care ("noise" search only) (0)

We screened all 37 results and included none for full text review.

Civil Aviation Authority
We searched www.caa.co.uk with the search terms ‘health’ and ‘aviation noise health’. In total, 37 results were screened and three included for full text review.

Airports
We searched the websites of the 5 busiest airports in the UK with the search terms "health" then "noise" and then "noise and health" (number of results screened in brackets):

- Heathrow airport www.heathrow.com (75)
- Gatwick airport www.gatwickairport.com (125)
- Luton airport www.london-luton.co.uk (103)
- Manchester airport www.manchesterairport.co.uk (150)
- Stanstead airport www.stanstedairport.com (150)

In total, 603 results were screened and none were included for full text review.

Chartered Institute of Environmental Health
We searched https://www.cieh.org/ with the search term 'aviation noise' and separately 'noise'. In total, 22 results were screened and none were included for full text review.

International Transport Forum
We searched https://www.itf-oecd.org/ with the search terms 'aviation noise' and 'noise and health'. In total, 61 results were screened, and none were included for full text review.

Strategic Aviation Special Interest Group
We searched https://www.sasig.org.uk/ with the search terms 'health', 'aviation noise and health' and 'noise and health'. In total, 15 results were screened and none were included for full text review.

Conference proceedings

In total there were 1309 papers from ICA Aachen and 893 from Internoise Madrid. In order to find all relevant papers we completed a two-stage screening process based on the conference sessions which were organised by topic.

ICA (Aachen, September 2019)
We screened all 182 session titles to decide which were likely to have relevant papers, including sessions which were directly and indirectly relevant. After this process, 11 sessions were selected for title and abstract screening of all papers (total 54).

Internoise (Madrid, June 2019)
We screened all 95 session titles to decide which were likely to have relevant papers, including sessions which were directly and indirectly relevant. After this process, 8 sessions were selected for title and abstract screening of all papers (total 69).

In total, 10 papers of relevance were identified for full text screening from both conferences and of these, six had already been included for full text screening from previous citation tracking. Of the four papers left for screening, all were from the ICA conference. In total, one of these papers was included in our update from the conference searches.

Appendix C Data extraction tables

Appendix table 2 Data extraction table template

Title	Country	Setting/population (e.g. age or social restrictions such as residents, students)	Study design (Longitudinal, case-control, cross-sectional, other)	Sample size (number of individuals)	Adjustment for confounders (Appropriate consideration of obvious potential confounders)	Health conditions summary	Health conditions included (all that are eligible)	Measurement of health conditions (e.g. self-report, individually measured in-study, ecologically measured)	Noise assessment (how measured/modelled)

Noise metrics used (e.g. Lden, LAeq,16h etc; dB levels/bands)	Effect size (Metric and direction of association or effect (odds ratio, risk ratio etc; harmful or protective)	Bias due to exposure assessment	Bias due to confounding	Bias due to selection of participants	Bias due to health outcome assessment	Bias due to not blinded outcome assessment	Total risk of bias	Notes

Appendix table 3   Data extraction (study characteristics)

Paper	Study characteristics	Adjustment for confounders
Basner et al. 2019 [27]	Cross-sectional study (n=80) of residents living both around Philadelphia airport and an area without air-traffic. Examined aviation noise and sleep quality measures through both objective and subjective methods.	Adjusted for various confounders (different models adjusted differently).
Baudin et al. 2019 [31]	Cross-sectional study (n=1300) of residents living near one of seven major European airports in seven countries (Italy, Greece, the Netherlands, Sweden, Germany, the UK, France). Data is from the HYENA and DEBATS studies. Examined aviation noise and saliva cortisol levels.	Adjusted for age, sex, BMI, country, smoking habits, alcohol consumption, physical activity and education level as a proxy for income.
Benz and Schreckenberg 2019 [35]	Panel study with three waves (n=3319). First wave was before and second and third wave following new runway and then night flight ban near Frankfurt airport. Data is from the NORAH study. Examined aviation noise and diagnosis of depression.	Adjusted for age, sex, BMI, migration background, period of residence, hours spent out of home, home ownership, socio-economic status, noise sensitivity, sports, railway noise exposure, road noise exposure, alcohol consumption and tobacco consumption.
Brink et al. 2019 [25]	Cross-sectional study (n=5592) of whole population in Switzerland (ages 19-75). Examined aviation noise and self-reported sleep disturbance.	Adjusted for age, sex, German language and postal mode.
Nassur et al. 2019 [32]	Cross-sectional study (n=92) of residents living near the Paris-Charles de Gaulle and Toulouse-Blagnac airports in France. Examined aviation noise and heart rate during sleep.	Adjusted for age, sex, BMI, physical exercise, smoking and alcohol consumption as well as the presence of cardiovascular or hypertensive problems. Models were also adjusted for time since onset of sleep.
Rocha et al. 2019 [26]	Cross-sectional study (n=268) of residents in households around Hartsfield-Jackson Atlanta international airport (ATL) which a minimum night noise of 35 dB. Examined aviation noise and self-reported sleep disturbance and quality.	Adjusted for age, sex, BMI, hearing problems, noise sensitivity and income.
Rojek 2019 [30]	Cross-sectional study (n=126) which compared residents of Krakow, Poland in areas exposed to high and low aircraft noise. Examined aviation noise and blood pressure, arterial hypertension and indices of asymptomatic organ damage.	Adjusted for age, sex, BMI, education, time spent at home, smoking status, alcohol consumption and antihypertensive treatment.
Smith et al. 2020 [28]	Cross-sectional study (n=34) of adult residents living around Atlanta Hartsfield Jackson international airport (ATL). Examined aviation noise and self-reported sleep disturbance.	Adjusted for age, sex, BMI, and time from sleep onset.
Spilski et al. 2019 [36]	Panel study with three waves (n=1200). First wave was before and second and third wave following new runway and then night flight ban near Frankfurt airport. Data is from the NORAH study. Examined aviation noise and health-related quality of life among children who were second-graders (mean age eight years, four months).	Adjusted for age, sex, socio-economic status, road-traffic and railway noise.

Paper	Study characteristics	Adjustment for confounders
Trieu et al. 2019 [29]	Cross-sectional study (n=755) of residents around Noi Bai airport, Vietnam. Examined aviation noise and cardiovascular disease (blood pressure and heart rate).	
Vienneau et al. 2019 [33]	Meta-analysis of five aircraft studies (accepted study designs were cohorts, case-control and small-area studies). Examined aviation noise and cardio-metabolic diseases (Ischemic Heart Disease and diabetes)	Various adjustments. Studies were only included if basic adjustments for socio-economic status were performed.
Weihofen et al. 2019 [34]	Systematic review and meta-analysis with seven studies were included in the meta-analysis. Examined aviation noise and incidents of stroke for residents of various countries.	Studies which were included were adjusted for various combinations of confounders including age, sex, ethnicity and socioeconomic status.

Appendix table 4 Data extraction (noise exposure and effect)

Paper	Noise assessment	Effect
Basner et al. 2019 [27]	Aviation noise measured through microphones set up near the participant's bed and also outside the participant's bedroom window. Noise metrics used were L_{night} for outside measurements and $L_{AS,max}$ and $L_{Aeq,1min}$ for inside measurements.	There was a significant exposure-response function (ERF) between the sound level of aircraft noise and the probability of awakening. In a random effect adjusted logistic regression model, the coefficient for $L_{AS,max}$ was positive and significant (0.0262, SE 0.0098, p=0.0117). The ERF for percent awakened increased with $L_{AS,max}$: visually, around 3.5% at $L_{AS,max}$ 50dB, 6% at 60 dB, 9% at 70 dB and 12% at 80 dB. Neither systolic nor diastolic morning blood pressure differed between the region with air traffic and the region without. It was also found that retrospective one-month sleep quality index measurements were significant and morning survey on last night's sleep not significant.
Baudin et al. 2019 [31]	Aviation noise modelled for each participant's home. For all countries except the UK, noise levels were provided from the Integrated Noise Model (INM) which is a computer model. In the UK, the Aircraft Noise Contour Model (ANCON v2) was used. Noise metrics used were $L_{Aeq,24hr}$, $L_{Aeq,16 h (06:00\text{-}22:00)}$, L_{den} and L_{night}.	There were statistically significant increases of evening cortisol levels in women with a 10 dB increase in aircraft noise exposure in terms of $L_{Aeq,16 h}$ (exp(β) = 1.08; CI95% = 1.00–1.16), L_{den} (exp(β) = 1.09; CI95% = 1.01–1.18), L_{night} (exp(β) = 1.11; CI95% = 1.02–1.20). Statistically significant association also found in women between a 10 dB increase in terms of L_{night} and the absolute cortisol variation per hour (exp(β) = 0.90; CI95% = 0.80–1.00). Statistically significant decreases in relative variation per hour in women were also shown, with stronger effects with the L_{night} (exp(β) = 0.89; CI95% = 0.83–0.96) than with other noise indicators. The morning cortisol levels were unchanged with all noise exposure indicators. No statistically significant association found between aircraft noise exposure and cortisol levels for men. Annoyance and noise sensitivity found not to modify the results when included as covariates.
Benz and Schreckenberg 2019 [35]	Aviation noise modelled for the most exposed façade of the participant's address. Noise metric used was L_{den}.	In the adjusted analysis the coefficient for L_{den} in t_1 (before the new runway and night flight ban) on prevalence of depression diagnosis in t_2 (after the new runway and night flight ban) was 0 (-0.03 to 0.03, p=0.89). In that adjusted analysis the coefficient for annoyance was -0.20 (-0.34 to -0.05, p<0.01). Structural equation modelling showed no significant direct effect of t_1 aircraft noise exposure on t_2 prevalence of depression but showed significant effects of the indirect path of exposure to annoyance and annoyance in t_1 to depression in t_2. It was shown that annoyance as a mediator from aviation noise to mental health conditions is very important. This relationship may be bi-directional, in that depression may also predict annoyance.
Brink et al. 2019 [25]	Aviation noise measured through one to three receiver points per façade segment and floor. The noise exposure assessment for each façade point comprised yearly averages of the 1-hour- L_{Aeq} and Intermittency Ratio. Based on this, source-specific L_{night} ($L_{Aeq, 23\text{-}07h}$) and IR were calculated and assigned to the dwelling units. Noise metrics used were L_{day} and L_{night}.	There was a statistically significant association between nighttime aviation noise level and the probability of reporting high sleep disturbance. There was an adjusted odds ratio of 1.1270 (p=<0.01) for high sleep disturbance (HSD) per 1 dB increase. Urbanization was an effect modifier, with aviation noise most sleep disturbing in rural areas. There were significant paired differences for %HSD due to aircraft noise between cities and towns/suburbs (-0.537 on the log odds ratio scale, p <2, Tukey-adjusted), and cities and rural areas (-0.914 on the log odds ratio scale, p <0.03). Season and temperature were found not to affect the relationship between aviation noise and HSD.

Paper	Noise assessment	Effect
Nassur et al. 2019 [32]	Aviation noise measured inside and outside of the participant's bedroom continuously for 8 days using a sound level meter on the outside wall of the bedroom and a second on the bedside table. An algorithm was then used to determine aircraft noise in the bedroom, taking account of the transfer between inside/outside as well as filtering out other acoustic events. Noise metrics used were $L_{Aeq,15s}$, $L_{Aeq,aero,15s}$, $L_{A90,15s}$ and $L_{Amax,1s}$.	Positive and significant associations were found between the energy indicators ($L_{Aeq,15s}$ and $L_{A90,15s}$) and the heart rate. A 10 dB increase in $L_{Aeq,15s}$ was associated with an increase of 0.71bpm in heart rate for all noise sources. However, there was no significant relationship between aircraft noise alone ($L_{Aeq,aero,15s}$) and heart rate in the multivariate models. A further model assessed aircraft noise exposure characterized by $L_{Amax,1s}$ and differences between heart rates recorded during or 15/30 seconds after the aircraft noise events. No significant relationships were found. In contrast, a positive association was found between $L_{Amax,1s}$ and the heart rate amplitude during an aircraft noise event. Heart rate amplitude was calculated as the maximum and minimum heart rate during an acoustic event, in beats per minute.
Rocha et al. 2019 [26]	Aviation noise modelled using the Integrated Noise Model (INM) to give noise levels for each aircraft over 84 nights. Noise metric used was L_{night}.	The adjusted OR (95%CI) for L_{night} (per dB) with sleep disturbance was 1.15 (1.10-1.23), overall sleep quality 1.04 (1.00-1.08), trouble falling asleep 1.06 (1.02-1.10), trouble sleeping at night 1.04 (1.00-1.08) use of sleep medication 0.98 (0.94-1.03) and trouble staying awake 1.05 (1.00-1.11). Noise sensitivity was also found to be highly associated with all sleep disturbance outcomes. L_{night} was also associated with a greater odds of using certain coping aids against noise when trying to sleep; alcohol (1.10, 1.00-1.21), TV (1.05, 1.01-1.10), music (1.07, 1.01-1.13) and closing windows (1.05, 1.01-1.09). After adjustments, L_{night} was not significantly associated with self-reported general health.
Rojek 2019 [30]	Aviation noise measuring using two groups. The groups included those who were more and less exposed to aircraft noise and lived in different areas of Krakow. One group were exposed to high aircraft noise (more than 60 dB L_{den}) and the other were exposed to low aircraft noise (less than 55 dB L_{den}). Noise metric used was L_{den}.	Long-term aircraft noise exposure was related to higher office and nighttime diastolic blood pressure (DBP) and more advanced arterial stiffness and unfavorable left ventricle diastolic function changes. Exposure to aircraft noise did not increase the prevalence of arterial hypertension (50%, both groups) but was associated with higher office (88.3 vs. 79.8 mmHg, p<0.001) and night-time DBP (66.6 vs. 63.6 mmHg, P<0.01). Participants exposed to higher aircraft noise level had a higher carotid–femoral pulse wave velocity (PWV) (10.3 vs. 9.4 m/s, p<0.01) and lower early mitral annulus velocity (e0) (8.4 vs. 9.2 cm/s, P=0.047). Accelerated arterial stiffening was also observed to a degree depending on noise annoyance.
Smith et al. 2020 [28]	Aviation noise measured using recording equipment shipped to participants. Equipment recorded raw audio data so that aircraft noise could be separated by trained research personnel who manually screened the audio recordings. Noise metrics used for indoor noise were $L_{AS,max,ANE}$, $L_{Aeq,sleep}$ and $L_{AS,max,sleep}$. L_{night} was used for outdoor noise.	Self-reported awakenings increased alongside the highest maximum aircraft noise level occurring during the sleep period. Adjusted model for random effect logistic regression was 0.0254 (SE 0.0126). This effect was of only borderline statistical significance (p=0.057), likely due to the low sample size of this pilot study. Comparisons of questionnaires and $L_{Aeq,sleep}$ and $L_{AS,max}$ were also made. No statistically significant effects of $L_{Aeq,sleep}$ were found. With increasing $L_{AS,max,sleep}$ there were significant increases in tiredness (β = 0.118, p=0.005) and self-reported awakenings (β=0.051, p=0.001). There was a significant effect of sleeping with open windows on awakenings in the $L_{AS,max,sleep}$ model.

Paper	Noise assessment	Effect
Spilski et al. 2019 [36]	Aviation noise modelled using spatial and urban planning data. Noise metric used was $L_{Aeq,16\ h\ (06:00-22:00)}$.	Authors hypothesized that increased aircraft noise exposure leads to increased stress responses in children and subsequently affects their well-being and health, mediated through annoyance at aircraft noise. They also tested urbanization as an effect modifier. A 10 dB $L_{Aeq,16\ h\ (06:00-22:00)}$ increase in aircraft noise was associated with an increase of 0.81 scale points for physical well-being which led to increases in headaches and stomach aches. The inclusion of annoyance as a mediator led to a non-significant direct effect (b=-0.003, SE =0.004, p>0.05, 95% CI: -0.011,0.006), indicating a mediation effect. Results for psychological well-being and aviation noise were not consistent and urbanization was found not to be a significant modifier.
Trieu et al. 2019 [29]	Aviation noise measured using noise monitors in each resident's house during the first phase. For the second phase, noise exposure was modelled using noise contour maps and operation data. Noise metrics used were L_{den} and $L_{Aeq,night}$.	The results suggested that although there was a high rate of high blood pressure around the airport there was no significant relationship with noise exposure levels (OR 1.024, 95% CI 0.969-1.082). However, a significantly higher rate of insomnia was found at survey phase 2 when the number of night flights had increased.
Vienneau et al. 2019 [33]	Various noise assessment techniques used but exposure had to be modelled or measured to be included in the meta-analysis. Noise metric used in the meta-analysis was L_{den}.	Authors concluded that the inclusion of the most recent studies into WHO findings is important. There were indications of associations with aircraft noise and IHD incidence but in the sample the current studies were heterogeneous indicating there was variation on study outcomes (relative risk [RR] 1.03, 95%CI, 0.98- 1.09 per 10 dB L_{den}). Risk of bias was also high. For diabetes, there was a higher but non-significant RR per 10 dB L_{den} rate of 1.20 (0.88-1.63) and risk of bias was low.
Weihofen et al. 2019 [34]	Various noise assessment techniques used in different papers included in the meta-analysis. Various noise metrics used in the different papers but L_{den} used in the meta-analysis.	The meta-analysis found a relative stroke risk of 1.013 (0.998-1.028) per 10 dB. Although the overall finding just fails to reach statistical significance the authors conclude that as the result is so close, an effect seems likely.

Appendix table 5 Risk of bias

Paper	Bias due to exposure assessment	Bias due to confounding	Bias due to selection of participants	Bias due to health outcome assessment	Bias due to not blinded outcome assessment	Total risk of bias
Basner et al. 2019 [27]	Low	Low	Unclear. Very low response to recruitment	Low	Low	Low
Baudin et al. 2019 [31]	Low	Low	Unclear. Participation rates from 30% to 78%	Low	Low	Low
Benz and Screnckenberg 2019 [35]	Low	Low	Low	Low	Low	Low
Brink et al. 2019 [25]	Low	Low	Moderate. 31% response	Low	N/A	Low
Nassur et al. 2019 [32]	Low	Low	Moderate. Self-selection into DEBATS follow-up	Low	Low	Low
Rocha et al. 2019 [26]	Low	Low	High. 8.5% response	Low	N/A	Moderate
Rojek 2019 [30]	Low	Low	Moderate. 42%/48% response	Low	Low	Low
Smith et al. 2020 [28]	Low	Low	Moderate. 10% response but attenuation of bias by selection into field study	Low	Low	Low
Spilski et al. 2019 [36]	Low	Low	Low	Low	Low	Low
Trieu et al. 2019 [29]	Low	Moderate. Not adjusted for gender	Unclear	High for blood pressure, low for insomnia	Low	High
Vienneau et al. 2019 [33]	N/A (meta-analysis)	N/A (meta-analysis)	N/A (meta-analysis)	N/A (meta-analysis)	N/A (meta-analysis)	Low. Most studies with low ROB and high quality)
Weihofen et al. 2019 [34]	N/A (meta-analysis)	N/A (meta-analysis)	N/A (meta-analysis)	N/A (meta-analysis)	N/A (meta-analysis)	Low. Most studies high quality (cohort/ case–control)

Appendix D Reason for exclusion at full text screening

Appendix table 6 Reasons for exclusion of papers

Paper	Source	Reason for exclusion
L. M. Argys, S. L. Averett and M. Yang, "Residential noise exposure and health: evidence from aviation noise and birth outcomes," *IZA DP,* vol. No. 12605, 2019.	Citation tracking	Aircraft noise measured by distance to airport
S. Bartels, J. Quehl and D. Aeschbach, "Effects of nocturnal aircraft noise on objective and subjective sleep quality in primary school children," in *Proceedings of the 23rd International Congress on Acoustics*, Aachen, Germany, 2019.	Citation tracking	Aircraft noise measured by number of noise events
T. W. Collins, S. E. Grineski and S. Nadybal, "Social disparities in exposure to noise at public schools in the contiguous United States," *Environmental Research,* vol. 175, pp. 257-265, 2019.	Database search	Doesn't look at health
T. W. Collins, S. Nadybal and S. E. Grineski, "Sonic injustice: disparate residential exposures to transport noise from road and aviation sources in the continental United States," *Journal of Transport Geography,* vol. 82, p. 102604, 2020.	Database search	Doesn't look at health
E. Generaal, E. J. Timmermans, J. E. C. Dekkers, J. H. Smit and B. W. J. H. Penninx, "Not urbanization level but socioeconomic, physical and social neighbourhood characteristics are associated with presence and severity of depressive and anxiety," *Psychological Medicine,* vol. 49, no. 1, pp. 149-161, 2019.	Database search	Aviation noise not separated from other noise sources
J. I. Halonen, "Transportation noise and cardiovascular health: role of multiple noise sources," *Occupational and Environmental Medicine,* vol. 76, pp. 199-200, 2019.	Database search	Paper is a commentary
H. Héritier, D. Vienneau, M. Foraster, I. C. Eze, E. Schaffner, K. de Hoogh, L. Thiesse, F. Rudzik, M. Habermacher, M. Köpfli, R. Pieren, M. Brink, C. Cajochen, J. M. Wunderli, N. Probst-Hensch and M. Röösli, "A systematic analysis of mutual effects of transportation noise and air pollution exposure on myocardial infarction mortality: a nationwide cohort study in Switzerland," *European Heart Journal ,* vol. 40, no. 7, pp. 598-603, 2019.	Database search	Already included in a previous review
S. Kleyn, I. May and D. Kiryanov, "Hygienic analysis of potential risks of health harm in the implementation of airport complexes activity," *Hygiene and Sanitation,* vol. 98, pp. 268-275, 2019.	Database search	Paper not accessible
D. Leger and C. Guilleminault, "Environmental open-source data sets and sleep-wake rhythms of populations: an overview," *Sleep Medicine,* vol. 11, no. 69, pp. 88-97, 2020.	Database search	No data

Paper	Source	Reason for exclusion
G. B. Marks, A. L. Hansell and F. H. Johnston, "The environment is a first order issue for lung health," *International Journal of Tuberculosis and Lung Disease,* vol. 23, no. 12, pp. 1239-1240, 2019.	Database search	Editorial
T. Munzel, S. Steven, O. Hahad and A. Daiber, "The sixth sense is involved in noise-induced stress responses and vascular inflammation: evidence for heightened amygdalar activity in response to transport noise in man," *European Heart Journal ,* vol. 41, no. 6, pp. 783-785, 2020.	Database search	Editorial
A. M. Nassur, D. Léger, M. Lefèvre, M. Elbaz, F. Mietlicki, P. Nguyen, C. Ribeiro, M. Sineau, B. Laumon and A. S. Evrard, "The impact of aircraft noise exposure on objective parameters of sleep quality: results of the DEBATS study in France," *Sleep Medicine ,* vol. 54, pp. 70-77, 2019.	Database search	Already included in a previous review
A. M. Nassur, M. Lefevre, B. Laumon, D. Leger and A. S. Evrard, "Aircraft noise exposure and subjective sleep quality: the results of the DEBATS study in France," *Behavioral Sleep Medicine,* vol. 17, no. 4, pp. 502-513, 2019.	Database search	Already included in a previous review
M. Oh, K. Shin, K. Kim and J. Shin, "Influence of noise exposure on cardiocerebrovascular disease in Korea," *Science of the Total Environment ,* vol. 651, no. 2, pp. 1867-1876, 2019.	Database search	Aviation noise not separated from other noise sources
M. T. Osborne, A. Radfar, M. Hassan, S. Abohashem, B. Oberfeld, T. Patrich, B. Tung, Y. Wang, A. Ishai, J. A. Scott, L. M. Shin, Z. A. Fayad, K. C. Koenen, S. Rajagopalan, R. K. Pitman and A. Tawakol, "A neurobiological mechanism linking transportation noise to cardiovascular disease in humans," *European Heart Journal ,* vol. 41, no. 6, pp. 772-782, 2020.	Database search	Aviation noise not separated from other noise sources
D. Pillay and B. L. Vieira, "Noise, screaming and shouting: classroom acoustics and teachers' perceptions of their voice in a developing coun," *South African Journal of Childhood Education,* vol. 10, no. 1, p. 681, 2020.	Database search	No health data
A. Pyko, N. Andersson, C. Eriksson, U. de Faire, T. Lind, N. Mitkovskaya, M. Ögren, Ö. C. G, P. N. L, D. Rizzuto, W. A. K and G. Pershagen, "Long-term transportation noise exposure and incidence of ischaemic heart disease and stroke: a cohort study," *Occupational and Environmental Medicine,* vol. 76, no. 4, pp. 201-207, 2019.	Database search	Already included in a previous review
C. Ribeiro, F. Mietlicki and P. Jamard, "Health impact of noise in Greater Paris Metropolis: assessment of health life years lost," in *Madrid Internoise 2019: noise control for a better environment*, Madrid, Spain, 2019.	Citation tracking	Aircraft noise measured by number of noise events
F. Z. Sakhvidi, M. J. Z. Sakhvidi, A. H. Mehrparvar and A. M. Dzhambov, "Environmental noise exposure and neurodevelopmental and mental health problems in children: a systematic review," *Current Environmental Health Reports ,* vol. 5, pp. 365-374, 2018.	Database search	Systematic review of papers pre-2019
Saucy, A. et al., "Aircraft noise exposure assessment for a case-crossover study in Switzerland" in *Madrid Internoise 2019: Noise control for a better environment*, Madrid, Spain, 2019.	Citation tracking	No health outcome

Paper	Source	Reason for exclusion
M. Schubert, J. Hegewald, A. Freiberg, S. K. R, F. Augustin, R.-H. S. G, H. Zeeb and A. Seidler, "Behavioral and emotional disorders and transportation noise among children and adolescents: a systematic review and meta-analysis," *International Journal of Environmental Research and Public Health,* vol. 16, no. 18, p. 3336, 2019.	Database search	Aviation noise not separated from other noise sources
A. Seidler, J. Hegewald, A. L. Seidler, M. Schubert and H. Zeeb, "Is the whole more than the sum of its parts? Health effects of different types of traffic noise combined," *International Journal of Environmental Research and Public Health ,* vol. 16, no. 9, p. 1665, 2019.	Database search	Aviation noise not separated from other noise sources
M. Sorensen and G. Pershagen, "Transportation noise linked to cardiovascular disease independent from air pollution," *European Heart Journal ,* vol. 40, no. 7, pp. 604-606, 2019.	Database search	Editorial
K. Wolf, U. Kraus, M. Dzolan, G. Bolte, T. Lakes, T. Schikowski, H. K. Greiser, O. Kuß, W. Ahrens, F. Bamberg, H. Becher, K. Berger, H. Brenner, S. Castell, A. Damms-Machado, B. Fischer, F. C. W, S. Gastell, G. K, B. Holleczek, L. Jaeschke, R. Kaaks, T. Keil, Y. Kemmling, L. Krist, N. Legath, M. Leitzmann, W. Lieb, M. Loeffler, C. Meinke-Franze, K. B. Michels, R. Mikolajczyk, S. Moebus, U. Mueller, N. Obi, T. Pischon, W. Rathmann, S. Schipf, B. Schmidt, M. Schulze, I. Thiele, S. Thierry, S. Waniek, C. Wigmann, K. Wirkner, J. Zschocke, A. Peters and A. Schneider, "Nightime transportation noise annoyance in Germany: personal and regional differences in the German National Cohort Study," *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz,* vol. 63, no. 3, pp. 332-343, 2020.	Database search	Health outcome is annoyance
Y. Yu, K. Paul, O. A. Arah, E. R. Mayeda, J. Wu, E. Lee, I. F. Shih, J. Su, M. Jerrett, M. Haan and B. Ritz, "Air pollution, noise exposure, and metabolic syndrome: a cohort study in elderly Mexican-Americans in Sacramento area," *Environment International,* vol. 134, p. 105269, 2020.	Database search	Paper does not include aviation noise

Appendix E GRADE for present REA outcomes only

Outcomes for which there was evidence from the present REA only

Self-reported diagnosis of sleep disorder

Appendix table 7 GRADE assessment for the effect of aviation noise on self-reported sleep disorder

Domain	**Criterion**	**Assessment**	**Quality & downgrading**
Starting level	Intervention/longitudinal	One cross-sectional study	Low
Study design	Majority of studies with low ROB	Moderate ROB	Downgrade
Inconsistency	Conflicting results, high I^2	NA – single study	No
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	CI fairly narrow	No
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – no effect**

Self-reported sleep coping behaviours

Appendix table 8 GRADE assessment for the effect of aviation noise on self-reported sleep coping behaviours

Domain	**Criterion**	**Assessment**	**Quality & downgrading**
Starting level	Intervention/longitudinal	One cross-sectional study	Low
Study design	Majority of studies with low ROB	Moderate ROB	Downgrade
Inconsistency	Conflicting results, high I^2	Mixed results within study	Downgrade
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	Unable to summarize	No
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – harmful effect**

Self-reported awakenings

Appendix table 9 GRADE assessment for the effect of aviation noise on self-reported awakenings

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	One cross-sectional study	Low
Study design	Majority of studies with low ROB	Low ROB	No
Inconsistency	Conflicting results, high I^2	NA – single study	No
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	No CI but p value (0.057) consistent with fairly narrow CI	No
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Low quality – harmful effect**

Self-reported sleep quality

Appendix table 10 GRADE assessment for the effect of aviation noise on self-reported sleep quality

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	Three cross-sectional studies	Low
Study design	Majority of studies with low ROB	Majority of studies with low ROB	No
Inconsistency	Conflicting results, high I^2	Some inconsistency	Downgrade
Indirectness	Direct comparison, same PECCO	Multiple different items	Downgrade
Precision	CI narrow	CI fairly narrow	No
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – harmful effect**

Arterial stiffness

Appendix table 11 GRADE assessment for the effect of aviation noise on arterial stiffness

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	One cross-sectional study	Low
Study design	Majority of studies with low ROB	Low ROB	No
Inconsistency	Conflicting results, high I^2	NA – single study	No
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	No CI but p<0.001 consistent with narrow CI	No
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Low quality – harmful effect**

Asymptomatic heart damage

Appendix table 12 GRADE assessment for the effect of aviation noise on asymptomatic heart damage

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	One cross-sectional study	Low
Study design	Majority of studies with low ROB	Low ROB	No
Inconsistency	Conflicting results, high I^2	Mixed results within study	Downgrade
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	Unable to assess	No
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – harmful effect**

Blood pressure in adults

Appendix table 13 GRADE assessment for the effect of aviation noise on blood pressure in adults

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	Three cross-sectional studies	Low
Study design	Majority of studies with low ROB	Low ROB	No
Inconsistency	Conflicting results, high I^2	Conflicting results across studies	Downgrade
Indirectness	Direct comparison, same PECCO	Some differences in exposure assessment	Downgrade
Precision	CI narrow	Unable to summarize	No
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – no effect**

Heart rate

Appendix table 14 GRADE assessment for the effect of aviation noise on heart rate

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	Two cross-sectional studies	Low
Study design	Majority of studies with low ROB	Low ROB	No
Inconsistency	Conflicting results, high I^2	Conflicting results within and across studies	Downgrade
Indirectness	Direct comparison, same PECCO	Some differences in population	Downgrade
Precision	CI narrow	Unable to summarize	No
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – harmful effect**

Cortisol levels

Appendix table 15 GRADE assessment for the effect of aviation noise on cortisol levels

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	One cross-sectional study	Low
Study design	Majority of studies with low ROB	Low ROB	No
Inconsistency	Conflicting results, high I^2	Mixed results within study	Downgrade
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	Unable to summarize	No
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – harmful effect**

Self-reported diagnosis of arrhythmia

Appendix table 16 GRADE assessment for the effect of aviation noise on self-reported diagnosis of arrhythmia

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	One cross-sectional study	Low
Study design	Majority of studies with low ROB	Moderate ROB	Downgrade
Inconsistency	Conflicting results, high I^2	NA – single study	No
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	CI fairly wide	Downgrade
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – no effect**

Self-reported diagnosis of diabetes

Appendix table 17 GRADE assessment for the effect of aviation noise on self-reported diagnosis of diabetes

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	One cross-sectional study	Low
Study design	Majority of studies with low ROB	Moderate ROB	Downgrade
Inconsistency	Conflicting results, high I^2	NA – single study	No
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	CI modest	No
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – no effect**

Self-reported diagnosis of heart disease

Appendix table 18 GRADE assessment for the effect of aviation noise on self-reported diagnosis of heart disease

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	One cross-sectional study	Low
Study design	Majority of studies with low ROB	Moderate ROB	Downgrade
Inconsistency	Conflicting results, high I^2	NA – single study	No
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	CI wide	Downgrade
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – no effect**

Self-reported diagnosis of hypertension

Appendix table 19 GRADE assessment for the effect of aviation noise on self-reported diagnosis of hypertension

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	One cross-sectional study	Low
Study design	Majority of studies with low ROB	Moderate ROB	Downgrade
Inconsistency	Conflicting results, high I^2	NA – single study	No

Domain	Criterion	Assessment	Quality & downgrading
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	CI includes 1	No
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – no effect**

Wellbeing of children

Appendix table 20 GRADE assessment for the effect of aviation noise on wellbeing of children

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	One panel study	Low
Study design	Majority of studies with low ROB	Low ROB	No
Inconsistency	Conflicting results, high I^2	NA – single study	No
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	CIs fairly wide	Downgrade
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – no effect**

Note: the design alone would give a starting point of Moderate quality, but as there is only one study we have downgraded this starting point to Low quality, consistent with the WHO review on cardiovascular and metabolic disorders [12].

Depression prevalence

Appendix table 21 GRADE assessment for the effect of aviation noise on prevalence of depression

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	One panel study	Low
Study design	Majority of studies with low ROB	Low ROB	No
Inconsistency	Conflicting results, high I^2	NA – single study	No
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	Narrow CI	No
Publication bias	Funnel plot indicates	Unable to assess	No

Domain	Criterion	Assessment	Quality & downgrading
Overall judgement			**Low quality – harmful effect through annoyance**

Note: the design alone would give a starting point of Moderate quality, but as there is only one study we have downgraded this starting point to Low quality, consistent with the WHO review on cardiovascular and metabolic disorders [12].

Self-reported general health

Appendix table 22 GRADE assessment for the effect of aviation noise on self-reported general health

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	One cross-sectional study	Low
Study design	Majority of studies with low ROB	Moderate ROB	Downgrade
Inconsistency	Conflicting results, high I^2	NA – single study	No
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	CI fairly narrow	No
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – no effect**

General physical health of children

Appendix table 23 GRADE assessment for the effect of aviation noise on general health of children

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	One panel study	Low
Study design	Majority of studies with low ROB	Low ROB	No
Inconsistency	Conflicting results, high I^2	NA – single study	No
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	No CIs but p>0.05 consistent with wide CIs	Downgrade
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – no effect**

Note: the design alone would give a starting point of Moderate quality, but as there is only one study we have downgraded this starting point to Low quality, consistent with the WHO review on cardiovascular and metabolic disorders [12].

Self-reported diagnosis of chronic headaches/migraine

Appendix table 24 GRADE assessment for the effect of aviation noise on self-reported diagnosis of chronic headaches/migraine

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	One cross-sectional study	Low
Study design	Majority of studies with low ROB	Moderate ROB	Downgrade
Inconsistency	Conflicting results, high I^2	NA – single study	No
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	CI fairly wide	Downgrade
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – no effect**

Self-reported diagnosis of stomach ulcer

Appendix table 25 GRADE assessment for the effect of aviation noise on self-reported diagnosis of stomach ulcer

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	One cross-sectional study	Low
Study design	Majority of studies with low ROB	Moderate ROB	Downgrade
Inconsistency	Conflicting results, high I^2	NA – single study	No
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	CI fairly wide	Downgrade
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – no effect**

Children's medication intake reported by the parent

Appendix table 26 GRADE assessment for the effect of aviation noise on children's medication intake reported by the parent

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	One panel study	Low

Domain	Criterion	Assessment	Quality & downgrading
Study design	Majority of studies with low ROB	Low ROB	No
Inconsistency	Conflicting results, high I^2	NA – single study	No
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	No CIs but p>0.05 consistent with moderate CIs	Downgrade
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – no effect**

Note: the design alone would give a starting point of Moderate quality, but as there is only one study we have downgraded this starting point to Low quality, consistent with the WHO review on cardiovascular and metabolic disorders [12].

Children's physical diseases reported by the parent

Appendix table 27 GRADE assessment for the effect of aviation noise on children's physical diseases reported by the parent

Domain	Criterion	Assessment	Quality & downgrading
Starting level	Intervention/longitudinal	One panel study	Low
Study design	Majority of studies with low ROB	Low ROB	No
Inconsistency	Conflicting results, high I^2	NA – single study	No
Indirectness	Direct comparison, same PECCO	No indirect comparison	No
Precision	CI narrow	No CIs but p>0.05 consistent with wide CIs	Downgrade
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Very low quality – no effect**

Note: the design alone would give a starting point of Moderate quality, but as there is only one study we have downgraded this starting point to Low quality, consistent with the WHO review on cardiovascular and metabolic disorders [12].

Appendix F GRADE for WHO/Defra and present outcomes combined

Outcomes for which there was evidence from the WHO or Defra reviews and from the present REA

Self-reported sleep disturbance in adults where noise was specified in the survey instrument GRADE assessment
For self-reported sleep disturbance in adults where noise was specified in the survey instrument, the WHO review concluded there was moderate evidence of a harmful effect of aviation noise. The Defra-RIVM review found 15 further studies on self-reported sleep disturbance; the authors did not report whether or not noise was specified in the survey instrument. The authors described the results as “not consistent, primarily due to methodological differences between the studies, nevertheless pointing in the same direction”, which we consider to be consistent enough with the finding of the WHO review. The present review found two further papers reporting on this outcome [26] [25], both of which were cross-sectional and one of which had moderate risk of bias [26]. Both papers found a harmful effect. We conclude that the quality of evidence remains moderate for a harmful effect of aviation noise on self-reported sleep disturbance in adults where noise was specified in the survey.

Appendix table 28 GRADE assessment for self-reported sleep disturbance in adults where noise was specified in the survey instrument

Existing evidence from WHO/Defra reviews		WHO review (6 studies) Defra-RIVM review (5 studies)	Moderate – harmful effect No GRADE conducted – Harmful effect
Additional evidence			
Domain	*Criterion*	*Assessment*	*Quality & downgrading*
Starting level	Intervention/longitudinal/meta-analysis	New evidence cross-sectional	Low
Study design	Majority of studies with low ROB	1 of 2 had low ROB	No
Inconsistency	Conflicting results, high I^2	Consistent results	No
Indirectness	Direct comparison, same PECCO	Did not make indirect comparison	No
Precision	CI narrow	Unable to summarize	No
Publication bias	Funnel plot indicates	Unable to assess	No

Overall judgement			**Moderate – harmful effect**

Physiologically measured awakenings GRADE assessment

For cortical awakenings measured by polysomnography, the WHO review concluded there was moderate quality evidence of a harmful effect. Polysomnography involves multiple monitors attached to the body to measure brain, eye, muscle and other signals. It is the state of the art for objective measures of sleep but is expensive and logistically difficult to implement and relatively invasive. The study by Basner et al. (2019) [27] involved using a less invasive single monitor of heart activity and movement that participants could apply themselves. Since the authors report that the agreement between this method and polysomnography was near perfect, we feel it is appropriate to consider this evidence together as "physiologically measured awakenings". The study by Basner et al. (2019) was a small cross-sectional study that on its own could only offer low quality evidence. Nonetheless, we conclude that given the strong result consistent with the finding of the WHO review it is appropriate to maintain the finding of moderate quality evidence of a harmful effect of aviation noise on physiologically measured awakenings.

Appendix table 29 GRADE assessment for physiologically measured awakenings

Existing evidence from WHO/Defra reviews		WHO review (1 study)	Moderate quality – harmful effect
Additional evidence			
Domain	*Criterion*	*Assessment*	*Quality & downgrading*
Starting level	Intervention/longitudinal/meta-analysis	1 cross-sectional study	Low
Study design	Majority of studies with low ROB	Low ROB	No
Inconsistency	Conflicting results, high I^2	Consistent results	No
Indirectness	Direct comparison, same PECCO	Difference in exposure assessment (as discussed)	No
Precision	CI narrow	No CI but low p value (0.012) consistent with narrow CI	No
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Moderate quality – harmful effect**

Incidence of IHD GRADE assessment

For incidence of IHD, the WHO review concluded there was very low quality evidence of a non-significant harmful effect. Its estimate was based on two ecological studies. The Defra-RIVM review concluded there was a small harmful effect but did not assess quality of evidence. The meta-analysis by Vienneau et al. (2019) [33] also concluded there was evidence of a non-significant harmful effect. Given the increased size of the evidence base and consistency of the results, on the one hand, and the high risk of bias in contributing studies on the other hand, we conclude that there is low quality evidence of a small harmful effect of aviation noise on the incidence of IHD.

Appendix table 30 GRADE assessment for incidence of IHD

Existing evidence from WHO/Defra reviews		WHO review (2 studies) Defra-RIVM review (4 studies)	Very low quality – harmful effect No GRADE conducted – harmful effect
Additional evidence			
Domain	*Criterion*	*Assessment*	*Quality & downgrading*
Starting level	Intervention/longitudinal/meta-analysis	Meta-analysis (of 5 studies)	High
Study design	Majority of studies with low ROB	Included studies had high risk of bias	Yes
Inconsistency	Conflicting results, high I^2	Inconsistency among large studies in meta-analysis	Yes
Indirectness	Direct comparison, same PECCO	Did not make indirect comparison	No
Precision	CI narrow	CI fairly narrow	No
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Low quality – harmful effect**

Incidence of diabetes GRADE assessment

For incidence of diabetes, the WHO review concluded there was low quality evidence of no effect of aviation noise. The Defra-RIVM review did not assess the quality of evidence but reported there was inconsistent evidence between high quality studies, with two cohort studies respectively indicating a harmful effect and no effect. Vienneau et al. (2019) [33] conducted a meta-analysis that included the studies from the WHO and Defra-RIVM reviews and concluded there was evidence of a harmful effect (a fairly large effect too, with a risk ratio of 1.20, 95% CI 0.88 to 1.63) but that this was not statistically significant. We made our GRADE assessment on the meta-analysis, considering that this was the most thorough treatment of the evidence available. As the contributing studies had high-quality designs (being all cohorts or case-control studies) the evidence started out at high quality, but was downgraded for inconsistency and lack of precision. We conclude that there is low quality evidence of a harmful effect of aviation noise on the incidence of diabetes.

Appendix table 31 GRADE assessment for incidence of diabetes

Existing evidence from WHO/Defra reviews		WHO review (1 study) Defra-RIVM review (2 studies)	Low quality – no effect No GRADE conducted – some evidence of harmful effect
Additional evidence			
Domain	*Criterion*	*Assessment*	*Quality & downgrading*
Starting level	Intervention/longitudinal/meta-analysis	Meta-analysis of 3 cohorts/case-control studies	High
Study design	Majority of studies with low ROB	Most studies had low ROB	No
Inconsistency	Conflicting results, high I^2	Highly conflicting results	Downgrade
Indirectness	Direct comparison, same PECCO	Did not make indirect comparison	No
Precision	CI narrow	CI wide	Downgrade
Publication bias	Funnel plot indicates	Unable to assess	No
Overall judgement			**Low quality – harmful effect**

Incidence of stroke GRADE assessment

For incidence of stroke, the WHO review concluded there was very low quality evidence of a non-significant effect of aviation noise. Weihofen et al. (2019) [34] conducted a meta-analysis that found a small (1.3%), marginally significant increased risk of stroke per 10 dB increase in aircraft noise exposure. We made our GRADE assessment on the meta-analysis, considering that this was the most thorough treatment of the evidence available. We considered meta-analysis to offer a high starting quality of evidence. The methodological quality of the included studies was low, for which we downgraded the quality of evidence. The confidence interval contained 1 (0.998 to 1.028) but counteracting this we note the authors' comments about the likelihood of confounding working toward underestimation of the association and we did not downgrade further. We conclude that there is moderate quality evidence of a small harmful effect of aviation noise on the incidence of stroke.

Appendix table 32 GRADE assessment for incidence of stroke

Existing evidence from WHO/Defra reviews		WHO review (2 studies)	Very low quality – harmful effect
Additional evidence			
Domain	*Criterion*	*Assessment*	*Quality & downgrading*
Starting level	Intervention/longitudinal/meta-analysis	Meta-analysis (of 7 studies)	High
Study design	Majority of studies with low ROB	Majority of studies with inadequate quality	Downgrade
Inconsistency	Conflicting results, high I^2	Consistent results	No
Indirectness	Direct comparison, same PECCO	Minor differences in health outcome assessment	No
Precision	CI narrow	CI modest with confounding toward underestimation	No
Publication bias	Funnel plot indicates	Assessed as low risk	No
Overall judgement			**Moderate quality – harmful effect**

Appendix G GRADE for WHO and Defra review findings combined

Outcomes for which there was evidence from both the WHO and Defra reviews

Unlike in the preceding sections, for these outcomes we did not conduct a formal GRADE process. This is because we did not have the original papers that went into the reviews. Here we narratively summarise the quality of evidence combining the findings of the WHO and Defra reviews.

Reading comprehension GRADE assessment
The WHO review included 14 studies and concluded that there was moderate quality evidence of a harmful effect of aviation noise on reading and oral comprehension. The Defra-Arup review included four studies and concluded there was very low quality evidence of a harmful effect. The authors of the latter wrote that they had made their assessment based on a smaller number of studies some of which had had methodological weaknesses leading to downgrading, and recommended that the finding of the WHO review stand. We consider therefore that the WHO finding stands and that there is moderate quality evidence of a harmful effect of aviation noise on reading comprehension.

Stroke mortality GRADE assessment
For stroke mortality, the WHO review included three studies and concluded that there was moderate quality evidence of no effect of aviation noise. The Defra-RIVM review included three studies and concluded there was a non-significant harmful effect but did not rate quality of the evidence. As the WHO review evidence was based on longitudinal studies, and the suggestion of effect in the Defra-RIVM review was of a small and non-significant effect, we consider the finding of the WHO review to stand and conclude there is moderate quality evidence of no effect on stroke mortality.

Incidence of hypertension GRADE assessment
The WHO review included one study and concluded that there was low quality evidence supporting an association between aviation noise and incidence of hypertension. The Defra-RIVM review added evidence from two cohort studies showing a harmful effect of aviation noise and one case-control study showing no effect. We conclude that given the finding of an effect in those two cohort studies, the evidence may point toward a harmful effect and that given the inconsistency, the quality of the evidence remains low.

Interview measures of depression and anxiety GRADE assessment
The WHO review included one study and concluded there was very low quality evidence of a harmful effect of aviation noise on interview measures of depression and anxiety. The Defra-Arup review included two studies and concluded that this should be upgraded to low quality evidence in light of new data from cohort studies. There is no new evidence in this update so the conclusion of low quality evidence of a harmful effect of aviation noise on interview measures of depression and anxiety stands.

Self-reported QOL or health GRADE assessment
The WHO review included seven studies and the Defra-Arup review included four studies. Both reviews concluded there was very low quality evidence of no effect of aviation noise on self-reported quality of life or health. There is no new evidence on this outcome so that conclusion stands.

Appendix H GRADE for WHO or Defra reviews alone

Outcomes for which there was evidence from only the WHO or Defra reviews

This table shows the GRADE assessments for the quality of evidence concluded in the WHO and Defra reviews for outcomes where no new evidence was available.

Appendix table 33 Summary of the quality of evidence for birth and reproductive health outcomes from the WHO and Defra reviews where no new evidence was available

Outcome	Quality of evidence – Direction of effect	Source of GRADE assessment
Congenital malformations	Very low quality – Not stated in GRADE but harmful effects reported	WHO review
Low birth weight	Very low quality – Not stated in GRADE but harmful effects reported	WHO review
Preterm birth	Very low quality – Not stated in GRADE but harmful effects reported	WHO review

Appendix table 34 Summary of the quality of evidence for cognition outcomes from the WHO and Defra reviews where no new evidence was available

Outcome	Quality of evidence – Direction of effect	Source of GRADE assessment
Assessments of student distraction	Very low quality – Harmful effect	Defra-Arup review
Attention	Low quality – No effect	WHO review
Executive function deficit (working memory capacity)	Very low quality – No effect	WHO review
Impairment assessed through SATs	Moderate quality – Harmful effect	WHO review
Short-term and long-term (episodic) memory	Moderate quality – Harmful effect	WHO review

Appendix table 35 Summary of the quality of evidence for sleep outcomes from the WHO and Defra reviews where no new evidence was available

Outcome	Quality of evidence – Direction of effect	Source of GRADE assessment
Self-reported sleep disturbance in adults (source not specified)	Very low quality – Harmful effect	WHO review

Appendix table 36 Summary of the quality of evidence for cardiovascular and metabolic outcomes from the WHO and Defra reviews where no new evidence was available

Outcome	Quality of evidence – Direction of effect	Source of GRADE assessment
Blood pressure in children	Very low quality – No effect	WHO review
Diabetes prevalence	Very low quality – No effect	WHO review
Hypertension prevalence	Low quality – No effect	WHO review
Incidence of central obesity	GRADE not conducted – Harmful effect	Defra-RIVM review
Ischaemic heart disease mortality	Low quality – No effect	WHO review
Ischaemic heart disease prevalence	Very low quality – No effect	WHO review
Obesity (change in BMI)	Low quality – No effect	WHO review
Obesity (change in waist circumference)	Moderate quality – Harmful effect	WHO review
Obesity (incidence of overweight)	GRADE not conducted – Harmful effect	Defra-RIVM review
Obesity (weight gain)	GRADE not conducted – Harmful effect	Defra-RIVM review
Stroke prevalence	Very low quality – No effect	WHO review

Appendix table 37 Summary of the quality of evidence for quality of life, mental health and wellbeing outcomes from the WHO and Defra reviews where no new evidence was available

Outcome	Quality of evidence – Direction of effect	Source of GRADE assessment
Emotional and conduct disorders in children	Low quality – No effect	WHO review
Hyperactivity	Low quality – Harmful effect	WHO review
Medication intake to treat anxiety and depression	Very low quality – Harmful effect	WHO review
Wellbeing	Very low quality – Harmful effect	Defra-Arup review

Appendix table 38 Summary of the quality of evidence for cancer and general health outcomes from the WHO and Defra reviews where no new evidence was available

Outcome	Quality of evidence – Direction of effect	Source of GRADE assessment
Incidence of breast cancer	Low quality – Harmful effect	Defra-Arup review