SOURCES, EFFECTS AND RISKS OF IONIZING RADIATION

United Nations Scientific Committee on the Effects of Atomic Radiation

UNSCEAR 2017
Report to the General Assembly, with Scientific Annexes

UNITED NATIONS
New York, 2018

NOTE

The report of the Committee without its annexes appears as *Official Records of the General Assembly*, Seventy-second Session, Supplement No. 46* (A/72/46*).

The designations employed and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The country names used in this document are, in most cases, those that were in use at the time the data were collected or the text prepared. In other cases, however, the names have been updated, where this was possible and appropriate, to reflect political changes.

UNITED NATIONS PUBLICATION
Sales No. E.18.IX.1
ISBN: 978-92-1-142322-8
eISBN: 978-92-1-362680-1

This publication has not been formally edited.

Information on uniform resource locators and links to Internet sites contained in the present publication are provided for the convenience of the reader and are correct at the time of issue. The United Nations takes no responsibility for the continued accuracy of that information or for the content of any external website.

ANNEX A
PRINCIPLES AND CRITERIA FOR ENSURING THE QUALITY OF THE COMMITTEE'S REVIEWS OF EPIDEMIOLOGICAL STUDIES OF RADIATION EXPOSURE

CONTENTS

I. INTRODUCTION

1. Evidence from radiation epidemiology studies forms an important part of the scientific evaluation of radiation effects regularly conducted and reported by the United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR). In many UNSCEAR evaluations, epidemiological studies are reviewed and their results used to assess health risks of radiation. As for any field of research, epidemiological studies have strengths and limitations, and each study requires careful and systematic assessment to gauge its contribution to the issue being addressed. Already in the UNSCEAR 1994 Report, it was stated that "Studies of disease in human populations must adhere strictly to epidemiological principles in order to achieve valid quantitative results. These include sound case ascertainment, an appropriate comparison group, sufficient follow-up, an accounting for confounding factors*[1] and well-characterized dosimetry." [U1]. In epidemiological studies of radiation effects, the key concerns are typically limitations in exposure assessment as well as the general issues for epidemiological studies, including study size, statistical power and chance, information bias* (comparability of information on health outcomes), selection bias* (potential selective inclusion or exclusion of subjects in relation to exposure or outcome) and control of confounding influences (impact of other determinants of disease risk than radiation). Many studies in radiation epidemiology deal with exposure at low doses* and low dose rates.* This makes the assessment of exposure-outcome associations, including the evaluation of dose responses, particularly challenging. Examples of such studies are used for illustration in this scientific annex. More examples are presented in more detail in annex B on the evaluation of epidemiological studies of cancer risk due to low-dose-rate radiation from environmental sources.

2. The Committee's evaluations commonly need to cover a broad array of epidemiological studies, often with widely differing objectives, designs and results. Clear and transparent criteria that define the processes and decisions for the inclusion or exclusion of individual studies are essential to ensure that its evaluations meet the key scientific norm of objectivity, i.e. to use sound evidence in an undistorted manner, regardless of the composition of the group of experts conducting the evaluation for UNSCEAR's scrutiny, with assessment unaffected by non-pertinent features of studies.

3. The Committee agreed at its sixty-second session (1–5 June 2015) that the use of epidemiological evidence to evaluate risks to the public and to workers from radiation exposure at low doses and low dose rates is a methodological challenge, and expressed the need for quality criteria to evaluate epidemiological studies in line with the UNSCEAR 2006 Report (annex A) [U3] and the UNSCEAR 2012 Report (annex A) [U7].

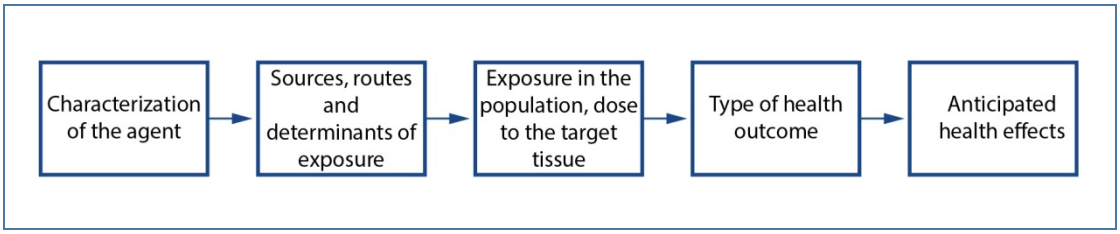
4. The key principles of systematic reviews are very clearly applicable and fundamental for the Committee's evaluations. Compiling, summarizing and comparing the results from all relevant and sound evidence is essential for achieving the goal of the Committee in producing high quality evaluations. As pointed out above, this requires that a question to be addressed is clearly defined, a search of all evidence is carried out, relevant studies fulfilling the inclusion criteria are identified, their results compiled and conclusions drawn from the evidence.

[1] Technical terms are explained in a detailed glossary, and are marked with an asterisk (*) the first time that they appear.

5. Systematic reviews have emerged as a key tool for providing guidelines in medicine, largely due to the evidence-based medicine (EBM) movement since the 1990s [E1, S1]. EBM offers broad principles for translating research into policies and practice. The need for it arose largely from conflicting guidelines prepared by professional societies and other bodies, and also from the explosion in the number of scientific literature available. A key tenet of EBM is the requirement that practice of medicine should be based on the best available scientific evidence. The concept of science-based medicine was not first introduced with EBM, but it emphasizes that the evidence must be directly applicable, i.e. based on real patients and outcomes of treatment. For health effects* of radiation exposure, this would translate into underscoring the importance of human studies.

6. Another parallel development has been the evolution of health risk assessment, combining several lines of evidence from toxicology, epidemiology and other disciplines for estimating the health impact of exposure to potentially harmful substances. In general, health risk assessment is a process consisting of hazard* identification, exposure assessment, dose-response assessment and risk characterization [I5] and it provides the basis for risk management and risk communication (figure I). Health risk assessment has been widely applied by expert committees mainly for chemical agents [I4]. Across the world, systematic reviews are increasingly becoming the basis for these assessments [F1].

Figure I. Health risk assessment [I5]



7. To satisfy the criteria of comprehensiveness and transparency, a thorough search for the relevant evidence must be performed as the basis for a systematic assessment of health risks. Scanning reference lists from studies, conducting dedicated hand/Internet searches, and actively contacting study authors and relevant organizations may additionally prove useful. Personal paper collections of experts in a working group can be a very good starting point on specific topics, but should be augmented by systematic searches. For example, key terms from identified key studies can be used in further searches to identify additional studies of relevance for the topic.

8. In critical appraisal of evidence derived from individual studies, evaluation criteria include relevance, validity* and precision. Relevance means that a study can provide information pertinent to the aims of the report. Validity means lack of bias (i.e. systematic error*) and capturing the essential features of study quality for epidemiological research. Validity can also be conceptualized as the degree to which a study reaches a correct conclusion. Precision refers to statistical power and amount of information, or limited presence of random and shared error.*

9. The Committee's evaluations synthesize studies of effects of radiation exposure on humans and the environment to guide decision-making, assist in preparing regulations and inform the scientific community and the public. Its reports comprise evidence summaries and aim to be comprehensive and systematic in order to conduct a balanced and exhaustive evaluation of the current scientific knowledge and provide evidence-based conclusions. However, the Committee has previously stated, in its UNSCEAR 2012 Report, that a balanced evaluation involves avoiding unjustified causal associations (false positives) and also unjustified dismissal of real health effects (false negatives) [U7]. The overall approach entails several sequential steps:

(a) Clearly define the topic and objectives of the specific evaluation;

(b) Perform a search that allows identification of all studies with potential to contribute to the evaluation and its defined objectives;

(c) Apply a uniform approach to scrutinizing the quality of the studies;

(d) Synthesize the available evidence from the studies by summarizing their results from those studies meeting criteria for inclusion (providing relevant, high-quality evidence);

(e) Develop overall conclusions drawn from the systematically retrieved, assessed and summarized studies.

10. This annex reconfirms and expands upon prior UNSCEAR Reports including guidelines to judge the quality of epidemiological studies [U2, U3]. Further, it aims to recommend procedures for assessing radiation epidemiological studies and criteria for conducting scientific reviews of epidemiological studies for the Committee's evaluations. These criteria make use of the experience of development of EBM, within Cochrane Collaboration and other settings, and of the extensive expertise specifically for radiation epidemiology available within the Committee. Furthermore, this annex also re-iterates the importance of adhering to the Governing Principles[2] of the Committee, notably to its process in the selection of experts according to defined terms of reference, qualifications and competence, and the declarations of potential conflicts of interest are essential steps in securing quality and scientific excellence of the Committee's evaluations. All conflict of interest statements collected in the framework of the Committee's work are available at the secretariat.

II. OVERVIEW OF EPIDEMIOLOGICAL STUDY TYPES

11. A commonly cited definition for epidemiology is "the study of distribution and determinants of disease in human populations" [M1], with some expansions of the definition more recently [P2]. Epidemiology has also been characterized as "occurrence research in medicine" [M3] and as the science of understanding the causes and distributions of population health with a view to intervening for disease prevention and health promotion [K2]. Its key features include a population perspective on human health and disease, with disease occurrence perceived as a stochastic* process and analysed with statistical tools. Epidemiology is largely an observational (i.e. non-experimental) science. Randomized, experimental or intervention designs can sometimes be used in epidemiology, for example screening or prevention trials, or post-intervention follow-up of persons included in a clinical trial, but generally cannot be applied to evaluations of radiation risk. A typical sequence of radiation risk evaluation using population-based epidemiological methods includes *(a)* assessment of radiation exposure, *(b)* assessment of presence or absence of a health effect of radiation, *(c)* determining the magnitude of the effect (if present), *(d)* studying dose-response patterns and variation of observed patterns in different populations/subgroups and *(e)* review the overall evidence with regard to causation* (e.g. by using Hill's considerations for the assessment of causality [H5]). The drawing of conclusions from scientific observations in the presence of uncertainty*—such as those obtained from well-designed studies of radiation epidemiology—is called inference. The two principal approaches—outlined in detail in the UNSCEAR 2012 Report (annex A) [U7]—are the frequentist inference* and the Bayesian inference.*

[2] http://www.unscear.org/unscear/en/about_us/governingprinciples.html

12. Unlike experimental studies (e.g. randomized controlled clinical trials), observational studies rely on data generated by the uncontrolled conditions of everyday life, i.e. research opportunities created by variations in exposure conditions within and between populations (sometimes slightly misleadingly called "natural experiments"). This exposure variation in the study population is the fundamental requirement that needs to be fulfilled for observational studies with the objective of evaluating radiation risks. Such settings rarely provide highly comparable study populations, and the effects of extraneous factors make interpretation of epidemiological results challenging despite the apparently straightforward data they produce, with temptingly oversimplified conclusions to be derived. However, as epidemiological findings relate directly to humans, the results are not (or are much less) affected by the limitations of generalizability that apply to extrapolations of the findings of experimental studies (e.g. using laboratory animals) to humans, where simplified and often highly specific model* settings do not represent real human disease, heterogeneity of susceptibility in the population, or variations in co-exposure.

13. Any epidemiological study has a defined scope, specified by the research questions to address one or several specific end points, or health outcomes. Descriptive studies provide estimates on features such as mortality or incidence rates* in populations and subgroups thereof and may be useful to generate hypotheses about possible associations based on the frequency* distributions observed. Studies that focus on risk factors/determinants such as ionizing radiation and their effect on health outcomes are generally termed analytical studies and tend to rely on individual, not aggregate data. In analytical studies the determinant (generically termed "exposure" in epidemiology), the effect of which is studied, has to be defined conceptually and operationally. Observational epidemiological studies, including cohort, case-control, cross-sectional and geographical/temporal correlation ("ecological") studies are non-experimental and thus, in addition to ever-present statistical uncertainties, always prone to bias or confounding, which can lead to spurious results.

14. High-quality epidemiological studies minimize bias and confounding and provide strong empirical evidence about health outcomes in human populations. This requires careful selection of study participants, valid and accurate assessment of exposure and outcome, and data on potential confounders, all of which have to be described in detail in a study protocol that also specifies in advance the study questions and statistical analyses to be conducted. Statistical power, i.e. the probability to detect an effect or association of a given strength if it really exists, can also be major issue as only studies with sufficient data (e.g. number of subjects, length of follow-up) allow precise estimation of effects. The Committee summarized the concept of statistical power in its 2000 Report [U2] and further expanded it in its 2006 Report [U3].

15. Calculation of sample size and statistical power in the planning stage is a feature of good quality epidemiological studies. When the sample is determined beforehand, for example by the inclusion of all exposed persons, sample size calculation is superfluous, but an estimate of statistical power—or, more generally, precision—of the study given the available study population is still informative. Sample size calculations are particularly relevant at the planning stage of studies with extensive and costly exposure and/or end point assessments, in order to achieve the optimum balance between scientific value and financial costs. The statistical power is, however, always estimated using simplifying assumptions such as no measurement error* and a particular underlying excess risk* and does not usually reflect other sources of uncertainty beyond random error* [P5]. Therefore, it is an imperfect measure of information that can be obtained in a study as a small high-quality study can provide more useful information than a large study with major systematic error.

16. The measure of effect for many epidemiological studies is some form of relative risk* (RR), such as the odds ratio* (OR) in case-control studies. The focus in most radiation epidemiology studies is on magnitude and precision of the central effect estimate of interest, often expressed as excess relative risk (ERR)* or excess absolute risk (EAR)* and usually accompanied by a confidence interval* as a

measure of statistical precision. Statistical methods that provide an estimate of the expected confidence interval (i.e. precision) are explained in Goodman and Berlin [G3]. Several realistic statistical power scenarios using the best available insight into the expected average dose or the dose distribution should be applied, if possible including uncertainty considerations as explained in the UNSCEAR 2012 Report (annex B) [U7]. As cohort studies commonly allow for the investigation of multiple end points, one or a few primary end points should be specified for which these calculations are done. The other health outcomes studied are then considered as secondary end points, which require particular caution in reporting and interpretation as they are not the primary focus of the study, and their statistical power may be too limited for them to be addressed. Guidelines on good epidemiological practice (GEP) such as those developed by the International Epidemiological Association (IEA) give further guidance on planning, conducting and interpretation of epidemiological studies [I1].

17. For the Committee's evaluations, in general, epidemiological studies are considered to be useful that address one of the following key study questions:

(a) Is there an increased frequency of disease (individual entities or disease groups) associated with radiation exposure?

(b) What is the magnitude of the effect associated with radiation exposure?

(c) What is the time-, sex- and age dependency and the shape of the dose-response curve?

It should be noted that the most valuable epidemiological studies on health effects of radiation exposure are those with quantitative dose-response results. This implies that the quality of quantitative exposure and disease data is central to the value of a given study.

18. Besides the overall effect, variation in effect due to factors that modify the exposure-disease associations such as age at exposure and time since exposure may be relevant for the Committee's evaluations. In epidemiological terms, the study of the existence of a causal effect (also called hazard identification in risk assessment) corresponds to hypothesis testing, the assessment of magnitude of effect to risk estimation, and the study of differences in risk within the study population refers to effect modification. The UNSCEAR 2006 Report (annex A) [U3] provides an overview of common epidemiological study types, and denotes two main types, both of which are based on individual data:

– Cohort studies collect data on exposure and other factors for a defined group—the cohort—before the outcome event such as cancer diagnosis or death. Cohort studies are useful for the study of rare exposure and allow investigation of multiple outcomes. The temporal relationship between exposure and outcome can be identified; exposure, outcome and confounders can be measured repeatedly over time. They are less useful for rare diseases, and in many instances time-consuming and relatively expensive. If—after baseline data for the cohort have been collected—further follow-up is conducted in the future (relative to the baseline), this is called a prospective cohort study. A prominent example of a prospective cohort study in radiation epidemiology is the life span study of the atomic bombing survivors, where individuals from Hiroshima and Nagasaki with individual exposure estimates have been followed-up over more than 60 years since the study began in 1950 [O1]. Historical (sometimes called retrospective) cohort studies are often used in studies of occupational exposure, but may also be used for population-wide exposure. Their particular feature is the fact that the study period is extended to the past, for instance, if occupational radiation exposure of employees first occurred decades before the data collection for the study was started [R1]. Thus, the cohort is defined and study entry occurs at a past date, while the conduct of the study, i.e. data collection on study subjects, exposure and outcomes takes place well after that date. All information on study

subjects needs to be reconstructed/retrieved without information bias, as well as without selection bias. Availability of historical exposure information is a core feature of these studies. The retrospective data collection necessary for these studies also frequently means that dose estimates have a higher uncertainty, for example, because exposure measurements from decades ago are likely to be fewer and less accurate than for more recent periods. For example, in European cohort studies of uranium miners [K3, V1] exposure assessment in the early years of mining was less accurate than in later years, leading to a higher degree of uncertainty. Similarly, reliable outcome information (e.g. from disease registries) may be available only for more recent periods. Thus, there are no or very limited outcome data for early time windows in cohort follow-up such that health effects in these early periods cannot be studied.

– Case-control studies collect retrospective exposure and other data for cases with the disease of interest and controls that do not have the disease under study and are representative of the underlying population from which the cases derive. Case-control studies are useful for rare diseases, and they allow investigation of multiple past exposures. They are often less time-consuming than cohort studies and, thus, require fewer resources. However, the case-control design is not suitable for the study of rare exposure, and it is vulnerable to numerous biases. The core difference from cohort studies is the fact that disease or, in certain cases, death has already occurred and thus defines case status whereas in prospective cohort studies, individuals from a defined group are followed up to determine whether disease or death occurs within a defined period (in historical cohort studies, the follow-up period starts in the past). For example, case-control studies have been used to investigate lung cancer and indoor radon. For cases with lung cancer and appropriately selected control persons without lung cancer, exposure histories based on residential radon measurements were constructed and compared, and individual information on other important risk factors, such as smoking history and occupational exposure, were used in the analysis. Case-control studies from different countries were later pooled to obtain more statistical power [D1, K4]. Other examples include case-control studies of thyroid cancer after the Chernobyl accident [A3, K1]. Case-control studies can also be nested within cohorts: detailed data on exposure and other relevant factors are collected on a subset of all cohort participants (e.g. all cases of a particular type of cancer, and a set of control persons draw from the cohort), which enhances efficiency as detailed information is collected only for a small subset of the cohort members. Thus, well-conducted nested case-control studies can be of similar value to that of cohort studies, and often there is only a moderate loss of statistical power. For example, Schubauer-Berigan et al. [S4] conducted a case-control study focused on leukaemia deaths in a larger cohort of nuclear shipyard workers. Information on benzene exposure as potential confounder was retrieved for all cases and the respective controls (but not for the whole cohort). Thus, the authors were able to include benzene as a confounder in their case-control analysis.

19. Other and related study designs include the following:

– Randomized trials use similar follow-up procedures to those of prospective cohort studies but form the groups (called trial arms) by randomization at baseline with one arm receiving an intervention and the control arm receiving another procedure (commonly no intervention except placebo/sham or, in treatment trials, the routine treatment). All individuals in both groups are then followed up for the outcome(s) under study as in cohort studies. In randomized controlled trials (RCTs), the intervention is administered by the researchers and these trials are experimental studies. Randomized trials are regarded as the gold standard for the study of effectiveness of medical interventions, including disease treatment, screening and prevention. This is so because the fundamental requirement of randomization is designed to remove or minimize bias and confounding, which is not possible in observational studies. It should be

noted that this study design is generally not feasible for the study of harmful exposure such as ionizing radiation in human populations, except in a therapeutic context or for radiation protection (where primarily the beneficial effects are assessed). For example, an RCT design can be used to assess the efficacy of occupational radiation protection measures [A2].

– In cross-sectional studies both exposure and outcome of interest are assessed at the same time. This study type is commonly used in health surveys, including the Fukushima Health Management Survey (FHMS), conducted among residents of Fukushima prefecture in Japan after the nuclear power plant accident [Y2]. Cross-sectional studies can serve as baseline investigation for a subsequent cohort follow-up (this is the case in the FHMS), but do not themselves allow inference about the cause and effect sequence. There are instances when a study that collects clinical/laboratory information on persons with previous radiation exposure is also described as cross-sectional. This type of study is conceptually closer to a cohort design and can provide evidence on cause-effect relationships or dose response if well conducted, with appropriate comparison groups. The thyroid disease survey of residents in eight villages near the Semipalatinsk Test Site in north-eastern Kazakhstan [L1] is an example of this approach. Individually reconstructed thyroid doses were used for the analysis of thyroid abnormalities detected by ultrasonographic screening and subsequent cytopathological examination of suspicious findings.

– Geographical/temporal correlation studies (at times called “ecological studies” in epidemiology) and time-series studies rely only on group-level (aggregated) data on exposure and outcome, without individual data. Therefore, they cannot show whether the exposure and outcome occur in the same subjects. The groups compared may also differ with regard to other relevant factors, for example lifestyle, occupation, medical care, which impairs comparability and can induce confounding factors. Studies comparing incidence or prevalence of disease between geographical areas related to distance from a nuclear power plant [H3, J1], to natural background radiation [N1] and to fallout from atmospheric nuclear weapon testing [D3] are examples for this study type. The potential bias of such studies has been detailed elsewhere [G5, P1].

20. It is not uncommon for epidemiological studies to involve some aggregate-level element, as some extent of grouping is commonly used in estimating exposure even in analytical studies such as those with a historical cohort design. For instance, typical exposure levels may be assigned for all subjects with a similar job title. This is distinct from ecological studies in that imputed exposure and health outcomes are being studied for individuals; in this case, the use of aggregate exposure data falls under the category of “grouped” or “Berkson”* measurement error, also called assignment error* [H2].

21. When assessing exposure-disease associations, cross-sectional and correlational studies generally provide evidence of markedly lower quality than cohort and case-control studies, and are considered to be hypothesis-generating or exploratory rather than answering questions of cause and effect.

22. Further epidemiological non-experimental (observational) study designs include the case-cohort, the case-only or further, less frequently used designs. Case-cohort studies are a variant of nested case-control studies with random selection of controls from the overall cohort at baseline. For example, Auvinen et al. [A4] studied the association of radon in drinking water with stomach cancer, using this design. All stomach cancer cases in a sample of the base cohort of people not connected to the municipal tap water system were compared with a set of controls defined from the same cohort at baseline, with focus on radon and other radionuclides in the drinking water of the study groups. No risk increases associated with ingestion of radon or other radionuclides were found. Case only studies compare subsets of cases in relation to a potential determinant. These simply provide an overview of all or a subset of cases reported, for example to a specific hospital or department. They are not usually

regarded as epidemiological studies and, more specifically, cannot provide information on exposure-disease relationships.

23. For many of the study designs mentioned here, mortality or disease registries, including cancer registries, are a major source of information for outcomes. Population registers are also an important source of information, as epidemiological studies require definition and enumeration of the population from which the cases arise, and there is a need to accrue information about vital status and place of residence in their study populations.

24. Of particular relevance for evidence synthesis are systematic literature reviews, with or without quantitative meta-analysis.* Systematic reviews are a key component of EBM. Well-conducted meta-analyses of randomized trials, if available, form the highest level of evidence in EBM or public health [S12]. Meta-analyses of non-randomized studies may also provide strong contributions to an overall evidence assessment and constitute the main method of pooling quantitative evidence for risk assessment. Comprehensive overviews of the scientific literature are also a cornerstone of many of the Committee's historical evaluations while meta-analyses of quantitative data have rarely been performed specifically in this regard. A systematic and comprehensive quality-oriented process of research synthesis is described in more detail in chapter IV.

III. MAIN FEATURES AFFECTING THE QUALITY OF EPIDEMIOLOGICAL STUDIES

25. Scientific assessment of research evidence needs to consider the quality of studies available. This chapter discusses the core features influencing study quality. All these issues are relevant when reviewing study quality, and should not be viewed separately. For studies of radiation health effects, however, radiation exposure assessment—i.e. a scientifically sound and validated approach to dosimetry—is a crucial issue when establishing how informative a single study is.

26. *Definition of study population.* Populations included in epidemiological studies (and also any exclusion criteria) need to be clearly defined with regard to age, sex, period of observation, exposure and other characteristics relevant for the particular study. For cohort studies, the data consists of person-years of observation and events (disease cases). Hence, the accumulated duration of follow-up is important, particularly as minimum latency* between radiation exposure and effect (onset of risk elevation) is typically several years.

27. *Exposure assessment*. Exposure assessment is a critical component of any epidemiological study. For radiation epidemiology, absorbed dose* to the relevant organ or tissue ("organ dose") is often considered the gold standard as an exposure indicator, in particular when exposure to an individual is a mixture of external and internal exposure. However, the choice of the most suitable exposure metric depends on the study question. For example, in cases where exposure measured by an instrument or device can be directly used for public health purposes, such as in radon decay products in homes, this exposure measure (e.g. long-term average Bq m^{-3} in air) may be the preferred indicator for epidemiological studies. The actual lung dose reconstruction requires information on breathing rate, equilibrium factor between radon and its decay products, inhaled particle size, deposition fraction, target cell location and other factors that may vary by individual and working conditions as well as over time. In practice, there are numerous constraints to the precise estimation of organ doses in the context of epidemiological studies including the fact that physical dose is not evenly distributed over large

organs/tissues such as lungs or skin. Nevertheless, useful and high quality exposure assessments can be conducted for epidemiological studies.

28. Dosimetry systems such as that developed for the Techa River cohort study provide organ dose estimates for external and internal doses, integrating information from different exposure situations and time dimensions and applying conversion factors, as described in detail in annex B. The equivalent dose to an organ or tissue is a radiation protection quantity and accounts for different radiation quality in terms of biological effectiveness. It is calculated using predefined factors for different types of radiation and thus represents a radiation-weighted absorbed dose. The radiation weighting factors used to calculate the quantity are not necessarily those to be used in a scientific study. This quantity cannot be measured directly [M2], which is the case also for other dose quantities—for example, for the absorbed dose received within the body from intakes of radionuclides. Effective dose, a derived dose quantity with application of predefined radiation and organ/tissue risk weighting factors, is not the preferred quantity for risk estimation. It is not a direct measure of dose, but a construct developed for radiation protection purposes [M2]. Further, changes over time, for example with regard to the tissue weighting factors, may limit the comparability of effective dose values from different periods.

29. Whole-body dose can be useful when the outcome is incidence of all cancers and the dose within the body is relatively uniform. Alternatively, colon or stomach dose has been used as an indicator of typical organ dose for several parts of the body, with the understanding that colon or stomach doses are representative of doses received deep within the body.

30. Epidemiological studies should be based on highly detailed and accurate dose data. Quantitative estimates of radiation doses are required because simple classification of the study population into exposed and non-exposed is inadequate for quantitative risk estimation. Ideally, radiation exposure should be assessed by measuring or recording the physical quantities at individual level, such as personal dosimetry of radiation workers or patient dosimetry [S9]. Regarding external irradiation, this is for example done using personal dosimetry of radiation workers or patient dosimetry. For environmental studies, because the individuals exposed are not expected to wear dosimeters during the period of exposure, external doses at individual or group level are derived from other measurements, when available, such as ambient dose rates* at the locations of exposure in the Karunagappally and in the Yangjiang studies of natural background radiation [M6, N2], or deposition densities of ^{137}Cs and other radionuclides in the localities contaminated as a result of the Chernobyl accident [M4]. When radiation measurements are not available during the period of exposure, reconstruction of the source term is used in the LSS studies to estimate ambient dose rates [C7] or environmental concentrations of radionuclides in the locations of interest affected by the early releases of ^{131}I at Hanford [N4]. With respect to internal exposures, ideal situations, in which empirical data are available at the individual level, usually occur in occupational studies and medical studies of patients, but are rare in environmental studies, with notable exceptions such as the thyroid scans performed on all subjects of the studies of thyroid disease in Belarus and Ukraine [D8] or ^{90}Sr body-burden measurements or in-vivo measurements of surface activity of anterior teeth in the Techa River study [D5].

31. Individual-based radiation measurements are not sufficient to determine the doses to specific organs and tissues. The steps involved in the derivation of the organ doses from the individual-based measurements vary according to the type of study and to the information available. Detailed information and reviews for a range of studies is provided in [T3]. Only two specific examples will be given in this annex. In the first example, related to personal dosimetry of radiation workers, care must first be taken that the recorded doses are available for each subject for the entire period of exposure and that the undetected and unmonitored doses are estimated in an appropriate manner. Undetected dose is defined as the dose received that was not measured by the dosimeter, because it fell below the minimum detectable response of the dosimeter. Since the undetected dose may in reality range from

zero to the minimum detectable, it is customary to assign some fraction of the minimum detectable dose (or some other clearly stated value) for each monitoring period in which the dosimeter read zero (i.e. less than the minimum detectable), although zeroes were customary in some historical occupational data. Unmonitored dose is that assumed to be received when a personal dosimeter was not worn, and often may be reconstructed from information about workplace activities or from co-worker data when others in the same location did wear dosimeters. Attention must also be paid to the fact that changes in technology may have affected the recorded dose values. Additional complications are: *(a)* the recorded doses may include a mixture of radiation types (e.g. photons and neutrons) or may have measured some radiation types poorly; *(b)* organs or tissues may be only partially irradiated, for example when medical personnel wear lead aprons; and *(c)* recorded doses must be collected from all facilities where the worker was exposed during her/his career. The main effort then consists in defining the exposure scenarios for the various tasks carried out by the workers in order to determine the irradiation geometry and the energy spectrum of the incident photons on the body. It is important to note that the relative geometrical relationship between the dosimeter's placement on the worker, the incident source direction, and the organ or tissue is needed to derive an organ dose from the recorded dose in a manner that is suitable for epidemiological purposes [B4].

32. In environmental studies, very different steps are taken to derive internal organ doses from individual-based measurements. In this second example, related to the derivation of internal thyroid doses derived from thyroid scans performed in Belarus following the Chernobyl accident, the exposure rate obtained from the radiation detector was first processed to subtract the background due to the contamination of body and clothes from the signal in order to obtain the exposure rate due to the ^{131}I activity in the thyroid at the time of measurement. The variation of the thyroidal ^{131}I activity before and after the measurement was then estimated using environmental transport models adjusted according to information on lifestyle and dietary habits provided by the subjects during a personal interview. The thyroidal ^{131}I activity was integrated over the time of exposure and divided by the age- and region-dependent thyroid mass of the subject to yield the integrated ^{131}I concentration in the thyroid, which is proportional to the thyroid dose [D8].

33. Owing to the fact that the exposure assessment is limited to quantities that are only indirect (proxy) indicators of the true dose* causing the effect, some uncertainty inevitably arises. When exposure is determined using ambient exposure measurement, but also in cases where a dosimeter is not placed close to the target organ or measurements from one person's dosimeter are extrapolated to a group, the magnitude of this uncertainty increases. Some examples of such situations include dosimeters worn on the chest for assessing eye lens doses [C2] and the epidemiological study of Chernobyl clean-up workers for whom shared dosimeters were used for exposure assessment [I6]. This error (called Berkson-type error) may result in too narrow confidence intervals of the dose-response relationship if standard statistical methodology is used. Epidemiological studies frequently need to reconstruct doses retrospectively, carefully identifying and evaluating all available dose exposure information. An example of such an approach is provided by nuclear worker studies where external doses in particular were reconstructed from dose monitoring registries and many other information sources. This approach is more difficult and less precise for neutrons and internal exposure occurring in some settings, and even the estimation of photon doses in early years was problematic in times [T2].

34. A crucial requirement is that to avoid information bias, the exposure assessment should be applied in a similar fashion regardless of outcome, i.e. whether a study participant develops the disease(s) of interest in the course of the study or not. For that reason, a standard algorithm applicable across exposure settings is preferred. Also, for combining data from several studies, comparability needs to be evaluated to ensure consistency as a requirement for meaningful pooling. Another requirement is to add the doses from external and internal irradiation, and to capture all sources of exposure, such as those from natural background and medical sources, particularly in low-dose studies.

35. The quality of the data is the main determinant of the precision and validity of the results. For instance, imprecise and non-specific data for exposure estimation result in inaccurate and uncertain dose estimates. Exposure measurement errors can bias dose-response relationships and the extent of bias depends on the amount of measurement error. Classical* non-differential random error in exposure estimation tends to dilute the dose response (bias towards null). However, if individual values are replaced with group means (Berkson-type error), no bias is expected (if the group means are unbiased) though precision is lost as the loss of information reduces statistical power. Good quality studies in radiation epidemiology include a scientifically sound and transparent scientific approach to dosimetry or dose estimation, preferably at the level of the individual study participant. Examining possible systematic measurement errors in dosimetry (e.g. with regard to occupational doses over time) is needed in analysis of uncertainties. Uncertainty due to measurement or estimation error should be considered in the main analysis or in sensitivity analyses [U7]. The effect of measurement error (or uncertainty) can be dealt with by using statistical error mitigation tools such as regression calibration,* Bayesian methods and simulations. Several such examples have been published [K5, L2, L6, S13]. Uncertainties are further discussed below.

36. Biodosimetry* and electron paramagnetic resonance (EPR) dosimetry can be used to validate dosimetric information but, at present, it generally applies to individual exposure at dose levels above about 0.1 Gy. However, there are reports of lower detection limits for unstable chromosomal aberrations related to doses from low dose gamma rays of around 0.02 Gy [I8]. For EPR, a detection limit of around 0.03 Gy has been described [S10]. Unstable chromosomal aberrations (dicentrics in lymphocytes detected in metaphase scoring) are likely to be the most sensitive such indicators and also reasonably specific for radiation exposure but they are eliminated relatively soon after exposure (with a half-life of around three years). Stable chromosomal aberrations (translocations that are retained in cell division) are, therefore, more useful for longer-term exposure in the past. However, the detection threshold of this technique is limited by the variation in the background levels, which are dependent on age, sex and lifestyle factors. Furthermore, stable translocations are not radiation specific and can be affected by other environmental agents that can cause chromosome damage (mutagenic agents or clastogens). As indicated above, a limitation of these methods of biodosimetry is the rather high detection threshold, roughly 0.1 Gy [E2, L4, S14, T1]. However, efforts to validate the dose estimates using biodosimetry or other techniques is encouraged, whenever feasible. An example is the use of personal thermoluminescent dosimeters in the Karunagappally and the Yangjiang studies of natural background radiation to validate ambient dose rates [M6, N2].

37. *Outcome definition*. The health outcome (or disease end point) under study should be defined and described very clearly, with explicit diagnostic criteria and coding with international standards (e.g. international statistical classification of diseases and related health problems (ICD)) [W1]. Large groupings of disease entities (all cancer, all cardiovascular disease) may be used to provide an overview on average effect for a group of diseases but generally information on more specific disease entities is preferred. Otherwise, the results indicate that the average effect (weighted by frequency of outcomes) and the possibly different effect on specific diseases are difficult to disentangle. Causes of death are the most widely available outcome data for assessment of health effects although incidence data are preferable, particularly for chronic diseases such as cancer. However, the accuracy of death certificates is likely to vary between different diseases and can also be affected by other factors such as time period. According to the ICD coding guidelines, the underlying cause of death should be distinguished from immediate and contributory causes.

38. Adjudicating the various diseases that may be involved in the process leading to death may be complicated and error-prone, particularly when no autopsy has been carried out. A potential source of misclassification* is a change in assigning and coding practices over time and, in some cases, across regions or institutions. In several industrialized countries, cancer registries provide comprehensive

high-quality incidence data on neoplastic disease. Quality parameters for cancer registries include the estimated completeness of ascertainment (e.g. Nordic countries: close to 100%), the percentage of cancer detected only retrospectively by death certificate (<2%), the percentage of cancer with unknown primary site (<5%), and the percentage with microscopic verification (>95%) [E6]. However, completeness varies between registries and, in some instances, over time or by type of malignancy (commonly less comprehensive for haematological malignancies). Medical records are a less ideal source of information as they are indicators of health-care use (not only disease risk), and often involve some variation in diagnostic criteria over time or between institutions and levels of health care. Medical examinations conducted systematically for research purposes can provide more standardized outcome data than medical records. For example, in the Adult Health Study such examinations were conducted periodically on the atomic bombing survivors in Japan [Y1]. Biomarkers of radiation effect are sometimes used in epidemiological studies. An example is thyroid nodules, which are correlated with radiation dose and are not a meaningful disease entity in themselves, but only when progressing to malignancy [A1].

39. *Bias*. As described in the UNSCEAR 2006 Report (annex A) [U3], bias is any process at any stage of study design and conduct that tends to produce results or conclusions that differ systematically from the true exposure-disease association. There are numerous types of bias, the major categories being selection and information bias. Publication bias is of particular relevance for evidence synthesis.

40. Selection bias occurs when the actual study group differs from the intended target group in a fashion that affects the results, i.e. study subjects differ from the target group in terms of exposure, disease risk or both. Different length of follow-up in cohort study can also result in selection bias. In case-control studies, participation among controls is often lower than among cases, and subjects who agree to participate as controls tend to be more motivated than non-participants—due to their interest in either the exposure or the outcome (being affected by these usually increases willingness to participate)—or, more generally, to have a higher level of education. Therefore, they may differ from the non-participating controls in terms of exposure or other determinants of disease risk. Particular care needs to be taken when controls are recruited in hospitals as they may represent a selected group with disease risk and exposure probabilities that differ from the general population.

41. In studies of occupational exposure, the exposed group is defined by a certain occupation, employer or tasks performed. Therefore, study participants have to be employed to be included in the exposed group. If the comparison group represents the entire population, including those outside employment due to existing health conditions, mortality rates tend to be lower in the employed group. This “healthy worker effect”, often observed in studies with standardized mortality ratio (SMR: observed mortality in the exposed group compared with expected mortality estimated from the entire source population) as the effect measure, is a classic example of selection bias. As employed persons tend to be healthier than those who are outside the workforce (the latter include those who are unable to work for health reasons), their mortality rate is generally lower than that in the entire population. Furthermore, long duration of employment, including extended work with radiation, requires that the good health status is retained for long periods. Such additional selection may result in selection effects for the subgroup of workers with highest cumulative doses. This means that the disease incidence or mortality rates in the population are not appropriate indicators of the expected rate in the exposed group (do not represent what would have happened among them in the absence of exposure). Hence, the healthy worker effect typically underestimates risks in the exposed population, i.e. causes a downward bias in results.

42. Information bias occurs if there are systematic differences in information between the groups being compared. Perhaps the most classic example of information bias is recall bias in case-control studies, which is due to the fact that the cases are generally more motivated to provide full exposure information and therefore provide different, more complete exposure histories (and sometimes even over-report it), while controls do not have a similar interest and are more likely to under-report their exposure. For example, this may at least partly explain the reportedly increased risk of meningioma associated with dental X-rays (despite very low doses) [C3]. For instance, more comprehensive ascertainment and recording of disease may occur for a group with known radiation exposure. If completeness of coverage of cancer registries is different in high background radiation areas and reference areas (control population), this would result in distortion of results due to information bias (cancer rates would differ between the areas even if the true cancer risks were identical). Chernobyl liquidators are an example of an exposed population that has received special medical examinations because of a specific exposure. This may improve diagnosis and registration of any disease detected and the group is, thus, difficult to compare with the general population not undergoing such checks.

43. As a further example, thyroid screening with ultrasound is likely to increase detection of small thyroid carcinomas [E5, I7, J2]. This will result in earlier diagnosis of asymptomatic disease and some cases might not have developed into a symptomatic stage, but have remained latent or even regressed and, hence, can be regarded as over-diagnosis. This can result in information bias if screening is more intense in groups with higher exposure. Examples of bias due to thyroid screening were discussed in the UNSCEAR 2008 Report (annex D) [U4]. To ensure uniform and comparable assessment, blinding can be used to avoid the effect of exposure outcome assessment that involves any subjective component (i.e. any degree of subjective judgement or interpretation of information). Also, to ensure comparability and minimize the effect of variability* between different assessors, the same people should conduct assessments in all study groups (with similar proportion of subjects in various exposure groups assigned to each assessor). Likewise, performing the assessments during the same period may be essential to avoid bias due to drift, i.e. change over time in assessment due to changes in equipment, calibration, reagents or procedures. Capturing records, for example by photography, may be a valuable approach to allow a standardized and systematic review at a later stage.

44. A specific type of bias related not to study design but to reporting of results is publication bias. It is a bias resulting from preferable publication of significant results. It can occur within a study with only positive results being reported. Publication bias can also affect the entire body of evidence, with results ending up completely unpublished, or only in the grey literature (e.g. conference abstracts). If an exploratory analysis is conducted, but only the significant findings are published, proper interpretation of the results is challenging as multiple testing can produce chance findings distorting statistical significance* unless all other (negative) test results are also revealed. It distorts the published evidence as null or non-significant results remain unreported and thus not retrievable even by systematic literature search. A related issue is “winner’s curse” due to very striking initial observations turning out much more modest in further research [I3].

45. *Confounding*. Confounding occurs when a third factor is associated with the exposure under study and also affects the outcome of interest, which leads to bias in the results pertaining to the exposure-disease association [U3]. Hence, it reflects the inability of the study to disentangle the effects of the exposure studied from the other determinants of disease risk. In principle, all known risk factors of the health outcome studied are potential confounders. They become actual confounders if they are also associated with the exposure of interest. This means that they can distort a radiation–disease association unless their effect is taken into account. An example of confounding in studies of high background radiation could be differences between high exposure and comparison areas in lifestyle, occupational or other environmental factors that affect the cancer risk. In such situations, the expected cancer rates

would not be similar between the areas even if background radiation levels were identical due to differences in other cancer risk factors. As a further example, smoking was considered a strong confounder in studies of nuclear workers as it is related both to cumulative radiation exposure and to cancer as the health outcome of interest [C1]. Confounding needs to be differentiated from effect modification where the effect of the exposure under study on an outcome varies by the level of the modifying factor.

46. Several methods can be used to control confounding. At the design stage, matching (e.g. selecting similar controls for cases in terms of age, sex or other potential confounders) or restriction (only including a strictly selected subgroup, such as never-smokers, for better comparability) address confounding to some extent. However, matching may also induce bias and is now regarded as an approach mainly to increase study efficiency. The basic concept of confounder control at the analysis stage is stratification* by the confounder. Beyond simple stratified analysis, adjustment* in multivariate analysis is one of the most widely used methods. However, collection of information on confounders is needed to control for their effects, which makes it equally important to obtain such data as to obtain information on the exposure and outcome. Some confounders may not be known at all, so control is not possible in an observational design. If information on confounders is incomplete or inaccurate, adjustment does not necessarily remove all confounding, but some of it may remain (known as residual confounding). Such inaccuracies arise when the confounder is assessed imperfectly, with measurement error or missing data (which is always the case to some extent). This implies that only some of the effect of the confounder(s) is actually removed, but some of it remains in the results. If the outcome under study is a broad group of diseases (such as all cancers or cardiovascular diseases), then various confounders show different effects in subtypes of disease. This means that effective control of confounding is challenging.

47. In order to illustrate the effect of confounding by a confounding factor, a simple example is shown in table 1. The dose distribution in this table is constructed so that it is similar to that of the Karunagappally cohort study of cancer risk in a high natural background radiation area in India [N2] (for more details see annex B). An important risk factor for cancers of the upper digestive tract is smokeless tobacco (in India predominantly tobacco chewing), and it may act as confounder when analysing the radiation-related cancer risk. The RR for the potential confounder with regard to solid cancer is assumed to be 3 for this example. The prevalence of tobacco chewers in the 50 mGy and 200 mGy groups is given as 40%. However, in the lowest dose group, the prevalence is set to be 20%, i.e. half of that in the higher dose groups. For the total sample, the ERR estimate is 0.1438 per 100 mGy. When stratifying the total sample by the confounder tobacco chewing, the ERR estimate is 0.05 per 100 mGy among both tobacco chewers and non-chewers. In other words, tobacco chewing positively confounded the ERR estimate per unit dose in this example by a factor of almost 3, since the overall ERR per unit dose was much higher than the ERR in the individual strata of smokeless tobacco use. As shown in this example, confounding can be caused irrespective of whether the radiation dose is linearly related to the prevalence of a confounder or not.

48. *Analysis.* An appropriate analysis is based on an a priori study plan. A clear and transparent data description is necessary, and a comprehensible choice of effect measures. To improve transparency of research findings, making the original non-identifiable data publicly available is encouraged for example by the open science framework and medical journal editors. Confidentiality requirements have to be kept in mind when making study data available to the scientific community.

Table 1. Illustration of confounding by tobacco chewing

ERRs estimated on theoretical cohort data, using model-based calculation software

Mean dose (mGy)	*Person-years*	*Cancer cases*	*Cancer incidence per 10^5*	*RR*
	TOTAL			
0	100 000	140	140	1.0 (referent)
50	400 000	738	184.5	1.32
200	100 000	198	198	1.36
			ERR=0.1438 per 100 mGy	
	WITH TOBACCO CHEWING			
0	20 000	60	300	1 (referent)
50	160 000	492	307.5	1.025
200	40 000	132	330	1.1
			ERR=0.05 per 100 mGy	
	WITHOUT TOBACCO CHEWING			
0	80 000	80	100	1 (referent)
50	240 000	246	102.5	1.025
200	60 000	66	110	1.1
			ERR=0.05 per 100 mGy	

49. In radiation epidemiology, the main purpose of the analysis is usually estimation of shape and slope of the dose response (particularly if the outcome is cancer), typically expressed as ERR per dose unit (sometimes as EAR per dose unit). As for cancer studies, for health effects of radiation other than cancer a quantitative risk estimate is desirable (besides statistical significance for hypothesis testing) as an obvious concern for application of well-established risks is radiation protection, including setting exposure limits. The confidence interval is a key feature reported beside the point estimate* and indicates comparability with other risk estimates, precision of the estimation procedure and assessment of consistency with previous results. Typically, 95% confidence intervals are estimated when presenting epidemiological results, but 99% and 90% confidence intervals are also reported in the literature, representing either broader or narrower coverage probabilities with respect to the true effect size. Ideally, the confidence interval should also incorporate uncertainties associated with the dose measurement errors, which are further discussed later on in this section, but it usually accounts only for statistical errors. Also, biologically based or mechanistic models have been developed to integrate epidemiological data with the understanding of disease mechanisms. They can supplement traditional risk models and provide information about consistency between biological understanding and disease occurrence in humans [E5].

50. The main and supplementary analyses should be clearly specified. The main analysis should generally cover all or a very extensive section of the study subjects. Sometimes, a subset with the most valid data can be defined as the primary analysis, but this should be specified a priori. The cut-point defining the categorizations also needs to be selected a priori; not on the basis of results obtained. Otherwise, the reported analysis can be seriously biased. Sensitivity analyses involving different cut-points can be informative for the interpretation of study results.

51. Several temporal factors are relevant for analysis. Radiation effect may vary in relation to time since exposure. Adequate latency needs to be considered and is often dealt with by lagging exposure in analyses. Age at exposure and age at observation (age attained) may modify the risk.

52. Additional sensitivity analyses (sometimes called analyses of influence) are often used to evaluate the robustness of the results, i.e. the impact of sources of error or choices made in the analysis plan (assumptions or definitions). Such analyses model the impact of changes in some or all factors contributing to the final results. Their results may help to strengthen and differentiate the results of the primary analysis. An example of a study with such sensitivity analysis is the recent INWORKS report [R1] where different lag times for cumulative doses, a restriction to lower dose ranges and the exclusion of women and workers with suspected or documented internal contamination were assessed in sensitivity analyses.

53. Adjustment for well-established risk factors to minimize confounding can be easily justified, but adjustment for factors not known as risk factors for the disease under study can also induce bias (adjustment does not always and necessarily reduce bias) [S3]. If several alternative measures or groupings of exposure/outcome are used, findings may arise just by chance. Multiple comparisons are sometimes dealt with by using significance level adjusted for the number of tests performed. It is preferable, however, to specify the main analyses beforehand.

54. Subgroup analyses can be used to address heterogeneity of effect (effect modification or effect measure modification) such as differences in risk by age at exposure or other features of specific subgroups. It is important to note that the criterion for presence of effect modification is the statistical significance of the interaction term. Such differences in effect may reflect susceptibility and subgroup analyses are then a secondary outcome worth evaluating and reporting, provided that the statistical power is sufficient. Other research questions involving effect modification include, for instance, joint effects of radiation and other exposure. Research focusing on joint effects is challenging as detailed information of good quality on all exposure considered in the respective analyses is necessary, and subgroups jointly exposed may be small. Nevertheless, evidence on joint effects may be highly informative for risk assessment and as guidance for future targeted studies.

55. *Dealing with uncertainties*. The evaluation of uncertainties begins by identifying the sources of uncertainty that are likely to have substantial impacts on conclusions of a single study or a systematic review. This depends on the topic and objective of the study/review as the following two examples illustrate:

(a) Certain types of shared dosimetry errors such as those resulting from uncertainty in a source term, could result in substantial bias in estimates of excess cancer risk associated with radiation, but would have no effect on calculations for testing whether the excess risk equals 0;

(b) Model uncertainty* regarding the choice and functional form used for effect modifiers is important regarding estimates of excess rates for parameter values outside the centre of the data, for example for doses lower that the average dose, specific age groups above or below the average age of the cohort, shorter or longer times after exposure than the average cohort value.

56. The various sources and types of uncertainties for observational studies were discussed in the UNSCEAR 2012 Report (annex B) [U7]. These include sources or uncertainty that result in random and/or systematic error in estimates of relative or excess risk:

(a) Random error is a major issue especially for small studies (particularly those with a limited number of cases by dose range) and studies of low doses. Two measures of the amount of information in a study that is affected by random error are statistical power and the precision of

relevant estimates of risk. Statistical power can be defined as the probability of the study being able to detect an association or effect of a given size if one exists [P2]. Precision is closely related and refers to how close an estimate might be expected to be to the target quantity if potential sources of systematic error could be properly adjusted for, and is indicated, for example by the width of confidence intervals. For cohort studies, both power and precision depend on the number of cases of the disease of interest, the functional relationship between the proportion of excess cancer cases (ERR) and dose, and the distribution of dose within the study population. The number of background (unexposed) cases is, of course, the product of the cancer background rate and the person-years at risk, which is closely related to the years of follow-up. For a linear dose response, the required study size to attain adequate statistical power is roughly inversely proportional to both (i) the square of the slope of the dose response, and (ii) the variance of the dose distribution. For simple comparisons between a large unexposed group and an exposed group, the required sample size to detect an effect in an exposed group is inversely proportional to the square of the mean effect. If the ERR is proportional to dose, it follows that the sample size will also be inversely proportional to the square of the mean dose. For example, it has been pointed out that, assuming an ERR proportional to dose, the sample size needed to detect an effect for a population-mean organ dose of 50 mGy would be about 16 times greater than that needed to detect an effect for 200 mGy [N3].

(b) Selection bias, information bias, confounding, the healthy worker effect, survivor bias,* diagnostic inaccuracy, follow-up losses, and some types of dosimetry error are among the potential sources of epidemiologic uncertainties that may lead to systematic error or bias in estimates of excess risk. Methods for adjusting estimates of excess risk for some potential sources of bias, such as classical non-differential dosimetry error for relatively simple situations, are well-developed. However, systematic error in effect estimates arising from many sources of uncertainties are especially difficult to quantify. Examples include biases associated with (i) residual confounding; (ii) differential dosimetry error in occupational studies where measurements may not have been performed in accordance with protocols (e.g. misplaced badge dosimeters); (iii) shared dosimetry error for exposure from internal emitters based on complicated biokinetic models of uncertain functional form, unrecorded intakes, and sparse bioassay data; and (iv) motivated personal recall bias in dose reconstructions for case-control studies. Furthermore, the amount and existence of bias depends on the risk metric used. Nevertheless, approaches have been proposed for estimating the bias in effect estimates. Typically, these involve a series of sensitivity analyses applied serially for each source of bias (as described by Lash et al. [L3]). A problem with this approach is that it does not account for the joint effect of multiple sources of bias. More complex methods are based on an expansion of models for the data likelihood* to include parameters that explicitly account for sources of bias. An example of this is a multiple-bias modelling approach outlined by Greenland [G6], which incorporates ideas from Bayesian inference. However, the underlying methodology requires a high degree of statistical sophistication to be properly implemented, and concerns have been raised about its applicability.

57. Special care is warranted when making inferences from small underpowered studies. In such studies, there is a high probability that even sizeable radiogenic health effects will not be detected, and negative results are at best weak indicators of null or negligible effects. Risk estimates associated with positive findings from small studies can also be misleading. Such estimates are likely to be gross overestimates of risk, as explained by the following example adapted from a recent NCRP report [N3]: suppose there are several underpowered studies, for which the expected ERR associated with the mean dose is about 0.2. However, suppose for each study, a statistically significant finding would require an estimate of ERR of at least 2. If there are enough such studies, one is likely to yield a significant result, based on an estimate of ERR which is at least 10 times greater than the “true” value.

58. Even large studies are often underpowered for making inferences about risks for specific categories of health effects or population subgroups. For example, a recent analysis by Ozasa et al. [O1] of solid cancer mortality in the Life Span Study (LSS) of the Japanese atomic bombing survivors yielded an estimate of 0.42 with 95% confidence interval for the sex-averaged ERR per Gy of (0.32, 0.53) at age 70 years and age at exposure 30 years. For specific cancers, random error is greater, and confidence intervals are noticeably wider, for example (0.21, 3.8) for bladder cancer, for which a relatively small number of 183 deaths had been observed and random error is especially large. For this cancer, the LSS is effectively a small study and the central (maximum likelihood) estimate of ERR per Gy of 1.5 is potentially a large overestimate of the "true" ERR, in addition to the lack of precision of the estimate noted above. Sample size requirements are, in general, even greater for making reliable inferences on how site-specific risks depend on effect modifiers such as age, sex and other factors such as tobacco use.

59. Measurement error in exposure assessment is an important source of both random and systematic error for estimates of dose and ultimately excess risk. Such error stems from both inaccuracies in measures of the specific exposure and doses received from other sources (e.g. environmental and medical exposure). Measurement error is often classified as differential (varying between groups) or non-differential (not varying), shared or unshared, and Berkson-type or classical. For some types of non-differential measurement error, standard methods for testing hypotheses of no effect are still valid and methods for adjusting for bias in estimates of excess risk are well established. However, differential dosimetry error is a major concern in epidemiological studies. While it may be possible to assess whether dose estimates have been underestimated—or overestimated—as a consequence of differential errors, quantifying the impact on the dose response is almost impossible without a detailed understanding of the structure and magnitude of such errors [U7]. Comprehensive data on all sources of radiation exposure is rarely available for epidemiological studies. Ignoring some sources of exposure causes non-differential misclassification if it is not correlated with the exposure of interest, or differential misclassification and bias if it differs between groups that are compared (e.g. across dose levels). The impact of dosimetry uncertainties on dose-response analyses is discussed by Gilbert [G2].

60. Some studies include complex correction approaches for non-differential measurement error that build on plausible assumptions about the distribution of the error. In the pooled European study of lung cancer and indoor radon [D2], such an approach was applied, leading to a substantial increase in the estimated lung cancer risk per 100 Bq m^{-3}. Medical doses were not included in the exposure estimation in this study but they were not likely to play a substantial role for lung doses. As described in annex B, studies in high background natural radiation areas such as the Karunagappally study [N2] do not include any information on medical exposure, and it is not easy to determine if this leads to differential or non-differential misclassification of exposure.

61. Assessment of outcome is also prone to measurement error, as in cohort studies there is typically some loss to follow-up (information on possible outcomes is not available due to emigration or other factors). Censoring (exit from follow-up) also occurs through death or withdrawal from study by persons originally included (loss to follow-up). Informative censoring indicates death (the deceased are no longer at risk of outcome), while non-informative censoring refers to loss to follow-up since the subjects remain at risk of the outcome but information on its occurrence is not available anymore. If the exposure studied is associated with the risk of other causes of death than the outcome of interest, such competing risks may need to be taken into account in the analysis. Misclassification of the type of outcome (exact diagnosis or cause of death) also has effects on study results similar to measurement error. As mentioned earlier, such measurement error adds to uncertainty in the results, as do missing data on the outcomes, exposure or confounder. Selective loss from follow-up (attrition) or selective participation (e.g. in medical examination or autopsy for establishing a diagnosis) can also cause bias.

62. Model uncertainty refers to the fact that models are approximations to reality based on assumptions on how dose and other factors impact risk. In general, model uncertainty can be reduced in well-designed studies with sufficiently rich information to evaluate underlying assumptions. For analysis of dose response, the exposure distribution is crucial: the dose response can be estimated only within the dose range covered with sufficient amount of observations. More information can be obtained from studies with a wide and even dose distribution (with subjects and cases along a range of exposure, though typically most study populations are skewed towards low exposure). An even distribution will facilitate, for example, evaluation as to whether a linear or linear-quadratic function provides an adequate description of the true dose response. Similarly, model assumptions for effect modifiers such as age-at-exposure can be evaluated only within ranges sufficiently covered by the data. Errors in the dose measurement (or other independent variables) also affect the modelling and can lead to biased results or confidence intervals that are too narrow. Model uncertainty is, in general, difficult to quantify although methods such as multimodel inference* have been developed [U7].

63. Evaluations of uncertainties should consider how results from individual studies might be incorporated within the review. For example, pitfalls in making inferences from a small study (with large sampling errors) might be mitigated if results from the study and others can be combined through a meta-analysis of published data from different studies or if original data from the study and others could be combined for a pooled analysis* (see also the section on synthesis of studies).

64. *Reporting*. While a study may have been conducted with great accuracy and care, the written report may not necessarily be of the same quality due to using selective reports or omitting details of the study plan and procedures). Reporting standards for observational studies in epidemiology (STROBE), systematic reviews (PRISMA) and non-randomized evaluations of interventions (TREND) provide guidelines for this stage of research [E8]. Most of the 22 items of the STROBE checklist are common to cohort, case-control and cross-sectional studies, while some are study design-specific. Improved study reporting following STROBE can facilitate critical appraisal and interpretation of studies.

65. The key items in STROBE include proper description of the setting, eligibility criteria, defining exposure, outcome and confounding variables, sources of information and methods of assessment, and also description of statistical methods and ways the investigators dealt with any missing information. Furthermore, the study subjects (including non-participants) and their key characteristics, and also numbers of outcome events should be described. The results section should cover the effects of adjustments and any subgroup and sensitivity analyses besides the main results. In the interpretation of the findings, the limitations of the study (such as biases and imprecision), and also generalizability and previous findings should be adequately reflected.

66. Similarly, PRISMA guidelines support transparent and comprehensive reporting of systematic reviews and meta-analyses. While the focus of PRISMA is on healthcare interventions, the checklist items can be applied with minor adaptation to systematic reviews of observational epidemiological studies [M5].

67. TREND guidelines promote transparent reporting of evaluations of behavioural and public health intervention studies with non-randomized designs. Similar to STROBE, the guideline includes 22 items in a checklist for standardized reporting [D6].

68. Beyond reporting, separate guidance has been developed to focus on the methodological quality of systematic reviews. For example, the AMSTAR checklist consists of a valid and reliable eleven-item scale for methodological quality assessment of systematic reviews [S6, S7].

IV. PROCESS OF SYNTHESIS OF RESEARCH RESULTS

69. For radiation risk assessment (as for other risks), epidemiological studies are instrumental as they provide the most directly applicable evidence, i.e. quantitative evidence on frequencies and determinants of disease and effects of interventions for human populations in real-life conditions. For traditional narrative reviews, no rigorous methodological standards have been established. They aim to provide assessment of weight or strength of evidence, i.e. judgement of value or contribution of studies commonly reporting (seemingly) contradictory results. However, the degree of credibility assigned to each study and formulation of summary measures does not tend to follow formal procedures. Hence, the appraisal of the balance of evidence is heuristic[3] and generally based on ad hoc criteria. This methodological "softness" makes such reviews susceptible to judgemental bias, however not intentionally.

70. For literature synthesis and evidence assessment, which are both essential components for the Committee's evaluation of epidemiological studies, appraising their quality of evidence is a key task. Two levels of quality are to be considered. The first is the quality of any individual study used for the topic at hand. As outlined in chapter III, several core aspects affect the quality of observational studies in radiation epidemiology: notably appropriate measurement of exposure and of outcomes, confounder control, and other design and analysis issues. Quality limitations influence the risk of bias, which may differ for the various outcomes considered. Guidelines for assessing risk of bias have been developed, focusing mainly on randomized controlled studies on treatment and diagnosis but also applicable to some extent for other study types [H4]. Furthermore, tools and guidelines to assess non-randomized studies are available elsewhere [M7, S11]. So far, there is limited consensus on the elements and details of quality assessment guidelines for observational epidemiological studies. Nevertheless, a common set of topics requiring consideration has emerged, and these are also applicable—with some modifications—to radiation epidemiology (see section IV.C).

71. The next step in the assessment of quality pertains to the totality of the evidence, i.e. all studies used for reporting on the topic of interest for the respective evaluation by the Committee. In the overall assessment of the available evidence, low-quality studies should be assigned less weight than high-quality studies, or even excluded if the quality limitations are critical. A qualitative categorization of quality is preferred over quantitative and semi-quantitative numerical scoring approaches because the relative weighting of bias related to individual quality items requires questionable assumptions [G8].

72. For a scientific evaluation of a defined research topic according to established scientific procedures and values (thus following Governing Principles[2] of the Committee), a systematic approach is recommended, including the following steps:

- Step 1: Transparent and systematic collection of information, based on a protocol;
- Step 2: Abstraction of relevant data from selected studies or other sources of information;
- Step 3: Assessment of individual study quality following unambiguous and consistent standards;
- Step 4: Synthesis of information;
- Step 5: Drawing of conclusions.

[3] Any approach to problem solving that uses practical methods not guaranteed to be optimal or perfect, but sufficient for the immediate goals. It is not formally derived.

73. This structured approach is being used in high-quality systematic reviews of radiation epidemiological studies. For example, the review by Little et al. [L5, L7] aimed to investigate a possible causal association between low-level ionizing radiation exposure and circulatory disease in a general unselected population. Clear descriptions of the information search and abstraction of data, quality assessment of included studies using clearly defined criteria and qualitative and quantitative synthesis in several meta-analyses led to evidence based conclusions provided in the review.

74. Well-conducted and reported studies should always form the main input into a scientific report (otherwise no informative conclusions can be drawn). Lower quality studies can and should be identified and evaluated with a transparent decision either to exclude them or to include them with a lower weight assigned to their results. Conclusions and summary quantitative risk assessments should rest primarily on the best quality studies. The low quality studies can assessed as to whether their results are generally consistent with or strongly deviate from the summary results.

75. The different steps in the process of research synthesis, including the definition of the study question (topic), systematic evidence search, study identification, quality assessment, synthesis and derivation of conclusions have been described [C5, H4, W2]. The following sections give an overview of the overall process.

A. Definition of the topic for which evidence is required

76. A crucial issue at the planning stage of a scientific evaluation is a clear and precise definition of the question or topic the evaluation seeks to cover. The scope of the planned work should be clearly outlined in a document plan, ideally with precise and answerable study questions organized using the population, intervention (exposure), comparison, and outcome (PICO or PECO) framework. For a review of epidemiological studies on cardiovascular diseases, such a question could be: is low dose and low-dose-rate exposure to ionizing radiation associated with increases in incidence and mortality from cardiovascular diseases in human populations? A protocol giving a clear definition and description of exposure, including dose levels of interest and further exposure details, should be produced. Likewise, the outcomes of interest need to be specified in advance, and eligible study designs and outcome measures outlined. Identification of relevant confounding factors that should be addressed in studies is also important. For example, in studies on ionizing radiation and cardiovascular risk, classical cardiovascular risk factors such as smoking, hyperlipidaemia and diabetes should be included as potential confounders, and confounder-adjusted regression analyses conducted. This preparatory work should result in a thorough and detailed work plan. Changes to the protocol and to the agreed scope/direction of the report should be avoided. Any necessary changes should be clearly documented and endorsed by the Committee or the scientific experts charged with an evaluation by the Committee.

B. Collection of information: searching the literature

77. Systematic reviews address a clearly defined objective, use transparent and reproducible methodology and include a systematic literature search for all studies that meet predefined eligibility criteria. The systematic literature search results in a set of research reports consequently used for the overall evidence assessment and research synthesis.

78. The substantial growth of new journals, particularly those with an open-access model, over recent years has changed the scientific publishing landscape. There are many (online) journals with non-transparent quality control measures that nevertheless claim to be peer-reviewed and reputable. This development makes research synthesis more difficult and calls for even more emphasis on strict quality assessment of publications that are to be included in evidence synthesis.

79. There are many ways to find relevant literature. A systematic search of at least two electronic databases for specified key words provides a way to ensure completeness of information and extend the knowledge base, minimizing the possibility of selective inclusion of research reports. Nowadays, several databases of scientific literature are used, such as Medline/PubMed, Scopus or EMBASE. These can provide a good coverage of the published literature, but augmentation by expert databases and expert knowledge of the respective literature is often warranted particularly for very specific topics that may not be identified through keyword searches. A standard procedure is to also cover the reference lists of included publications and of related research overviews. Through its revision and discussion process, the Committee additionally ensures that further literature contained only in expert databases can be screened for relevance and inclusion. Unpublished research ("grey literature") is more difficult to find, and while EBM systematic reviews nowadays do include unpublished research, the Committee usually does not use unpublished literature. However, reference lists of unpublished work may hint to further published research, and at times such work may be useful for hypothesis generation or for highlighting previously unconsidered facts. Even if English is the predominant language in scientific publishing, studies published in other languages can also contribute to the knowledge base and should, as a rule, be identified and evaluated. Expert groups often comprise members from several different countries and can thus potentially provide for expertise in different languages.

80. For searches in electronic databases, a translation of the review topic into the language of the database (e.g. in terms of specific keywords) is required, and also a clear and concise use of search terms for the population(s) and exposure(s) of interest, for the comparisons (e.g. disease rates according to different dose levels) and the outcomes (e.g. specific cancer types, ischaemic heart disease, lens cataract). Searches may be limited to particular designs or publication types such as cohort studies, using index terms. Spelling variants and natural language variants of text terms should be considered [E7]. Overlap between publications because of several publications deriving from the same material needs to be identified and taken into account.

81. A documentation of the literature search and selection process is needed so that the search can be repeated or updated at a later stage. Methods used for retrieval of information and data in the report should be described, including evidence abstraction (e.g. was it based on titles and abstracts, review of full texts?), criteria for inclusion, and quality assessment. This part is not necessarily to be published in an UNSCEAR report, but should be made available to reviewers upon request and archived at the secretariat. An information flow chart is recommended by PRISMA [M5] as a useful tool to provide an overview of the process, which could be included in an UNSCEAR report, as supplementary information or attachments. An example of a flow diagram from a systematic review on cataracts and low doses of ionizing radiation using an adaptation of the PRISMA template is given below in figures II and III [H1]. The Committee has applied this approach—with some modification—in its evaluation on the Fukushima accident and its follow-up white papers [U5, U6, U8]. In general, updates of systematic reviews may be necessary when new evidence and new methods become available. Frameworks for decisions on updating systematic reviews are available [G1].

Figure II. Information flow chart for systematic reviews [M5]

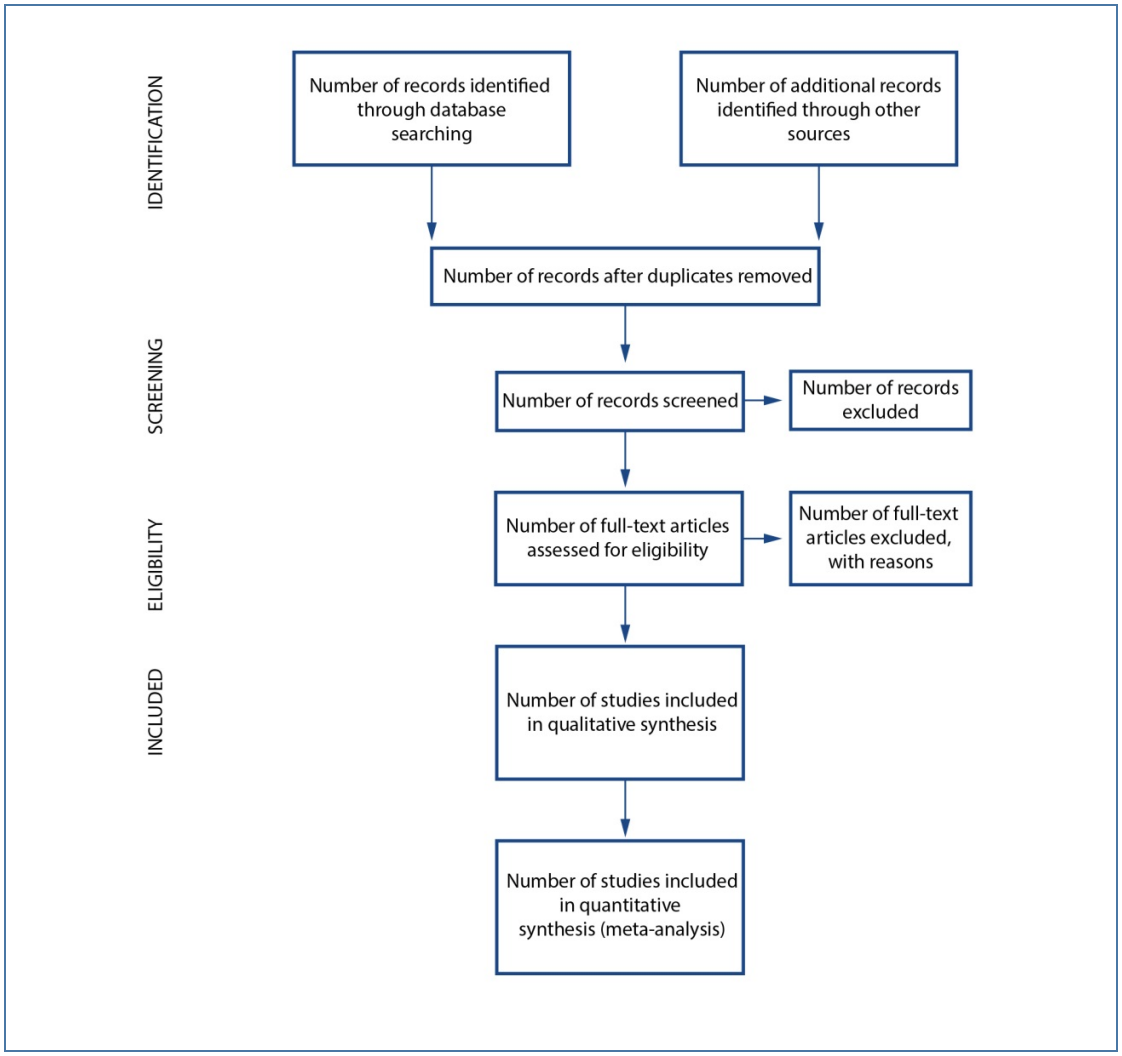
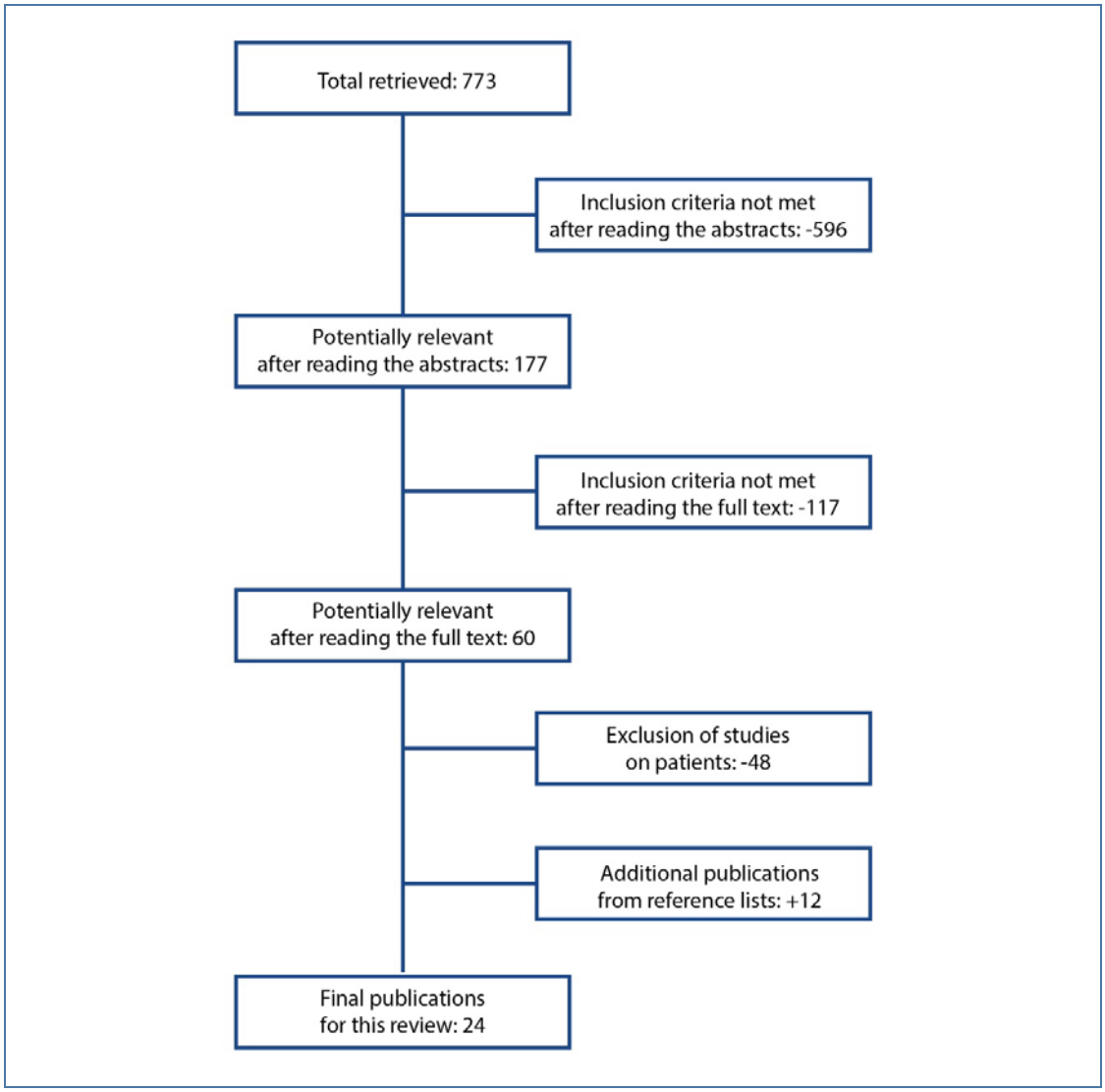


Figure III. Flow chart based on PRISMA from a systematic review on cataract development and low doses of ionizing radiation [H1]



C. Assessment of study quality

82. Once the studies to be included have been identified, core features are abstracted for further assessment, for example in tabular form as presented in table 2. These features usually include study identifier (e.g. author, year), study design, characteristics of the population studied and duration of follow-up (for cohort studies), type of exposure, exposure assessment methods, dose distribution, core results such as ERR/unit dose or EAR/unit dose per outcome (with confidence interval), and main confounding variables included in the analysis. Further comments or details relevant to the respective study can be added. Next, study quality and relevance are assessed. Study quality assessment aims to provide a transparent judgement on the risk of bias and other methodological issues of the study at hand. Study quality assessment forms an integral part of evidence synthesis, and must be performed in a transparent and reproducible fashion. There are numerous tools available to assess individual study quality, and they are frequently designed for specific study types [S2].

83. Risk of bias or quality assessments are focused on the methodological quality of the study, and not on issues such as magnitude of effect or applicability of results. It should be noted that study size or statistical power to identify effects (if existing) are not included in the quality assessment, but need to be considered in the overall assessment of the evidence. Several core domains addressed in quality assessment have been described while for the Committee's work, the following domains should be addressed:

(a) Method of selection of study participants;

(b) Methods for assessing exposure;

(c) Methods for measuring outcome;

(d) Sources of bias specific to the study design;

(e) Methods to control for confounding;

(f) Statistical methods;

(g) Study reporting;

(h) Statement on conflict of interest of authors.

84. For the quality assessment, these domains are reviewed and appraised individually. The appraisal results in a domain being judged as having low, moderate, serious or critical risk of bias (or no information for judging risk of bias). A critical risk of bias judgement would be evident if a study is considered too problematic in the respective domain to provide any useful information. From the judgement on bias risk in the individual domains, an overall risk of bias—the overall quality assessment for the specific study—is derived, using the same categories. As indicated earlier, there is a critical discussion on the use of numerical summary scores as these involve some weighting of components and this may be not justified or applicable across studies [S2].

85. For the appraisal of these domains in radiation epidemiological studies, the following quality issues—framed as questions—are of particular importance, as explained in chapter III. To provide linkage to concepts used by the Cochrane Collaboration [H4], biases addressed using the Cochrane terminology are given in brackets. It should be noted, however, that each study requires its own critical assessment, taking account of specific study objectives, design and methods, and available data:

(a) Study participants (selection bias)
- For cohort studies: was the selection of participants unrelated to outcome? Do start of follow-up and start of exposure coincide for most study subjects?
- For case-control studies: was the selection of cases and controls unrelated to exposure?
- All study types: is there a clear description of inclusion and exclusion criteria? Is the degree of participation (response proportion) specified? How high is it?
- All study types: are the comparison groups appropriate?

(b) Exposure (performance bias)
- Are individual organ doses available and used for risk estimation?
- How extensive is the problem of missing exposure data? Were reasonable steps taken to address this problem?
- Were temporal changes in technology, frequency of measurement, and reporting procedures of yearly doses considered while preparing the dose estimates—if applicable?
- Were dose estimation uncertainties taken into account appropriately? Are the uncertainties substantial? How much could they affect the results?

(c) Outcomes (detection bias)
- Was there an objective measure of outcome, such as mortality, cancer incidence from registries or reliable medical sources?
- How were changes in disease classification (ICD coding) handled over time?
- Are the outcomes appropriately justified and selected, i.e. specific entities rather than large groupings of outcomes that provide only averaged overviews?
- Were methods of outcome assessment identical across all exposure groups? Was the same effort expended on case identification in (suspected) high versus low or non-exposed groups?
- Were there any systematic outcome measurement errors?

(d) Design-specific bias (attrition bias, other biases)
- Cohort study: is the follow-up reasonably complete? Is follow-up related to exposure or outcome?
- Case-control study: does case recruitment focus on incident cases? Do controls represent the reference population, i.e. the source population from which the cases arise? Is there potential for recall bias? Is the study nested in a cohort?

(e) Confounder control (other biases)
- Are all important confounders assessed in the study? For example, in cancer studies are other carcinogens considered that can be associated with radiation exposure. However, some confounders may not be known.
- How likely is residual confounding?

(f) Statistical methods (other biases)
- Was there a pre-specified main analysis?
- Are the statistical methods appropriate for the available data (e.g. regression models for cohort (longitudinal) data or conditional logistic regression for matched case-control data)?
- Are confounder-adjusted or stratified analyses available?
- Does the study provide dose-response analyses and measures such as ERR or EAR per unit dose?
- Are sensitivity analyses and analyses incorporating uncertainty available?

(g) Reporting (other biases)
- Is the reporting of the study complete and unbiased, as assessed for example via STROBE guidelines?
- Are the reported results unlikely to be selected from different multiple outcome measures, multiple subgroups, multiple analyses?

(h) Conflict of interest
- Have all contributors provided conflict of interest statements in line with international recommendations?[4]
- Is any potential influence of conflict of interest appropriately dealt with?

86. The Cochrane Collaboration provides a risk of bias tool for randomized clinical trials [H4] and a separate tool for non-randomized studies (ACROBAT-NRSI, now called ROBINS-I) [S11]. This later tool is being developed to provide a consistent approach for assessing risk of bias in non-randomized studies of interventions. It can also be applied to observational studies of exposure-disease associations [S11] including radiation epidemiological studies. However, given the specific issues that need to be

[4] For example: International Committee of Medical Journal Editors (www.icjme.org).

taken account of in these studies, adapted tools and approaches are useful in this context. Therefore, the Committee recommends working with the study domain approach for quality assessment as outlined here, individually addressing the domains *(a)–(h)* (see above).

87. A risk of bias assessment of radiation epidemiological studies can also be based on quality assessment instruments, tools or approaches such as Newcastle-Ottawa Scale but they generally cover similar aspects as the ROBINS-I and the related approach applied by the Committee. Generally, assessment of bias is an area of ongoing development. Monitoring of initiatives such as the Cochrane Collaboration for assessment of the risk of bias in non-randomized studies of exposure (ROBINS-E) is recommended. A practical way to provide an overview of study quality is the use of study-specific tables noting strengths and limitations, as applied by the Committee [U3]. These tables can be organized according to the general domains described above, noting all issues that contribute to low risk of bias (i.e. high quality = strength) and moderate or (very) serious risk of bias (moderate to very low quality = limitation) [U3]. Annex B on epidemiological studies of cancer risk due to low-dose-rate radiation from environmental sources provides an overview of study quality arranged according to the study-specific domains. The specific aspects of study quality in radiation epidemiology are also summarized in table 3.

88. In addition, a setting-specific description of key quality features is provided. Some of these features for setting-specific studies can be used for the risk of bias assessment as described above. The overall assessment of study quality is then based on the quality assessment across domains, by using the quality categories:

- High;
- Moderate;
- Low;
- Very low.

89. Observational epidemiological studies can be of high quality if no or very few limitations, none of them serious, are noted in the quality assessment. Vice versa, if there are serious or very serious quality limitations in one or more domains, the overall study quality is judged as low or very low. Most studies will fall in between. A transparent domain-specific assessment and overall judgement allows for discussion and justification of further (non-) use of the study and its results.

90. The result of the quality assessment for each study is documented, and the review should then describe how the quality rating approach is used in the further steps of the review. There is little use in assessing quality without utilizing its results. Studies with critical risk of bias or no information on risk of bias from the evidence synthesis should generally be excluded. This also pertains to studies with seriously inadequate dosimetry or outcome assessment. Sensitivity analyses in quantitative meta-analyses are used to investigate results in pre-specified subgroups of studies, for example those with a similar design or those with high quality. The results of these sensitivity analyses can provide a perspective on the interpretation of the main results whether results are robust across carefully selected subgroups.

91. Applying the approach recommended by the Committee will ensure that research overviews will be more informative as they identify and summarize the best and most relevant studies on a given topic and put less or no weight on flawed or non-relevant investigations. Providing transparency in quality assessment as outlined here will form a good scientific basis for the Committee’s discussions on the merits of study-specific scientific evidence used for the Committee’s evaluations.

Table 2. Example of an extraction table of epidemiological studies dealing with diagnostic exposure (based on table 15, annex A [U3])

Study	*Type of study, end point*	*Population studied*		*Follow-up (years)*	*Total person-years*[a]	*Type of exposure*	*Type of dosimetry*	*Cancers studied*
		Characteristics	*National origin*					
EXTERNAL HIGH-DOSE-RATE EXPOSURE								
Diagnostic examinations								
TB fluoroscopy: Massachusetts [B2, S8]	Cohort, incidence	2 367 exposed women 2 427 unexposed women Age: 12–50 (26)	United States	0–>50	54 609 (11.4)	Multiple X-ray chest fluoroscopies	Individual exposure from medical records and doses from phantom measurements and computer simulations	Breast, skin
TB fluoroscopy: Massachusetts [D4]	Cohort, mortality	6 285 exposed persons 7 100 unexposed persons 49% females Age: 12–50 (26)	United States	0–>50	331 206 (24.7)	Multiple X-ray chest fluoroscopies	Individual exposure from medical records and doses from phantom measurements and computer simulations	Breast, oesophagus, lung, leukaemia
TB fluoroscopy [H6, H7]	Cohort, mortality	25 007 exposed persons 39 165 unexposed persons 50% females Age: <20–>35 (28)	Canada	0–57	1 608 491 (25.1)	Multiple X-ray chest fluoroscopies	Individual exposure from medical records and doses from phantom measurements	Lung, breast
Diagnostic X-rays (U.S. health plans) [B1]	Case-control 565 leukaemia 318 non-Hodgkin's lymphoma 208 multiple myeloma 1 390 controls	2 203 exposed persons 278 unexposed persons 39% females Age: 15–>50	United States	n.a.	n.a.	Diagnostic X-rays	Average dose based on number and type of procedures and estimated doses from published literature	Leukaemia, non-Hodgkin's lymphoma, multiple myeloma
Medical and dental X-rays: Los Angeles [P3]	Case-control 408 cases 408 controls	62% females	United States	2–64	n.a.	Medical and dental diagnostic X-rays	Average dose based on number and type of procedures and estimated doses from published literature	Parotid gland
Diagnostic X-rays: Los Angeles [P4]	Case-control 130 cases 130 controls	39% females	United States	3–20	n.a.	Diagnostic X-rays	Average dose based on number and type of procedures and estimated doses from published literature	Chronic myeloid leukaemia

Study	*Type of study, end point*	*Population studied*		*Follow-up (years)*	*Total person-years*[a]	*Type of exposure*	*Type of dosimetry*	*Cancers studied*
		Characteristics	*National origin*					
Diagnostic X-rays: Sweden [I2]	Case-control 484 cases 484 controls	736 exposed persons 232 unexposed persons 77% females Age: <20–>60	Sweden	5–>50	n.a.	Diagnostic X-rays	Average dose based on number and type of procedures and estimated doses from published literature	Thyroid
Scoliosis (U.S. Scoliosis Cohort Study) [D7]	Cohort, mortality	4 822 exposed women 644 unexposed women Age: <3–≥10 (10.6)	United States	3–>60	218 976 (40.1)	Diagnostic X-rays	Average dose based on number of treatments and estimated doses from published literature	Breast

[a] *Mean per person in parentheses.*

Table 3. Overview of specific aspects of study quality in radiation epidemiology, based on selected domains

Positive	*Negative/Questionable*
1. STUDY POPULATION/PARTICIPANTS	
Individual level data	Only aggregate level data (ecological or geographical/temporal correlation study)
Comprehensive identification of the subjects	Poorly defined study population
Wide range of ages at exposure	Age at exposure not specified, or very narrow age-range (unless this age group is the target group, e.g. children)
In case-control studies, controls representative of exposure in source population	Inadequately-defined or absent control group; in case-control studies hospital-based controls with more potential for bias
In cohort studies, non-exposed reference group representative of disease risk in absence of exposure	Inadequately-defined, unsuitable or absent control group
In cohort studies, completeness of participant selection (or the selection sampling frame) comparable across the dose range	Indications of differential completeness of participant selection according to level of dose (e.g. higher completeness in high dose/exposure groups)
2. EXPOSURE	
Individual dosimetry, assessed for accuracy	Exposure indicator not organ dose, but indirect proxy indicator, e.g. in diagnostic exposure administered amount of activity, number of examinations
Organ doses calculated	Inaccuracy in dose estimation due to, e.g. lack of detailed information
Exposure assessment comprehensive, includes all relevant radiation exposure sources	Assessment does not cover all sources or periods of radiation exposure
Detailed dose data, e.g. for relevant organ or tissue compartment such as red bone marrow for leukaemia	Self-reported exposure data, e.g. number of X-ray examinations, food consumption data for estimating intake
Wide range of doses	Narrow and crude dose range description, e.g. high in radiotherapy, low in environmental exposure
Detailed documentation of exposure features	Only mean level of exposure assessed (e.g. from a sample of the exposed subjects)
Comprehensive, individual dose monitoring, with proper assessment and treatment of potential deficiencies	Retrospective dose reconstruction with uncertain data
Other relevant exposure sources considered, e.g. medical radiation	No or unreliable information of other radiation exposure
3. OUTCOME	
Comprehensive and uniform disease ascertainment not varying by exposure, e.g. through high-quality population-based cancer registry	Self-reported outcome through active follow-up, without verification; different outcome assessment for different exposure groups
Pathological review of cases for diagnostic verification	Abstraction of data from non-comprehensive medical records
Disease incidence, plus mortality	Mortality only
Large number of cases with specific disease	Unclear or heterogeneous diagnostic criteria

Positive	*Negative/Questionable*
Long-term follow-up	Potential for confounding by indication or reverse causality in diagnostic medical exposure
High percentage of histologically verified cancer diagnoses	Disease ascertainment or screening varies by exposure
4. BIAS AND 5. CONFOUNDING	
Detailed information on confounders	Missing data on key confounders, e.g. smoking or other behavioural factors
Attempts made to indirectly assess confounding potential, e.g. through adjustments for sociodemographic variables, job type, duration of employment	Indications of uncontrolled or residual confounding
Low likelihood of important unknown confounders	Exposure associated with other disease risk factors, e.g. both radiation and chemotherapy in studies of second cancers, co-exposure to chemical agents in occupational studies
Risk factors similar across exposure levels	Potential for recall and other information bias in exposure assessment
Systematic and unbiased exposure assessment	Low/differential participation in groups being compared
Selection bias minimized	No consideration of potential selection bias, inclusion of (highly) selected study participants; effect of migration not considered
6. STATISTICAL METHODS/ANALYSIS	
Sample size/power calculations presented and discussed	No consideration of study size and power
Adjustment for confounders	Comparison to disease occurrence in general population only (SMR, SIR)
Attempts to indirectly assess confounding potential, e.g. through adjustments for sociodemographic variables, job-type, duration of employment	No or very limited assessment potential for confounding through indirect approaches
Analysis of several dose-response models, including linear, linear-quadratic and quadratic, and non-parametric dose-response curves	Analysis limited to one model, or no dose-response considerations; insufficient number of categories to allow dose-response assessment
Use of pre-specified cut-points in Poisson regression	Opportunistic or post-hoc cut points
Sensitivity analyses	No sensitivity analyses reported
Pre-specified analysis plan	Unclear if analysis plan existed
7. REPORTING	
Complete, clear and impartial study reporting	Incomplete report, report outside range of reputable peer-reviewed journals
Conscientious search for unidentified bias or confounding, non-selective outcome reporting	Suspicion of selective reporting
Adherence to reporting guidelines, e.g. STROBE	Omission of important sections or issues in the report
8. CONFLICT OF INTEREST	
Statement is included	No statement
No conflict is reported	Significant conflict reported or likely

D. Synthesis of studies: meta-analysis and narrative approaches

92. The subsequent step of the literature review process is the actual study synthesis and interpretation. Narrative synthesis involves description and summary of the core findings from the included studies of varying structure and method. This involves grouping of the studies, for example according to study design, population or outcome, and providing a verbal account of the body of evidence relevant for the review question. On the basis of focused description of the individual studies, the evidence across the different studies is summarized and jointly assessed. When few studies are available or are very diverse in terms of important characteristics, a narrative review may be more appropriate than meta-analysis of data that are not truly compatible. An example of a table describing study results in a narrative approach is given in table 4. The issues described here correspond closely to those described in table 2.

93. An important concern is the potential for publication bias, i.e. the selective reporting of studies, often those with statistically significant findings. Publication bias can seriously affect the synthesis of evidence, and several methods to detect and evaluate publication bias have been developed for use in meta-analyses [E3, J3].

94. A meta-analysis involves a formal statistical analysis of pooled results from at least two studies, yielding a joint effect estimate for the outcome. There are numerous meta-analytic techniques, and the approach chosen depends on aspects such as the study designs and the outcomes/effects to be analysed. Guidance can be found in the Cochrane handbook for systematic reviews [H4], and in textbooks for example by Borenstein et al. and Egger et al. [B3, E4]. Briefly, the minimum requirement for meta-analysis is abstraction of point estimates for the effect measures (e.g. rate ratios or hazard ratios) and their confidence intervals from each study. Study design and procedures can also contribute useful information. In pooling the results for meta-analysis, the summary estimate is generally obtained as weighted average, with inverse variance weighting. This technique assigns more weight to studies with more information, as indicated by the standard errors or confidence intervals of the effect estimates. It may be useful to examine the weight of studies as only a small number of studies may drive the overall result. Influential studies can be identified in sensitivity analyses that omit single or several studies [H4]. Meta-analyses of individual study data require that researchers obtain raw data from all studies to be included, which are subsequently analysed jointly. This approach (also called pooled analysis) allows more analytic consistency and provides flexibility for the analysis (e.g. in studying dose response across the combined data set). It is best performed as a prospectively planned pooled analysis, since retrospectively obtaining study data from many researchers may be cumbersome.

95. Assessment of consistency of the results is needed (analysis of heterogeneity). If the consistency of results is adequate, the studies can be taken to estimate a single entity i.e. their estimates are compatible and can be interpreted in a similar fashion. Alternatively, significant differences between study results indicate that their study setting, participants, measurements or other methodological issues differ to the extent that they may affect the findings. The factors underlying the differences should be explored to assess what causes the heterogeneity. Two types of statistical methods can be applied in the analysis. In the absence of major heterogeneity, a fixed effect model can be used, while heterogeneous study results require application of a random effect model. Fixed effect models are appropriate when a common effect estimate is assumed across the different studies. A random effect model is applicable when the effect estimate is considered to vary across studies, i.e. there is statistical heterogeneity of the effect estimate [B3], which is usually assessed by the Der Simonian & Laird Q statistic and the I^2 – statistic quantifying the percentage of variation across the included studies that is due to heterogeneity

rather than chance. Both statistics have limitations and alternatives have been proposed [C4, C6]. Reports on meta-analyses of several outcomes may include both models, depending on the heterogeneity between studies on the respective outcomes. For example, Little et al. in their meta-analysis of circulatory disease and exposure to low-level ionizing radiation applied both fixed and random-effect models for the computation of common effect estimates [L5].

96. The advantage of the joint estimate is an increased precision of the central estimate, i.e. improved stability and reduced random error by virtue of combining data from several studies. So far, meta-analyses have rarely been conducted specifically for the Committee's evaluations. This approach requires specific statistical and methodological expertise, and is not feasible or recommended in cases of substantial heterogeneity between studies in terms of scope (exposure or outcome) or procedures, or for studies of low quality.

Table 4. Example of narrative approach to describe strengths and limitations of studies dealing with diagnostic exposure (based on table 17 of annex A [U3])

Study	*Strengths*	*Limitations*
EXTERNAL HIGH-DOSE-RATE EXPOSURE		
Diagnostic examinations		
TB fluoroscopy: Massachusetts [B2, D4, S8]	Incidence study with long-term follow-up (50 years) Individual dosimetry based on patient records and measurements Unexposed TB patients comparison group Fractionated exposure occurred over many years Dose–response analyses	Uncertainty in dose estimates related to fluoroscopic exposure time and patient orientation Questionnaire response probably under ascertained cancers Debilitating effect of TB may have modified radiation effect for some sites, e.g. lung
Diagnostic X-rays (U.S. health plans) [B1]	Information on diagnostic X-rays abstracted from medical records Surveillance bias unlikely, since cases and controls were at equal risk for having X-ray procedures recorded and malignancy diagnosed	Potential for ascertainment bias, e.g. through early diagnosis of a malignancy Analyses based on number of X-ray procedures rather than actual doses
TB fluoroscopy: Canada [H6, H7]	Large number of patients Unexposed TB patients comparison group Individual dosimetry for lung and female breast Fractionated exposure occurred over many years Dose–response analyses	Mortality limits comparisons with breast cancer incidence series, e.g. time response Uncertainties in dosimetry limit precise quantification of risk Different dose responses for female breast cancer between one sanatorium and the rest of Canada may indicate errors in dosimetry, differential ascertainment or differences in biological response
Diagnostic medical and dental X-rays: Los Angeles [P3, P4]	Dosimetry attempted on the basis of number and type of examinations	No available records of X-rays Potential for recall bias in dose assessment Doses likely to have been underestimated
Diagnostic X-rays: Sweden [I2]	Information on diagnostic X-rays over many years abstracted from medical records	Analyses based on number and type of X-ray procedures rather than actual doses
Scoliosis (U.S. Scoliosis Cohort Study) [D7]	Adolescence possibly a vulnerable age for exposure Dosimetry undertaken on the basis of number of films and breast exposure Dose–response analysis	Comparison with general population potentially misleading, since scoliosis associated with several breast cancer risk factors (e.g. nulliparity) Dose estimates may be subject to bias and to random error

E. Reaching an overall conclusion from the evidence synthesis

97. The final part of the evaluation of epidemiological studies should provide insight into the confidence of the overall conclusions and risk estimates obtained. Previously, a hierarchy of evidence (or evidence pyramid) was used by scientists working in EBM to provide a simplified approach to classify evidence quality according to study type. Meta-analyses of several well-conducted randomized controlled trials were considered by the scientific EBM community to provide the best empirical evidence. At the bottom of the hierarchy, expert opinions and case series were located. This pyramid was primarily geared towards evidence from therapies and other interventions.

98. Nowadays, a more articulated approach with wider assessment of various aspects of evidence has been developed within the Grading of Recommendations Assessment, Development and Evaluation (GRADE) collaboration for the Cochrane Library and has also been adopted by some international institutions involved in guideline development, including the World Health Organization [W2]. The GRADE categorization [G4] is usually specific to an outcome, for example cancer site-specific incidence. GRADE methodology provides a formalized way to categorize the quality of evidence for this outcome, leading to a four-level quality grading as high, moderate, low or very low.

99. For GRADE, the initial categorization of the evidence considers randomized (start with: high quality) or non-randomized designs (start with: low quality). Following this, several additional issues are assessed, which can lead to either a down- or an upgrading of the overall quality. Risk of bias, inconsistency across studies (heterogeneity), imprecision (small outcome numbers, wide confidence intervals) and indirectness (i.e. the available evidence covers issues that are distant from the actual topic of interest for the review) and also suspicion of publication bias all lead to a downgrading, whereas a large effect size, a dose-response relationship and indications of a likely underestimation of the effect (plausible direction of bias) support an evidence quality upgrading. For example, this allows for the rating of evidence from well-conducted observational cohort studies on radiation-associated leukaemia risk as high quality if several strengths of the studies on this outcome are demonstrated and lead to upgrading. Details of the GRADE methodology are provided by the GRADE working group [G4] and numerous articles [G7, S5].

100. Development of GRADE, particularly its application to and development for non-clinical areas, including environmental or other exposure, is ongoing. However, GRADE can be used with some adaptations for many different study designs, notably observational evidence as typical in radiation epidemiology. For the Committee, the GRADE framework can provide orientation and guidance. However, as randomized controlled trials are generally absent from the evidence body in radiation epidemiology, the focus in the scientific field of relevance for the Committee's evaluations is in detailed and specific documentation on critical assessment of the specific aspects of studies of radiation effects, as described in section IV.C.

101. A GRADE-informed UNSCEAR evidence synthesis approach should include the following:

(a) The evidence on radiation effects is assessed per health outcome. For cancer, if the case numbers for specific cancer types are sufficiently large, it is preferable to focus on these types and less on large groupings, for example all cancers, all solid cancers; however, smaller groups of cancer based on pathophysiological similarities may be useful. For other health outcomes, reasonably non-heterogeneous outcome groups should be specified. This is consistent with the Committee's approach to provide outcome-specific assessments of radiation effects [U1, U3].

(b) The following core criteria for the synthesis are used (with associated GRADE terms in brackets):
- Strengths and limitations (risk of bias) across studies, using the quality rating for included studies as described in section IV.C;
- Heterogeneity (inconsistency) between the various studies with regard to the radiation effect assessed;
- Precision of effects and uncertainty (imprecision);
- Applicability to specific topic of interest for the Committee (indirectness);
- Publication bias.

(c) The overall evidence (e.g. concerning the presence of a radiation effect) is rated as either high, moderate or low. For cohort and case-control studies, an initial rating as moderate is used, and for geographical/temporal correlation studies, time-series and cross-sectional studies a ranking as low is the starting point.

(d) If the individual studies forming the evidence base have been found to have particular strengths and no or few limitations, the joint evidence from these studies is upgraded to moderate (for geographical/temporal correlation and cross-sectional studies) or high quality (cohort, case-control studies). Upgrading can also be considered if a large effect size is found by the available studies (or the majority of them) and if a dose response from studies with sound dosimetry can be demonstrated.

(e) Conversely, if few strengths and several, clearly serious, limitations are noted for the included studies, the joint evidence is downgraded to low quality (for those with initially moderate quality). Suspicion of publication bias, preferably supported by formal analysis of this bias type, always leads to downgrading.

102. The Committee requires a clear justification of evidence synthesis judgements for upgrading or downgrading the overall quality of evidence from the body of studies included in the assessment. It is worth noting that a joint framework—such as the one presented here, and also the GRADE approach—does not ensure consistency of conclusions on evidence, but provides for transparent and explicit judgements on the research results assessed and summarized. An explicit and transparent approach helps to reduce errors, facilitates critical review of the evidence, and improves communication of information. However, this transparency relies on careful documentation [G8].

103. Transparent organization and reporting of evidence is also an important step towards reducing differences of understanding and subsequent interpretation between expert evidence synthesis, the way the Committee uses this synthesis, and public understanding of the topic. To support transparency and clarity, standardized vocabulary and terminology with clear definitions are required. This would also help to ensure that reports can be translated accurately into other languages.

104. The Committee's draft evaluations, including syntheses of epidemiological evidence, have always been subject to close scrutiny and discussion at its annual sessions. This is an important step in continuous quality control and encompasses all steps of scientific report development, with critical revisions from experts with different scientific backgrounds. The Committee also assesses the consistency and plausibility of epidemiological evidence synthesis with results from biology and toxicology, and indicates research needs arising from the overall assessment. The identification of uncertainties regarding the respective topic and possible approaches to their reduction will usually be part of the Committee's research recommendations.

V. CONCLUSIONS

105. Individual epidemiological studies provide the basic evidence on radiation effects, and this annex provides guidance on assessing the quality of individual studies and of the synthesis of evidence from several studies. Methods of evidence synthesis have evolved considerably during recent decades. The current methodological standards define procedures for literature search, evaluation of quality, combining research results and grading the overall strength of evidence. The current methods of evidence synthesis are systematic reviews, which are regarded as the state-of-the-art scientific standards for pooling research evidence and superior to traditional narrative reviews. These developments have been applied in EBM, risk assessment and other fields.

106. The Committee will benefit from adopting a framework that is informed by these scientific developments. However, the specific nature and scientific contents of radiation epidemiological studies speak against a mechanistic application of generic quality criteria. Therefore, this annex provides an UNSCEAR approach to radiation epidemiological study quality assessment and to synthesis of findings across studies. The approach provides for increased methodological rigour, which could enhance the degree of coherence, transparency and objectivity in assessments. Although the methodological guidelines generally developed for systematic reviews are not always applicable for reasons including lack of study on specific exposure, overall paucity of evidence, and lack of comparability between available studies, the Committee seeks, nevertheless, to ensure that quality considerations guide their scientific assessment of the information provided by the respective evaluation.

107. The evidence synthesis approach outlined in this annex is focused on evidence from epidemiology. For a full assessment of scientific areas of interest to the Committee, more evidence—from, for example, radiobiology, radiation dosimetry and physics—is required. In principle, similar criteria apply to the selection and inclusion of literature from other sciences into the respective UNSCEAR evaluations. However, different scientific approaches, study designs, publication traditions and many other variations characterize these sciences, calling for a measured approach in transferring insights from this annex to other fields.

108. Overall, the Committee aims to work with a quality-oriented systematic review approach for its evaluation wherever applicable, based on the concepts described in this annex. While the focus on study quality and the explicit review of strengths and limitations of radiation epidemiology studies is a long-standing feature of the Committee's work, using quality criteria in a systematic way as outlined here has not been always applied to UNSCEAR reports that rely on epidemiological data. In summary, a high degree of transparency and a systematic approach to collecting, analysing and interpreting information for the Committee's evaluations and assessments will help to maintain the high scientific standard necessary for its widely appraised reports.

ACKNOWLEDGEMENTS

The Committee wishes to acknowledge with gratitude the work of the consultants and the Expert Group on Epidemiological Quality involved in preparing this annex. In particular, the Committee would like to thank H. Zeeb and A. Auvinen (consultants) who elaborated the manuscript. The views expressed in the annex remain those of the Committee and do not necessarily represent the views of individual experts, the United Nations or its Member States.

Members of the Expert Group on Epidemiological Quality

P. Smeesters (Belgium); Q. Sun (China); A. Auvinen and S. Salomaa (Finland); W.-U. Müller, R. Michel and H. Zeeb (Germany); S. Akiba, K. Kodama and S. Saigusa (Japan); M. Tirmarche (France), A. Akleyev and V. Ivanov (Russian Federation); R. Wakeford (United Kingdom); and J. Boice, A. Bouville, B. Napier, D. Pawel and R. Shore (United States).

REFERENCES

A1 Ajmal, S., S. Rapoport, H. Ramirez Batlle et al. The natural history of the benign thyroid nodule: what is the appropriate follow-up strategy? J Am Coll Surg 220(6): 987-992 (2015).

A2 Alazzoni, A., C.L. Gordon, J. Syed et al. Randomized controlled trial of radiation protection with a patient lead shield and a novel, nonlead surgical cap for operators performing coronary angiography or intervention. Circ Cardiovasc Interv 8(8): e002384 (2015).

A3 Astakhova, L.N., L.R. Anspaugh, G.W. Beebe et al. Chernobyl-related thyroid cancer in children of Belarus: a case-control study. Radiat Res 150(3): 349-356 (1998).

A4 Auvinen, A., L. Salonen, J. Pekkanen et al. Radon and other natural radionuclides in drinking water and risk of stomach cancer: a case-cohort study in Finland. Int J Cancer 114(1): 109-113 (2005).

B1 Boice, J.D., Jr., M.M. Morin, A.G. Glass et al. Diagnostic x-ray procedures and risk of leukemia, lymphoma, and multiple myeloma. JAMA 265(10): 1290-1294 (1991).

B2 Boice, J.D., Jr., D. Preston, F.G. Davis et al. Frequent chest X-ray fluoroscopy and breast cancer incidence among tuberculosis patients in Massachusetts. Radiat Res 125(2): 214-222 (1991).

B3 Borenstein, M., L.V. Hedges, J.P.T. Higgins et al. Introduction to Meta-Analysis. Wiley, New York, 2008.

B4 Bouville, A., R.E. Toohey, J.D. Boice, Jr. et al. Dose reconstruction for the million worker study: status and guidelines. Health Phys 108(2): 206-220 (2015).

C1 Cardis, E., M. Vrijheid, M. Blettner et al. The 15-Country Collaborative Study of Cancer Risk among Radiation Workers in the Nuclear Industry: estimates of radiation-related cancer risks. Radiat Res 167(4): 396-416 (2007).

C2 Chumak, V.V. Physical dosimetry of Chernobyl cleanup workers. Health Phys 93(5): 452-461 (2007).

C3 Claus, E.B., L. Calvocoressi, M.L. Bondy et al. Dental x-rays and risk of meningioma. Cancer 118(18): 4530-4537 (2012).

C4 Cornell, J.E., C.D. Mulrow, R. Localio et al. Random-effects meta-analysis of inconsistent effects: a time for change. Ann Intern Med 160(4): 267-270 (2014).

C5 CRD. Systematic reviews. CRD's guidance for undertaking reviews in health care. Centre for Reviews and Dissemination. [Internet] Available from (http://www.york.ac.uk/crd/SysRev/!SSL!/WebHelp/SysRev3.htm) on 02.02.2016.

C6 Crippa, A., P. Khudyakov, M. Wang et al. A new measure of between-studies heterogeneity in meta-analysis. Stat Med 35(21): 3661-3675 (2016).

C7 Cullings, H.M., S. Fujita, S. Funamoto et al. Dose estimation for atomic bomb survivor studies: its evolution and present status. Radiat Res 166(1 Pt 2): 219-254 (2006).

D1 Darby, S., D. Hill, A. Auvinen et al. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. Br Med J 330(7485): 223 (2005).

D2 Darby, S., D. Hill, H. Deo et al. Residential radon and lung cancer--detailed results of a collaborative analysis of individual data on 7,148 persons with lung cancer and 14,208 persons without lung cancer from 13 epidemiologic studies in Europe. Scand J Work Environ Health 32(Suppl 1): 1-83 (2006).

D3 Darby, S.C., J.H. Olsen, R. Doll et al. Trends in childhood leukaemia in the Nordic countries in relation to fallout from atmospheric nuclear weapons testing. Br Med J 304(6833): 1005-1009 (1992).

D4 Davis, F.G., J.D. Boice, Jr., Z. Hrubec et al. Cancer mortality in a radiation-exposed cohort of Massachusetts tuberculosis patients. Cancer Res 49(21): 6130-6136 (1989).

D5 Degteva, M.O., M.I. Vorobiova, V.P. Kozheurov et al. Dose reconstruction system for the exposed population living along the Techa River. Health Phys 78(5): 542-554 (2000).

D6 Des Jarlais, D.C., C. Lyles, N. Crepaz et al. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health 94(3): 361-366 (2004).

D7 Doody, M.M., J.E. Lonstein, M. Stovall et al. Breast cancer mortality after diagnostic radiography: findings from the U.S. Scoliosis Cohort Study. Spine (Phila Pa 1976) 25(16): 2052-2063 (2000).

D8 Drozdovitch, V., V. Minenko, V. Khrouch et al. Thyroid dose estimates for a cohort of Belarusian children exposed to radiation from the Chernobyl accident. Radiat Res 179(5): 597-609 (2013).

E1 Eddy, D.M. Evidence-based medicine: a unified approach. Health Aff (Millwood) 24(1): 9-17 (2005).

E2 Edwards, A.A., C. Lindholm, F. Darroudi et al. Review of translocations detected by FISH for retrospective biological dosimetry applications. Radiat Prot Dosim 113(4): 396-402 (2005).

E3 Egger, M., G. Davey Smith, M. Schneider et al. Bias in meta-analysis detected by a simple, graphical test. Br Med J 315(7109): 629-634 (1997).

E4 Egger, M., G. Davey-Smith and D.G. Altman. Systematic Reviews in Health Care: Meta-Analysis in Context, 2nd edition. Wiley, New York, 2008.

E5 Eidemuller, M., E. Ostroumova, L. Krestinina et al. Analysis of solid cancer mortality in the Techa river cohort using the two-step clonal expansion model. Radiat Res 169(2): 138-148 (2008).

E6 Engholm, G., J. Ferlay, N. Christensen et al. NORDCAN--a Nordic tool for cancer information, planning, quality control and research. Acta Oncol 49(5): 725-736 (2010).

E7 EPOC. EPOC Resources for review authors. Effective Practice and Organisation of Care (EPOC). Norwegian Knowledge Centre for the Health Services, Oslo. [Internet] Available from (http://epoc.cochrane.org/epoc-specific-resources-review-authors) on 02.02.2016.

E8 Equator Network. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Enhancing the Quality and Transparency of Health Research. [Internet] Available from (http://www.equator-network.org/reporting-guidelines/prisma) on 02.02.2016.

F1 Fenner-Crisp, P.A. and V.L. Dellarco. Key elements for judging the quality of a risk assessment. Environ Health Perspect 124(8): 1127-1135 (2016).

G1 Garner, P., S. Hopewell, J. Chandler et al. When and how to update systematic reviews: consensus and checklist. Br Med J 354: i3507 (2016).

G2 Gilbert, E.S. The impact of dosimetry uncertainties on dose-response analyses. Health Phys 97(5): 487-492 (2009).

G3 Goodman, S.N. and J.A. Berlin. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. Ann Intern Med 121(3): 200-206 (1994).

G4 GRADE. Grading of Recommendations Assessment, Development and Evaluation. [Internet] Available from (www.gradeworkinggroup.org) on 02.02.2016.

G5 Greenland, S. and H. Morgenstern. Ecological bias, confounding, and effect modification. Int J Epidemiol 18(1): 269-274 (1989).

G6 Greenland, S. Multiple-bias modelling for analysis of observational data (with discussion). J R Stat Soc Ser A 168(2): 267–306 (2005).

G7 Guyatt, G.H., A.D. Oxman, H.J. Schunemann et al. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol 64(4): 380-382 (2011).

G8 Guyatt, G.H., A.D. Oxman, G. Vist et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). J Clin Epidemiol 64(4): 407-415 (2011).

H1 Hammer, G.P., U. Scheidemann-Wesp, F. Samkange-Zeeb et al. Occupational exposure to low doses of ionizing radiation and cataract development: a systematic literature review and perspectives on future studies. Radiat Environ Biophys 52(3): 303-319 (2013).

H2 Heid, I.M., H. Kuchenhoff, J. Miles et al. Two dimensions of measurement error: classical and Berkson error in residential radon exposure assessment. J Expo Anal Environ Epidemiol 14(5): 365-377 (2004).

H3 Heinavaara, S., S. Toikkanen, K. Pasanen et al. Cancer incidence in the vicinity of Finnish nuclear power plants: an emphasis on childhood leukemia. Cancer Causes Control 21(4): 587-595 (2010).

H4 Higgins, J. and S. Green. Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. [Internet] Available from (http://handbook.cochrane.org/) on 02.02.2016.

H5 Hill, A.B. The environment and disease: association or causation? Proc R Soc Med 58: 295-300 (1965).

H6 Howe, G.R. Lung cancer mortality between 1950 and 1987 after exposure to fractionated moderate-dose-rate ionizing radiation in the Canadian fluoroscopy cohort study and a comparison with lung cancer mortality in the Atomic Bomb survivors study. Radiat Res 142(3): 295-304 (1995).

H7 Howe, G.R. and J. McLaughlin. Breast cancer mortality between 1950 and 1987 after exposure to fractionated moderate-dose-rate ionizing radiation in the Canadian fluoroscopy cohort study and a comparison with breast cancer mortality in the atomic bomb survivors study. Radiat Res 145(6): 694-707 (1996).

I1 IEA. Good epidemiological practice. International Epidemiological Association. [Internet] Available from (http://ieaweb.org/good-epidemiological-practice-gep) on 02.02.2016.

I2 Inskip, P.D., A. Ekbom, M.R. Galanti et al. Medical diagnostic x rays and thyroid cancer. J Natl Cancer Inst 87(21): 1613-1621 (1995).

I3 Ioannidis, J.P., L. Belbasis and E. Evangelou. Fifty-year fate and impact of general medical journals. PLoS One 5(9): e12531 (2010).

I4 IPCS. IPCS risk assessment terminology. Harmonization project document no. 1. International Programme on Chemical Safety Harmonization Project of WHO, UNEP and ILO, Geneva, 2004.

I5 IPCS. WHO human health risk assessment toolkit: chemical hazards. Harmonization project document no. 8. International Programme on Chemical Safety Harmonization Project of WHO, UNEP and ILO, Geneva, 2010.

I6 Ivanov, V.K., A.I. Gorski, A.F. Tsyb et al. Solid cancer incidence among the Chernobyl emergency workers residing in Russia: estimation of radiation risks. Radiat Environ Biophys 43(1): 35-42 (2004).

I7 Ivanov, V.K., V.V. Kashcheev, S.Y. Chekin et al. Radiation-epidemiological studies of thyroid cancer incidence in Russia after the Chernobyl accident (estimation of radiation risks, 1991-2008 follow-up period). Radiat Prot Dosim 151(3): 489-499 (2012).

I8 Iwasaki, T., Y. Takashima, T. Suzuki et al. The dose response of chromosome aberrations in human lymphocytes induced in vitro by very low-dose gamma rays. Radiat Res 175(2): 208-213 (2011).

J1 Jablon, S., Z. Hrubec and J.D. Boice, Jr. Cancer in populations living near nuclear facilities. A survey of mortality nationwide and incidence in two states. JAMA 265(11): 1403-1408 (1991).

J2 Jacob, P., T.I. Bogdanova, E. Buglova et al. Thyroid cancer risk in areas of Ukraine and Belarus affected by the Chernobyl accident. Radiat Res 165(1): 1-8 (2006).

J3 Jin, Z.C., X.H. Zhou and J. He. Statistical methods for dealing with publication bias in meta-analysis. Stat Med 34(2): 343-360 (2015).

K1 Kesminiene, A., A.S. Evrard, V.K. Ivanov et al. Risk of thyroid cancer among Chernobyl liquidators. Radiat Res 178(5): 425-436 (2012).

K2 Keyes, K.M. and S. Galea. Epidemiology Matters: A New Introduction to Methodological Foundations. Oxford University Press, Oxford, 2014.

K3 Kreuzer, M., B. Grosche, M. Schnelzer et al. Radon and risk of death from cancer and cardiovascular diseases in the German uranium miners cohort study: follow-up 1946-2003. Radiat Environ Biophys 49(2): 177-185 (2010).

K4 Krewski, D., J.H. Lubin, J.M. Zielinski et al. A combined analysis of North American case-control studies of residential radon and lung cancer. J Toxicol Environ Health A 69(7): 533-597 (2006).

K5 Kwon, D., F.O. Hoffman, B.E. Moroz et al. Bayesian dose-response analysis for epidemiological studies with complex uncertainty in dose estimation. Stat Med 35(3): 399-423 (2016).

L1 Land, C.E., Z. Zhumadilov, B.I. Gusev et al. Ultrasound-detected thyroid nodule prevalence and radiation dose from fallout. Radiat Res 169(4): 373-383 (2008).

L2 Land, C.E., D. Kwon, F.O. Hoffman et al. Accounting for shared and unshared dosimetric uncertainties in the dose response for ultrasound-detected thyroid nodules after exposure to radioactive fallout. Radiat Res 183(2): 159-173 (2015).

L3 Lash, T.L., M.P. Fox and A.K. Fink. Applying Quantitative Bias Analysis to Epidemiologic Data. Springer, Dordrecht, 2009.

L4 Leonard, A., J. Rueff, G.B. Gerber et al. Usefulness and limits of biological dosimetry based on cytogenetic methods. Radiat Prot Dosim 115(1-4): 448-454 (2005).

L5 Little, M.P., T.V. Azizova, D. Bazyka et al. Systematic review and meta-analysis of circulatory disease from exposure to low-level ionizing radiation and estimates of potential population mortality risks. Environ Health Perspect 120(11): 1503-1511 (2012).

L6 Little, M.P., A.G. Kukush, S.V. Masiuk et al. Impact of uncertainties in exposure assessment on estimates of thyroid cancer risk among Ukrainian children and adolescents exposed from the Chernobyl accident. PLoS One 9(1): e85723 (2014).

L7 Little, M.P. Radiation and circulatory disease. Mutat Res 770(Pt B): 299-318 (2016).

M1 MacMahon, B. and T.F. Pugh. Epidemiology: Principles and Methods. Little, Brown & Company, Boston, Massachusetts, 1970.

M2 Menzel, H.G. and J. Harrison. Effective dose: a radiation protection quantity. Ann ICRP 41(3-4): 117-123 (2012).

M3 Miettinen, O.S. Theoretical Epidemiology: Principles of Occurrence Research in Medicine. Wiley, New York, 1985.

M4 Minenko, V.F., A.V. Ulanovsky, V.V. Drozdovitch et al. Individual thyroid dose estimates for a case-control study of Chernobyl-related thyroid cancer among children of Belarus--part II. Contributions from long-lived radionuclides and external radiation. Health Phys 90(4): 312-327 (2006).

M5 Moher, D., A. Liberati, J. Tetzlaff et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 339: b2535 (2009).

M6 Morishima, H., T. Koga, K. Tatsumi et al. Dose measurement, its distribution and individual external dose assessments of inhabitants in the high background radiation areas in China. J Radiat Res 41(Suppl): 9-23 (2000).

M7 Morton, S.C., M.R. Costlow, J.S. Graff et al. Standards and guidelines for observational studies: quality is in the eye of the beholder. J Clin Epidemiol 71: 3-10 (2016).

N1 Nair, M.K., K.S. Nambi, N.S. Amma et al. Population study in the high natural background radiation area in Kerala, India. Radiat Res 152(6 Suppl): S145-148 (1999).

N2 Nair, R.R., B. Rajan, S. Akiba et al. Background radiation and cancer incidence in Kerala, India-Karanagappally cohort study. Health Phys 96(1): 55-66 (2009).

N3 NCRP. Uncertainties in the estimation of radiation risks and probability of disease causation. NCRP Report No. 171. National Council on Radiation Protection and Measurements, Bethesda, 2012.

N4 NRC. The Hanford Environmental Dose Reconstruction Project: A review of four documents. National Research Council (US) Committee on an Assessment of CDC Radiation Studies, National Academies Press, Washington D.C., 1994.

O1 Ozasa, K., Y. Shimizu, A. Suyama et al. Studies of the mortality of atomic bomb survivors, Report 14, 1950-2003: An overview of cancer and noncancer diseases. Radiat Res 177(3): 229-243 (2012).

P1 Piantadosi, S. Invited commentary: ecologic biases. Am J Epidemiol 139(8): 761-764; discussion 769-771 (1994).

P2 Porta, M. A Dictionary of Epidemiology. Sixth edition. Oxford University Press, Oxford, 2014.

P3 Preston-Martin, S., D.C. Thomas, S.C. White et al. Prior exposure to medical and dental x-rays related to tumors of the parotid gland. J Natl Cancer Inst 80(12): 943-949 (1988).

P4 Preston-Martin, S., D.C. Thomas, M.C. Yu et al. Diagnostic radiography as a risk factor for chronic myeloid and monocytic leukaemia (CML). Br J Cancer 59(4): 639-644 (1989).

P5 Preston, R.J., J.D. Boice Jr., A. Bertrand Brill et al. Uncertainties in estimating health risks associated with exposure to ionising radiation. J Radiol Prot 33(3): 573-588 (2013).

R1 Richardson, D.B., E. Cardis, R.D. Daniels et al. Risk of cancer from occupational exposure to ionising radiation: retrospective cohort study of workers in France, the United Kingdom, and the United States (INWORKS). Br Med J 351: h5359 (2015).

S1 Sackett, D. Evidence-based medicine. Lancet 346(8983): 1171 (1995).

S2 Sanderson, S., I.D. Tatt and J.P. Higgins. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. Int J Epidemiol 36(3): 666-676 (2007).

S3 Schisterman, E.F., S.R. Cole and R.W. Platt. Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiology 20(4): 488-495 (2009).

S4 Schubauer-Berigan, M.K., R.D. Daniels, D.A. Fleming et al. Risk of chronic myeloid and acute leukemia mortality after exposure to ionizing radiation among workers at four U.S. nuclear weapons facilities and a nuclear naval shipyard. Radiat Res 167(2): 222-232 (2007).

S5 Schunemann, H., S. Hill, G. Guyatt et al. The GRADE approach and Bradford Hill's criteria for causation. J Epidemiol Community Health 65(5): 392-395 (2011).

S6 Shea, B.J., J.M. Grimshaw, G.A. Wells et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 7: 10 (2007).

S7 Shea, B.J., C. Hamel, G.A. Wells et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol 62(10): 1013-1020 (2009).

S8 Shore, R.E. Overview of radiation-induced skin cancer in humans. Int J Radiat Biol 57(4): 809-827 (1990).

S9 Simon, S.L., A. Bouville, R. Kleinerman et al. Dosimetry for epidemiological studies: learning from the past, looking to the future. Radiat Res 166(1 Pt 2): 313-318 (2006).

S10 Simon, S.L., I. Bailiff, A. Bouville et al. BiodosEPR-2006 consensus committee report on biodosimetric methods to evaluate radiation doses at long times after exposure. Radiat Meas 42(6): 948-971 (2007).

S11 Sterne, J., J. Higgins and B. Reeves. A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI). Edited on behalf of the development group for ACROBAT-NRSI. [Internet] Available from (http://www.riskofbias.info) on 02.02.2016.

S12 Sterne, J.A., M. Egger and G.D. Smith. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. Br Med J 323(7304): 101-105 (2001).

S13 Stram, D.O., D.L. Preston, M. Sokolnikov et al. Shared dosimetry error in epidemiological dose-response analyses. PLoS One 10(3): e0119418 (2015).

S14 Swartz, H.M., B.B. Williams and A.B. Flood. Overview of the principles and practice of biodosimetry. Radiat Environ Biophys 53(2): 221-232 (2014).

T1 Thierens, H. and A. Vral. The micronucleus assay in radiation accidents. Ann Ist Super Sanita 45(3): 260-264 (2009).

T2 Thierry-Chef, I., D.B. Richardson, R.D. Daniels et al. Dose estimation for a study of nuclear workers in France, the United Kingdom and the United States of America: Methods for the International Nuclear Workers Study (INWORKS). Radiat Res 183(6): 632-642 (2015).

T3 Till, J.E., H.L. Beck, H.A. Grogan et al. A review of dosimetry used in epidemiological studies considered to evaluate the linear no-threshold (LNT) dose-response model for radiation protection. Int J Radiat Biol 93(10): 1128-1144 (2017).

U1 UNSCEAR. Sources and Effects of Ionizing Radiation. UNSCEAR 1994 Report. United Nations Scientific Committee on the Effects of Atomic Radiation, 1994 Report to the General Assembly, with scientific annexes. United Nations sales publication E.94.IX.11. United Nations, New York, 1994.

U2 UNSCEAR. Sources and Effects of Ionizing Radiation. Volume II: Effects. UNSCEAR 2000 Report. United Nations Scientific Committee on the Effects of Atomic Radiation, 2000 Report to the General Assembly, with scientific annexes. United Nations sales publication E.00.IX.4. United Nations, New York, 2000.

U3 UNSCEAR. Effects of Ionizing Radiation. Volume I: Report to the General Assembly, Scientific Annexes A and B. UNSCEAR 2006 Report. United Nations Scientific Committee on the Effects of Atomic Radiation. United Nations sales publication E.08.IX.6. United Nations, New York, 2008.

U4 UNSCEAR. Sources and Effects of Ionizing Radiation. Volume II: Effects. Scientific Annexes C, D and E. UNSCEAR 2008 Report. United Nations Scientific Committee on the Effects of Atomic Radiation. United Nations sales publication E.11.IX.3. United Nations, New York, 2011.

U5 UNSCEAR. Sources, Effects and Risks of Ionizing Radiation. Volume I: Report to the General Assembly and Scientific Annex A. UNSCEAR 2013 Report. United Nations Scientific Committee on the Effects of Atomic Radiation. United Nations sales publication E.14.IX.1. United Nations, New York, 2014.

U6 UNSCEAR. Developments since the 2013 UNSCEAR Report on the Levels and Effects of Radiation Exposure due to the Nuclear Accident following the Great East-Japan Earthquake

and Tsunami. A 2015 white paper to guide the Scientific Committee's future programme of work. United Nations Scientific Committee on the Effects of Atomic Radiation. United Nations, New York, 2015.

U7 UNSCEAR. Sources, Effects and Risks of Ionizing Radiation. Report to the General Assembly and Scientific Annexes A and B. UNSCEAR 2012 Report. United Nations Scientific Committee on the Effects of Atomic Radiation. United Nations sales publication E.16.IX.1. United Nations, New York, 2015.

U8 UNSCEAR. Developments since the 2013 UNSCEAR Report on the Levels and Effects of Radiation Exposure due to the Nuclear Accident following the Great East-Japan Earthquake and Tsunami. A 2016 white paper to guide the Scientific Committee's future programme of work. United Nations Scientific Committee on the Effects of Atomic Radiation. United Nations, New York, 2016.

V1 Vacquier, B., S. Caer, A. Rogel et al. Mortality risk in the French cohort of uranium miners: extended follow-up 1946-1999. Occup Environ Med 65(9): 597-604 (2008).

W1 WHO. International statistical classification of diseases and related health problems. 10th revision, Volume 2: Instruction manual. 2010 Edition. World Health Organization, Geneva, 2011.

W2 WHO. Handbook for guideline development. 2nd edition. World Health Organization, Geneva, 2014.

Y1 Yamada, M., F.L. Wong, S. Fujiwara et al. Noncancer disease incidence in atomic bomb survivors, 1958-1998. Radiat Res 161(6): 622-632 (2004).

Y2 Yasumura, S., M. Hosoya, S. Yamashita et al. Study protocol for the Fukushima Health Management Survey. J Epidemiol 22(5): 375-383 (2012).