Mutation Research-Reviews in Mutation Research 798 (2026) 108601

Contents lists available at ScienceDirect

Mutation Research-Reviews in Mutation Research

journal homepage: www.elsevier.com/locate/mutrev

The comet assay as a tool for human biomonitoring of exposure to environmental and occupational agents – A summary of systematic reviews and meta-analyses

Peter Møller [a,*], Goran Gajski [b], Marko Gerić [b], Amaya Azqueta [c,d], Lisa Giovannelli [e], Anja Haveric [f], Helga Stopper [g], Ezgi Eyluel Bankoglu [g,h], Andrew Collins [i], Carina Ladeira [j,k,l]

[a] Department of Public Health, Section of Environmental Health, University of Copenhagen, Copenhagen, Denmark
[b] Division of Toxicology, Institute for Medical Research and Occupational Health, Zagreb, Croatia
[c] Department of Pharmaceutical Sciences, School of Pharmacy and Nutrition, University of Navarra, Pamplona, Spain
[d] University of Navarra, BIOMA Institute for Biodiversity and the Environment, Pamplona, Spain
[e] Department NEUROFARBA, Section Pharmacology and Toxicology, University of Florence, Florence, Italy
[f] University of Sarajevo, Institute for Genetic Engineering and Biotechnology, Sarajevo, Bosnia and Herzegovina
[g] Institute of Pharmacology and Toxicology, University of Würzburg, Würzburg, Germany
[h] German Federal Institute for Risk Assessment (BfR), Department Food and Feed Safety in the Food Chain, Berlin, Germany
[i] Department of Nutrition, University of Oslo, Oslo, Norway
[j] H&TRC-Health & Technology Research Center, ESTeSL-Escola Superior de Tecnologia da Saúde, Instituto Politécnico de Lisboa, Lisbon, Portugal
[k] NOVA National School of Public Health, Public Health Research Centre, Universidade NOVA de Lisboa, Lisbon, Portugal
[l] CHRC – Comprehensive Health Research Centre, Portugal

ARTICLE INFO

Keywords:
Human biomonitoring
Comet assay
DNA strand breaks
Meta-analysis
Public health
Systematic review

ABSTRACT

This paper compiles results from six systematic reviews and meta-analyses on associations between environmental and occupational exposure to chemicals and levels of DNA strand breaks in human leukocytes, measured by the comet assay. There are no differences in effect sizes when using different comet descriptors. However, lower central tendencies are obtained by using a non-parametric test as compared to the standard parametric analysis, indicating that the standard meta-analyses tend to overestimate the effect size. The compiled results indicate that exposures can be sorted into three groups with decreasing effect size: high (pesticides), moderate (volatile organic compounds, heavy metals and antineoplastic drugs), and low (anaesthetic gases and air pollution). Interestingly, studies from middle-income countries have higher effect sizes than those seen in studies from high-income countries. This may be related to higher exposures or lower socioeconomic status in middle-income countries. However, there is also some co-variability between studies from middle-income countries and the risk of comet assay measurement bias, assessed as information provided in published papers. Lack of information on assay controls and blinded/coded sample analysis appears to be a general issue in studies on comet assay results. Risk of exposure misclassification is mainly related to the type of exposure; there are good biomarkers for some exposures (e.g. heavy metals), whereas other exposures are more challenging to assess with biomarkers (e.g. pesticides). In conclusion, all examined exposures result in significant increases in DNA strand breaks at the population level, though to varying degrees.

1. Introduction

The comet assay was developed approximately four decades ago as a relatively simple technique for the detection of DNA damage in cells [1]. The standard version of the comet assay detects DNA strand breaks, alkali-labile sites and transient DNA repair sites [2]. In this manuscript, the term *DNA strand breaks* will be used to refer to these types of lesions and to distinguish them from lesions (e.g. enzyme-sensitive sites) that

* Corresponding author.
E-mail addresses: pemo@sund.ku.dk (P. Møller), ggajski@imi.hr (G. Gajski), mgeric@imi.hr (M. Gerić), amazqueta@unav.es (A. Azqueta), lisa.giovannelli@unifi.it (L. Giovannelli), anjahaveric@gmail.com, anja.haveric@ingeb.unsa.ba (A. Haveric), helga.stopper@uni-wuerzburg.de (H. Stopper), Eyluel.Bankoglu@bfr.bund.de (E.E. Bankoglu), collinsand@gmail.com (A. Collins), carina.ladeira@estesl.ipl.pt (C. Ladeira).

https://doi.org/10.1016/j.mrrev.2026.108601
Received 6 July 2025; Received in revised form 25 March 2026; Accepted 2 June 2026
Available online 6 June 2026

are detected as additional strand breaks by modified comet assay procedures. Since the 1990s, the standard alkaline comet assay has become increasingly popular among researchers in genetic toxicology and human biomonitoring. It has essentially replaced the alkaline elution and alkaline unwinding assays, which detect the same kinds of DNA lesions with similar sensitivity [3]. The success of the comet assay can be attributed to the easy-to-use procedure and the relatively low costs of analysing samples. However, the major limitation has been, and still is, whether the difference in levels of DNA strand breaks is associated with the risk of disease. Many studies have found higher levels of DNA strand breaks in patients with non-communicable diseases as compared to healthy controls [4,5]. However, reverse causality cannot be ruled out in that kind of cross-sectional study. A retrospective pooled analysis of results from several prospective studies has shown that high levels of DNA strand breaks are associated with increased risk of mortality [6,7]. The study did not have sufficient statistical power to assess associations for specific diseases because of the relatively small combined size of the populations studied (about 2400) and the low number of subjects developing disease or dying. In general, there is a paucity of prospective studies that investigate associations between levels of DNA strand breaks in healthy subjects and the risk of disease in the future. Meanwhile, it is desirable to assess whether exposure to genotoxic agents is associated with increased levels of DNA strand breaks in humans, as this would represent an early step in disease development and possible value as a predictive marker of disease risk.

A common approach in biomonitoring studies is to investigate DNA strand breaks in subjects with external exposures as compared to unexposed subjects. It has been popular to include the comet assay in studies on environmental, occupational, medical and lifestyle exposures. A cursory overview of the literature indicates that many exposures influence the background level of DNA strand breaks in the comet assay. However, it is unclear whether different external exposures produce the same or different levels of DNA strand breaks, or even if these exposures cause greater DNA damage than that due to intrinsic host factors such as age or sex. To the best of our knowledge, no study has previously attempted to compare the effect sizes of external exposures on DNA strand breaks, measured by the comet assay, using the same meta-analysis procedure.

The overall aim of this project is to examine associations between environmental or occupational exposure to chemicals, and DNA strand breaks measured by the comet assay in human leukocytes, lymphocytes, or peripheral blood mononuclear cells. Fig. 1 outlines the general structure of the project and the connection between the different papers (described below). First, we conducted a scoping review with the aim of systematically analysing evidence on the use of the comet assay in human biomonitoring studies assessing genotoxic effects from environmental or occupational exposures [8]. We assessed the literature, key concepts, types of evidence and knowledge gaps. The scoping exercise served as an instrumental precursor for a series of systematic reviews and meta-analyses, in line with the groups of exposures that were previously addressed. Specifically, the scoping review focused on air pollutants, anaesthetic gases, antineoplastic drugs, heavy metals, pesticides, and volatile organic compounds (VOCs) [8]. Subsequently, we have performed systematic reviews and meta-analyses on environmental/occupational exposure to air pollution [9], anaesthetic gases [10], antineoplastic drugs [11], heavy metals [12], pesticides [13], and VOCs [14]. The present paper is a summary of the six systematic reviews and meta-analyses. It resembles an umbrella review on tailored (own) systematic reviews, investigating whether the standard alkaline comet assay detects different levels of DNA strand breaks for different types of exposures. As such, the statistical null hypothesis is that the six types of genotoxic agents generate the same magnitude of DNA strand breaks. Thus, the number of subjects will only affect the statistical power of the study; larger studies will be able to detect a smaller effect size. In sequence, the paper is structured to (i) describe the scope of the project, (ii) elaborate on common challenges in the separate systematic reviews, and (iii) compare effect sizes of DNA strand breaks in humans exposed to different environmental and occupational agents. The unique strength of the compilation of systematic reviews lies in the fact that all associations have been assessed by identical qualitative and quantitative procedures. Thus, it is possible to compare effect sizes across chemical exposures. Nevertheless, there are special issues in each of the six exposure groups, which are addressed in the separate papers. The present paper introduces the project and demonstrates a summary of the genotoxic effects of the most commonly investigated environmental and occupational exposures, covering air pollution, anaesthetic gases, antineoplastic drugs, heavy metals, VOCs, and pesticides. In addition, the paper elaborates on background information on various challenges in the systematic reviews and meta-analyses of comet assay results from human biomonitoring studies.

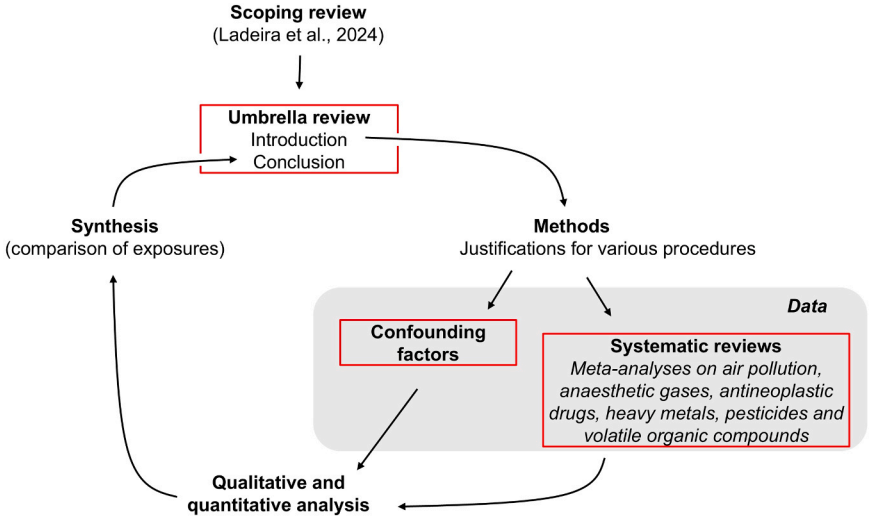

Fig. 1. Outline of the project. The overarching aim of the project is to assess whether exposures to occupational and environmental agents have different effects on the levels of DNA strand breaks, measured by the comet assay, in human leukocytes and subpopulations thereof. The first stage was a scoping review by Ladeira et al., 2024 [8], which identified six different exposure groups that were eligible for systematic review and meta-analyses [9–14]. Effects of confounding factors were assessed in a separate review [23]. The present umbrella review closes a circle of systematic reviews, serving both the purpose of introducing the project, conducting statistical analyses on the full dataset of studies, and summarising the outcomes in a general discussion.

2. Methods

In the method section, we describe the procedures used in the systematic reviews of specific exposure groups (Section 2.1.) and statistical analyses in the present umbrella review (Section 2.2.).

An umbrella review compiles results from systematic reviews on different predictors for diseases of interest, such as results from randomized clinical trials on different treatments on the same disease, comparison of the strength of association of different exposures on a specific outcome in epidemiology. The umbrella review entails the same steps as systematic reviews, including a systematic literature search, study selection, data extraction, statistical analysis, grading of evidence, and interpretation of results [15]. Our study is slightly different from the traditional umbrella review because it is based on our own standardized systematic reviews (i.e. there is not heterogeneity in methodology between reviews to consider). We have not performed a specific literature search on systematic reviews related to effects of environmental/occupational exposures on DNA strand breaks because our systematic reviews are the most recent articles. The results of the meta-analyses are discussed in each of the six systematic reviews on air pollution [9], anaesthetic gases [10], antineoplastic drugs [11], heavy metals [12], pesticides [13], and VOCs [14].

2.1. Systematic reviews and meta-analyses of specific exposures

The systematic review of exposure types has been carried out in different steps (Table 1). The overarching aim has been to include as many studies as possible in the meta-analysis (step 3) and to assess the robustness of the central tendency by subgroup and sensitivity analyses (steps 5 and 6). In addition, the generalizability of the results in the meta-analysis has been assessed by comparing the proportion of studies with positive associations in the meta-analysis to the proportion in the dataset of studies that are not included in the meta-analysis, but still included in the review (step 4). Below are descriptions of each step in the systematic reviews and the various challenges (and considerations) we have encountered in the process.

Table 1
Description of steps in the systematic reviews.

Step	Description	Outcome
1	Identification of studies	None
2	**Inclusion in review** *Exclusion due to suspected risk selection bias, double publication of results, lack of exposure gradient, larger exposure gradient of co-exposures, cell type, insufficient comet assay description*[a]*, etc.*	Proportion of statistically significant positive and null results
3	**Inclusion in meta-analysis** *Exclusion due to insufficient information on the mean (or variability) of central tendency, geometric mean/variation, etc.*	Standardised mean difference (SMD)
4	Generalizability of the studies in meta-analysis to all of the studies that are included in the reviw	Proportion of studies with positive associations in meta-analysis compared to all included studies (Z-test)
5	Subgroup analysis (stratification of studies into groups with different exposures)	Comparison of SMD between subgroups
6	Sensitivity analysis (comet assay description, blinding, assay controls and exposure assessment)[b]	Comparison of SMD between groups with different risk of bias and exposure misclassification

[a] Refers to papers where there is uncertainty about whether it is a neutral or alkaline version of the comet assay. It may also involve studies on the enzyme-linked comet assay, where there is uncertainty as to whether the results are DNA strand breaks or a total number of lesions after incubation with enzymes. Lastly, there can be uncertainty as to whether the statistical analysis is based on all the comets scored or the central tendency of comets in each gel.

[b] Refers to risk of measurement bias due to incomplete information of comet assay procedures, uncertainty about concealment (blinded/coded analysis) and inclusion of assay controls. Exposure misclassification refers to subjects categorised as exposed even though they are unexposed or vice versa.

2.1.1. Identification of studies (step 1)

The studies have been identified by searches in at least two different databases (PubMed and EMBASE/Web of Science). The database searches cover the period starting from the year 2000. The reasons for choosing the year 2000 as the starting point were because it is the year that (i) the comet assay was adopted as Medical Subject Heading (MeSH) in PubMed, (ii) the first guideline for using the comet assay in human biomonitoring studies was published [16], and (iii) the first focused reviews on comet assay results in human biomonitoring studies were published [17,18]. Although human biomonitoring studies on comet assay results existed before 2000, they have been obtained during a time when the comet assay was still relatively new to many researchers, and its technical procedures, data analysis and reporting were still undergoing development and refinement. It can be argued that sufficient experience with the comet assay had been gained by 2000 to use this year as a starting point for systematic searches in databases. Certain studies (published in 1999) have been identified as citations in the papers retrieved through the systematic searches. We have included these studies in the systematic reviews by a case-by-case approach. Vodicka et al., 2002 [19] compiled results from papers published before 2000, but it seems more relevant to include the original results even though they were published in 1999. A study on antineoplastic drugs, published in 1999 [20], has been cited in several papers after 2000, and it is included in a systematic review [21]. In this case, it has been appropriate to include it in our meta-analysis to compare the results from different types of meta-analysis (i.e. combining comet descriptors to one value *vs.* meta-analysis on separate comet descriptors in each sample).

Fig. 2 shows a flow diagram of the identification of studies in two different databases, screening of studies (after removal of duplicates, papers in non-English language, conference abstracts and other grey literature) and inclusion of studies in the individual systematic reviews and the present umbrella review. Information on reasons for excluding studies are reported in flow diagrams of the original systematic reviews. In total, the systematic reviews cover meta-analyses of 255 studies, which are also included in the present umbrella review. A few studies have investigated complex mixtures involving more than one relevant genotoxic agent and could potentially fit into multiple exposure groups. These have been allocated to the exposure group to which they belong the most. For example, one study has used benzene metabolites as a biomarker of air pollution exposure [22]; thus, this study is allocated in the group of air pollution rather than VOCs in the overall analysis.

2.1.2. Inclusion of studies in the review (step 2)

In each systematic review, at least two independent reviewers have assessed whether studies were eligible for inclusion in the review and meta-analyses. In case of uncertainty about the quality of the study or eligibility of results in meta-analyses, the project leaders (PM and CL) were consulted. Inter-rater agreement was not assessed. Studies have been included in the review if they were not suspected to have selection bias, to have inadequate information on population characteristics (i.e. uncertainty about the type of subjects or relevant confounding factors), or to be influenced by effects of confounders, lack of exposure gradient, much larger exposure gradient of co-exposures, double publication of results, lack of information on the cell type used for the comet assay, or to have results that do not follow the state-of-the-art for studies on the comet assay (e.g. all scored comets used in statistical analyses rather than the mean/median of the comet descriptors per individual).

We have included studies on leukocytes and their subpopulations in the meta-analysis. Studies on buccal cells, sputum cells, epithelial cells, sperm or tissue cells were excluded from meta-analyses, although they may have been included in the systematic reviews as background

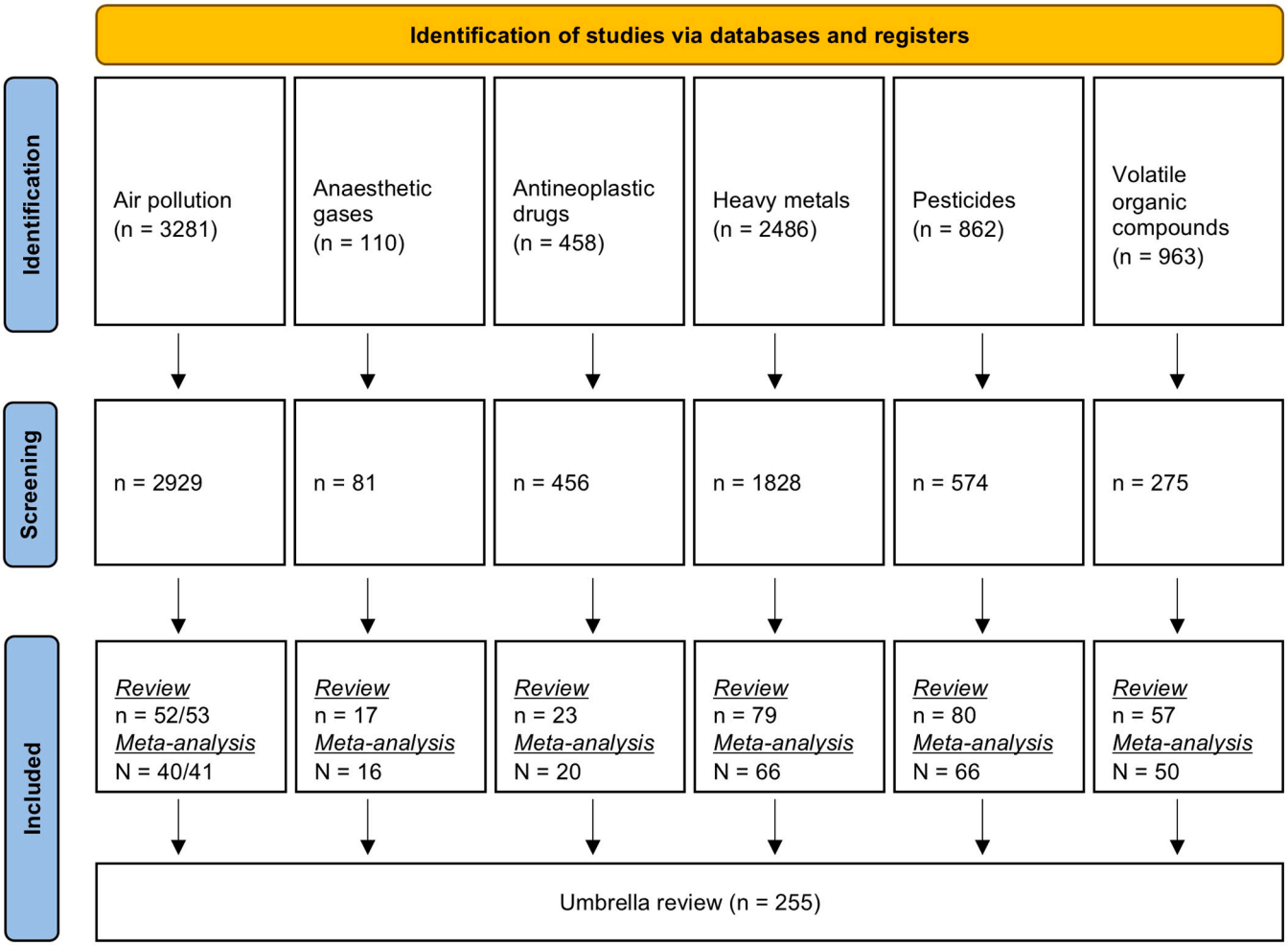

Fig. 2. Flow diagram of literature search and selection of studies for the systematic reviews and meta-analyses. In each of the systematic reviews, studies have been identified by searches in two different databases (PubMed and EMBASE/Web of Science) and other sources (e.g. literature lists of identified studies, unstructured Google searches). Duplicates, papers in the non-English language, and conference abstracts were excluded before the studies were screened for inclusion in the systematic reviews. In the systematic review on air pollution, one study contains two datasets (indicated by a slash). For each systematic review, studies were included in meta-analyses if they reported raw or group-level results. Some studies had sufficient quality to be included in the review, although the results were not applicable for inclusion in meta-analysis (e.g. correlation studies). The umbrella review covers all studies that were included in the meta-analyses in the original systematic reviews.

information or supporting evidence for genotoxic effects.

The effect of confounding factors on levels of DNA strand breaks has been assessed in a separate paper [23]. In the context of biomonitoring studies that assess the effects of exposure (predictor) on DNA strand break levels (outcome), confounders can be seen as genotoxins in association with the predictor. Many factors might affect associations between exposure and DNA strand breaks in the comet assay. However, it is not possible to pinpoint a certain set of confounders that should be controlled in human biomonitoring studies. In general, there are no strong associations or consistent results for even the most evident confounders (e.g. smoking). Researchers may have controlled for confounding by excluding certain populations (e.g. smokers) or using subjects as their own control (e.g. controlled exposure studies or pre-post exposure study design). It should be noted that studies with subjects as their own controls are susceptible to confounding by period effects (e.g. diurnal or seasonal variation), although traditional confounders such as age and sex are controlled. Many studies have controlled for the effect of confounders by matching subjects on age, sex and smoking, or composite variables such as socioeconomic factors. In general, most biomonitoring studies based on comet assay results have used some way of controlling for confounding in the study design. By default, we have not challenged the original efforts to control for traditional confounders as reported in the original papers. As a general note, confounding is an issue when assessing specific agents in complex mixtures. For instance, assessment of genotoxic effects of benzene (a VOC) in urban outdoor air is likely to be confounded by co-exposure to ultrafine particles from vehicle exhaust.

Fig. 3 shows the two most important sources of selection bias in human biomonitoring studies on comet assay results. Selection bias may arise as a consequence of uneven recruitment of subjects or uncontrolled confounding. It is difficult to assess selection bias related to sampling of subjects in biomonitoring studies with comet assay results because the original papers may have insufficient information on the selection of

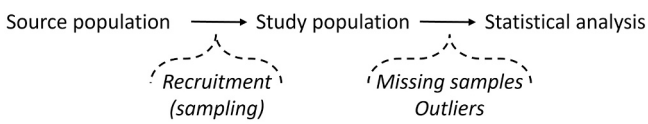

Fig. 3. Types of events that might cause selection bias in human biomonitoring studies with comet assay results. The top row shows the transitions in a dataset from a source population, through a study population, to the dataset used for statistical analysis. During each step (arrow), data vanish from the dataset, causing the final dataset to be a smaller and potentially biased set of results.

subjects from the source population to the study population. Typically, information might consist of the number of exposed and unexposed subjects in the study (i.e. study population) without relevant information on the number of subjects in the source population (e.g. how many workers from the workforce in a plant were recruited and whether they are representative of the exposure situation, including effects of co-exposures and confounders). In addition, selection bias may also occur by statistical analysis if the analysed samples are not representative of the recruited subjects because of drop-outs (i.e. not all subjects from the study population are actually recruited) or loss of data (e.g. due to experimental errors or removal as outliers before statistical analysis) from the sampling step to the statistical analysis. Ideally, studies with missing data should adjust for confounding in statistical analysis, even for factors that are used for matching subjects. However, this is rarely the case as many studies use univariate analysis (often Student's *t*-test or Mann-Whitney *U* test).

A priori, we have assumed that selection bias is not an outstanding problem in human biomonitoring studies on comet assay results because the studies typically use selection criteria (e.g. restriction to non-smokers or sex-matched exposure groups). In cases where uncertainty about confounding has been flagged, it has been evaluated by at least two independent reviewers.

2.1.3. Meta-analysis (step 3)

We have included studies in the meta-analysis if they have sufficient information on central tendency and variability results. This includes group mean and standard deviation (SD) or standard error of the mean (SEM), or median and quartiles or inter-quartile range (IQR). In addition, the number of subjects in each group is necessary information for a meta-analysis. There are fewer studies in the meta-analysis because some of the studies included in the review lack crucial information.

2.1.3.1. Diversity of exposure situations. The studies cover a wide range of exposure scenarios, from cross-sectional studies with little information on specific chemicals to controlled exposure studies (e.g., subjects exposed in a chamber). Controlled exposure studies have well-defined exposure periods and post-exposure sampling time points. Cross-sectional studies on environmental exposures typically entail ongoing exposure (e.g., subjects living in areas with a high concentration of arsenic in drinking water). It is not possible to adjust for time of sampling in the meta-analyses. Studies on occupational exposure tend to collect samples at specific timepoints (e.g., end of work-shift). For studies on occupational exposures, room ventilation, use of closed systems (e.g., for antineoplastic drugs), adherence to safety instructions and use of personal protection equipment are factors that affect the level of exposure. However, these factors are difficult to standardize in exposure assessments across studies. For instance, the importance of personal protection equipment is immediately recognized for antineoplastic drugs, whereas welders may use helmets, gloves and goggles to shield against sparks, but may not protect the worker against inhalation of welding fumes. Personal protection may not even be considered relevant for certain exposures (e.g. traffic-generated air pollution). The difference in exposure situations is a source of inter-study heterogeneity in meta-analyses, especially when exposure levels are based on concentrations of chemicals in air or drinking water. Biomarkers are preferable for the ascertainment of exposure, and they can be used to describe the biologically effective dose. Unfortunately, only for studies on lead has it been possible to assess the relationship between a biomarker of exposure and DNA strand breaks [12]. This analysis demonstrated no relationship between blood lead levels and effect sizes, indicating that co-exposures and confounder factors cause inter-study heterogeneity in effect sizes.

2.1.3.2. Different types of central tendencies. In the present project, we use central tendencies (i.e. mean and median) on three different levels of data. It is important to keep in mind that the central tendencies are used in different levels of a hierarchy of data, described as items (i) to (iii) below.

(i) *Central tendency of the distribution of primary comet descriptors in each sample*. There has been a long tradition among comet assay researchers to discuss how one can best obtain a reliable value of DNA migration in the 50–100 comets that are typically scored in each gel. The typical approach is to calculate the mean or median of the primary comet descriptor results. Other comet descriptors have also been reported, but typically connected to a more or less exploratory assessment of the primary comet descriptor.

(ii) *Central tendency of DNA strand breaks in exposure groups*. Differences in DNA strand breaks between groups with different levels of exposure are the main results in studies. Depending on the distribution of these results, effects could be analysed by either parametric or non-parametric tests. The former would usually imply that results are reported as group mean and SD (or SEM), whereas median and quartiles (or IQR) constitute the central tendency and variability result of choice for non-parametric tests.

(iii) *Standardized mean difference (SMD) between exposure groups in meta-analysis*. The meta-analysis in the present project has used group mean/median and SD/IQR in the individual studies to calculate the SMD for each study. These are subsequently combined into an overall SMD using a parametric test. In the systematic reviews and meta-analysis, we have also calculated the overall median of SMDs from individual studies. This is an unusual – possibly unprecedented - way of reporting the overall central tendency of a meta-analysis. We have used this procedure as an alternative to the traditional parametric test of the overall SMD because the datasets are typically skewed by a few outliers with very high effect size.

2.1.3.3. Geometric mean and geometric standard deviation factor. Certain studies report the geometric mean and geometric SD factor. However, this information cannot be used in meta-analysis because the geometric SD factor is different from the ordinary SD. The geometric SD is a factor that is multiplied/divided by the mean, whereas the ordinary SD is added/subtracted from the mean. In the present project, studies with geometric means of comet results are not included in the meta-analysis.

2.1.3.4. Assessment of effect size. We have used SMD as a measure of the effect size. This descriptor is useful for studies that assess the same outcome in different ways. In this case, DNA migration is measured by different comet descriptors. It describes the effect size in each study relative to the variability in the study (formally, the SMD is a difference between means of the exposed and control groups, divided by the common SD). In other words, the SMD is the effect size in SD units.

We have calculated SMDs using a random-effects model. This model assumes that studies have different effects that follow a normal distribution. It implies that differences in results are due to chance or variation in the unexplained effects. Expressed in other words, if the variation between studies is due to random effects, the variation between studies would follow a normal distribution, and there would not be heterogeneity in effect sizes between studies. Heterogeneity in effect sizes between studies might be due to publication bias or unrecognised dissimilarity between studies. One example of the latter could be personal protection equipment in work situations, which might differ in quality, be used/not used, or be available/not available in different exposure situations.

To calculate SMD for studies with image analysis of comets, we have only used results on tail length, tail intensity (also called percentage of DNA in the comet tail) and tail moment (i.e. product of tail length and tail intensity). Other comet descriptors can also be measured by image analysis, but they are rarely reported or used in biomonitoring studies.

For visual scoring, we have included studies that have used three or more categories. Most studies use five categories for visual scoring. It should be noted that the same level of inter-investigator variation is obtained by scoring the same samples by visual scoring and tail intensity by image analysis [24]. In addition, there is a linear relationship between visual scoring (five-class system) and tail intensity (image analysis) [24]. A few publications report numbers of DNA strand breaks (e.g. lesions per million base pairs) after calibration of the primary comet descriptor with ionising radiation. Calibration of the comet assay using ionising radiation has been described in detail elsewhere [25], including studies on the variation in results when different laboratories use the same calibration samples of irradiated cells [26,27] or use samples from their own irradiation experiments [28].

We have calculated SMD and 95% confidence interval (CI) in a random effects model with inverse-variance weighting (Review Manager 5.4). The program provides measures of inter-study heterogeneity (τ^2, I^2 and Q statistics), whereas tests for small study effect (publication bias) are not included in the software package. We have used Stata 15 (StataCorp LCC, College Station, TX, USA) to assess small study effects by Egger's test, using *esize(cohensd)* and *meta bias, egger* commands. The trim-and-fill method was used to assess publication bias [Stata command: *meta trimfill, estimator(run)*]. In order to assess the robustness of the effects, we have used non-parametric tests (median and 95% CI) of the overall SMD in addition to the traditional calculation of the pooled SMD by parametric test.

2.1.3.5. Standardisation of comet descriptors. Several publications report results with more than one comet descriptor, but the same comet descriptor has not appeared in all publications. For a meta-analysis, this creates a challenge because it is unknown whether certain comet descriptors are better proxy-measures of the level of DNA damage in samples. In fact, discussions on comet descriptors have typically focused on whether their numerical values reflect the visual appearance of comets. Nowadays, tail intensity seems to be the preferred comet descriptor. However, it has to be stressed that primary comet descriptors are not direct descriptors of DNA damage unless they are converted to lesions per million base pairs or similar by calibration with ionising radiation [29].

Previous meta-analyses on comet assay results have typically dealt with the different comet descriptors by analysing effect sizes in separate strata of comet descriptors. However, this introduces a problem in the interpretation of the results if the effect sizes of different comet descriptors differ. In the extreme case, one comet descriptor may show a statistically significant difference between exposed and controls, whereas another descriptor may not show the effect on the same populations.

In addition to reporting results with different comet descriptors, there are other ways in which publications vary in their expression of comet assay results. An example is the use of either the mean or median value of the primary comet distribution of typically 50–100 comets per sample. The median value is typically lower than the mean for the same samples, emphasising that primary comet descriptors do not directly correspond to levels of DNA damage [30]. This aligns with studies showing that means, or medians of the same individual comet distribution usually do not affect the statistical result (i.e. P-value) between exposed and controls [31]. In addition, some studies have assessed DNA strand breaks in more than one cell type. Some studies may even have reported data from both fresh and cryopreserved cells from the same subjects [32].

In a meta-analysis, it is impossible to objectively select a single comet descriptor that represents all others within a study without risking a selective reporting or "cherry picking" situation. Therefore, we have standardised comet descriptors whenever results with more than one comet descriptor have been reported. Table 2 shows an example of the standardisation of comet descriptors [33]. In this case, the original results have been reported as tail length and tail moment values. As can be seen, there is a statistically significant difference in effect size for tail length values (SMD = 0.61, 95% CI: 0.20, 1.01). Using tail moment values gives rise to a statistically non-significant effect (SMD = 0.19, 95% CI: −0.20, 0.58). The average effect size (SMD = 0.33, 95% CI: −0.07, 0.72) is the mean of the two standardised tail length and tail moment results.

Table 2
Calculation of combined (standardised) comet descriptor.

Comet descriptor	Exposed			Controls			SMD
	Mean	SD	N	Mean	SD	N	
TL (original)	21.04	7.30	50	17.57	3.39	50	0.61
TM (original)	0.58	0.40	50	0.51	0.19	50	0.19
TL (stand.)	1.20	0.42	50	1.00	0.19	50	0.61
TM (stand.)	1.14	0.78	50	1.00	0.63	50	0.19
Combined	*1.17*	*0.60*	*50*	*1.00*	*0.41*	*50*	*0.33*

Comet descriptors are tail length (TL) and tail moment (TM) as either original results or standardised against the values in the controls (i.e. 17.57 and 0.51 for tail length or tail moment units, respectively). Standardised mean difference (SMD) has been calculated in RevMan version 5.4. Original results have been published by Rozgaj et al., 2009 [33].

2.1.4. Generalizability of results in the meta-analysis (step 4)

The studies in the meta-analysis only include a fraction of all studies on specific exposures. Therefore, it is relevant to compare genotoxic effects between studies that have been included and excluded from the meta-analysis. For this assessment, we compare the proportion of studies that have reported statistically significant effects in the two strata of studies, using the Z-test. It should be noted that the assessment only extends the generalizability from the meta-analysis to the included studies in the review. It is not a generalizability from the study populations in the included studies to the general population. For instance, studies on genotoxic effects of occupational lead exposure cannot be extrapolated to effects at the whole population level because there could be co-exposures in the occupational settings (i.e. confounding) and there might be vulnerable groups in the whole population, such as children (i.e. effect modification).

2.1.5. Subgroup analysis (step 5)

The subgroup analysis is designed to stratify the dataset into groups with similar characteristics. In the meta-analyses, we have used subgroup analysis to assess whether there are populations with higher exposure and/or susceptibility to DNA damage. This includes populations in countries with different economies and subjects with exposure to complex mixtures. The subgroup analysis serves the purpose of assessing the robustness of the results and/or demonstrating the effect of collecting dissimilar studies in the same meta-analysis.

Preliminary analyses in the first systematic reviews (air pollution and heavy metals) suggested a high degree of inter-study heterogeneity in effect sizes. We also noted that studies on populations from high-income countries tended to have a lower effect size than populations in middle-income countries. It led us to speculate that subjects in high-income countries might be less exposed to the investigated genotoxic agent due to stricter regulation or better protective measures. Moreover, it suggested that the studies could be segregated into different subgroups, leading to less unexplained inter-study heterogeneity in effect sizes.

The World Bank categorisation of income has been used as a proxy measure of countries' socioeconomic status, regulation of environmental and occupational exposures (reflecting economic means to uphold control over emissions) and research infrastructure (determining availability of equipment for exposure assessment). The comet assay in its simplest form is an inexpensive technique, whereas exposure measurement of chemicals may require special equipment that is not available in middle-income countries. All studies in the meta-analyses have been

grouped into populations from high-income and middle-income countries, according to the World Bank classification (https://data.worldbank.org/, downloaded March 2025). The World Bank uses four categories of income (low, lower middle, upper middle and high), which are updated annually. It is our impression that economies have improved in countries over the period that is covered in the present review (years 2000–2025). We have used the highest ranking for the country rather than the rank before the publication of the study. Especially, several European countries have been upgraded from the upper middle-income to the high-income group in the period after the year 2000. This is to some extent tied to desires to join the European Union, which typically has been associated with the implementation of legislation and regulation before formal membership. We have merged lower-middle and upper-middle countries into one category. There are no studies from low-income countries in the meta-analyses, described in the systematic reviews.

Most of the biomonitoring studies deal with complex exposures to chemicals from environmental or occupational settings. For some types of exposure, there are studies that have focused on a particular compound in a complex mixture and studies that have considered the whole exposure situation. An example is lead exposure studies; approximately half of the studies have only assessed exposure to lead (typically blood lead concentration), whereas the rest of the studies have assessed a panel of heavy metals, including lead [12]. However, all lead exposure studies appear to have been conducted in settings with exposure to many heavy metals or other genotoxic agents. Another example is studies on biomass combustion particles, where one group of studies are controlled exposures to clean fuel and optimal combustion conditions, whereas another group is field studies with combustion of mixed biomass under seemingly poor combustion conditions [9].

2.1.6. *Sensitivity analysis (comet assay description and exposure assessment) (step 6)*

The purpose of the sensitivity analysis is to assess the robustness of the overall meta-analysis, considering various potential uncertainties in comet assay experiments and exposure misclassification. In the meta-analysis, we have assessed effects related to the risk of measurement bias in the comet assay procedure, blinding/coding of samples, inclusion of positive controls and exposure assessment. The risk of bias in each study is typically visualized in plots with colours indicating the level of information, such as green, yellow and red indicating low, moderate and high risk of bias, respectively. In this review, the risks of bias and exposure misclassification categories are visually presented as greyscale background colour (low = light grey, moderate = middle grey, high = dark grey).

2.1.6.1. Comet assay description. Only a few tools among the various risk-of-bias-assessment tools for systematic reviews specifically focus on biomarkers, including STrengthening the Reporting of OBservational studies in Epidemiology – Molecular Epidemiology (STROBE-ME) [34] and Risk Of Bias In Non-randomized Studies - of Exposures (ROBINS-E) [35]. However, even these tools, which are dedicated to studies on biomarkers, provide little guidance for systematically reviewing technical issues related to biomarkers of exposure or early biologically effect, including comet assay protocols. Consistent and coherent reporting of comet assay protocols and results should be mandatory, although often it is not the case in published papers. The lack of information on comet assay protocols might have two consequences in systematic reviews: (i) causing inter-study heterogeneity in effect sizes because of differences in comet assay protocols [i.e. a non-differential type of error], and (ii) overrepresentation of significant effects from research groups with insufficient comet assay expertise [i.e. a systematic type of error or bias, assuming that a higher proportion of significant effect sizes are published by researchers with little comet assay experience because they have a higher tendency to exclude null effect results of biomarker assays that are not their primary tools].

A detailed description of the comet assay procedure is desirable in biomonitoring studies. Views on essential information have probably changed over the years. For instance, many early publications highlight that comet assay procedures were carried out under dim or red light to avoid adventitious DNA damage from ultraviolet light in the laboratory, which might have been due to a recommendation of this procedure in an early review of comet assay protocols [36]. However, information on laboratory light is rarely included in publications nowadays. Moreover, to the best of our knowledge, the effect of laboratory light on background levels of DNA strand breaks has not been formally demonstrated to bias comet assay results [37]. A recent report on technical recommendations on the use of the comet assay in human biomonitoring studies did not address laboratory light but emphasized the importance of standardizing a number of steps to reduce variation between experiments. These steps include: (i) the concentration of agarose in which the cells are embedded, (ii) the duration of cell lysis, (iii) the duration of alkaline unwinding, (iv) the duration of electrophoresis, (v) the voltage gradient applied, and (vi) the method used for scoring comets [38]. In a related effort, a group of comet assay researchers have suggested a number of recommendations for describing comet assay procedures and results, called Minimum Information for Reporting on the Comet Assay (MIRCA) recommendations [39]. The recommendations differentiate between desirable and essential information on comet assay procedures and results, which should be described in future publications. The MIRCA recommendation has not been developed as a checklist for systematic reviews, although the items are obviously relevant for the assessment of comet assay procedure descriptions.

In the systematic reviews, we have used a comet assay description scoring system, based on the MIRCA recommendations, rated as

Table 3
The comet assay procedure scoring system used in the present meta-analysis of environmental and occupational exposures.

Item	Max
Type of cells (including cell type and isolation procedure)	1 (0.5)[a]
Temperature and time conditions prior to processing of samples	1 (0.5)[a]
Storage (if specimens have been cryopreserved)[b]	1
Final low melting point agarose concentration	1
Lysis buffer composition	1
Alkaline solution composition	1
Temperature of the solution	1
Duration of alkaline treatment	1
Electrophoresis solution composition	1
Electrophoresis condition (V/cm)	1
Temperature during electrophoresis	1
Duration of electrophoresis	1
Type of stain used	1
Type of primary comet assay descriptor	1
Number of comets scored per gel and number of gels scored	1 (0.5)[a]
Type of software for image analysis[c]	1
Mean/median of comets (when image analysis systems have been used for analysis of DNA migration)[d]	1
Statistical analysis of results	1
Sum	18[e]

The scoring system is based on essential information, described in the Minimal Information for Reporting Comet Assay (MIRCA) results and procedure recommendations [39].
[a]The item contains two types of information, each of 0.5 points.
[b]The item is not relevant for studies on fresh samples (i.e. the item is not included for studies on fresh samples).
[c]Only relevant for studies with image analysis systems (i.e. the item is not included for studies on visual scoring).
[d]The visual scoring system is equivalent to the product of the mean and number of comets in each class (i.e. studies on visual score is *per se* 1 point).
[e]The maximal score can be 16–18 points, depending on storage condition and the comet scoring system.

essential information for the standard comet assay. Table 3 outlines the essential information of the MIRCA recommendations. Certain MIRCA items contain two types of information, which count as 0.5 point each.

At first sight, many papers have clear descriptions of the comet assay procedure, but a scrutiny of the text opens up different interpretations. Table 4 shows some examples of texts where uncertainty arises in otherwise well-written method descriptions (the examples are anonymised). A lack of precision creates uncertainty in the comet assay description scoring system because it is based on the interpretation of texts. We have ameliorated the uncertainty by having two persons assess each paper in the meta-analysis. In addition, we have categorised studies as having low ($\leq$ 25% missing information), moderate ($>$ 25% to $\leq$ 50% missing information) or high ($>$50% missing information) risk of bias. The thresholds take into consideration that some interpretation of the comet assay procedure description occurs, and there are items where the reporting practice has changed over the years (e.g. it is now recommended to report V/cm for electrophoresis, whereas the voltage alone used to be sufficient information in earlier papers). The present review describes the risk of bias in clusters of studies, based on a similar type of exposure or country-income category. We have visualized the risk of bias in graphs by as greyscale background colour (light grey = low risk of bias).

Table 4
Examples of comet assay procedure descriptions.

Context	Description in the paper	Interpretation
The final concentration of agarose is not specified, only the percentage of agarose in the stock solution	*Example 1*: "…a second layer containing the whole blood sample mixed with 0.5% low melting point (LMP) agarose (Sigma) was placed on the slide".	It is not possible to infer the final concentration (percentage) of the final concentration of agarose. However, it is possible that the authors have used a very small volume of blood (i.e. the concentration is essentially 0.5%) or blood cells have been centrifuged and resuspended in 0.5% LMP agarose
	Example 2: "Between 10,000 and 30,000 cells were embedded in 0.6% low melting agarose on a fully frosted microscope slide, precoated with 1.5% normal melting agarose".	The text only mentions "cells" as shown in the quote. It is not specified whether the cells are leukocytes (whole blood) or isolated cells. Without information on the volume of cell suspension, it is impossible to assess the final percentage of agarose
The temperature of the alkaline solution can be interpreted from the text, although not specified	*Example 3*: "The tank was filled with fresh electrophoresis solution (1 mM Na_2EDTA and 300 mM NaOH, pH 13) to a level approximately 0.25 cm above the slides. Before electrophoresis, the slides were left in the solution for 20 min to allow unwinding of the DNA and expression of alkali labile damage. Electrophoresis was conducted at low temperature (4 °C) for 20 min…"	The description does not specify the temperature of the alkaline unwinding. However, it seems odd that the authors would first do the alkaline unwinding step in the electrophoresis tank with a solution, which is not 4 °C, and then replace this with a solution at 4 °C
	Example 4: "The slides were removed from the lysing solution and placed in a horizontal gel electrophoresis tank filled with fresh alkaline buffer (1 mM Na_2EDTA and 300 mM NaOH, pH 13) for 20 min at 4°C to allow denaturing and unwinding of the DNA, and the expression of alkali-labile sites. Electrophoresis was performed in the same buffer at 20 V and 300 mA for 20 min…"	The description does not specify the temperature of the electrophoresis, although it seems odd that the authors would allow the temperature to rise during the relatively short electrophoresis

2.1.6.2. Blinding (coding) samples. Comet assay slides should be analysed blindly to avoid measurement bias. Alternatively, samples should be coded/blinded when isolated from humans in a biomonitoring study. Blinding/coding samples before analysis is a fundamental aspect of conducting and reporting scientific studies. As blinding/coding is a binary factor, the risk of measurement bias is low (light grey) and high (dark grey).

2.1.6.3. Positive (assay) controls. Biomonitoring studies do not have positive controls, as it would be unethical to expose humans to genotoxic and potentially carcinogenic agents. Nevertheless, it is important to assess inter-assay variation and demonstrate that the comet assay can detect DNA damage under conditions used in the laboratory. As these control samples should be stable over time, it is recommended to use frozen cells [40]. For the assessment of inter-assay variation, it is best to use samples that resemble the test samples in the level of DNA damage. As positive assay controls, any type of direct-acting genotoxic agent can be used (e.g. hydrogen peroxide or methyl methanesulfonate) to expose cells, including cell lines or human blood cells. In the present review, we have regarded both types of controls as sufficient information for methodological assessment. We have segregated the risk of bias into low (light grey; results on assay controls are reported), moderate (middle grey; assay controls are mentioned, but the results are not shown) and high (dark grey; no report of assay controls).

2.1.6.4. Exposure ascertainment. There is no universal tool to assess exposure, which applies to all types of exposures, samples and study designs in this review. The studies have used a variety of exposure assessment methods, from exposure categories without measurement of chemicals to clinically used biomarkers such as blood lead concentrations. We have considered exposures to be different between groups if there has been a statistically significant difference, or if the information on study design has been sufficiently convincing to believe the subjects had been exposed, and confounding by co-exposures was controlled. In studies of complex exposures, we have applied expert judgment to distinguish between the sizes of exposure contrasts of chemicals. For instance, certain papers may have a minor difference in benzene exposure and a larger effect with other VOCs. In this case, the benzene exposure contrast would be considered to be confounded by other VOCs, and the study would be categorised as "other VOCs" rather than a benzene exposure study.

We have categorized the exposure assessment into three levels with increasing certainty of exposure: (i) grouping based on work categories or geographical areas, (ii) environmental monitoring assessment, such as pollutants in air, contaminants on surfaces in a room, or personal dosimeters, and (iii) biomarkers of exposure (internal and/or effective dose). The categories align with the assessment of exposure from the population level (1st category) to individual level (3rd category). Nevertheless, biomarkers are rarely specific indicators of sources of exposure. For certain exposures, it is necessary to combine biomarkers of exposure with other categories. For instance, urinary excretion of 1-hydroxypyrene is considered a biomarker of polycyclic aromatic hydrocarbon (PAH) exposure, which may originate from inhalation of combustion products or dietary components. Thus, urinary excretion of

1-hydroxypyrene can only provide high certainty of exposure in cases with suspected inhalation exposure (e.g. occupational exposure or controlled environmental exposure studies) or when confounders are controlled (e.g. avoiding smoked food). In certain cases, biomarkers of early biological effect can be regarded as a proxy measure of the actual exposure (e.g. acetylcholine esterase activity in blood is a biomarker of exposure to certain pesticides). Nevertheless, we have not used biomarkers of early biological effect as exposure markers in our systematic reviews because of uncertainty about confounding (e.g. acetylcholine esterase activity in blood may be confounded by fruit/vegetable intake).

2.2. Statistical analysis of results in the present umbrella review

In the present umbrella review, we summarize results from previously published meta-analyses on air pollution, anaesthetic gases, antineoplastic drugs, heavy metals, pesticides and VOCs [9–14]. Specifically, for the umbrella review, we have done a number of analyses on the full dataset to assess the effects of exposures, locations, comet assay measurement bias and exposure misclassification. For these analyses, we have used SMDs from original studies as raw data (n = 255). Parametric analyses have been done on cube root-transformed results because of heterogeneity in effect sizes, and some studies have SMDs with negative values. The distribution of cube root-transformed data looks more like a normal distribution than untransformed (results not shown). This is also seen in the Shapiro-Wilk W test, where the W-values are 0.69 and 0.90 for raw and cube root-transformed results, respectively. The W-value has a range from 0 to 1, where high values suggest normally distributed data. We have also used logistic regression where the predictor has been dichotomized according to below/above the mean or median in the whole dataset (n = 255) or datasets of separate exposure groups. Statistical analyses, including linear or logistic regression, Spearman's and Kendall's rank correlation tests, and ANOVA, were done in Stata 15 (StataCorp LCC, College Station, TX, USA).

The umbrella review assesses whether different environmental/occupational exposures are associated with differences in the magnitude of effect (exposure/dose to outcome relationship). However, we cannot use the same dosimetry for comparing diverse exposure situations, such as occupational exposure to anaesthetic gases and environmental exposure to arsenic. Inference of genotoxic potency by different exposure/chemicals should be supported by animal or *in vitro* studies with well-defined doses and exposure time points.

3. Results and discussion

3.1. Inter-study heterogeneity and publication bias

Overall, the meta-analyses cover 255 studies. For the meta-analyses on air pollution, heavy metals and VOCs, we have pooled exposure groups because of relatively small sample sizes within each of the exposure groups. Table 5 shows the number of studies, number of subjects and inter-study heterogeneity measures (τ^2, I^2 and Q statistics). Table 6 shows results from two indicator tests of publication bias (Egger's test and trim-and-fill method). In the systematic reviews, most meta-analyses have inter-study heterogeneity (τ^2, I^2 and Q statistics). This is typically seen as an unsymmetric Funnel plot. Fig. 4 shows a typical example of the Funnel plots, based on studies on styrene exposure (original results are reported by Giovannelli et al. [14]). As can be seen, the Funnel plot is skewed, showing an inverse relationship between the effect size (x-axis) and precision (y-axis). This is also supported by Egger's test, which assesses Funnel plot asymmetry; the intercept value (β) differs from zero (β ± SE: 9.88 ± 2.31, z = 4.27, P < 0.001). Egger's test is statistically significant on datasets on anaesthetic gases, antineoplastic drugs, heavy metals, pesticides and VOCs (Table 6). The trim-and-fill method estimates the number of missing studies in unsymmetric Funnel plots and calculates the adjusted effect size by imputing the estimated effects of the missing studies into the meta-analysis. The trim-and-fill method indicates that many of the meta-analysis have missing studies and there is lower effect size than the unadjusted SMDs (Table 6). It should be emphasized that it is an estimated rather than actual number of missing studies, assuming that publication bias is the cause of Funnel plot asymmetry. It is likely that the trim-and-fill method overestimates the magnitude of publication bias. For instance, the trim-and-fill method estimates that there are 12 missing studies in the meta-analysis on antineoplastic drugs; all studies are predicted to show lower levels of DNA strand breaks in exposed subjects (estimated SMD in the range of −3.9 to −0.5). Moreover, according to the trim-and-fill method there should be 75 missing studies, which seems relatively high as it suggests that approximately 25% of the results from biomonitoring studies are not published [75/(75 +255) ≈ 25%].

The combination of Funnel plot asymmetry, Egger's test and trim-and-fill method suggests a risk of publication bias, although it is not definitive proof that human biomonitoring studies with positive comet assay results have a higher chance of being published. For instance, DNA strand breaks may be one endpoint among other biomarkers in a human biomonitoring study. Therefore, a null effect result on DNA strand breaks might not be critical if, for example, results with the enzyme-linked comet assay or other DNA damage endpoints show statistically

Table 5
Number of studies and subjects, and tests for inter-study heterogeneity and Funnel plot asymmetry.

Exposure	Number (studies)	Number (subjects)	Test for heterogeneity			Egger's test for Funnel plot asymmetry	
			τ^2	I^2	Cochran's Q (P-value)	Intercept (β) ± SE	Z-value (P-value)
Air pollution (combined)	40	4983	0.77	95%	780.7 (p < 0.001)	2.78 ± 2.11	1.32 (P = 0.19)
Traffic-derived air pollution	21	2481	0.33	89%	148.5 (P < 0.001)	0.90 ± 1.60	0.57 (P = 0.75)
Biomass	10	891	2.51	97%	308.1 (P < 0.001)	4.07 ± 5.29	0.77 (P = 0.44)
Coke oven	10	1622	0.70	96%	204.7 (P < 0.001)	1.42 ± 6.35	0.22 (P = 0.82)
Anaesthetic gases	16	1238	0.79	93%	200.7 (P < 0.001)	9.30 ± 2.87	3.24 (P < 0.01)
Antineoplastic drugs	20	1502	1.29	95%	375.3 (P < 0.001)	9.53 ± 1.96	4.87 (P < 0.001)
Heavy metals (combined)	66	9558	1.46	97%	2431 (P < 0.001)	11.4 ± 1.71	6.64 (P < 0.001)
Lead	35	4913	2.37	98%	1694 (P < 0.001)	16.4 ± 2.71	6.05 (P < 0.001)
Arsenic	14	2187	0.57	95%	232.2 (P < 0.001)	7.65 ± 1.96	3.91 (P < 0.001)
Chromium/welding fumes	20	2664	1.49	97%	647.0 (P < 0.001)	11.1 ± 2.40	4.62 (P < 0.001)
Other heavy metals	6	716	0.15	78%	22.6 (P < 0.001)	1.37 ± 2.10	0.65 (P = 0.52)
Pesticides	66	8325	1.73	97%	2375 (P < 0.001)	10.2 ± 0.99	10.3 (P < 0.001)
Volatile organic compounds (combined)	50	4896	1.48	96%	1361 (P < 0.001)	11.4 ± 1.03	11.0 (P < 0.001)
Benzene	18	2068	1.99	98%	822.1 (P < 0.001)	12.2 ± 1.68	7.27 (P < 0.001)
Formaldehyde	8	807	0.53	92%	89.3 (P < 0.001)	−0.71 ± 5.82	−0.12 (P = 0.90)
Other VOCs	13	840	1.33	93%	108.1 (P < 0.001)	11.6 ± 1.41	8.22 (P < 0.001)
Styrene	11	717	1.05	93%	146.2 (P < 0.001)	9.88 ± 2.31	4.27 (P < 0.001)

Table 6
Number of studies and subjects, and tests for inter-study heterogeneity and Funnel plot asymmetry.

Exposure	Number (studies)	Egger's test for Funnel plot asymmetry		Trim-and-fill method	
		Intercept (β) ± SE	Z-value (P-value)	Missing studies	SMD (95% CI), % change compared to unadjusted effect size[a]
Air pollution (combined)	40	2.78 ± 2.11	1.32 (P = 0.19)	2	0.70 (0.49, 1.20), −32%
Traffic-derived air pollution	21	0.90 ± 1.60	0.57 (P = 0.75)	4	0.39 (0.07, 0.72), −52%
Biomass	10	4.07 ± 5.29	0.77 (P = 0.44)	0	1.76 (0.30, 3.23), 0%
Coke oven	10	1.42 ± 6.35	0.22 (P = 0.82)	1	0.73 (0.67, 0.89), 0%
Anaesthetic gases	16	9.30 ± 2.87	3.24 (P < 0.01)	5	0.53 (−0.14, 1.21), −55%
Antineoplastic drugs	20	9.53 ± 1.96	4.87 (P < 0.001)	12	0.13 (−0.53, 0.80), −90%
Heavy metals (combined)	66	11.4 ± 1.71	6.64 (P < 0.001)	30	0.73 (0.68, 0.77), −35%
Lead	35	16.4 ± 2.71	6.05 (P < 0.001)	7	1.14 (0.17, 2.10), −44%
Arsenic	14	7.65 ± 1.96	3.91 (P < 0.001)	4	0.87 (0.24, 1.50), −38%
Chromium/welding fumes	20	11.1 ± 2.40	4.62 (P < 0.001)	9	0.84 (−0.03, 1.69), −59%
Other heavy metals	6	1.37 ± 2.10	0.65 (P = 0.52)	NA	NA
Pesticides	66	10.2 ± 0.99	10.3 (P < 0.001)	23	0.83 (0.09, 1.57), −62%
Volatile organic compounds (combined)	50	11.4 ± 1.03	11.0 (P < 0.001)	3	1.02 (−0.04, 2.08), −37%
Benzene	18	12.2 ± 1.68	7.27 (P < 0.001)	1	1.21 (−0.89, 3.14), −36%
Formaldehyde	8	−0.71 ± 5.82	−0.12 (P = 0.90)	NA	NA
Other VOCs	13	11.6 ± 1.41	8.22 (P < 0.001)	1	2.56 (0.84, 4.28), 0%
Styrene	11	9.88 ± 2.31	4.27 (P < 0.001)	0	0.95 (0.01, 1.88), 0%

[a] The trim-and-fill method estimates the number of missing studies in the dataset and estimate the standardized mean difference (SMD) in a dataset with the missing studies included. The results are not analysed (NA) on meta-analysis with fewer than 10 studies. Negative percent change indicates the adjusted SMD is lower than the unadjusted SMD.

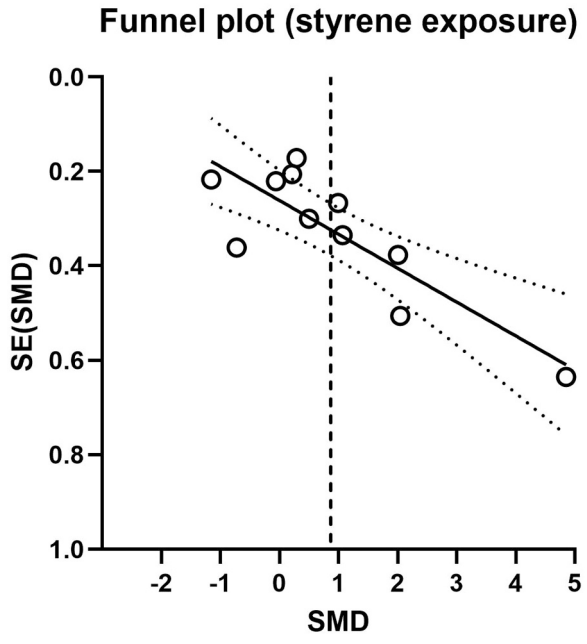

Fig. 4. Example of a Funnel plot with unsymmetrical distribution of results. Each symbol in the plot is one study (n = 11). The graph depicts the precision, i.e. standard error (SE) of the standardised mean difference (SMD) as a function of the effect size (SMD). As can be seen, there is asymmetry around the central tendency of the effect size (dotted line). There appears to be a tendency where the studies (symbols) spread from the upper left-hand (i.e. studies with low effect and high precision) to the lower right-hand corner (i.e. studies with high effect and low precision), as also indicated by the regression line ($r = 0.84$, $P < 0.01$, linear regression). The graph has been reproduced from a Funnel plot in RevMan.

significant effects. On the contrary, it can be speculated that there is a selective reporting of statistically significant comet descriptors, which might lead to inter-study heterogeneity in the reported results. Taking the issues with the reporting of comet assay results into consideration, we think it is likely that inter-study heterogeneity is related to differences in study designs (e.g. cross-sectional *vs.* controlled exposure), exposure circumstances (e.g. environmental *vs.* occupational exposure), populations (e.g. workers, general population or susceptible groups) and uncontrolled confounding.

Another source of inter-study heterogeneity in the meta-analyses is a difference in descriptive statistics, reported as either parametric (mean and SD) or non-parametric (median and IQR). Many studies do not report descriptive results on distributions (histograms or Q-Q plots) or statistical tests for normal distribution of data to justify the reporting of mean/SD or median/IQR results. This introduces an uncertainty as datasets may not have normal distributions even though the results have been reported as mean and SD. In the meta-analyses, we have used median and IQR as proxy for mean and SD in studies that have reported non-parametric descriptive statistics. There are various computations to estimate the SD from IQR, which typically converges to SD ≈ IQR/1.35 on model datasets [41,42]. In theory, a direct conversion of IQR to SD introduces a slight overestimation of the variation in studies with non-parametric descriptive statistics. This is preferable because results reported as median and IQR may have a higher proportion of outliers than datasets that have a mean and SD. In the full dataset, 22 out of the 255 studies (8.7%) have reported median and IQR values. As the studies with median/IQR values are not equally represented in all exposure and country-income groups, we have assessed both raw SMDs, and SMDs that have been normalised within each exposure group [$SMD_{normalised} = (X_{i,j} - \mu_{i,j}) + \mu_{all}$; $X_{i,j}$ is the SMD of the *i*'th exposure group and the *j*'th country-income group, $\mu_{i,j}$ is the average value of studies that have used mean/SD in the *i*'th exposure group and the *j*'th country-income group, and μ_{all} is the average of all results in the dataset]. Fig. 5 shows histograms of the datasets with mean/SD and median/IQR as raw SMDs, and results that have been normalised for differences in the distribution of exposure and country-income groups. The effect size in the studies with median/IQR is smaller (raw SMD ± SD: 0.95 ± 1.33, n = 22) than the effect size in the studies with mean/SD (raw SMD ± SD: 1.75 ± 2.38, n = 233), although this difference is not statistically significant (P = 0.12, Student's *t*-test). Adjustment for inequal distribution of studies in different exposure and country-income groups shows a borderline statistical significance (P = 0.06, Student's *t*-test) for lower effect size in studies with median/IQR (normalised SMD ± SD: 0.79

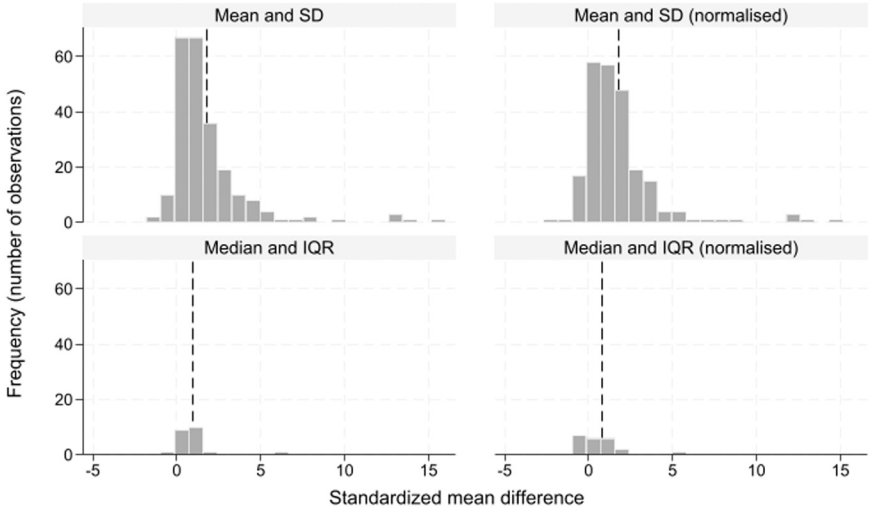

Fig. 5. Histograms of standardized mean difference (SMD) results in studies that have reported mean and standard deviation (SD) or median and inter-quartile range (IQR). The results are raw values as reported in the publications (left) or normalised values (right). The results have been normalised within each exposure group (air pollution, anaesthetic gases, antineoplastic drugs, heavy metals, pesticides and volatile organic compounds). Vertical dotted lines are mean values of the distributions.

± 1.42, n = 22) as compared to studies with mean/SD (normalised SMD ± SD: 1.75 ± 2.24, n = 233). It should be noted that the studies with median/IQR values have much smaller effect sizes (raw SMD = 0.95 and normalised SMD = 0.79) than the expected size from a conversion of IQR to SD, using a factor of 1.35 (expected SMD = 1.75/1.35 = 1.30). Combined, the studies that have reported comet descriptors as median/IQR values have a slightly lower effect size than the studies that have reported mean/SD. However, this difference in effect sizes is not only attributable to systematic error of using SDs as a proxy-measure of IQRs.

Overall, the assessment of inter-study heterogeneity and publication bias indicates that the central tendency of the SMD in each of the exposure groups is possibly overestimated. In general, all of the meta-analyses have inter-study heterogeneity, right-hand skewed Funnel plots, and lower effect after adjustment for Funnel plot asymmetry. It is the same direction in all datasets, suggesting the comparison of effect sizes between exposure groups is probably not affected by differences in inter-study heterogeneity and publication bias.

3.2. Assessment of standardised mean differences between different comet descriptors

The studies in the review and meta-analysis have used diverse comet descriptors. An important question is whether these comet descriptors give rise to differences in effect sizes. In principle, the comet descriptors on the same samples should give rise to the same SMD if they reflect the level of DNA damage in the samples. However, to the best of our knowledge, this has not previously been tested. In the present study, we have collected information on tail length, tail intensity, tail moment and visual score (damage index) from the 76 publications where the meta-analysis is based on more than one comet descriptor (original results are available in references [9–14]). Some image analysis programs provide two different values of tail moments (typically called tail moment and Olive tail moment). For this analysis, we have used the average SMD in studies with more than one tail moment value. Fig. 6**A** shows the SMDs of the four comet descriptors. Table 7 shows results from pair-wise comparisons of the comet descriptors. There is no difference between paired samples of tail length, tail intensity and tail moment ($P > 0.05$, paired-sample Student's *t*-test), whereas there are statistically significant relationships between the comet descriptors ($P < 0.001$, linear regression and non-parametric Spearman and Kendall rank correlation tests). There are only seven studies that have reported results from both visual scoring and image analysis. It is not meaningful to assess pair-wise effect sizes on each of the comet descriptors. However, a comparison of SMDs of visual scoring (2.08 ± 0.90) and image

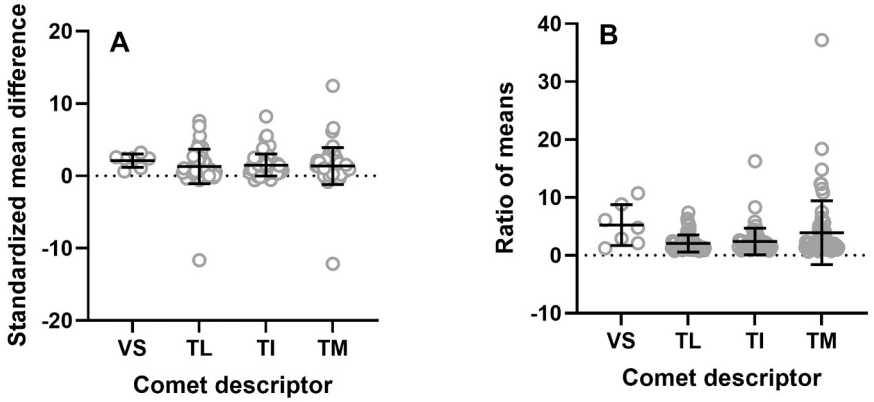

Fig. 6. Standardised mean difference (SMD (A)) and ratio of means (RoM (B)) between exposed and unexposed subjects in studies that have assessed DNA migration by more than one comet descriptor. These cover visual score (VS, n = 7), tail length (TL, n = 61), tail intensity (TI; percentage of DNA in the comet tail, n = 58) and tail moment (TM, n = 66). The symbols represent individual studies. The lines and whiskers are the mean and standard deviation.

Table 7
Effect sizes as standardized mean difference or ratio of means in studies that have reported more than one comet descriptor.

Comparison	N	Tail length (TL), mean ± SD	Tail intensity (TI), mean ± SD	Tail moment (TM), mean ± SD	Linear regression (*r*-value, 95% CI)	Rank correlation (ρ/τ values)
Standardized Mean Difference						
TL vs TI	46	0.95 ± 2.36	1.40 ± 1.49$^{\#}$	NA	0.70 (0.52, 0.82)***	0.87/0.72***
TL vs TM	53	1.30 ± 2.48	NA	1.24 ± 2.67	0.92 (0.86, 0.95)***	0.80/0.64***
TI vs TM	50	NA	1.49 ± 1.50	1.27 ± 2.76	0.76 (0.61, 0.86)***	0.88/0.74***
Ratio of Means						
TL vs TI	46	2.06 ± 1.61	2.36 ± 2.35	NA	0.60 (0.38, 0.76)***	0.67/0.51***
TL vs TM	53	2.06 ± 2.51	NA	4.26 ± 6.06**	0.69 (0.52, 0.81)***	0.79/0.61***
TI vs TM	50	NA	2.51 ± 2.44	4.29 ± 6.09**	0.85 (0.74, 0.91)***	0.86/0.70***

Standardized Mean Difference (SMD) and Ratio of Means (RoM) values cannot be compared. P-values are based as paired-sample Student's *t*-tests ($^{\#}P = 0.08$, $^{**}P < 0.01$). Correlation coefficients (*r*-values) are based on linear regression ($^{***}P < 0.001$). Spearman (ρ) and Kendall (τ) values are based on rank correlation tests ($^{***}P < 0.001$ for both Spearman and Kendall tests).

analysis (2.30 ± 1.45) does not indicate a difference ($P > 0.05$, paired-sample Student's *t*-test, mean ± SD, average SMDs have been used in studies with more than one comet descriptor for image analysis).

In meta-analysis studies, SMD has been the method of choice for continuous outcomes which are not expressed on the same scale. However, the dependency on SD for the calculation of SMD is not desirable because a large effect can, in principle, be obtained by a small SD, reflecting technical variability rather than biological variability. This may also be a hidden source of inter-study heterogeneity if SEMs have mistakenly been reported as SDs. It might also be difficult to interpret the size of SMDs because it refers to a difference in SD units. The Ratio of Means (RoM) method was developed as an alternative to the SMD. It has the advantage that the calculation of RoM is independent of SD, and the outcome is contextually interpreted in a similar way to the odds ratio or relative risk. The disadvantage of RoM is that it can only be calculated for values with the same sign (e.g. zero value in the control group cannot be used). Fig. 6**B** shows the distribution of RoM results. The analysis of RoM results suggests a larger heterogeneity as compared to SMD results, especially for tail moment and visual score. In addition, effect sizes in terms of tail moment are higher than the effect sizes by tail length and

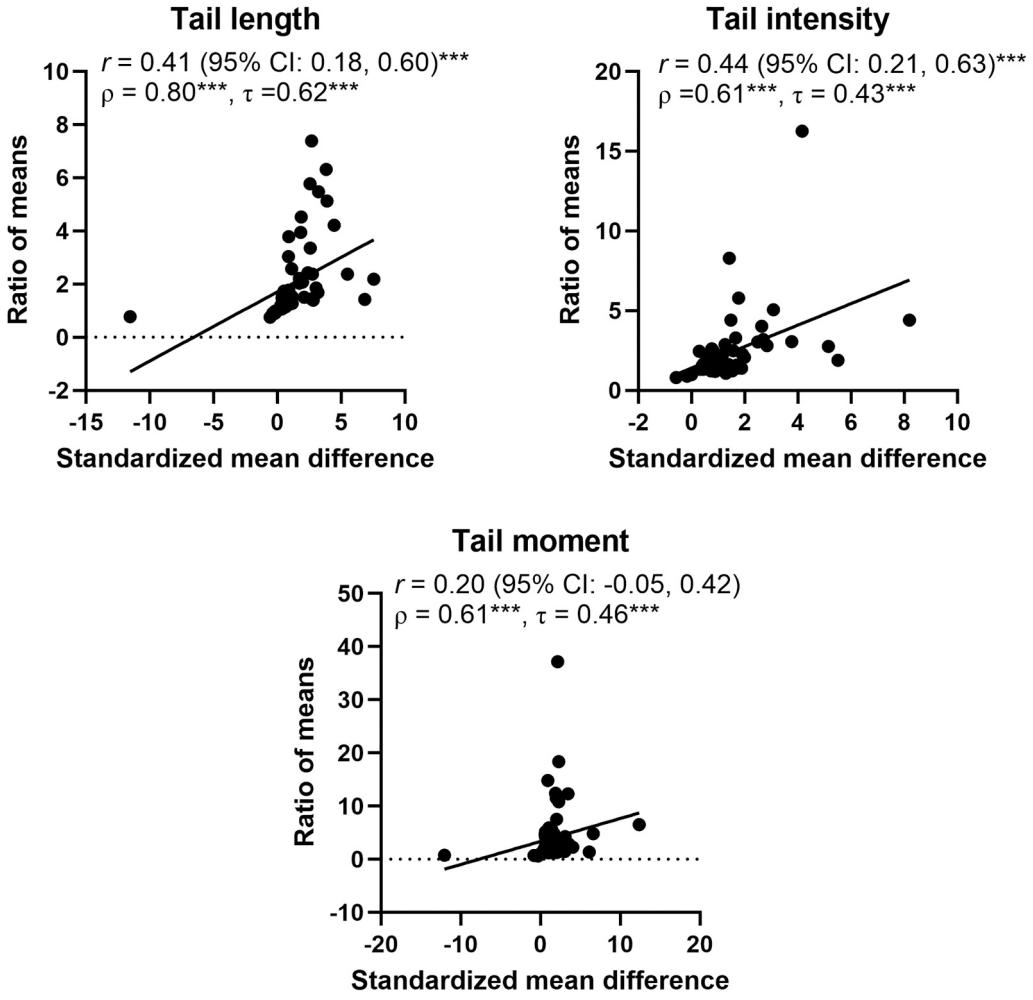

Fig. 7. Correlations between standardized mean difference (SMD) and ratio of means (RoM) values in datasets of tail length, tail intensity and tail moment. Symbols are individual studies (n = 61, 58 and 66 for tail length, tail intensity and tail moment, respectively). Correlations are reported as *r*-values and 95% confidence intervals (linear regression, $^{***}P < 0.001$, slope of the relationship) as well as Spearman (ρ) and Kendall (τ) values of rank correlation tests ($^{***}P < 0.001$).

tail intensity on the same datasets (lower part of Table 7). In general, there are strong linear relationships between the comet descriptors for both RoM and SMD (*r*-values between 0.60 and 0.92, Table 7). There are also statistically significant relationships between RoM and SMD values in each of the comet descriptor datasets, but *r*-values between 0.20 and 0.44 (Fig. 7) indicate that the common source of variability only explains between 4% (tail moment) and 19% (tail intensity) of the relationship between RoM and SMD. **Figure 7** suggests that linear regression is not an optimal parametric analysis for this type of results. Still, assessment of association by non-parametric rank correlation tests indicates that Spearman (ρ) and Kendall (τ) values – equivalent to correlation coefficients - are not substantially different from *r*-values from parametric tests. The Kendall rank correlation is more robust for small sample sets, and it is not sensitive to outliers. The parametric and non-parametric analyses do not indicate that a meta-analysis on RoM produces more reliable results than the SMD results.

Collectively, the results indicate that effect sizes are similar between different comet descriptors for SMD results. Importantly, the results imply that meta-analysis, using a random effect model on SMD results, is a reliable tool for assessing central tendencies of comet assay results in terms of tail length, tail intensity, tail moment and visual score. The inconsistent results on the RoM approach for tail moment values compared to tail length and tail intensity values suggest a need for further analyses on datasets with equal representation of all three comet descriptors, and assessment of how different ways of calculating the tail moment affect the distribution of results.

3.3. Generalizability

A number of studies have been excluded from the meta-analysis, although included in the systematic reviews. The main reasons for excluding studies have been (i) results reported as geometric means and the corresponding type of variability for parametric tests or (ii) variability reported as range for non-parametric tests. Table 8 shows a summary of the proportions of positive test results in the six exposure groups and the pooled results from all exposures. Overall, there is no difference between reported effects in studies that have been included or excluded from the meta-analyses. There is no reason to suspect that the included studies constitute a selected group with overrepresentation of results in one direction.

Table 8
Proportion of positive test results in studies included in the meta-analysis and other studies in the review.

Exposure	Meta-analysis	Review	Statistics (Z-value)
Air pollution	68% (28/13)	75% (9/3)	0.09 (ns)
Anaesthetic gases	63% (10/6)	100% (1/0)	NA
Antineoplastic drugs	65% (13/7)	100% (3/0)	NA
Heavy metals	90% (61/7)	70% (14/6)	1.61 (ns)
Volatile organic compounds	86% (43/7)	100% (7/0)	1.36 (ns)
Pesticides	89% (59/7)	86% (12/2)	< 0 (ns)
Overall	82% (214/47)	84% (53/11)	< 0 (ns)

The table shows the proportions of positive results (in percent). The numbers of positive and negative (or null) results are shown in brackets, separated by a slash (/). Studies have been counted as positive in cases where different results have been obtained on different comet descriptors. The overall number of results in the meta-analyses (n = 261) is slightly higher than the total number of studies (n = 255) because a few studies have been included in more than one exposure group. Statistical analysis is based on differences in proportions, using binominal distribution and correction for continuity (critical Z-value for statistical significance is 1.96).

3.4. Summary of effect sizes

Fig. 8 shows the difference in effect sizes by exposure to different occupational or environmental agents (actual results are available in the original systematic reviews [9–14]). The SMD ranges from 0.39 (95% CI: −0.15, 0.92) to 2.14 (95% CI: 1.48, 2.81) (Fig. 8**A**). In general, there are asymmetric (non-Gaussian) distributions, with some studies having very high effect sizes. Using a non-parametric distribution (SMD_{median}) gives a smaller range between the lowest (0.47, 95% CI: 0.24, 0.79) and highest effect (1.48, 95% CI: 0.71, 3.52) (Fig. 8**B**). Smaller effect sizes are typically obtained with a non-parametric test than when parametric tests are used (48% lower effect, according to the slope of the regression line in Fig. 8**C**). There are probably many reasons for the heterogeneity in effects between studies, including differences in exposure levels, use of personal protection equipment and uncontrolled confounding, as also previously reported in other studies [21,43,44]. However, heterogeneity in effect size and variation between studies may also be the result of publication/reporting bias and erroneous reporting of SEM as SD.

In the systematic reviews, we have noted that studies from high-income countries (mainly European) have a lower effect size than studies from middle-income countries. Fig. 9 shows the effect sizes of pooled exposure groups, stratified into subgroups according to the World Bank ranking of the income level. There is a relatively large difference in effect sizes between middle-income and high-income countries. In fact, this difference is larger than the difference between exposure groups within strata of middle- and high-income countries. As judged by the overlapping 95% CIs, statistical tests indicate no difference between exposures. However, it is important to note that a somewhat similar ranking of effect sizes is seen within each stratum of middle- and high-income countries. As a further test, we have assessed differences in rankings in terms of critical values (i.e. difference between the highest and lowest effect size, divided by intervals). Based on this evaluation, we find that effect sizes fall into three groups for both high-income and middle-income countries. For the high-income countries, the groups are air pollution, anaesthetic gases and antineoplastic drugs (lowest effect), heavy metals and VOCs (middle effect), and pesticides (highest effect). For the middle-income countries, the groups are air pollution and anaesthetics (lowest effect), antineoplastic drugs, heavy metals and VOCs (middle effect), and pesticides (highest effect) (Fig. 10).

In datasets on anaesthetic gases and pesticides, there appears to be a lower effect size in studies from Brazil and/or the Americas than in other middle-income countries (mainly Asian) [14]. The same tendency is also seen in other exposure groups, although it is not as clear because of smaller effects or more variation. Table 9 shows the effect sizes segregated as high-income, middle-income (Americas) and middle-income (other countries). The combined analysis of all studies shows a clear incremental difference from high-income countries (SMD = 0.27, 95% CI: 0.15, 0.45), middle-income/Americas (SMD = 0.81, 95% CI: 0.52, 1.19) and middle-income/other countries (1.90, 95% CI: 1.46, 2.40).

3.5. Risk of bias and exposure misclassification

The risks of comet assay measurement bias and exposure misclassification have been assessed for all studies in the meta-analyses. Fig. 11 shows the average level of information on comet assay procedures, blinding/coding of samples, assay controls, and exposure misclassification in the datasets of exposure groups. The information is expressed as risk of bias, segregated into three groups (low/light grey, moderate/middle grey and high/dark grey). It should be emphasised that the risk of bias related to comet assay experiments refers to the information in the publications. The descriptions of comet assay procedures, blinding/coding of samples and inclusion of assay controls may have been inadequate, but studies were still performed according to high-quality procedures. It should also be noted that a lack of blinding/coding before scoring is expected to augment the effect size, as investigators might be

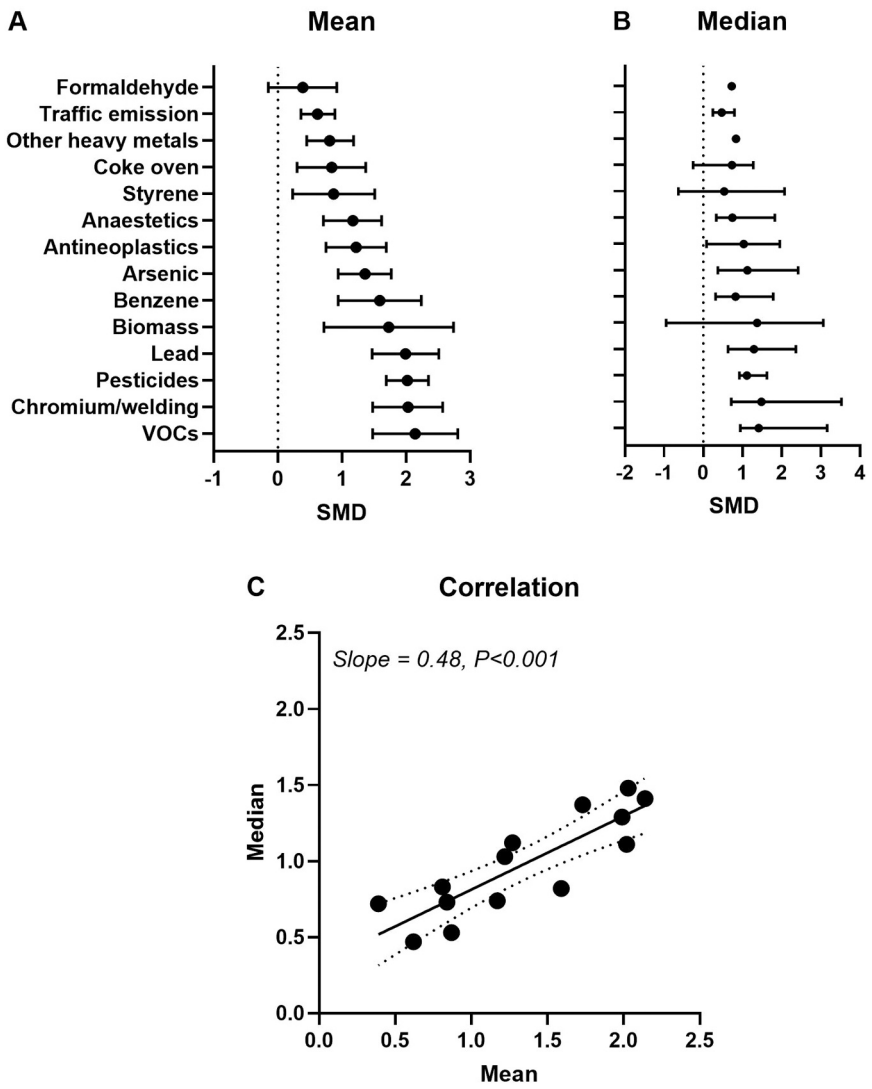

Fig. 8. Summary of overall effect sizes of formaldehyde (n = 8), traffic emissions (n = 21), other heavy metals (n = 6), coke oven (n = 10), styrene, (n = 11), anaesthetic gases (n = 16), antineoplastic drugs (n = 20), arsenic (n = 14), benzene (n = 18), biomass smoke (n = 10), lead (n = 35), pesticides (n = 66), chromium/welding fumes (n = 20) and other volatile organic compounds (n = 13). The results include: standardised mean difference with 95% confidence intervals in a regular meta-analysis (A), an analysis using the median as the measure of central tendency (B), and the correlation between the two approaches (C). The slope and P-value are based on linear regression analysis.

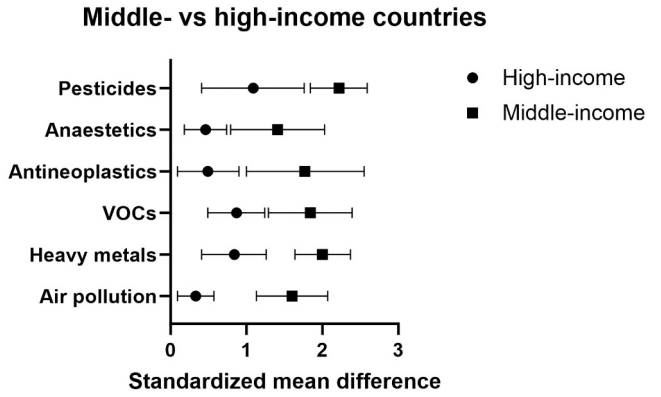

Fig. 9. Overall effect size stratified by country-income group. The results are the mean and 95% confidence intervals. The results are based on meta-analysis of studies on air pollution (n = 22/18), anaesthetic gases (n = 4/12), antineoplastic drugs (n = 9/11), heavy metals (n = 15/51), pesticides (n = 11/55), and volatile organic compounds (n = 22/28) (high/middle income countries).

biased if they know the exposure status of the samples. Information on assay controls is mainly an assurance of the quality of the comet assay experiments. In particular, assay controls are important in studies with null effect findings, although they should be included/reported in all types of biomonitoring studies. In addition, Fig. 11 shows the average risk of exposure misclassification. The highest risk of exposure misclassification occurs in studies where subjects have been segregated into groups according to a work category or area of residence. This is most likely associated with non-differential misclassification, which is expected to attenuate the effect size.

Interestingly, the risk of comet assay measurement bias and exposure misclassification has not changed over the study period (Fig. 12). In fact, the risk-of-bias scores appear to have remained relatively constant over the period: comet assay measurement (0.50, 95% CI: 0.41, 0.60), exposure misclassification (0.94, 95% CI: 0.83, 1.06), blinded/coded analysis (1.10, 95% CI: 0.98, 1.22) and assay controls (1.67, 95% CI: 1.58, 1.75). The comet assay procedures are typically well-described (i.e. low or moderate risk of bias). Inclusion of assay controls and blinded/coded analysis is clearly an area of concern (high risk of bias). The exposure assessment appears to depend on the type of compound and

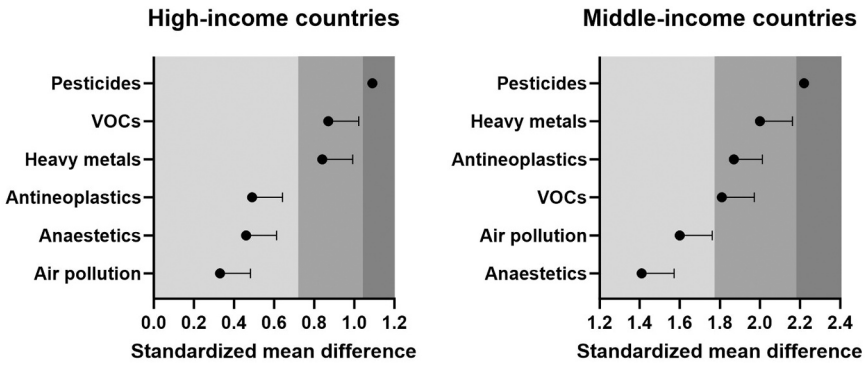

Fig. 10. Overall effect size and critical range for groups of exposures, stratified into high-income and middle-income countries. The critical range is the difference between the highest and lowest effect size, divided by the number of intervals between exposures. The graphical presentation is based on the assumption that there is a real difference in exposures, and the central tendency is an accurate estimate of the effect of each exposure. Central tendencies (circles) that overlap with critical ranges (whiskers) of other exposures are considered to have the same effect size. The exposures are categorised into three different groups (grey background shading), based on the lack of overlapping critical ranges. The results are based on meta-analysis of studies on air pollution (n = 22/18), anaesthetic gases (n = 4/12), antineoplastic drugs (n = 9/11), heavy metals (n = 15/51), pesticides (n = 11/55), and volatile organic compounds (n = 22/28) (high/middle income countries).

Table 9
Analysis of effect sizes (standardized mean difference (SMD) and 95% confidence interval) in different areas.

Exposure	High-income	Middle-income (Americas)	Middle-income (other)
Air pollution (n = 22/3/14)	0.33 (0.09, 0.57)	1.21 (0.57, 1.85)	1.71 (1.15, 2.28)
Anaesthesia gases (n = 4/6/6)	0.46 (0.18, 0.79)	0.47 (0.23, 0.71)	2.32 (1.55, 3.09)
Antineoplastic drugs (n = 9/4/7)	0.53 (0.10, 0.97)	1.35 (0.29, 2.41)	2.15 (1.08, 3.21)
Heavy metals (n = 15/15/36)	0.84 (0.41, 1.26)	1.07 (0.73, 1.42)	2.39 (1.91, 2.88)
Volatile organic compounds (n = 22/10/16)	0.87 (0.49, 1.24)	1.28 (0.01, 1.66)	2.19 (1.40, 1.81)
Pesticides (n = 12/30/24)	1.09 (0.41, 1.76)	1.22 (0.98, 1.46)	3.22 (2.41, 4.02)
Combined (unadjusted)[a]	0.22 (0.12, 0.37)	0.94 (0.63, 1.34)	1.91 (1.48, 2.41)
Combined (adjusted)[a]	0.27 (0.15, 0.45)	0.81 (0.52, 1.19)	1.90 (1.46, 2.40)

[a]Based on cube root-transformed standardized mean differences as either unadjusted analysis or adjusted for types of exposure. For the combined dataset, each of the country-income groups is statistically significant from the other groups ($P < 0.05$, one-way ANOVA).

whether it is actually possible to make an adequate description of a complex exposure. Studies on heavy metals typically have little risk of exposure misclassification, since it is relatively straightforward to measure heavy metals in biological fluids. Exposure to VOCs is difficult to assess because (i) they are usually a mixture of chemical agents; (ii) some VOCs share metabolic paths; and (iii) certain VOCs, such as formaldehyde, do not have a specific metabolite. Assessment of outdoor air pollution exposure is relatively simple, but it requires special equipment to measure mass or number concentration of particles.

It is a mystery why so many publications on comet assay in biomonitoring studies do not contain information on assay controls. The lack of reporting and/or analysis of positive assay controls has been a concern, which has been voiced increasingly loudly during the last ten years. For instance, in 2014, a paper on the topic only considered it "helpful to include…reference standards" in comet assay experiments [45]. A stronger statement was put forward in a meeting report from the 11th International Comet Assay Workshop (published in 2017), which highlighted the necessity to analyse and report assay controls [46]. This extends to versions of the comet assay for measurement of oxidatively damaged DNA [40,47,48]. Recent reviews and recommendations emphasise that assay controls are a necessity in human biomonitoring studies [2,49]. In the MIRCA recommendations, blinding of samples is considered a fundamental aspect of conducting scientific studies, not limited to the comet assay [39]. Scoring samples without blinding is likely to bias results, especially if the difference between exposure groups is small, as is usually the case in samples from human biomonitoring studies. There appears to be little information on the magnitude of bias that might occur in scoring without blinding. An attempt to assess this type of bias in an inter-laboratory ring study showed no difference when laboratories tested the same samples,

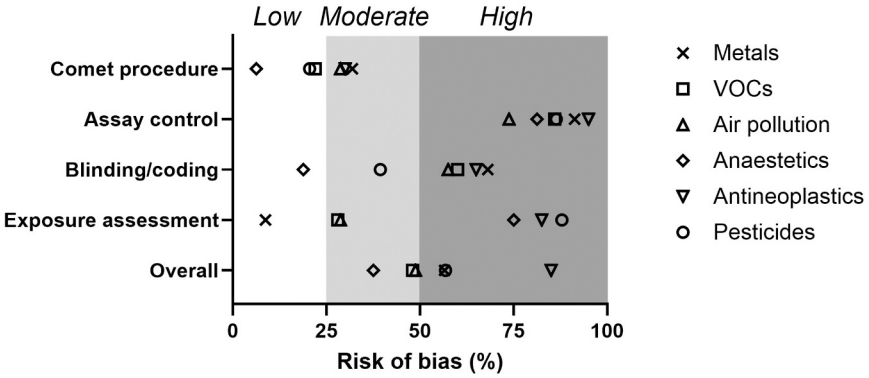

Fig. 11. Risk of measurement bias and exposure misclassification. Background shading indicates risk of bias categories; low (white), moderate (light grey) and high (dark grey). Symbols are the mean of all studies in the meta-analysis.

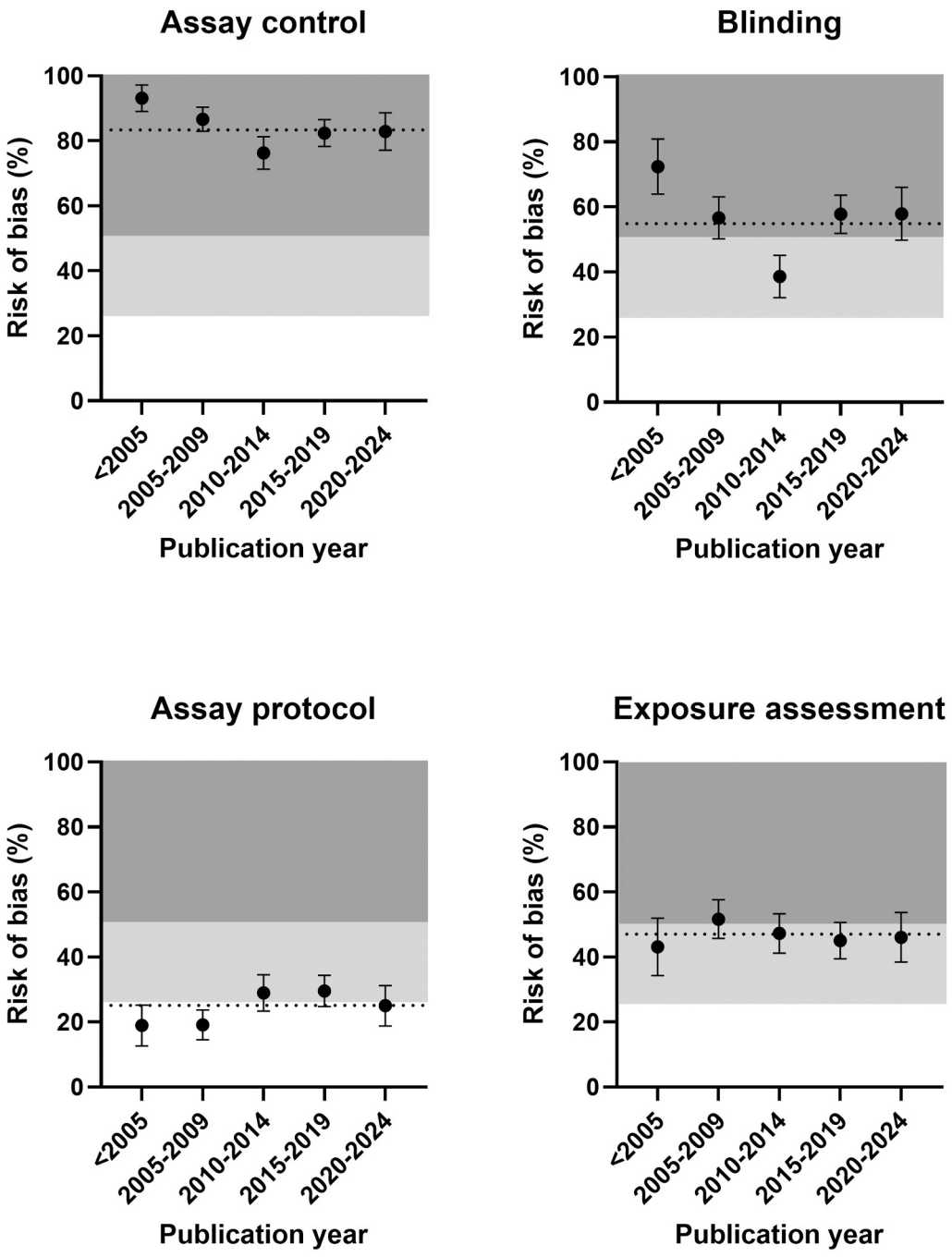

Fig. 12. Risk of comet assay measurement bias and exposure misclassification. The results are segregated into groups of 5 years. Results are reported as means (circles) and standard error of the mean (whiskers). The dotted line is the mean of all studies. Background shading indicates risk of bias categories: low (white), moderate (light grey) and high (dark grey).

delivered as non-blinded calibration curve samples, and coded samples [26].

3.6. Assessment of the effect of predictors for effect sizes

The preceding sections have described individually a number of predictors for the effect sizes, including the type of exposure, location of the study and experimental factors (i.e. measurement bias and exposure misclassification). Obviously, it is highly interesting to determine whether the comet assay can detect a difference in DNA strand breaks between exposures. Unfortunately, there are some challenges to consider when answering that simple question. Below, we describe a sequence of statistical tests to shed light on the effect of different predictors on the magnitude of effect sizes.

3.6.1. Risk of biases in studies from middle-income and high-income countries

From the previous sections, it has become clear that location is a major determinant of effect size, but experimental factors also have some influence on the effect size of DNA strand breaks. Table 10 shows an analysis of the co-variability between the predictors (country-income group *vs* measurement biases/exposure misclassification). The analysis indicates a co-variability between the country-income group and risk of comet assay measurement bias/exposure misclassification. In particular,

Table 10
Proportion of studies in country-income groups and risk of comet assay measurement bias and exposure misclassification.

Item	High-income	Middle-income (Americas)	Middle-income (other)	Statistics
Comet assay description	77, 13, 10[$]	82, 15, 3[$]	**50, 17, 33**[§]	34.4 (P < 0.001, O *vs.* A)
Assay controls	13, 10, 77[§]	15, 24, 62[$]	**7, 5, 88**[§]	19.5 (P < 0.001, O *vs.* H/A)
Blinding	48, 52	56, 44[$]	**36, 64**[§]	6.9 (P < 0.05, O *vs.* A)
Exposure assessment	58, 18, 24[§$]	**34, 9, 57**	44, 15, 41	18.0 (P < 0.01, A *vs.* H)
Overall	26, 62, 12[$]	26, 50, 24[$]	**6, 58, 36**[§]	26.1 (P < 0.001, O *vs.* H)

Values are percentages of studies with low (left), moderate (middle) and high (right) risk of bias. Statistics are reported as χ^2-value and P-value for differences in proportion between country-income groups. Post-hoc analyses are pairwise χ^2-tests ([§]Statistically different from proportion in middle-income (Americas) countries, [$]Statistically different from proportion in middle-income (other) countries). Group with the highest proportion in high-risk group is listed first (slash indicates that country-income groups have not statistically significant proportions). Abbreviations: Americas (A), high-income (H), not applicable (NA), other middle-income countries (O).

the overall analysis of both comet assay measurement and exposure misclassification shows that studies from high-income countries and the Americas have a lower risk of bias than do studies from other middle-income countries ($\chi^2 = 26.1$, $P < 0.001$). Statistical analysis of individual items indicates that studies from other middle-income countries exhibit the highest risk of comet assay measurement bias, whereas studies from the Americas have the highest risk of exposure misclassification. It should be emphasised that the analysis only demonstrates a co-variability, which should not be interpreted as a causal relationship, and it does not demonstrate that studies from middle-income countries are unreliable. However, it does flag a necessity to be cautious when speculating on the reasons why effect sizes seem to be higher in middle-income countries.

3.6.2. *Parametric analysis on cube root-transformed results*

It is evident that the distribution of raw results is skewed, and assumptions for parametric tests are violated, especially the homogeneity of variance in groups being compared. Therefore, we have used the cube root-transformation of the results because there are null and negative values. The transformation of raw data does not produce a normal distribution, and the parametric tests should be interpreted with caution. Nevertheless, in line with previous results, there is a difference in effect sizes between the six exposure groups in univariate tests on cube root-transformed data ($P < 0.01$, ANOVA). Effect sizes with increasing value are as follows: air pollution (0.33, 95% CI: 0.14, 0.67), antineoplastic drugs (0.55, 95% CI: 0.19, 1.23), VOCs (0.72, 95% CI: 0.39, 1.18), anaesthetic gases (0.91, 95% CI: 0.33, 1.96), heavy metals (1.18, 95% CI: 0.78, 1.18) and pesticides (1.44, 95% CI: 0.97, 2.04). The sequence is not identical to the results using critical ranges (Fig. 10) because of the difference between untransformed and transformed data. Adjustments for country-income group, measurement bias and exposure misclassification reduce the effects of exposure differences to statistically non-significant effects, whereas there are statistically significant effects of country-income group and comet assay description bias (results not shown).

3.6.3. *Logistic regression on cube root-transformed results*

In order to avoid the effect of outliers, we have assessed the effect of predictors, using logistic regression with effect sizes (i.e. outcome) dichotomized as below/above the mean or median of the distribution of cube root-transformed effect sizes. It is easy to compare regression coefficients from different statistical tests because the outcome variable has the same range (0 or 1) in all logistic regression tests. In this analysis, we have reported regression coefficients and standard error (SE). In total, we have analysed the results by four different types of logistic regression tests, because of the differing stratification of the raw data into below/above median or cube root-transformed mean, and whole dataset or separate datasets on the six types of exposures. Table 11 shows results from all four logistic regression analyses, which are described in detail below. We have only reported effects of exposures, country-income group and comet assay description bias because they are the strongest predictors. Assay controls, blinded/coded analysis of samples and exposure misclassification are not statistically significant in univariate tests, and there have been only sporadic/inconsistent effects in multivariate tests (effects of these predictors are omitted from the description of results from adjusted analyses in Table 11).

Logistic regression (including exposure variable). The purpose of this analysis is to assess whether the six different exposures are associated

Table 11
Effects of exposure, country-income group and comet assay measurement bias on effect sizes.

Group (country-income group/comet procedure)	Stratification of effect based on mean		Stratification of effect based on median	
	Univariate	Adjusted	Univariate	Adjusted
Stratification of full dataset				
Exposure				
-Air pollution	Base (0)	Base (0)	Base (0)	Base (0)
-Anaesthetic gases	0.18 ± 0.62	−0.03 ± 0.81	0.42 ± 0.63	0.36 ± 0.80
-Antineoplastic drugs	0.89 ± 0.56	0.36 ± 0.73	0.93 ± 0.57	0.72 ± 0.73
-Heavy metals	**1.46 ± 0.43**[a]	**1.16 ± 0.53**[a]	**1.12 ± 0.43**[a]	*0.95 ± 0.53*
-Volatile organic compounds	*0.78 ± 0.45*	0.83 ± 0.52	**0.93 ± 0.46**[a]	**1.13 ± 0.54**[a]
-Pesticides	**1.25 ± 0.42**[a]	0.35 ± 0.58	**1.18 ± 0.43**[a]	0.65 ± 0.59
Country-income group				
-Middle-income (Americas)	**0.94 ± 0.34**[b]	0.58 ± 0.39	**0.74 ± 0.34**[b]	0.36 ± 0.40
-Middle-income (other countries)	**1.58 ± 0.32**[b]	**1.40 ± 0.37**[b]	**1.62 ± 0.32**[b]	**1.55 ± 0.37**[b]
Comet assay procedure				
-Moderate risk	0.45 ± 0.36	0.32 ± 0.41	0.08 ± 0.35	−0.06 ± 0.39
-High risk	*0.60 ± 0.35*	0.49 ± 0.44	*0.64 ± 0.35*	0.45 ± 0.42
Stratification of exposure-specific dataset				
Country-income group				
-Middle-income (Americas)	0.10 ± 0.33	0.01 ± 0.36	0.32 ± 0.34	0.12 ± 0.37
-Middle-income (other countries)	**1.27 ± 0.31**[b]	**1.20 ± 0.34**[b]	**1.59 ± 0.32**[b]	**1.45 ± 0.34**[b]
Comet assay procedure				
-Moderate risk	−0.01 ± 0.35	−0.06 ± 0.38	0.12 ± 0.35	0.04 ± 0.39
-High risk	**1.09 ± 0.38**[c]	*0.72 ± 0.42*	**1.09 ± 0.36**[c]	0.68 ± 0.41

Logistic regression with exposure, country-income group and comet assay measurement bias as single variables (univariate) or fully adjusted for effects of assay control, blinded/coded analysis and exposure misclassification. Results are shown as regression coefficient and standard error. Statistically significate effects ($P < 0.05$) are indicated in bold text. Certain effects with statistical borderline effects ($P < 0.10$) are indicated in italic text to give an impression of the effect size. [a] Statistically significant difference compared to air pollution exposure ($P < 0.05$). [b] Statistically significant difference compared to high-income countries (Base (0); $P < 0.05$). [c] Statistically significant difference compared to low risk of comet assay measurement bias group ((Base (0); $P < 0.05$).

with different levels of effect size. The studies have been segregated into below/above mean or median value of cube root-transformed SMDs in the whole dataset (results in the upper half of Table 11). The dataset with stratification of studies according to the median has equal representation of results in groups of studies with effects below and above the median; the mean of cube root-transformed results has more studies (n = 142) with high effect size as compared to the group with low effect size (n = 113). The four analyses show slightly different results, although they generally show that studies on heavy metals, VOCs and pesticides have a higher effect size than air pollution (i.e. weakest exposure and defined as base). In addition, there is a weak positive association between effect size and comet assay description bias, and a strong association with country-income group. Fully adjusted analyses generally show attenuation of effects (i.e. lower coefficients), although they remain statistically significant. There is some co-variability between types of investigated exposures, location of the study and comet assay description bias.

Logistic regression (without exposure variable). The purpose of this analysis is to assess whether comet assay measurement biases and/or exposure misclassification are associated with different levels of effect size. We have dichotomized effect sizes (outcome) as below/above the mean or median of the distribution of cube root-transformed effect sizes within each of six overall exposure groups. This can be interpreted as a dataset on exposure-standardized effect sizes. The results show a strong effect of country-income group (lower half of Table 11). Studies from high-income countries and the Americas have similar effect size of exposures, whereas studies from other middle-income countries have shown higher effects. For the comet assay description bias, studies with high risk have a stronger effect than studies with moderate or low risk (base = 0, not shown in the table) in univariate analyses. This effect is attenuated in analyses with adjustment for other predictors. There are no effects of bias related to assay controls and blinded/coded analysis of samples, or to exposure misclassification in the statistical analysis (results not shown).

Collectively, the assessment of the effect of predictors for effect sizes indicates that country-income is the co-variate with the strongest influence on the magnitude of SMDs in studies on environmental and occupational exposure to air pollution, anaesthetic gases, antineoplastic drugs, heavy metals, pesticides and VOCs. The statistical analyses indicate small independent effects related to differences in exposures and comet assay measurement.

3.6.4. Limitations

This umbrella review of meta-analysis in six different systematic reviews has certain limitations; at least, these should be noted, although we think that the procedure is sound and unprecedented in the effort to compare effects of external exposures on levels of DNA strand breaks in human leukocytes. The limitations are listed and shortly described below.

(i) The studies describe complex exposures *per se*, even though individual studies may have focused on specific chemicals. It can be argued that the major concern is not heterogeneity in exposure assessment between individual studies or among the exposure groups we have created, but rather the unaccounted or overlooked exposures in the original studies, which are inevitably carried forward in the summary of meta-analysis. In other words, this summary of meta-analyses of the present umbrella review suffers from layers of heterogeneity. One way forward to assess whether the comet assay really detects true differences in DNA strand breaks in human leukocytes is to focus on studies where the same research group has assessed more than one type of exposure. In theory, it would be possible to standardise such data with separate laboratories, but in reality, it is unlikely to work as most laboratories focus on specific exposures.

(ii) The summary of meta-analyses in the present umbrella review demonstrates a very consistent effect of the location of the studies (i.e. effect sizes from studies in middle-income countries are larger than effects from high-income countries). We have used the World Bank income ranking, which is not ideal and may even be insulting to researchers from middle-income countries who rightfully think they are not doing inferior research. The statistical analyses indicate that the difference between country-income groups depends on both technical issues (i.e. high risk of comet assay measurement bias/exposure misclassification) and exposure-response relationships (i.e. high exposure or lower socioeconomic status). We find these explanations too simplistic and unsatisfactory, although we recognise it is not possible to come closer to a more convincing explanation at the moment. Information on assay controls would have been especially valuable in a further assessment of the effect size between locations. Even better would have been reporting of results as DNA strand breaks per million base pairs using calibration with ionising radiation. At this point, we can only encourage researchers to include and report results on assay controls. It is highly recommended to use ionising radiation for positive assay controls because it offers the opportunity to express results as lesions per million base pairs.

(iii) There is an unequal distribution of studies from different parts of the world in the exposure groups. Some types of exposures are mainly assessed in middle-income countries (e.g. pesticides and heavy metals), whereas other exposures seem to have a more equal representation of studies from middle-income and high-income countries (e.g. air pollution). With respect to a meta-analysis, this might create bias if the country-income group determines the type of exposure being investigated. Following this chain of reasoning, if exposure to chemical *X* is mainly *relevant* in middle-income countries and these tend to have *biased* comet assay analysis, it would skew the meta-analysis toward a high effect size. Again (and unfairly), the chain of reasoning assumes that high-income countries have *unbiased* comet assay analysis.

4. Conclusions

The present review covers a total of 255 studies from which it has been possible to collect results that are useful for meta-analyses. Overall, there is a relatively large difference in effect sizes, and grouping of studies in exposure categories suggests that overall effects are susceptible to various biases and/or differences between areas. The compiled results indicate that exposures can be segregated into three groups with decreasing effect size: high (pesticides), moderate (VOCs, heavy metals and antineoplastic drugs), and low (anaesthetic gases and air pollution). However, the review also raises concerns about the risk of measurement bias and exposure misclassification in human biomonitoring studies on DNA strand breaks, measured by the comet assay.

The individual systematic reviews and the present umbrella review show a substantial heterogeneity in effect sizes, which could be due to differences in (i) vulnerability to genotoxic effects related to socioeconomic status, (ii) exposure levels, and (iii) sensitivity of the comet assay to detect DNA strand breaks. Items (i) and (ii) are rooted in the biological domain as exposure level and effect modification, whereas item (iii) is a technical challenge. Hopefully, the recently published compendium of comet assay protocols and recommendations for reporting comet assay procedures and results will reduce such technical concerns in the future [2,39]. The variability in comet assay results from the use of standardized protocols has been assessed in European multi-laboratory ring studies, including the *European Committee on Oxidative DNA damage* (ESCODD) [50–52], *European Comet Assay Validation Group* (ECVAG) [26,27,53–57], *The comet assay as a human biomonitoring tool* (hCOMET) [24,30,47,48,58], and *Comet assay and cell array for fast and efficient genotoxicity testing* (COMICS) [28].

Unfortunately, the ring studies have not convincingly demonstrated that using standardized comet assay protocols reduces inter-laboratory variation in reported levels of DNA damage. There appear to be knowledge gaps on which procedures in the comet assay protocol create the most variation in results between laboratories, which can only be assessed by more inter-laboratory ring studies.

To the best of our knowledge, this umbrella review and summary of meta-analyses is the first attempt to assess whether the standard alkaline comet assay detects different levels of DNA strand breaks by exposure to different chemicals. The topic is relevant to environmental and occupational toxicology, and it may also have implications for public health and regulatory programs. It is encouraging that the results indicate that the comet assay detects different levels of DNA strand breaks, related to both different exposures and different locations. However, there are too many uncertainties regarding the effects of confounding factors, technical variability within and between laboratories, and the paucity of studies demonstrating a predictive value of comet assay results in terms of risk of disease, to recommend the measurements of DNA strand breaks in human leukocytes as a technique for assessment of risks at the individual level. The results from this summary of meta-analysis indicate that the standard alkaline comet assay may be used to assess a potential hazard of DNA-damaging effects by specific exposures.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the affiliated institutions, European Regional Development Fund project KK.01.1.1.02.0007 (Rec-IMI), the Croatian Science Foundation (HUMNap project #1192), the Horizon Europe (EDIAQI project #101057497), the European Union—Next Generation EU 533–03–23–0006 (BioMolTox), FCT/MCTES UIDP/05608/2020 (10.54499/UIDP/05608/2020) and UIDB/05608/2020 (10.54499/UIDB/05608/2020), and the International Comet Assay Working Group (ICAWG).

Data availability

Data will be made available on request.

References

[1] O. Östling, K.J. Johanson, Microelectrophoretic study of radiation-induced DNA damage in individual mammalian cells, Biochem. Biophys. Res. Commun. 123 (1984) 291–298.

[2] A. Collins, P. Møller, G. Gajski, S. Vodenkova, A. Abdulwahed, D. Anderson, E. E. Bankoglu, S. Bonassi, E. Boutet-Robinet, G. Brunborg, C. Chao, M.S. Cooke, C. Costa, S. Costa, A. Dhawan, J. de Lapuente, C. Del Bo', J. Dubus, M. Dusinska, S. J. Duthie, N.E. Yamani, B. Engelward, I. Gaivao, L. Giovannelli, R. Godschalk, S. Guilherme, K.B. Gutzkow, K. Habas, A. Hernandez, O. Herrero, M. Isidori, A. N. Jha, S. Knasmuller, I.M. Kooter, G. Koppen, M. Kruszewski, C. Ladeira, B. Laffon, M. Larramendy, L.L. Hegarat, A. Lewies, A. Lewinska, G.E. Liwszyc, A.L. de Cerain, M. Manjanatha, R. Marcos, M. Milic, V.M. de Andrade, M. Moretti, D. Muruzabal, M. Novak, R. Oliveira, A.K. Olsen, N. Owiti, M. Pacheco, A.K. Pandey, S. Pfuhler, B. Pourrut, K. Reisinger, E. Rojas, E. Runden-Pran, J. Sanz-Serrano, S. Shaposhnikov, V. Sipinen, K. Smeets, H. Stopper, J.P. Teixeira, V. Valdiglesias, M. Valverde, F. van Acker, F.J. van Schooten, M. Vasquez, J.F. Wentzel, M. Wnuk, A. Wouters, B. Zegura, T. Zikmund, S.A.S. Langie, A. Azqueta, Measuring DNA modifications with the comet assay: a compendium of protocols, Nat. Protoc. 18 (2023) 929–989.

[3] P. Møller, The comet assay: ready for 30 more years, Mutagenesis 33 (2018) 1–7.

[4] M. Milic, A. Frustaci, B.A. Del, J. Sanchez-Alarcon, R. Valencia-Quintana, P. Russo, S. Bonassi, DNA damage in non-communicable diseases: A clinical and epidemiological perspective, Mutat. Res. 776 (2015) 118–127.

[5] P. Møller, H. Stopper, A.R. Collins, Measurement of DNA damage with the comet assay in high-prevalence diseases: current status and future directions, Mutagenesis 35 (2020) 5–18.

[6] S. Bonassi, M. Ceppi, P. Møller, A. Azqueta, M. Milic, N. Monica, G. Brunborg, R. Godschalk, G. Koppen, S.A.S. Langie, J.P. Teixeira, M. Bruzzone, J. Da Silva, D. Benedetti, D. Cavallo, C.L. Ursini, L. Giovannelli, S. Moretti, P. Riso, C. Del Bo', P. Russo, M. Dobrzynska, I.A. Goroshinskaya, E.I. Surikova, M. Staruchova, M. Barancokova, K. Volkovova, A. Kazimirova, B. Smolkova, B. Laffon, V. Valdiglesias, S. Pastor, R. Marcos, A. Hernandez, G. Gajski, B. Spremo-Potparevic, L. Zivkovic, E. Boutet-Robinet, H. Perdry, P. Lebailly, C.L. Perez, N. Basaran, Z. Nemeth, A. Safar, M. Dusinska, A. Collins, Cp h, DNA damage in circulating leukocytes measured with the comet assay may predict the risk of death, Sci. Rep. 11 (2021) 16793.

[7] M. Milic, M. Ceppi, M. Bruzzone, A. Azqueta, G. Brunborg, R. Godschalk, G. Koppen, S. Langie, P. Møller, J.P. Teixeira, A.A. Alija, D. Anderson, V. Andrade, C. Andreoli, F. Asllani, E.E. Bangkoglu, M. Barančoková, N. Basaran, E. Boutet-Robinet, A. Buschini, D. Cavallo, C.C. Pereira, C. Costa, S. Costa, J. Da Silva, C. Del Bo', V.D. Srećković, N. Djelićw, M. Dobrzyńska, Z. Duračková, M. Dvořáková, G. Gajski, S. Galatiz, O.G. Lima, L. Giovannell, I.A. Goroshinskaya, A. Grindel, K. B. Gutzkowe, A. Hernández, C. Hernández, K.B. Holven, I. Ibero-Baraibar, I. Ottestad, E. Kadioglu, A. Kažimirová, W. Kuznetsova, C. Ladeira, B. Laffon, P. Lamonaca, P. Lebailly, H. Louro, T.M. Cardoso, F. Marcon, R. Marcos, M. Moretti, S. Moretti, M. Najafzadeh, Z. Nemeth, M. Neri, B. Novotna, I. Orlowaw, Z. Paduchova, S.P., H. Pastor, B. Spremo-Potparevic, D. Ramadhani, P. Riso, P. Rohr, E. Rojas, P. Rossner, A. Safar, S. Sardas, M.J. Silva, N. Sirota, B. Smolkova, M. Staruchova, R. Stetina, H. Stopper, E.I. Surikova, S.M. Ulven, C.L. Ursini, V. Valdiglesia, M. Valverde, P. Vodicka, K. Volkovova, K.H. Wagner, L. Zivković, M. Dušinská, A.R. Collins, S. Bonassi, The hCOMET project: International database comparison of results with the comet assay in human biomonitoring. Baseline frequency of DNA damage and effect of main confounders, Mutat. Res. 787 (2021) 108371.

[8] C. Ladeira, P. Møller, L. Giovannelli, G. Gajski, A. Haveric, E.E. Bankoglu, A. Azqueta, M. Geric, H. Stopper, J. Cabeda, F.S. Tonin, A. Collins, The comet assay as a tool in human biomonitoring studies of environmental and occupational exposure to chemicals - a systematic scoping review, Toxics 12 (2024) 270.

[9] P. Møller, G. Gajski, M. Geric, A. Haveric, H. Stopper, E.E. Bankoglu, A. Azqueta, L. Giovannelli, A. Collins, C. Ladeira, The comet assay as a tool in human biomonitoring of exposure to combustion-derived air pollution - a systematic review and meta-analysis, Mutat. Res. Rev. Mutat. Res. 797 (2026) 108583.

[10] G. Gajski, A. Haveric, P. Møller, A. Azqueta, L. Giovannelli, M. Geric, H. Stopper, E. E. Bankoglu, A. Collins, C. Ladeira, The comet assay as a tool in human biomonitoring of exposure to anaesthetic gases - a systematic review and meta-analysis, Mutat. Res. Rev. Mutat. Res. 797 (2026) 108586.

[11] C. Ladeira, A. Azqueta, L. Giovannelli, G. Gajski, M. Geric, A. Haveric, H. Stopper, E.E. Bankoglu, A. Collins, P. Møller, The comet assay as a tool in human biomonitoring of exposure to antineoplastic drugs - a systematic review and meta-analysis, Mutat. Res. Rev. Mutat. Res. 797 (2026) 108590.

[12] P. Møller, E.E. Bankoglu, H. Stopper, G. Gajski, M. Geric, A. Haveric, A. Azqueta, L. Giovannelli, A. Collins, C. Ladeira, The comet assay as a tool in human biomonitoring of exposure to heavy metals - a systematic review and meta-analysis, Mutat. Res. Rev. Mutat. Res. 796 (2025) 108567.

[13] M. Geric, A. Azqueta, P. Møller, E.E. Bankoglu, G. Gajski, L. Giovannelli, A. Haveric, H. Stopper, A. Collins, C. Ladeira, The comet assay as a tool in human biomonitoring of exposure to pesticides - a systematic review and meta-analysis, Mutat. Res. Rev. Mutat. Res. 796 (2025) 108565.

[14] L. Giovannelli, P. Møller, G. Gajski, H. Stopper, A. Azqueta, E.E. Bankoglu, A. Haveric, M. Geric, A. Collins, C. Ladeira, The comet assay as a tool in human biomonitoring of exposure to volatile organic compounds - a systematic review and meta-analysis, Mutat. Res. Rev. Mutat. Res. 797 (2026) 108584.

[15] L. Belbasis, V. Bellou, J.P.A. Ioannidis, Conducting umbrella reviews, BMJ Med. 1 (2022) e000071.

[16] R.J. Albertini, D. Anderson, G.R. Douglas, L. Hagmar, K. Hemminki, F. Merlo, A. T. Natarajan, H. Norppa, D.E.G. Shuker, R. Tice, M.D. Waters, A. Aitio, IPCS guidelines for the monitoring of genotoxic effects of carcinogens in humans, Mutat. Res. 463 (2000) 111–172.

[17] F. Kassie, W. Parzefall, S. Knasmüller, Single cell gel electrophoresis: a new technique for human biomonitoring studies, Mutat. Res. 463 (2000) 13–31.

[18] P. Møller, L.E. Knudsen, S. Loft, H. Wallin, The comet assay as a rapid test in biomonitoring occupational exposure to DNA-damaging agents and effect of confounding factors, Cancer Epidemiol. Biomark. Prev. 9 (2000) 1005–1015.

[19] P. Vodicka, R. Stetina, M. Koskinen, P. Soucek, L. Vodickova, P. Hlavac, M. Kuricova, R. Necasova, K. Hemminki, New aspects in the biomonitoring of occupational exposure to styrene, Int. Arch. Occup. Environ. Health 75 (2002) S75–85 (Suppl).

[20] Ü. Ündeger, N. Basaran, A. Kars, D. Güc, Assessment of DNA damage in nurses handling antineoplastic drugs by the alkaline COMET assay, Mutat. Res. 439 (1999) 277–285.

[21] V. Gianfredi, D. Nucci, C. Fatigoni, T. Salvatori, M. Villarini, M. Moretti, Extent of Primary DNA Damage Measured by the Comet Assay in Health Professionals Exposed to Antineoplastic Drugs: A Systematic Review and Meta-Analysis, Int. J. Environ. Res. Public. Health 17 (2020) 523.

[22] P.H. Avogbe, L. Ayi-Fanou, H. Autrup, S. Loft, B. Fayomi, A. Sanni, P. Vinzents, P. Møller, Ultrafine particulate matter and high-level benzene urban air pollution in relation to oxidative DNA damage, Carcinogenesis 26 (2005) 613–620.

[23] P. Møller, G. Gajski, M. Geric, L. Giovannelli, A. Azqueta, A. Haveric, H. Stopper, E. E. Bankoglu, A. Collins, C. Ladeira, The comet assay as a tool in human biomonitoring studies: effcets of confounding factors, Mutat. Res. Rev. Mutat. Res. 796 (2025) 108566.

[24] P. Møller, A. Azqueta, J. Sanz-Serrano, T. Bakuradze, E. Richling, E.E. Bankoglu, H. Stopper, V.C. Bastos, S.A.S. Langie, A. Jensen, F. Scavone, L. Giovannelli, M. Wojewodzka, M. Kruszewski, V. Valdiglesias, B. Laffon, C.T. Costa, S. Costa, J. P. Teixeira, M. Marino, C. Del Bo, P. Riso, C. Zheng, S. Shaposhnikov, A. Collins, Visual comet scoring revisited: a guide to scoring comet assay slides and obtaining reliable results, Mutagenesis 38 (2023) 253–263.
[25] P. Møller, Measurement of oxidatively damaged DNA in mammalian cells using the comet assay: Reflections on validity, reliability and variability, Mutat. Res. Genet. Toxicol. Environ. Mutagen. 873 (2022) 503423.
[26] L. Forchhammer, C. Johansson, S. Loft, L. Möller, R.W. Godschalk, S.A. Langie, G. D. Jones, R.W. Kwok, A.R. Collins, A. Azqueta, D.H. Phillips, O. Sozeri, M. Stepnik, J. Palus, U. Vogel, H. Wallin, M.N. Routledge, C. Handforth, A. Allione, G. Matullo, J.P. Teixeira, S. Costa, P. Riso, M. Porrini, P. Møller, Variation in the measurement of DNA damage by comet assay measured by the ECVAG inter-laboratory validation trial, Mutagenesis 25 (2010) 113–123.
[27] R.W. Godschalk, C. Ersson, M. Stepnik, M. Ferlinska, J. Palus, J.P. Teixeira, S. Costa, G.D. Jones, J.A. Higgins, J. Kain, L. Möller, L. Forchhammer, S. Loft, Y. Lorenzo, A.R. Collins, F.J. van Schooten, B. Laffon, V. Valdiglesias, M. Cooke, V. Mistry, M. Karbaschi, D.H. Phillips, O. Sozeri, M.N. Routledge, K. Nelson-Smith, P. Riso, M. Porrini, C.A. Lopez de, A. Azqueta, G. Matullo, A. Allione, P. Møller, Variation of DNA damage levels in peripheral blood mononuclear cells isolated in different laboratories, Mutagenesis 29 (2014) 241–249.
[28] G. Brunborg, D.M. Eide, A. Graupner, K.B. Gutzkow, S. Shaposhnikov, M. Kruszewsk, N. Sirota, G.D.D. Jones, G. Koppen, F. Vanhavere, P. Møller, R. Stetina, H. Dahl, A. Collins, Calibration of the comet assay using ionizing radiation, Mutat. Res. Genet. Toxicol. Environ. Mutagen. 885 (2023) 503560.
[29] P. Møller, S. Loft, C. Ersson, G. Koppen, M. Dusinska, A. Collins, On the search for an intelligible comet assay descriptor, Front. Genet. 5 (2014) 217.
[30] P. Møller, A. Azqueta, M. Collia, T. Bakuradze, E. Richling, E.E. Bankoglu, H. Stopper, V.C. Bastos, S.A.S. Langie, A. Jensen, S. Ristori, F. Scavone, L. Giovannelli, M. Wojewodzka, M. Kruszewski, V. Valdiglesias, B. Laffon, C. Costa, S. Costa, J.P. Teixeira, M. Marino, C. Del Bo, P. Riso, C. Zhang, S. Shaposhnikov, A. Collins, Inter-laboratory variation in measurement of DNA damage by the alkaline comet assay in the hCOMET ring trial, Mutagenesis 38 (2023) 283–294.
[31] P. Møller, S. Loft, Statistical analysis of comet assay results, Front. Genet. 5 (2014) 292.
[32] K. Matkovic, M. Geric, L. Kazensky, M. Milic, V. Kasuba, A. Cvitkovic, M. Sankovic, A. Sumanovac, P. Møller, G. Gajski, Comparison of DNA damage in fresh and frozen blood samples: implications for the comet assay in human biomonitoring studies, Arch. Toxicol. 98 (2024) 3467–3476.
[33] R. Rozgaj, V. Kasuba, G. Brozovic, A. Jazbec, Genotoxic effects of anaesthetics in operating theatre personnel evaluated by the comet assay and micronucleus test, Int. J. Hyg. Environ. Health 212 (2009) 11–17.
[34] V. Gallo, M. Egger, V. McCormack, P.B. Farmer, J.P. Ioannidis, M. Kirsch-Volders, G. Matullo, D.H. Phillips, B. Schoket, U. Stromberg, R. Vermeulen, C. Wild, M. Porta, P. Vineis, STrengthening the Reporting of OBservational studies in Epidemiology–Molecular Epidemiology (STROBE-ME): an extension of the STROBE statement, Mutagenesis 27 (2012) 17–29.
[35] J.P.T. Higgins, R.L. Morgan, A.A. Rooney, K.W. Taylor, K.A. Thayer, R.A. Silva, C. Lemeris, E.A. Akl, T.F. Bateson, N.D. Berkman, B.S. Glenn, A. Hrobjartsson, J. S. LaKind, A. McAleenan, J.J. Meerpohl, R.M. Nachman, J.E. Obbagy, A. O'Connor, E.G. Radke, J. Savovic, H.J. Schunemann, B. Shea, K. Tilling, J. Verbeek, M. Viswanathan, J.A.C. Sterne, A tool to assess risk of bias in non-randomized follow-up studies of exposure effects (ROBINS-E), Environ. Int. 186 (2024) 108602.
[36] V.J. McKelvey-Martin, M.H.L. Green, P. Schmezer, B.L. Pool-Zobel, M.P. De Meo, A. Collins, The single cell gel electrophoresis assay (comet assay): A European review, Mutat. Res. 288 (1993) 47–63.
[37] M. Karbaschi, Y. Ji, A.M.S. Abdulwahed, A. Alohaly, J.F. Bedoya, S.L. Burke, T. M. Boulos, H.G. Tempest, M.S. Cooke, Evaluation of the major steps in the conventional protocol for the alkaline comet assay, Int. J. Mol. Sci. 20 (2019).
[38] A. Azqueta, D. Muruzabal, E. Boutet-Robinet, M. Milic, M. Dusinska, G. Brunborg, P. Møller, A.R. Collins, Technical recommendations to perform the alkaline standard and enzyme-modified comet assay in human biomonitoring studies, Mutat. Res. Genet. Toxicol. Environ. 843 (2019) 24–32.
[39] P. Møller, A. Azqueta, E. Boutet-Robinet, G. Koppen, S. Bonassi, M. Milic, G. Gajski, S. Costa, J.P. Teixeira, P.C. Costa, M. Dusinska, R. Godschalk, G. Brunborg, K. B. Gutzkow, L. Giovannelli, M.S. Cooke, E. Richling, B. Laffon, V. Valdiglesias, N. Basaran, C. Del Bo', B. Zegura, M. Novak, H. Stopper, P. Vodicka, S. Vodenkova, V.M. de Andrade, M. Sramkova, A. Gabelova, A. Collins, S.A.S. Langie, Minimum Information for Reporting on the Comet Assay (MIRCA): recommendations for describing comet assay procedures and results, Nat. Protoc. 15 (2020) 3817–3826.
[40] P. Møller, K. Jantzen, M. Løhr, M.H.G. Andersen, D.M. Jensen, M. Roursgaard, P. H. Danielsen, A. Jensen, S. Loft, Searching for assay controls for the Fpg- and hOGG1-modified comet assay, Mutagenesis 33 (2018) 9–19.
[41] X. Wan, W. Wang, J. Liu, T. Tong, Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range, BMC Med. Res. Methodol. 14 (2014) 135.
[42] K.Y. Chi, M.Y. Li, C. Chen, E. Kang, T. Cochrane, Ten circumstances and solutions for finding the sample mean and standard deviation for meta-analysis, Syst. Rev. 12 (2023) 62.
[43] G. Malta, S. Matera, F. Plescia, A. Calascibetta, A. Argo, E. Cannizzaro, Occupational accidents and the use of PPE: a global meta-analysis, Front. Public. Health 12 (2024) 1368991.
[44] V. Quintero Santofimio, A.F.S. Amaral, J. Feary, Occupational exposures in low- and middle-income countries: A scoping review, PLOS Glob. Public. Health 4 (2024) e0003888.
[45] A.R. Collins, Y.N. El, Y. Lorenzo, S. Shaposhnikov, G. Brunborg, A. Azqueta, Controlling variation in the comet assay, Front. Genet. 5 (2014) 359.
[46] G. Koppen, A. Azqueta, B. Pourrut, G. Brunborg, A.R. Collins, S.A.S. Langie, The next three decades of the comet assay: a report of the 11th International Comet Assay Workshop, Mutagenesis 32 (2017) 397–408.
[47] P. Møller, D. Muruzabal, T. Bakuradze, E. Richling, E.E. Bankoglu, H. Stopper, S.A. S. Langie, A. Azqueta, A. Jensen, F. Scavone, L. Giovannelli, M. Wojewodzka, M. Kruszewski, V. Valdiglesias, B. Laffon, C. Costa, S. Costa, J.P. Teixeira, M. Marino, C. Del Bo', P. Riso, S. Shaposhnikov, A. Collins, Potassium bromate as positive assay control for the Fpg-modified comet assay, Mutagenesis 35 (2020) 341–348.
[48] P. Møller, A. Azqueta, A. Rodriguez-Garraus, T. Bakuradze, E. Richling, E. E. Bankoglu, H. Stopper, V.C. Bastos, S.A.S. Langie, A. Jensen, S. Ristori, F. Scavone, L. Giovannelli, M. Wojewodzka, M. Kruszewski, V. Valdiglesias, B. Laffon, C. Costa, S. Costa, J.P. Teixeira, M. Marino, C. Del Bo, P. Riso, C. Zhang, S. Shaposhnikov, A. Collins, Long-term cryopreservation of potassium bromate positive assay controls for measutement of oxidatively damaged DNA by the Fpg-modified comet assay, Mutagenesis 38 (2023) 264–272.
[49] A. Azqueta, C. Ladeira, L. Giovannelli, E. Boutet-Robinet, S. Bonassi, M. Neri, G. Gajski, S. Duthie, C. Del Bo', P. Riso, G. Koppen, N. Basaran, A. Collins, P. Møller, Application of the comet assay in human biomonitoring: An hCOMET perspective, Mutat. Res. 783 (2020) 108288.
[50] ESCODD, Comparative analysis of baseline 8-oxo-7,8-dihydroguanine in mammalian cell DNA, by different methods in different laboratories: an approach to consensus, Carcinogenesis 23 (2002) 2129–2133.
[51] ESCODD, Measurement of DNA oxidation in human cells by chromatographic and enzymic methods, Free. Radic. Biol. Med. 34 (2003) 1089–1099.
[52] ESCODD, C.M. Gedik, A. Collins, Establishing the background level of base oxidation in human lymphocyte DNA: results of an inter-laboratory validation study, FASEB J. 19 (2005) 82–84.
[53] P. Møller, L. Möller, R.W. Godschalk, G.D. Jones, Assessment and reduction of comet assay variation in relation to DNA damage: studies from the European Comet Assay Validation Group, Mutagenesis 25 (2010) 109–111.
[54] C. Johansson, P. Møller, L. Forchhammer, S. Loft, R.W. Godschalk, S.A. Langie, S. Lumeij, G.D. Jones, R.W. Kwok, A. Azqueta, D.H. Phillips, O. Sozeri, M. N. Routledge, A.J. Charlton, P. Riso, M. Porrini, A. Allione, G. Matullo, J. Palus, M. Stepnik, A.R. Collins, L. Möller, An ECVAG trial on assessment of oxidative damage to DNA measured by the comet assay, Mutagenesis 25 (2010) 125–132.
[55] L. Forchhammer, C. Ersson, S. Loft, L. Möller, R.W. Godschalk, F.J. van Schooten, G.D. Jones, J.A. Higgins, M. Cooke, V. Mistry, M. Karbaschi, A.R. Collins, A. Azqueta, D.H. Phillips, O. Sozeri, M.N. Routledge, K. Nelson-Smith, P. Riso, M. Porrini, G. Matullo, A. Allione, M. Steepnik, M. Komorowska, J.P. Teixeira, S. Costa, L.A. Corcuera, A.L. de Cerain, B. Laffon, V. Valdiglesias, P. Møller, Inter-laboratory variation in DNA damage using a standard comet assay protocol, Mutagenesis 27 (2012) 665–672.
[56] C. Ersson, P. Møller, L. Forchhammer, S. Loft, A. Azqueta, R.W.L. Godschalk, F. J. van Schooten, G.D.D. Jones, J.A. Higgins, M.S. Cooke, V. Mistry, M. Karbaschi, D.H. Phillips, O. Sozeri, M.N. Routledge, K. Nelson-Smith, P. Riso, M. Porrini, G. Matullo, A. Allione, M. Stepnik, M. Ferlinska, J.P. Teixeira, S. Costa, L. A. Corcuera, A.L. de Cerain, B. Laffon, V. Valdiglesias, A.R. Collins, L. Möller, An ECVAG inter-laboratory validation study of the comet assay: inter- and intra-laboratory variation of DNA strand breaks and FPG-sensitive sites in human mononuclear cells, Mutagenesis 28 (2013) 279–286.
[57] R.W. Godschalk, C. Ersson, P. Riso, M. Porrini, S.A. Langie, F.J. van Schooten, A. Azqueta, A.R. Collins, G.D. Jones, R.W. Kwok, D.H. Phillips, O. Sozeri, A. Allione, G. Matullo, L. Möller, L. Forchhammer, S. Loft, P. Møller, DNA-repair measurements by use of the modified comet assay: an inter-laboratory comparison within the European Comet Assay Validation Group (ECVAG), Mutat. Res. 757 (2013) 60–67.
[58] P. Møller, A. Azqueta, A. Rodriguez-Garraus, T. Bakuradze, E. Richling, E. E. Bankoglu, H. Stopper, V.C. Bastos, S.A.S. Langie, A. Jensen, F. Scavone, L. Giovannelli, M. Wojewodzka, M. Kruszewski, V. Valdiglesias, B. Laffon, C. T. Costa, S. Costa, J.P. Teixeira, M. Marino, C. Del Bo, P. Riso, C. Zheng, S. Shaposhnikov, A. Collins, DNA strand break levels in cryopreserved mononuclear blood cell lines measured by the alkaline comet assay: results from the hCOMET ring trial, Mutagenesis 38 (2023) 273–282.