Protocol: Evaluation of Human Cancer Studies on Exposure to *ortho*-Toluidine for the Report on Carcinogens

Background information

ortho-Toluidine (CASRN 95-53-4) is an arylamine used (either directly or as an intermediate) to make dyes, rubber chemicals, and herbicides. It has been listed as *reasonably anticipated to be a human carcinogen* since 1983 in the *Report on Carcinogens* (RoC) based on sufficient evidence of carcinogenicity from studies in experimental animals and significant U.S. exposure (12th RoC, NTP 2011).[1] Since that time, several cancer studies in humans have been published in the peer-reviewed literature. *ortho*-Toluidine has been selected as a candidate substance for possible change in listing status based on an adequate database of cancer studies and significant U.S. exposure.

The cancer evaluation of *ortho*-toluidine, captured in the draft RoC monograph, will focus on the human cancer studies and mechanistic data. The concept for the proposed approach for conducting the cancer evaluation – including the literature search strategy, the scope and focus of the monograph, and the approaches to obtain scientific and public input to address the key scientific questions and issues – is available at http://ntp.niehs.nih.gov/go/37803. This protocol is limited to the procedures used to prepare the human cancer studies section of the draft RoC monograph and describes (1) the RoC listing criteria used in the evaluation, (2) the key scientific issues, and (3) the procedures and guidelines for each step in the cancer evaluation process.

RoC Listing Criteria for Evaluating Carcinogenicity from Studies in Humans

Sufficient evidence of carcinogenicity from studies in humans: indicates a causal relationship between exposure to the agent, substance, or mixture, and human cancer.

Limited evidence of carcinogenicity from studies in humans: causal interpretation is credible, but alternative explanations, such as chance, bias, or confounding factors, could not adequately be excluded.

Key Scientific Questions for the Cancer Evaluation of *ortho*-Toluidine

Based on the findings of animal studies, early case reports, and case series, the principal cancer endpoint of interest in humans is urinary bladder cancer. A number of epidemiologic studies of cohort or case-control designs have examined urinary bladder cancer in workers employed in the rubber chemical and dye industries. Besides *ortho*-toluidine, there is potential for co-exposure to several other known or suspected urinary bladder carcinogens.

The key question for the evaluation is:

- What is the level of evidence (sufficient, limited, or inadequate) of the carcinogenicity of *ortho*-toluidine from studies in humans?

Secondary questions are as follows:

- What are the methodological strengths and limitations of the studies?
- What are the potential confounders for urinary bladder cancer risk in these studies?

[1] http://ntp.niehs.nih.gov/go/roc12.

- Is there a credible association between exposure to *ortho*-toluidine and cancer of the urinary bladder?
- If so, can the relationship between urinary bladder cancer and exposure to *ortho*-toluidine be explained by chance, bias, or confounding?

Steps in the Cancer Evaluation Process

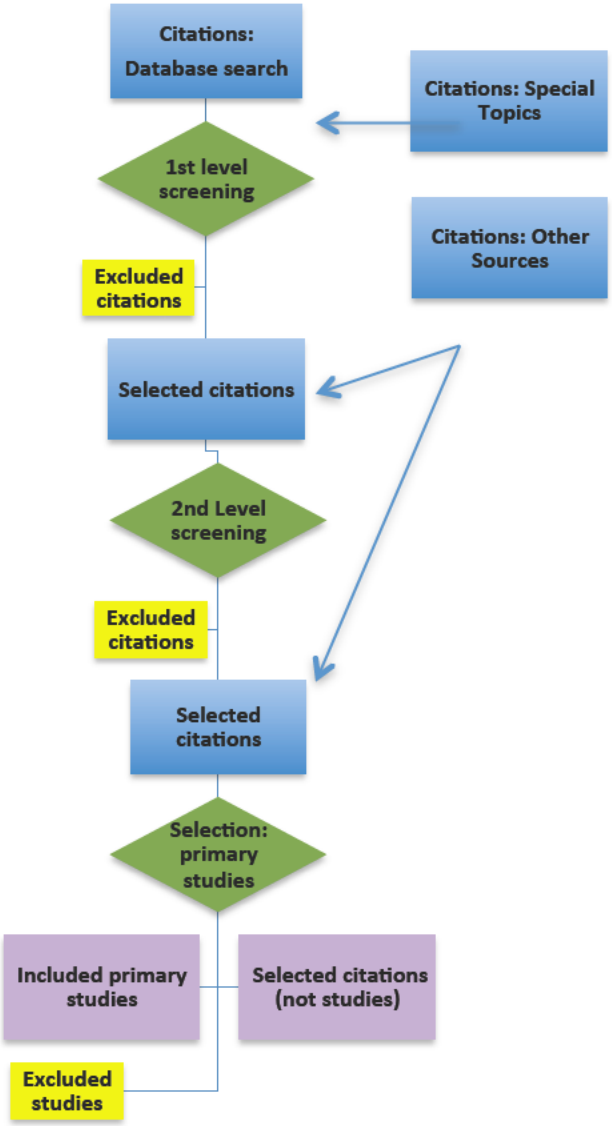
The steps for conducting the human cancer evaluation are outlined below. The procedures and guidelines for conducting each step are described in Appendices 1 through 5.

1. Selection of the literature included in the cancer evaluation (Appendix A)
2. Systematic extraction of data from the epidemiologic studies (Appendix B)
3. Assessment of the quality of the individual epidemiologic studies (Appendix C)
4. Approach for the evaluation of potential confounding (Appendix D)
5. Assessment of the level of evidence of carcinogenicity (sufficient, limited, or inadequate) of *ortho*-toluidine from studies in humans (Appendix E)

Appendix A: Selection of the literature included in the human cancer evaluation

This section discusses procedures to identify and select literature relevant for the human cancer evaluation, including the literature search strategy and inclusion and exclusion criteria. This literature includes the primary epidemiologic studies, which form the basis for cancer evaluation, and supporting literature that may be relevant for the interpretation of the studies. The first step in the process is to develop a literature search strategy and associated inclusion/exclusion criteria to identify the relevant literature, and the second step is to select the primary epidemiologic studies from this database. Figure 1 is a schematic of the process, which is described in detail below.

Figure 1. Literature identification and selection process



1 Identification of relevant citations for cancer evaluation

The identification of the relevant literature for the cancer evaluation includes strategies for searching for citations and inclusion/exclusion questions for selecting the relevant citations from the searches.

1.1 Literature search strategy

Potential exposure to *ortho*-toluidine may occur in the following occupational settings: (1) dye manufacturing in general, (2) magenta manufacturing, (3) aniline-based dye manufacturing, (4) production of rubber chemicals, (5) chloro-*ortho*-toluidine manufacturing, and (6) production of the herbicides metolachlor or acetochlor. These exposure scenarios are used to develop search terms in the literature search strategy. The following approaches for identifying literature will be employed.

1. Literature searches of three scientific databases – PubMed, Scopus, and Web of Science – using a pre-determined range of search terms. Search terms for potential *ortho*-toluidine exposure, e.g., terms related to exposure scenarios and terms specific for *ortho*-toluidine are combined (using "and") with search terms for epidemiologic studies and combined (using "and") with search terms for the outcome, cancer. The specific search terms are described in the table below.

Substance-specific search terms	**Cancer search terms**	**Epidemiologic search terms**
dyestuff OR (dye AND (manufacturing OR manufacture)) OR rubber chemicals OR ortho toluidine OR o-toluidine OR chloro-o-toluidine OR chloro-*ortho* toluidine OR aniline OR ((manufacture OR manufacturing OR production) AND magenta) OR metolachlor OR acetochlor)	cancer OR tumors	epidemiolog* OR case-control OR cohort OR case-report OR case-series OR workers OR workmen or meta-analysis [publication type]

2. Full-text searches of a Quosa-based database of case-control studies on occupational exposure (general) using the term *"ortho*-toluidine" or its synonyms.[2]
3. Searches of a pre-determined standard list of general sources including U.S. and international government agency reports, authoritative reviews and related reports (e.g., International Agency for Research on Cancer, U. S. Environmental Protection Agency, Agency for Toxic Substances and Disease Registry, European Union, National Academy of Sciences) to identify any additional primary epidemiologic studies together with supporting reviews and material that may be relevant for the interpretation of the primary studies.
4. Citation searches from articles, reports and reviews identified above to identify any additional primary studies or other relevant literature.

Additional literature searches may be conducted on special topics or issues.

[2] toluidine" or "*ortho*-toluidine" or "95-53-4" or "1-amino-2-methylaniline" or "2-amino-1-methylaniline" or "*o*-aminotoluene" or "*ortho*-aminotoluene" or "2-methyl-1-aminobenzene" or "2-methylaniline" or "*o*-methylaniline" or "*ortho*-methylaniline" or "*o*-methylbenzenamine" or "*ortho*-methylbenzenamine" or "2-methylbenzenamine" or "2-methylphenylamine" or "*o*-tolylamine" or "*ortho*-tolylamine" or "2-tolylamine"

The scientific database search strategies will be saved in Scopus, Web of Science, and PubMed, respectively, which automatically send out weekly notifications concerning newly identified citations using the saved search strategy.

1.2 Selection of relevant literature

Citations retrieved from literature searches will be uploaded to web-based systematic review software and screened using pre-defined inclusion and exclusion criteria (see below). Multi-level screening of the literature identified from the searches is conducted (see Figure 1); the initial screening is based on titles and abstracts only (Level 1), and subsequent screening is based on full-text PDFs (Level 2 and 3).

Literature is screened at each level by two reviewers using inclusion/exclusion criteria for each level as listed below. The objective of Level 1 and 2 is to identify literature that is useful for the cancer evaluation section, including primary research studies, reviews, and studies on relevant issues (such as confounders) related to cancer evaluation of *ortho*-toluidine. In general, the exclusion and inclusion criteria are similar at each level, but because the screening of the literature at Level 1 is done using titles and abstracts, the "bar" for excluding literature is very high; a more detailed review of the studies for inclusion/exclusion is conducted at Level 2 using the full text article. The objective of the Level 3 review is to select the primary epidemiologic studies that will be discussed in the cancer review, as described in Section 3, below. (Note that human studies pertaining to ambient or biomarker measures of exposure, or to genotoxic or mechanistic data on *ortho*-toluidine and which are not part of the primary epidemiologic studies of cancer outcomes, are considered under separate sections of the monograph.)

Inclusion/exclusion questions: Level 1 (titles and abstracts)

(1) Does this publication appear to contain information on potential exposure to ortho-toluidine (including exposure inferred from knowledge of an exposure scenario) and human cancer? Relevant information includes, but is not limited to, epidemiologic studies, descriptive studies, pooled analyses, meta-analyses, reviews, letters to editors, exposure-assessment studies (for use in epidemiologic studies) and information on co-exposures or potential confounders and other special topics of relevance to the evaluation.

- ❑ Yes
- ❑ No

(2) If the response to Question 1 is "No," identify all reasons that apply from the list below for excluding this publication from the Human Cancer section.

- ❑ (a) No information is provided on potential exposure to *ortho*-toluidine in this study.
- ❑ (b) The study is not a study in humans or related to an issue or information relevant to interpreting the epidemiologic data.

Inclusion/exclusion questions: Level 2 (full text)

(1) Does this publication contain relevant information (as defined above) on potential exposure to ortho-*toluidine (including exposure inferred from knowledge of an exposure scenario) and human cancer?*

- ❑ Yes
- ❑ No

(2) If the response to Question 1 is "No," identify all reasons that apply from the list below for excluding this publication from the Human Cancer section.

- ❑ (a) No information is provided on potential exposure to *ortho*-toluidine in this study.
- ❑ (b) Potential exposure to *ortho*-toluidine is likely in the study or is mentioned in the review or publication, but the publication is not one of the following:
  - (i) an epidemiologic study (such as cohort, case-control, ecological, pooled, meta-analysis) or descriptive study (such as case report or case-series) that provides information on human cancer.
  - (ii) a review, letter to the editor, or abstract, or other type of study provided relevant information related to *ortho*-toluidine and human cancer.
  - (iii) a study or other source of data that provides information, such as on exposure assessment, relevant to evaluating the epidemiologic studies.
- ❑ (c) The publication does not provide information on co-exposures or potential confounders or other special topic(s) relevant to the evaluation.

2 Selection of primary epidemiologic studies

Primary literature to be used in the human cancer evaluation will be selected from the Level 2-selected references and must meet the criteria listed below in Questions 1 to 3. Analytical studies (such as case-control studies, cross-sectional studies, cohort studies or pooled analyses) in which potential exposure to *ortho*-toluidine can be established by quantitative or qualitative exposure assessment, by the authors' report, or reliably inferred using relevant knowledge of industrial processes, will be included in the cancer evaluation. Technical assistance from experts in dye and rubber chemical manufacturing processes may be used to establish the likelihood of potential exposure to *ortho*-toluidine (including relative levels) and co-exposures over time in studies where qualitative or quantitative exposure data are not available. In the majority of cohort or case-control studies, subjects are typically exposed to multiple chemical agents in addition to the substance of concern. In certain studies, where the predominant exposure is to another suspect or known carcinogen and where the levels of *ortho*-toluidine are low or unknown, it may not be possible to evaluate the possibility of an independent association (or effect modification) between observed cancers and exposure to *ortho*-toluidine. In such cases, studies will be noted but may be excluded from the full cancer evaluation.

Primary epidemiologic studies (such as descriptive studies) on potential exposure to *ortho*-toluidine that were retrieved from the literature search strategy and not included in the cancer evaluation, and the reason for their exclusion, will be identified in the monograph. Information from multiple publications relating to the same study population may be included in the draft monograph, but the publications will be counted as one study.

Exclusion/inclusion questions

(1) Is the publication a peer-reviewed, primary research study on potential exposure to ortho-toluidine and human cancer?

- ❑ Yes
- ❑ No

(2) If the answer to question 1 is yes, does the study report a risk estimate (or information to calculate a risk estimate) for cancer?

- ❑ Yes
- ❑ No

(3) Are there exposure analyses specific for ortho-toluidine? If not, is there evidence that exposure to ortho-toluidine is likely to be substantial or predominant in relation to exposure to known or suspect carcinogenic co-exposures?

- ❑ Yes
- ❑ No

Appendix B: Systematic extraction of data from the epidemiologic studies

Two independent reviewers will extract data (such as methods and findings) from the individual studies into a database in a systematic manner using standardized instructions and questions. The database contains "fields" that are specific for the different types of extracted information (such as study population characteristics, exposure assessment, analytical methods and results). The instructions (questions or guidelines) describe the specific type of information that should be summarized or entered into each "field." The fields will be used to populate tables used in the monograph. The assessment of study quality (see Appendix C) will also be entered into the database.

For the cohort and case-control studies, the reviewer will usually extract data from the latest published follow-up or update for each cancer endpoint included in the study. Other relevant information (such as exposure data or re-analyses) from earlier and related publications on the same or overlapping study population will also be included in the review.

Quality assurance and quality control of data extraction and database entry will be accomplished by (a) double-checking of each data entry by the two independent reviewers and (b) flagging any discrepant entries and resolving them by mutual discussion in reference to the original data source.

Appendix C: Assessment of the quality of the individual studies

Each primary study will be systematically evaluated for its ability to inform hazard identification. Studies that will be given the most weight in the evaluation are those that provide the most valid (low risk of systematic error or biases) and precise (low risk from random error) risk estimates, and that have adequate ability (e.g., sufficient power and adequate range of exposure) to detect an effect. In addition, studies should accurately report their findings and apply appropriate analytical methods for calculating risk estimates. The procedures (checklists and guidelines) for evaluating the different components of study quality are described in Sections 1.0 to 4.0. The guidelines state characteristics of the ideal studies; however, most epidemiologic studies typically do not meet these ideals, and the evaluation of their ability to identify a cancer hazard is performed in the context of how close or far they approach these ideals.

1 Reporting quality checklist

- Is there adequate documentation and reporting of the (1) description of the selection and follow-up of the population, (2) methods to assess exposure and disease, (3) analytical methods, and (4) cancer findings?

2 Analyses of biases: *a priori* guidelines and checklists

The application of the RoC listing criteria to the body of studies on *ortho*-toluidine include an analysis of whether any association observed between exposure to *ortho*-toluidine and cancer can be explained by chance, bias, or confounding. The first step in the assessment is to evaluate the study methods to determine whether there is a potential bias. Biases in observational studies are often classified into three major categories: (1) selection bias, (2) information bias, and (3) confounding (Rothman *et al.* 2008[3]). Studies with lower potential for bias are generally considered to be the most informative for cancer evaluation. However, the presence of a potential bias in a study does not necessarily mean that the findings of the study should be disregarded. Therefore, an important step in the process of evaluating biases is to determine the probable impact of the described biases on study results—that is, the magnitude of distortion and the direction in which each bias is likely to affect the outcome of interest. This step is reflected in the second part of the checklist questions and is analyzed in the assessment of the level of evidence (Appendix D).

Checklists and guidelines for evaluating methods used to select the study population and obtain information on exposure and disease are provided in Section 2.1.1 and 2.1.2. The approach for evaluating confounding, which is a key issue in the cancer evaluation of *ortho*-toluidine, is discussed in Appendix D.

2.1.1 Selection and attrition

Checklist for evaluating selection and attrition-related biases: cohort or case-control studies

- Are the non-exposed subjects and exposed subjects from the same underlying population? If not, what information is available to estimate the potential direction and relative magnitude of distortion from the bias? Is there any evidence of a healthy worker hire effect? If so, what is the direction and relative magnitude of the distortion from the bias on the risk estimate?

[3] Rothman K, Greenland S, Lash T. 2008. *Modern Epidemiology, 3rd Edition*. New York: Lippincott, Williams, and Wilkins, 851 pp.

- In case-control studies, are controls selected from the same underlying population as the cases using similar inclusion/exclusion criteria? If not, what is the likely direction and relative magnitude of distortion from the bias?
- In case-control studies, is there any evidence that the methods used to identify and select the controls and/or cases are related to exposure to *ortho*-toluidine? If so, what information is available to estimate the direction and magnitude of distortion from any potential bias?
- Is there any evidence for self-selection or that refusal to participate in the study is related to both exposure and disease status? If so, what information is available to estimate the direction and magnitude of distortion from the bias?
- Is the ascertainment of vital status at the end of the follow-up period in cohort studies adequate? Does it differ between the exposed and unexposed subjects? Is there any evidence to suggest that completeness of follow-up is related to both exposure and disease (e.g., urinary bladder cancer mortality or incidence) status? If so, is it possible to predict the direction and magnitude of distortion from the bias?
- Is there any evidence of a healthy worker survival effect or left truncation[4] in cohort studies? If so, were appropriate analyses performed to control for the potential bias? Is it possible to predict the direction and relative magnitude of distortion from any uncontrolled (residual) bias?

Guidelines

In cohort studies, the exposed and unexposed groups should ideally be similar in all respects except for exposure to *ortho*-toluidine. Occupational cohorts would consist of all potentially exposed employees within a given plant (employed over a specified period of use of *ortho*-toluidine) compared with similar unexposed employees from within the same plant (i.e., internal controls) to minimize the healthy worker effect and other differences between exposed and unexposed groups. When external referents are used in e.g., SMR or SIR studies, local (or regional) mortality or incidence rates are generally, but not always, preferable to national rates.

Systematic biases may be introduced if the length and completeness of follow-up differ between exposed and non-exposed groups and are related to the outcome of interest. Ideally, the total loss to follow-up should be less than approximately 5% over the duration of the study observation period. Overall, studies should have more than 80-90% total follow-up, although incidence studies may have greater loss to follow-up than mortality studies.

In case-control studies, controls and cases are selected from the same underlying population and representative of the population from which they were selected. Controls should be free of any diseases related to *ortho*-toluidine exposure; the use of controls with diseases related to *ortho*-toluidine exposure would bias towards the null hypothesis. Ideally, participation rates should be high and should be similar for cases and controls, although it is recognized that participation rates in population-based case-control studies are often lower than those in a hospital-based study.

2.1.2 Information (observation) bias

Studies will be evaluated for their adequacy in measuring exposure and disease endpoints, including missing data and the probability of misclassification of exposure and disease.

[4] Healthy worker effect can also be considered as a confounder.

Checklist for information bias

- What is the method used to assign exposure to subjects according to their potential for *ortho*-toluidine exposure? Does the method permit exposure assessment for individual subjects or only for exposure groups? Is the exposure measure qualitative, semi-quantitative, or quantitative? Are errors (if any) in classifying exposure similar (non-differential misclassification) or different (differential misclassification) across study groups? If there is evidence for misclassification of exposure, what is the direction and relative magnitude of distortion from the bias?
- Are exposure data missing for the cases and controls or cohort members? Were missing data imputed, and if so, how was this done and are these methods adequate?
- What is the level of confidence that the study was able to identify and classify subjects accurately and completely with respect to cancer endpoints? Was disease assessed similarly across study groups? If disease was misclassified, is it possible to predict the direction and relative magnitude of distortion from the bias?

Guidelines: exposure assessment

One of the most important aspects of a study is the ability to characterize exposure at the individual level. The ideal is to have quantitative estimates of exposure to *ortho*-toluidine and relevant co-exposures for each individual that are based on a job-exposure matrix or expert assessment that link the subject's occupation history (e.g., job or department titles, task descriptions, duration of employment, calendar years worked) with data on relevant industrial processes that are calendar-year specific. Ambient air data that validate estimates of exposure intensity, relevant biomonitoring data, and exposure estimates using multiple metrics (such as cumulative, peak, average) improve the quality of the assessment. Studies inferring exposure to *ortho*-toluidine via industrial processes alone have a higher probability of nondifferential misclassification of exposure. In general, exposure is better characterized in occupational cohort studies than in population-based case-control studies, but in both types of studies misclassification is typically non-differential, usually biasing towards the null. Recall bias in case-control studies in which occupational exposure is assigned based on job titles is less likely to be a concern than in studies using self-assessment of chemical-specific exposures (e.g., use of questionnaires with exposure check-list).

Exposure to *ortho*-toluidine may occur via inhalation, oral, or dermal routes. Potential exposure via the dermal route has not been well characterized but may be measured indirectly using biological monitoring. Urinary excretion of *ortho*-toluidine provides an integrated measure of short-term absorbed dose via all routes of exposure (most is excreted within 48 to 72 hours). Urine levels of *ortho*-toluidine have been found to correlate adequately with ambient air levels of *ortho*-toluidine. *ortho*-Toluidine hemoglobin adducts also appear to provide a valid short-term indicator of absorbed *ortho*-toluidine dose.

Guidelines: outcome assessment

Incidence data from population-based cancer registry sources generally provide more detailed and accurate diagnostic data and more accurate population (comparison) cancer rates than death certificate-based mortality data, while being less easy to access. In addition, cancer incidence data may be considerably more informative than mortality data (depending on ascertainment, reporting, and diagnostic accuracy) for cancers with longer survival times, as in the case of urinary bladder cancer. There is also evidence that urinary bladder cancer may be under-ascertained in some mortality studies due to reporting as a contributory rather than underlying cause of death on some

death certificates (Axtell *et al.* 1998[5]). Length of follow-up is also critical in identifying cases or deaths from long latency cancer endpoints such as urinary bladder cancer (see Section 2.2).

3 Ability of the studies to detect an effect

Factors that influence the ability of a study to detect an effect (if present) include the statistical power of the study, the level or duration of exposure to *ortho*-toluidine, the exposure range studied, and the length of follow-up in cohort studies (or follow-back in case-control studies). In general, study characteristics that increase the potential for random error (such as high loss to follow-up in cohort studies or low participation rates in case-control studies that are not related to exposure or disease status) will bias the study findings towards the null hypothesis (see Section 2.1.1 for questions and guidelines).

Checklist

- Is there adequate statistical power to detect an effect in the exposed population or subgroups of the exposed population?
- What were the levels of exposure and the duration of exposure of the cohort population?
- Were risk estimates calculated for subgroups of workers with higher levels or longer durations of exposure?
- Was the follow-up or follow-back period adequate to allow for a cancer induction period of 20 years or greater?
- Was the relevant window of exposure taken into account in the exposure assessment in the case-control study of exposure to *ortho*-toluidine and childhood cancer?
- Were any analyses of exposure lagging adequate for detecting cancers with long latency such as urinary bladder cancer?

Guidelines

The overall number of study participants and numbers in each exposure or case group will be evaluated in terms of power to detect given elevations of risk for specific endpoints while controlling, if necessary, for potential confounding. In the United States urinary bladder cancer has age-adjusted annual incidence rates of 37.0 and 8.9 per 100,000 among males and females, respectively, and corresponding annual mortality rates of 7.7 and 2.2 per 100,000 (NCI, SEER 2012[6]). In the European Union, age-adjusted incidence rates in 2008 were 27.4 and 5.6 per 100,000 in males and females, respectively[7], and the corresponding mortality rates were approximately 6 and 1 per 100,000 in the early part of the decade (Ferlay *et al.* 2008[8]). Ideally, studies should have at least 80% power to detect a 2-fold increased relative risk.

Studies of workers in industries with higher levels of exposure or workers with longer duration of exposure, and sufficient variability in exposure are generally more informative for evaluating cancer risk. Studies evaluating exposure groups in which the majority of workers classified as "exposed" have in fact very low exposure, very short duration of employment, or limited evidence of actual exposure may be inadequate to detect an effect due to a dilution effect. Groups of workers

[5] Axtell CD, Ward EM, McCabe GP, Schulte PA, Stern FB, Glickman LT. 1998. Underlying and multiple cause mortality in a cohort of workers exposed to aromatic amines. *Am J Ind Med* 34(5): 506-511.
[6] http://seer.cancer.gov/statfacts/html/urinb.html; years 2005 to 2009.
[7] http://info.cancerresearchuk.org/cancerstats/types/bladder/incidence/uk-bladder-cancer-incidence-statistics#geog.
[8] Ferlay J, Randi G, Bosetti C, Levi F, Negri E, Boyle P, La Vecchia C. 2008. Declining mortality from bladder cancer in Europe. *BJU Int* 101(1): 11-19.

who are classified as definitely exposed should include only workers for whom there is adequate evidence of exposure to *ortho*-toluidine.

Urinary bladder cancer has been reported as having a latency ranging from 18 to over 40 years (Matanoski and Elliot, 1981[9]) and relatively long survival time from diagnosis (ranging from a median of approximately 40 to over 100 months, depending on cell type, stage, demographic, and other factors). The median age at diagnosis of urinary bladder cancer is 73 years of age and the expected 5-year relative survival rate is 78% in the United States (both sexes combined; NCI SEER 2012[10]). This suggests that follow-up times, particularly for mortality analysis, ideally should be 20 or more years from first exposure. Inadequate follow-up (or follow-back in case-control studies) may bias findings toward the null.

4 Analytical methods

The use of appropriate analytical methods will be evaluated. Analysis of exposure-response relationships (using either linear models or exposure categories) and calculation of trends using quantitative exposure assessments adds more information than analysis by simple binary exposure categories, as does analysis of tumor site by average, cumulative, peak, or duration of exposure, time since first exposure, calendar periods of exposure, and exposure lags. Without *a priori* knowledge, it is difficult to know which exposure metric is most appropriate for evaluating causality, so a positive relationship observed with any exposure metric is a concern. In addition, all analyses should examine and, if necessary, adjust for demographic variables and other potential confounders of a priori interest, if not done so in the study's design. Evidence of under- or over-controlling for confounding, multicollinearity, and residual confounding will also be evaluated. In the absence of information on confounders, analyses using internal referents who are similar to the exposed subjects can help reduced potential confounding (see Appendix D for further discussion on the evaluation of methods to assess potential confounding). Ideally, analytical methods should also identify and consider potential modifying variables; however, many studies do not have adequate statistical power to adequately evaluate effect modification.

[9] Matanoski GM, Elliott EA. 1981. Bladder cancer epidemiology. *Epidemiologic Rev* 3: 203-229.

[10] http://seer.cancer.gov/statfacts/html/urinb.html.

Appendix D: Approach for evaluating confounding

A key question in the evaluation of the level of evidence from human studies is whether an association (if any) between exposure to *ortho*-toluidine and cancer can be explained by confounding. Potential confounders include any exposures or risk factors that could be associated with both exposure to *ortho*-toluidine and the disease outcome(s) of interest, i.e., urinary bladder cancer, and that are not part of the disease pathway.

The evaluation of potential confounding will take into account the following:

- Identification of potential confounders and evidence related to their carcinogenicity potency
- Assessment of the level of the potential confounder(s) the workers were exposed to compared with *ortho*-toluidine exposure
- Assessment of analytical or statistical methods to control for variables with evidence of confounding
- The magnitude of the risk estimate for exposure to *ortho*-toluidine and cancer

1 Identification and characterization of potential confounders

In the occupational cohorts included in the studies under review*, ortho*-toluidine-exposed workers are typically exposed to a large number of co-exposures. These may have been quantified or noted by the study authors or may be inferred from expert knowledge of the industrial processes described by the authors. Whether or not a given co-exposure should be considered as a potential confounder depends on whether there is *a priori* evidence that the co-exposure is potentially associated with urinary bladder cancer. While a full secondary review of the literature on each co-exposure is beyond the scope of the evaluation monograph, authoritative reviews or original toxicological and epidemiologic studies on each co-exposure specified or expected to be present among *ortho*-toluidine-exposed workers will be consulted, where available, to evaluate the likelihood that it is potentially carcinogenic. In general, the guidelines used to categorize a co-exposure as potentially carcinogenic would include one or more of:

- Classification as a Group 1, 2A, or 2B carcinogen or equivalent category by IARC, NTP, EPA or comparable government agency;
- Evidence of carcinogenic activity from animal data (irrespective of tissue site concordance);
- Strong evidence of carcinogenic potential from mechanistic data or structure-activity relationships.

2 Assessment of analytical methods to evaluate confounding

Studies will be evaluated for their adequacy in measuring potential confounders, such as occupational co-exposures, age, and lifestyle factors, and the appropriateness of the analytical methods and models used to control for confounding,

Checklist

- How well were co-exposure, age, or lifestyle factors (such as cigarette smoking) measured in the study? If there are no actual data on confounders, are surrogate data on potential confounders available?
- Does the design or analysis control or account for important confounding through matching, stratification, multivariable analysis, or other approaches?

- Are the models used to control for confounding appropriate? What strategy was used to determine whether the variable belongs in the models? Is there evidence for under- or over-controlling for confounding, residual confounding, or negative confounding?

Guidelines

Ideally, all potential confounders should be quantified and considered for inclusion in the statistical analysis for confounding, using appropriate statistical models. Final statistical models should only include "actual" confounders and not variables that have no effect on the risk estimate. Guidelines for evaluating methods to assess exposure to confounding are as follows:

Occupational co-exposures: Ideally, studies would provide quantitative exposure data for each potential confounder as part of a job-exposure matrix or expert assessment for each worker, but this is rare. However, some studies provide quantitative or qualitative data on co-exposures for subsets of workers, which can be used to evaluate potential confounding. In addition, knowledge of industrial processes may also be helpful in providing relative estimates of the ratio of exposure to *ortho*-toluidine and exposure to the potential confounder.

Lifestyle factors: Ideally, quantitative information on smoking and other non-occupational exposures should be assessed, and preferably by in-person interview by interviewers blinded to the status of the respondent in cancer incidence studies, rather than via proxy respondents, work records, or other indirect methods. For smoking, information on level (such as number of packs smoked), duration, and status (e.g., current or ex-smoker) is desirable. Residual confounding is more likely when only limited qualitative information (such as yes or no) is available. Typically, few or no data are available on non-occupational risk factors in historical cohort studies.

3 Impact of potential confounders on study findings

Ideally, all potential confounders should be both quantified and subject to consideration for analysis for confounding, using appropriate statistical models, or confounding should be controlled for using other methods such as in the selection of the study participants. In many cohort studies, there is a paucity of quantitative co-exposure data at the individual level; however, an indirect evaluation of the impact of confounding from co-exposures may be conducted by considering (1) the relative levels of exposure to *ortho*-toluidine compared with exposure to the confounder, (2) the relative carcinogenicity potency (such as in experimental animals) of the confounder compared with *ortho*-toluidine, and (3) the magnitude of the risk estimate for *ortho*-toluidine and cancer. Indirect information on the relationship between the estimated levels of exposure (albeit based on crude approximations) to the confounder compared with the estimated level of exposure to *ortho*-toluidine may be available from air monitoring or biomonitoring studies of subsets of workers or from expert knowledge of industrial processes.

Typically, few or no data are available on non-occupational risk factors in historical cohort studies. Internal comparison groups and analyses can help reduce confounding from non-occupational risk factors. In the case of urinary bladder cancer, the major non-occupational exposure of *a priori* concern is smoking. Smoking data may not be available at all or available only for a subset of exposed and/or non-exposed cohorts or case and control subjects. Where available, such data should be taken into consideration as noted above. Indirect evidence of confounding, such as risk estimates for smoking-related non-malignant disease, such as respiratory disease or cardiovascular disease, can also be used to evaluate whether smoking was associated with exposure to *ortho*-toluidine.

Appendix E: Synthesis

This section outlines the approaches for synthesizing the findings across the body of studies for each endpoint (primarily urinary bladder cancer) and making a recommendation on the level of evidence (e.g., sufficient, limited, or inadequate) of the carcinogenicity of *ortho*-toluidine from studies in humans. Studies with the lowest risk of bias and greatest sensitivity to detect an effect will be identified by using the guidelines and checklist described in Appendix C, and these studies are given the most weight in the assessment. The application of the RoC listing criteria to the body of studies on *ortho*-toluidine includes evaluating (1) whether there is credible evidence for an association between exposure to *ortho*-toluidine and cancer, and (2) whether such an observed association can be explained by chance, bias, or confounding. Several existing guidelines – strength of the association, consistency across studies, evidence of an exposure-response gradient, and temporality of exposure (Hill 1965[11]) – are used to help guide the evaluation of these questions. It should be noted that they are not criteria, and with the exception of temporality, each and every element is not required to demonstrate causality. Emphasis should be placed on evaluating the extent to which biases, or confounding by co-exposures that may also cause urinary bladder cancer could explain observed increases in cancer risk.

The cancer assessment will evaluate the following:

- *Temporality*.
- *The consistency of findings across studies with the most adequate methodologies*, as evaluated according to the guidelines described in Appendix C. Consistency needs to be evaluated in the context of variations in outcome definitions, exposure assessment methodologies, exposure levels or duration of exposure of the population, exposure windows, length and completeness of follow-up or other differences in population characteristics or study methodologies.
- *The strength of observed associations between ortho-toluidine exposure and cancer*. The strength of the association is important in evaluating whether specific confounders or biases can explain the observed association; however, the fact that an association is weak does not necessarily rule out a causal relationship.
- *Evidence for an exposure-response gradient*. A positive exposure-response relationship generally provides more convincing evidence of a causal association than a simple excess of disease. However, there may be biological or methodological reasons for not observing a gradient, and the absence of evidence for an exposure-response relationship is not strong evidence *per se* for the absence of a causal association.
- *Evidence for associations with appropriate latency*.
- *Alternative explanations of chance, bias, or confounding*. The process for identifying potential biases was discussed in Appendix C, and that for evaluating potential confounding was outlined in Appendix D. As noted in Appendix C, the presence of bias in a study does not mean that the study should be disregarded; the potential for the bias should be analyzed to determine its impact (including the direction and magnitude) on the study findings (e.g., risk estimates for *ortho*-toluidine and cancer). The finding of consistent, elevated, positive associations across studies in different populations, with different study designs and in different occupational settings, reduces the likelihood that specific biases or potential confounders in individual studies can explain the associations observed across the body of studies.

[11] Hill AB. 1965. The environment and disease: association or causation? *Proc R Soc Med* 58 (5): 295-300.

- *Evidence from methodologically limited studies*, including an evaluation of whether such limitations can help explain inconsistent findings.