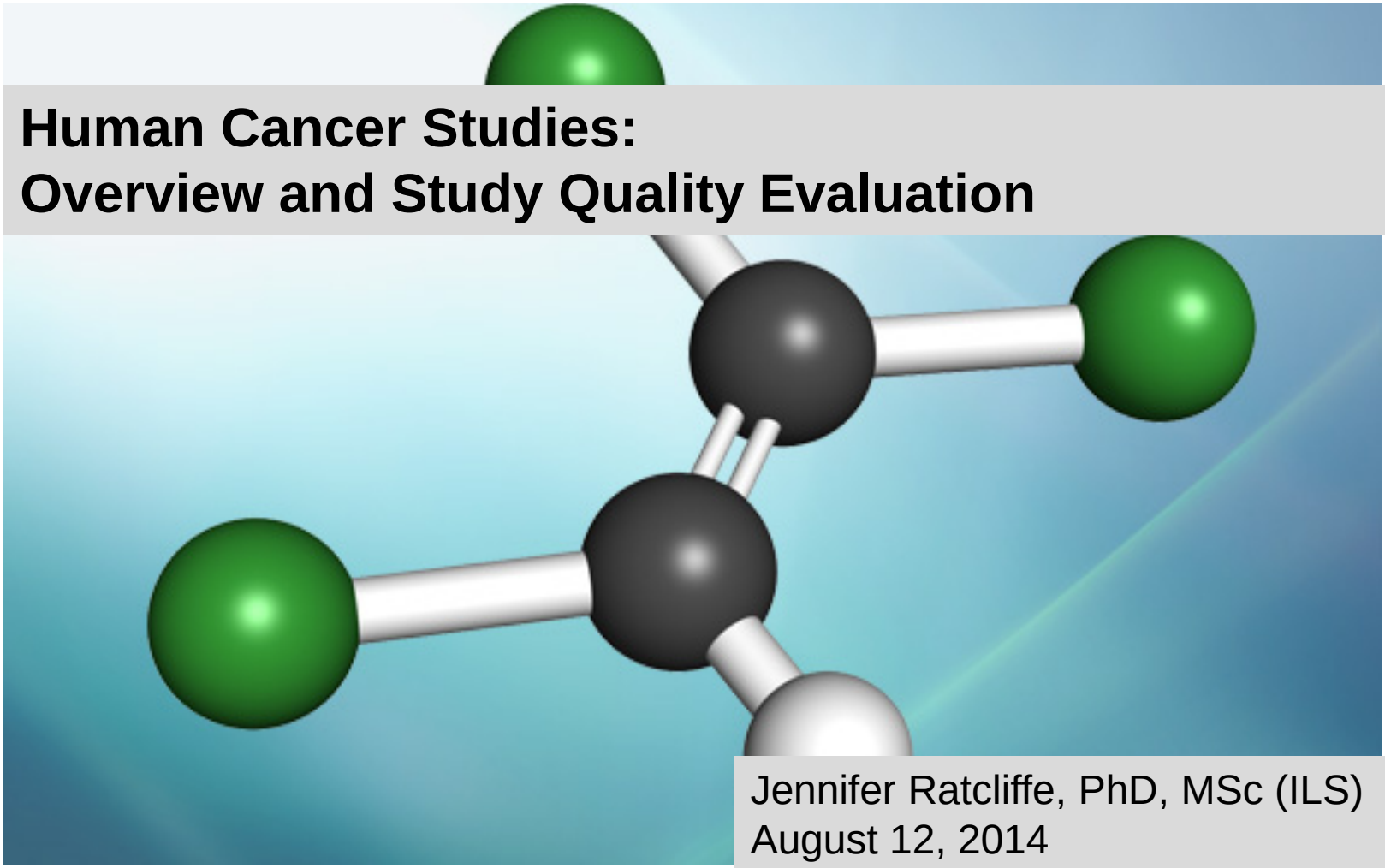
# Human Cancer Studies: Overview and Study Quality Evaluation



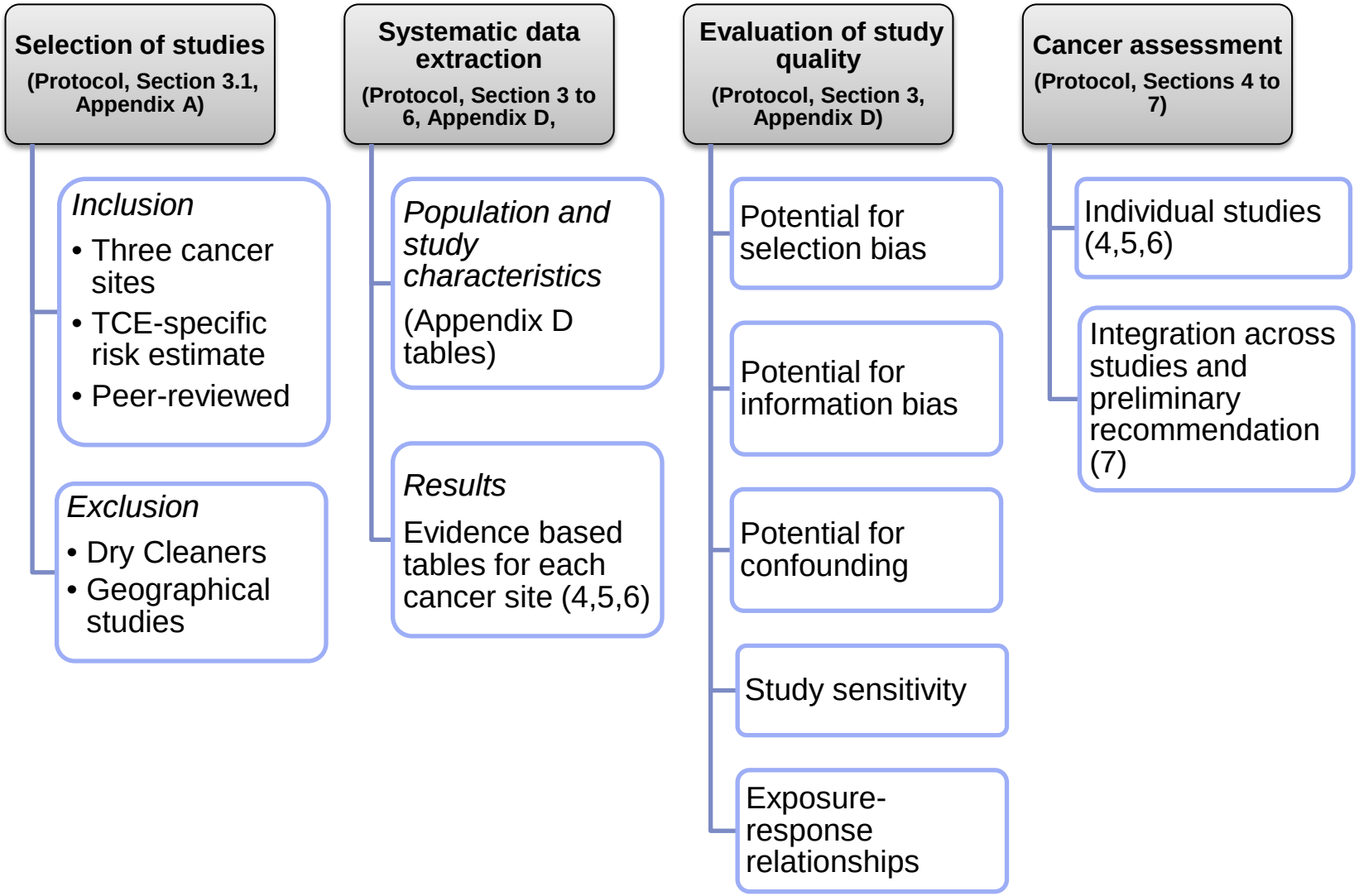
Jennifer Ratcliffe, PhD, MSc (ILS)
August 12, 2014

# Outline

- Steps in the evaluation of human cancer studies
- Studies included in the evaluation of study quality
- Overview of the evaluation of study quality
- Summary of guidelines for the evaluation of study quality elements and examples of studies
- Integration of study quality elements (cohorts example)
- Overall strengths and limitations of the database

Reviewer comments and panel discussion

# Evaluation of human cancer studies



# TCE cohort and nested case-control studies (16)

| Population (# studies) | Endpoints |
|---|---|
| Nordic cohorts (various occupations) (3) | Incidence (3)<br>External + internal (3)<br>**Kidney, NHL, MM, liver (3)** |
| US Aerospace/aircraft workers cohorts (5) | Mortality only (3)<br>Mortality + Incidence (2)<br>External/internal (5)<br>**Kidney, NHL, MM, liver** |
| Other industry cohorts (7)<br>(US uranium workers [3], US electronic workers [2], US rubber mfg. (1), German cardboard mfg.[1]) | Mortality only (2)<br>Mortality + incidence (2)<br>External + internal (3)<br>Internal only (1)<br>Nested case-control (3)<br>**Kidney (4) NHL (4) MM (1) liver (2)** |
| US environmental (drinking water) exposure (1) | Mortality<br>External + internal<br>**Kidney, NHL, liver** |

# TCE case-control studies (15)

| Population (# studies) | Endpoints |
|---|---|
| **Kidney or liver cancer (7)** | |
| Studies in specific industrial areas (4)<br>US population (1)<br>German population (1) | **Renal cell carcinoma** |
| Montreal population (1) | **Renal cell carcinoma**<br>**Liver cancer** |
| **NHL and related subtypes (8)** | |
| Pooled study (of 4 studies Europe and US) | **NHL and subtypes** |
| Population-based (US, Canada, Nordic) (5) | **NHL and subtypes (1)**<br>**NHL, (4), HCL (1)** |
| US population-based (1) | **MM** |
| Italy population-based (1) | **MM, CLL** |

# Overview of study quality evaluation

- Each primary study was systematically evaluated according to guidelines laid out in the TCE protocol (website)
- Input from webinar and technical advisors
- Studies given most weight are generally those that provide the most valid (low risk of systematic biases) and precise (low risk of random biases) risk estimates and have adequate sensitivity to detect effects, adequate methods to evaluate potential confounding and appropriate analytical methods and reporting
- The impact of potential biases (direction and magnitude) and potential for confounding on study findings is evaluated in the cancer risk assessment (Sections 4,5,6)

# Overview of study quality evaluation

- Terms used to evaluate selection bias and some information biases:
  - *Unlikely:* Information from design and methods indicate potential for bias is unlikely and study is close to ideal study characteristics
  - *Possible:* Study design or methods are close to but less than ideal – there is some potential for certain types of bias
  - *Probable:* Study designs or methods suggest that potential for specific type of bias is likely

Note: The presence of a specific risk of bias does not necessarily mean that the magnitude of the bias is sufficient to strongly affect observed positive risk estimates

# Selection and attrition bias

- Selection bias (unlikely, possible, probable):

Unlikely if:

  - Cohorts or cases/controls represent the same underlying population
  - Little or no evidence of healthy worker hire or survival effect
  - Cases and controls are selected by similar criteria not related to TCE exposure
  - Participation in case-control studies is high and not related to exposure or disease status
  - Loss to follow-up < approx. 5% and similar in both groups?

*Example:*

| Selection bias | Study characteristics |
|---|---|
| Unlikely (Zhao 2005) | All workers + potential TCE exposure selected for cohort |
| Probable (Henschler 1995) | Cohort based on cluster |

# Quality of exposure assessment

- Input from webinar
- Ranking (good, adequate, limited, inadequate): Good if many of the following were present:
    - Industrial hygiene or biomonitoring data for TCE
    - Individual detailed job-exposure matrices with expert assessment
    - Job or task descriptions
    - Consideration of frequency, confidence and intensity of exposure
    - Calendar period-specific exposure data

*Example:*

| **Exposure assessment** | **Characteristics** |
|---|---|
| Adequate to good (Zhao 2005) | Semi-quantitative JEM (no quantitative exposure)<br>Multiple metrics<br>Calendar year-specific<br>Co-exposures |
| Limited (Henschler 1995) | Exposure based on job location<br>Level, duration not measured |

# Misclassification of exposure

- Includes quality of exposure assessment and exposure setting
- Consider ever exposure and exposure intensity separately
- Consider whether non-differential or differential

*Example:*

| Exposure misclassification | Study characteristics |
|---|---|
| Not a concern (Charbotel 2006) | Adequate to good exposure assessment<br>Exposure prevalence high |
| A concern (Hardell 1994) | Self-reported exposure<br>Minimal (1 wk) ever exposure |

# Study sensitivity

- **Sensitivity (good, adequate, limited):** depends on statistical power, levels and duration of exposure, length and completeness of follow-up, and misclassification of exposure
- A ranking of good was given if many of the following were present:
  - Larger numbers of exposed subjects or cases
  - Adequate length and completeness of follow-up (cohorts)
  - High levels of exposure or long exposure duration
  - Large groups or subgroups with a range of exposures
  - Little concern about exposure misclassification

*Example:*

| Study sensitivity | Characteristics |
|---|---|
| Adequate (Cocco 2013) | Adequate exposure assessment<br>Little exposure misclassification among subjects with high confidence<br>Large numbers of cases and controls |
| Limited (Vlaanderen 2013) | Exposure low<br>Misclassification a concern |

# Disease assessment and misclassification

- Quality of disease assessment (good, adequate, limited) was considered good if many of the following were present:
  - Multiple verified sources of case/death ascertainment (e.g., cancer registries, SSA, vital records, hospital records)
  - Consistent coding/classification (may be issue for NHL)
  - Histological or pathologist- confirmed cases
  - Almost complete follow-up and sufficient length (based on latency)
  - No evidence of bias in case ascertainment

*Example:*

| Disease assessment | Characteristics |
|---|---|
| Adequate (Zhao 2005) | Cancer registries<br>National death index<br>Missing data NR |
| Limited (Henschler 1995) | Different methods used for exposed cohort vs. general population |

# Assessment of analytical methods for evaluating confounding

- Internal and external analyses in cohort studies (all)
- Adjustment for co-exposures or any other potential confounders in design or analysis
  - Few studies adjusted for co-exposures
  - Many case-control studies for lifestyle factors
- Matching for age, sex, race, calendar period (most case-control)
- Indirect methods of evaluating confounding (e.g., lung cancer rates) in cohort studies

# Overall study quality – High

Each of the key elements are close to ideal

- Little evidence of selection or information biases
- Misclassification of exposure and disease are not a concern
- Study sensitivity is adequate or good
- Potential for confounding appears minimal or is considered in design or analysis

# Study quality: Cohort studies

| Study quality | Cohort studies |
|---|---|
| High | Zhao 2005 |
| Moderate | Hansen 2013 |
| | Radican 2008 |
| | Morgan 1998 |
| Low/moderate | Lipworth 2011 |
| | Yiin 2009 |
| | Boice 2006 |
| | Raaschou-Nielsen 2003 |
| Low | Silver 2014 |
| | Bove 2014 |
| | Vlaanderen 2013 |
| | Greenland 1994 |
| | Henschler 1995 |
| | Ritz 1999 |
| | Bahr 2011 |

Grey: Study quality; Blue = sensitivity (light = high); Lilac = bias (positive Henschler (kidney), Ritz (liver); Peach = other concerns

# Overall strengths and limitations of database

- Strengths:
  - Large database with studies of different occupations in different geographical locations
  - Several cohort studies, and several case-control studies of kidney cancer and NHL, are of high or moderate quality
  - Studies of liver cancer are more limited, primarily due to lack of case-control studies
  - Cohort studies include internal analyses and some of specific industries control for co-exposures
  - Many case-control studies control for lifestyle factors

# Overall strengths and limitations of database

- Limitations:
  - Some studies considered to be of low or moderate quality due to low sensitivity (limited power or low levels of TCE), although methodologies may be adequate
  - Observed selection or information biases in the majority of the studies are nondifferential and tend to bias toward the null
  - In a few studies biases or potential for confounding were towards an overestimate of the risk estimate (Hardell 1994, Vamvakas 1998, Henschler 1995, Ritz 1999)
  - Few studies have adequate data to evaluate exposure-response relationships

# Evaluation of human cancer studies: Reviewer questions

- Comment on the overall approach for preparing the cancer assessment of the epidemiologic studies.
- Comment on whether the methods for evaluating study quality and other related issues are systematic and transparent.
- Comment on whether the assessment of the utility of the studies for informing the cancer evaluation is reasonable and clearly presented.