

Documentation for Disease-Lifestyle Relations Annotation

Annotation scope

The scope of this annotation is to detect lifestyle factors which affect the risk of disease onset and development. Below are some general examples of the guidelines that will be used in the annotation.

General guidelines

- Annotations should be made according to the annotator's best understanding of the **author's intended meaning in context**.
- Annotators should treat named entities as being **masked**, i.e. they shouldn't annotate relationships between entities just based on their names, when they would be unable to make the same annotations for two other entities.

What to annotate:

1. Causal relationship: *LSF causes disease / disease causes LSF*. Examples:

- A **LSF** causes a **disease**.
- A **LSF** contributes to the development of a **disease**.
- A **LSF** developed a **disease**.
- A **LSF** is a recognized/reversible/know/common cause of a **disease**
- **LSF-induced disease**
- A **LSF** may induce a **disease**.
- A **LSF** increases a **disease** mortality.
- A **disease** is a result/side-effect of a **LSF**.
- A **disease** is attributed to / transmitted by / determined by a **LSF**.
- The consequences of a **LSF** is a **disease**.

2. Statistically associated relationship: *LSF is statistically associated with disease*.

Examples:

- A **LSF** is associated with a **disease**.
- A **LSF** is associated with the development of **disease**.
- A **LSF** is associated with the risk of a **disease**.
- A **LSF** is associated with **disease** treatment and control.

- A **LSF** has a role in/effect on a **disease**.

2.1 Positive statistical association:

- A **LSF** increases the risk of a **disease**.
- A **LSF** is the risk factor a **disease**.
- A **LSF** carries a risk of a **disease**.
- A **LSF** is a predictor of a **disease**.
- A **disease** is characterized by a **LSF**.
- **disease** patients were more likely to practice **LSF** than controls.
- A **LSF** increases incidence of a **disease**.
- A **LSF** increases the risk of developing a **disease**.
- A **LSF** increases prevalence of a **disease**.
- A **LSF** contributes to the burden of a **disease**.
- A **LSF** is linked/connected to a **disease**. → (implies positive direction and is always used as such)
- A **LSF** should be considered during risk stratification for a **disease**.
- The prevalence of **LSF** was higher in **disease** patients than in controls. → (such a **comparative** mention in abstracts implies significance in the majority of cases even when significance is omitted. Comparative prevalence numbers (x % vs x'%) can also work).

2.2 Negative statistical association:

- A **LSF** decreases the risk of a **disease**.
- A **LSF** decreases incidence of a **disease**.
- A **LSF** decreases prevalence of a **disease**.
- A **LSF** could be critical in the current fight against **disease**.
- A **LSF** is inversely associated with a **disease**.
- The prevalence of **LSF** was lower in **disease** patients than in controls.

3. Controls: *LSF controls disease*. Examples:

- A **LSF** play a regulatory role in **disease**.
- A **LSF** has beneficial effect for the control of **disease**.
- A **LSF** improves survival outcomes in **disease**.
- A **LSF** decreases/reduces/attenuates **disease** → (not the prevalence of **disease**).

3.1 Prevents relationship: *LSF prevents disease / disease prevents LSF*. Examples:

- A **LSF** is therefore imperative for preventing a **disease** morbidity and mortality.
- A **LSF** is chemopreventive in a **disease**.
- A **LSF** is protected against a **disease**.
- A **LSF** reduce a **disease** mortality.

3.2 Therapeutic relationship: *LSF treats disease*. Examples:

- The treatment of a **disease** includes a **LSF**.
- A **LSF** is essential for treating a **disease**.
- A **LSF** is used for the therapy of a **disease**.
- The efficacy of a **LSF** in a **disease**.
- A **LSF** is the relief of a **disease**.

- A **LSF** is was effective in a **disease**.
 - A **disease** were eliminated by a **LSF**.
 - A **disease** were improved after (using) a **LSF**.
4. No statistical association: *LSF is not associated with disease*. Examples:
- A **LSF** is **not** associated with a **disease**.
 - There is no association between **LSF** and **disease**.

What **NOT** to annotate:

1. Hypothetical statements:
 - “Here we study the link between LSF and disease”.
 - “It is possible to suspect a relationship between ESRD and insecticides or pesticides”.
 - “LSF might be involved in Dis”. → Take context into consideration in case this is no more than a hypothesis.
2. Tendency but no statistical significance :
 - “We **did not** find the relationship between LSF and disease to be statistically significant” / “There is an association between **LSF** and **disease**, but **no significance**”.
3. No statistical test implied/ no control group comparison:
 - “A majority percentage of HIV-positive MSM engage in unprotected sexual behavior”.
 - other individuals without HIV could have the same behavior.
 - “A total of 45% of children receiving **LSF** had no symptom recurrence of **disease**”.
 - “In our study, **disease** was very common in **LSF** practitioners”.
 - “In our study, 54% of cancer patients suffer from poor sleep and 34% of low energy”.
 - What is problematic is not the lack of a significance report but the **absence of a control group** implication in all above cases. We cannot make an assumption that a statistical test was actually performed.

or

Observation:

 - “Cannabis is the most widely used illicit substance in the United States with especially high prevalence of use among those with psychiatric disorders.”
4. LSF that is a part of a bigger Named entity: sleep in multiple sleep latency test (MSLT)
5. Do NOT annotate “... in/among LSF/DIS”:
 - Dairy farmers** is not part of relation: Among **dairy farmers**, moreover, **lung cancer** SMRs showed a significant downward trend across the quartiles of increasing **length of work**.
6. Statistical associations should not be annotated if p-value is greater than 0.05.
7. Statistical associations should not be annotated if CI encompasses 1.0.
8. Inconsistent/Debatable evidence → do not annotate.

Special rules for relationships:

1. Across sentence boundaries should be annotated.
2. "Is believed" should be annotated.
"Air pollutants are believed to induce or exacerbate a range of inflammatory diseases (atopic dermatitis...)"
3. In cases where "Limited/Weak/Poor/Little evidence" is mentioned → Judge the author's intention: If the author implies that the evidence is inadequate, do not annotate. In the opposite case, annotate.
"Results provide limited evidence for an association of early-life mobile source air pollution with childhood asthma incidence ..."
4. Animal experiments should be annotated, as they are supposed to be a model for a human disease.
5. Be careful with an occupation + a clause with LSFs.
In the following examples, farmers **should not** be linked with acute lymphatic or chronic lymphatic leukemia.\
"Farmers from major corn-producing, hog- and chicken-raising, and pesticide- and fertilizer-using counties tended to be at higher risk of acute lymphatic".\
"Farmers from counties with large cattle inventories and significant dairy activity were at higher risk of chronic lymphatic leukemia".\
6. Annotate what the sentence says, even if there are contradictory statement. Example:
Previous study shows A causes B... : annotate A Cause B.
In contrast to the previous study, A causes C, or C causes B or no relation... :
annotate either A causes C, C causes B or nothing
7. Mentions of "the X-Y association" between an LSF X and a disease Y should be annotated as statistically associated relationships.
8. Indirect relations should be annotated. However, in cases like "**LSF** was not independently associated with **disease**" do not annotate unless it specifically mentions that the LSF was dependently associated.
9. Relationships like the following:
LSF1 and **LSF2** when present together cause disease **disease**, but when **LSF1** is present alone it **does not** cause **disease**.
Annotate as *LSF1 causes disease, LSF2 causes disease* (for the first sentence) and no annotations between LSF1 and disease in the second sentence.
10. Compared/comparison: a) if the comparison is between LSF/not having LSF or Dis/not having Dis then annotate only LSF to Dis with the appropriate relationship (not the not-LSF or not-Dis).

"The seizure rate was significantly higher in cocaine users (37 [26%] of 142 patients) than in non-cocaine users (151 [15.2%] of 992 patients, $p = 0.001$)". → Annotate Positive SA between cocaine and seizure. b) if the comparison is between separate things (eg one type of cancer to another) then annotate based on the direction you would assume could be applied if the comparison were cancer/healthy controls or **do not annotate at all** if assuming is not possible.\

“The proportion of patients working in professions with exposures to known carcinogens was 33.5% for lung cancer, and 17.1% for large bowel cancer (p=0.000)”. → carcinogens cause cancer, so since lung cancer patients were more likely to work in carcinogen exposed professions than LB cancer patients then it is safe to say that lung cancer is positively associated with carcinogens (no annotation for LB cancer)”.

1. Annotate relationships even when they are not independent.
2. Also consider numbers in ORs, HRs, RRs for the direction of the association, even if not specifically written in words as “positive” or “negative” (eg OR>1 means positive association and OR<1 means negative association).

“**SO2** was also significantly associated with **birth defects** in the second month before the pregnancy (aOR = 1.31; 95% CI: 1.20 ~ 3.22)”.

1. A **LSF** used as defense for a **disease**. → annotate either as treats or prevents according to context.
2. Annotate LSF and **Disease mortality** associations as LSF and **Disease**.

Detailed guidelines

For information on Annodoc, see <http://spyysalo.github.io/annodoc/>.