

Multimedia Appendix 1: Annotation Guideline

1. Annotation of Disease Name

The annotation standard of disease name is the same with the NCBI disease corpus [1]. First, diseases are automatically annotated by using a commonly-used disease recognizer, DNorm tool [2], and then only human disease entities are reserved for annotating disease-disease association (DDA). Finally, our annotators will manually remove incorrect disease entity annotations generated from DNorm.

2. The Scope of Disease-Disease Association (DDA)

In our disease-disease association extraction (DDAE) corpus, the annotated DDAs are categorized into two types: **Positive** and **Negative** associations. Their definitions are as follows:

- **Positive association:** include *comorbidity*, *complication*, *physical association*, and *risk factor*, if there are any of the mentioned terms linking a pair of diseases, the type will be annotated as Positive association.
 - **comorbidity:** A sentence uses *comorbidity* keyword to link the DDA. Eg.

Diagram illustrating a comorbidity association: A central green box labeled "Association" is connected to two blue boxes labeled "Disease". The left "Disease" box is connected to the "Association" box via a green arrow labeled "Cause". The right "Disease" box is connected to the "Association" box via a green arrow labeled "Theme".

Furthermore, with DLB cases excluded, comorbidity of PD and AD was significantly more prevalent among subjects + VH than subjects -VH ($p < .001$).
[PD \(MeSH:D010300 \)](#) and [AD \(MeSH:D000544 \)](#) were comorbidity [PMID:25027359]

- **complication:** *Complication* is mentioned as a bridge of DDA. Eg.

Diagram illustrating a complication association: A central green box labeled "Association" is connected to two blue boxes labeled "Disease". The left "Disease" box is connected to the "Association" box via a green arrow labeled "Cause". The right "Disease" box is connected to the "Association" box via a green arrow labeled "Theme".

HZ and its most common complication, termed postherpetic neuralgia (PHN), often cause long-term psychological distress and physical disabilities leading to profoundly negative impacts on the quality of patients' lives.
[postherpetic neuralgia \(MeSH:D051474 \)](#) was a complication of [HZ \(MeSH:D006562 \)](#) [PMID:23088666]

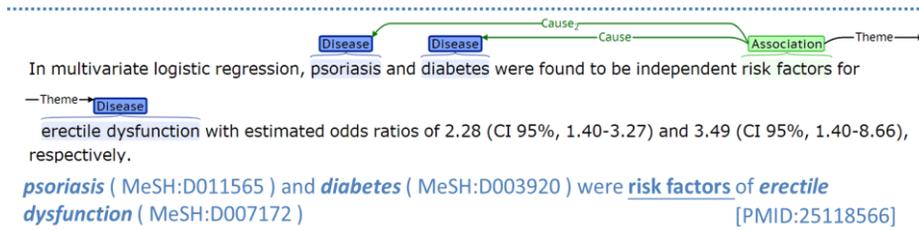
- **physical association:** The regular co-occurrence of DDA without additional information is assigned *physical association*. Eg.

Diagram illustrating a physical association: A central green box labeled "Association" is connected to two blue boxes labeled "Disease". The left "Disease" box is connected to the "Association" box via a green arrow labeled "Cause". The right "Disease" box is connected to the "Association" box via a green arrow labeled "Theme".

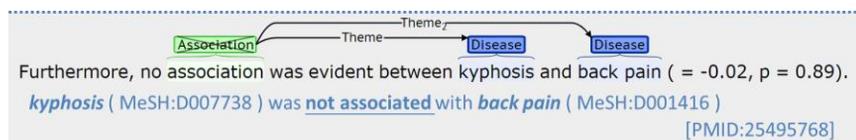
BACKGROUND: Menstrual disorders are common among women with schizophrenia, particularly when they are being treated with antipsychotics.
[Menstrual disorders \(MeSH:D008599 \)](#) was associated with [schizophrenia \(MeSH:D012559 \)](#) [PMID:23074030]

- **risk factor:** *Risk factor* is used for DDA when the target disease raises the

possibility of getting another one. For example:



- **Negative association:** Negative association is picked when the text clearly states that there is no association between two diseases. Eg.



3. Annotation of General DDA:

- **Annotation of Relation Keyword:** annotators have to annotate the evidence keyword(s) of DDA in a given sentence. For example, “*comorbidity*”, “*complication*”, “*symptom*” and “*risk factor*”. If there is no such keyword, the main verb which connects the disease relationship is the next available option. Moreover, once there is no proper verb, the annotators then consider the main adjective or noun which connects the relationship between two diseases.
- **Annotations of Cause and Theme:** If a DDA contains causal relationship, annotators should annotate one disease as Cause and another as Theme. Otherwise, both diseases will be labeled as Themes.

4. Annotation of Special DDA Cases:

- **The relationship between Disease_A and Disease_B in/among Disease_C (or ... with Disease_C):** If a sentence describes the relationship of Disease_A and Disease_B in Disease_C, annotators will consider Disease_C as a condition statement and treat it as an additional state. Then only the relationship between Diseases_A and Diseases_B is annotated as DDA. Eg.

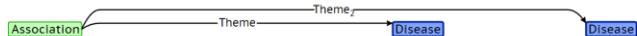


<PMID:26536829>

Above sentence can be simplified as among women who have “*schizophrenia*”, investigating the relationship between “*hyperprolactinemia*” and “*menstrual*”

disorders". Therefore annotators will consider "schizophrenia" as an additional state and is not related to two diseases.

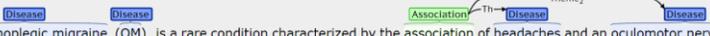
Similar, the following case:


 [The relationship between antipsychotic-induced hyperprolactinemia and menstrual disorders in women with schizophrenia; a systematic review].

<PMID:23074030>

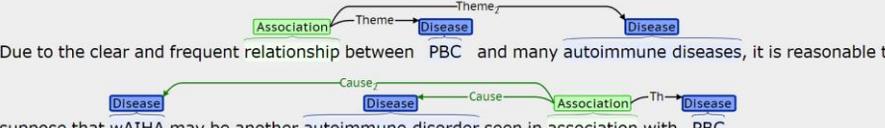
The "short sleep duration" is considered to be an additional state of "insomnia", and "short sleep duration" did not directly participate in the association of "insomnia" and "hypercholesterolemia".

- **Disease_A is characterized by ... Disease_B,**
Disease_A is classified by ... Disease_B, or
Disease_A is ... Disease_B: If a sentence describes the Disease_A is ... Disease_B, it usually indicates Disease_A/Disease_B is a symptom or subcategory of another one Disease_B/Disease_A. Therefore the association is not taken into consideration. Eg.


 BACKGROUND: Ophthalmoplegic migraine (OM) is a rare condition characterized by the association of headaches and an oculomotor nerve palsy.

<PMID:23150890>

In above example. "headaches" and "oculomotor nerve palsy" are symptoms of "Ophthalmoplegic migraine". Therefore, there is no association between "Ophthalmoplegic migraine" and two diseases. For more examples:


 Due to the clear and frequent relationship between PBC and many autoimmune diseases, it is reasonable to suppose that wAIHA may be another autoimmune disorder seen in association with PBC.

<PMID:26630456>

The phrase "... wAIHA may be ... autoimmune disorder ...". "wAIHA" belongs to a member of "autoimmune disorder" so the association between them is neglected.

Reference

- [1] R. I. Doğan, R. Leaman, and Z. Lu, "NCBI disease corpus: A resource for disease name recognition and concept normalization," *Journal of Biomedical Informatics*, vol. 47, pp. 1-10, 2014/02/01/ 2014.

- [2] R. Leaman, R. I. Dogan, and Z. Lu, "DNorm: disease name normalization with pairwise learning to rank," in *Bioinformatics*, ed, 2013.