

# **SOCIAL DETERMINANTS OF HEALTH FOR CANCER PATIENTS ANNOTATION GUIDELINES**

Written by Marco Guevara, MS and Danielle Bitterman, MD with input from Idalid Franco, MD MPH,  
Madeleine Goldstein, and Susan Harper

## **Table of Contents**

<b>1. Introduction and Purpose Statement</b>	<b>2</b>
<b>2. Annotation Task</b>	<b>2</b>
<b>2.1 SDOH Categories: Definitions and Attributes</b>	<b>3</b>
2.1.1 Housing	3
2.1.2 Transportation	4
2.1.3 Relationship Status	4
2.1.4 Parental Status	5
2.1.5 Employment Status	6
2.1.6 Social Support	7
<b>3. Annotation Process</b>	<b>8</b>
<b>3.1 Annotation Tool</b>	<b>8</b>
Example Document	9
<b>3.2. SDOH Selection and Challenging Situations</b>	<b>9</b>
<b>APPENDIX A - Fully Annotated Examples</b>	<b>11</b>
<b>APPENDIX B – SDOH Tag List</b>	<b>12</b>

## 1. Introduction and Purpose Statement

Social Determinants of Health (“SDOH”) are a major driver of cancer disparities, but our ability to study and act on them is limited by the fact that many are almost only documented in the free text of clinic notes. This project aims to address cancer disparities by developing natural language processing technologies that can identify and extract SDOH data from clinical texts in order to improve the care of radiotherapy (RT) patients at risk for adverse SDOH.

These guidelines describe specific SDOH we wish to capture in the notes of patients undergoing RT (“RadOnc notes”), and the methodology for annotation. In addition to identifying the presence of SDOH, each SDOH will have certain attributes to further enrich our datasets.

The annotations developed in this project will subsequently be used to train natural language processing systems that can “read” clinical documents, identify the SDOH of interest, and extract them from the texts. NO EXAMPLES IN THIS DOCUMENT ARE PHI - THEY ARE NOT REAL CLINICAL TEXT.

**Annotators should carefully read these guidelines in their entirety prior to starting annotations.**

Cross-references to different sections in this document are linked for ease of use.

A brief orientation to this document:

- [Annotation Task](#): Defines the overall annotation task for this project, and then defines each SDOH Category/attribute to be annotated. Example annotations are provided.
- [Annotation Process](#)
  - [Annotation Tool](#): Instructions on how to use the MAE that will be used in this project.
    - [Example Document](#): Annotated example document in MAE using fake patient data
  - [3.2. SDOH Selection](#): Detailed instructions on how to carry out the annotation process and how to approach difficult cases. Common ambiguous situations and how to approach them are described.
- [APPENDIX A - Fully Annotated Examples](#): Fully annotated examples using fake patient data
- [Appendix B – SDOH Tag List](#): Complete list of all SDOH tags

## 2. Annotation Task

Each RadOnc note may have sentences that represent one or more of the SDOH that we have defined below. Annotators are tasked with identifying the sentences that describe these SDOH concepts. If a sentence describes an SDOH concept, annotators then assign an attribute of that SDOH as further detailed in the specific subsections below. Assigning the label is interchangeably referred to as “annotating” or “tagging” the sentence in these guidelines.

During the annotation process, each RadOnc note is pre-split into sentences, where each annotated data point will be a sentence within a document. Most annotations will not require annotators to take in any context other than what is in the boundaries of each sentence. For instances where context outside of the sentence boundaries is needed to make an annotation judgment, we ask annotators to use their best judgment, and to defer to surface level annotation without the use of textual context.

Moreover, annotators should take in the context of the note-writing event, as well as common sense logic. Within reason, an annotator should consider the intent of the note-writer for a particular sentence. The thought process one should consider is, “why would a physician / resident / social

worker write this, if it was not important for the reader to consider”? More examples of implicit SDOH are detailed below, and we ask annotators to take into account common sense (not solely explicit textual evidence) when tagging sentences.

Finally, a sentence can describe more than one different category of SDOH, and in some cases a sentence can describe more than one attribute within a single category (example: Notes and Edge Cases in [2.1.3](#)). We ask annotators to please label every instance of an SDOH per relevant sentence.

Annotators should annotate strictly according to these guidelines. We acknowledge that there are many different ways to define protective, adverse, and neutral SDOH that our guidelines may not always capture. However, the goal of this project is to develop **consistent** annotations for the specific definitions of SDOH that are detailed below. Sentences may describe a SDOH or SDOH attribute that may seem at first glance to align with one of our SDOH categories and attributes, but do not meet our specific definitions. **In these cases, please adhere only to the definitions below.** Consistency is key because consistent labels are paramount to being able to train models that can identify and extract SDOH.

## 2.1 SDOH Categories: Definitions and Attributes

### 2.1.1 Housing

The status of a patient’s housing is a critical SDOH, known to affect the outcome of treatment. For the purposes of this annotation task, a sentence will be annotated as housing if it expresses **a challenge relating to the place of residence of the patient**. Please note that references to cities and towns, without mention of specific housing should NOT be considered an SDOH annotation. E.g., *“Pt lives in Arlington.”*

**Attributes:** {Poor, Undomiciled, Other}

**Poor:** The patient is housed with governmental assistance, or there is evidence that they are struggling to afford their housing due to resource limitations.

*“Pt came from Assisted Living Corp. and complained about rent increase.”*

*“He says he is worried about making his mortgage payments.”*

**Undomiciled:** There is evidence that the patient does not have stable housing and/or is living in a shelter.

*“Pt is staying with a friend and does not have a mailing address”*

*“Pt currently staying at Barbara McInnis shelter.”*

**Other:** The sentence expresses a housing issue that is not necessarily financial/resource-limited in nature.

*“Pt is staying at the Motel for the time being, while on the waitlist for the Hope Lodge”*

### **Notes and Edge Cases**

Please do not assume housing status simply based on type of housing, unless the type of housing clearly indicates housing status. For example, a description that the patient lives in a trailer is not sufficient to know that the patient has poor housing. By contrast, as in the examples above,

descriptions that the patient lives in a shelter or in government housing is sufficient to know that the patient is undomiciled and has poor housing, respectively.

### 2.1.2 Transportation

This SDOH pertains to **a patient's inability to get to/from their healthcare visits..** A patient being present at the treatment location, even if explicitly textually represented, or discussions of transportation unrelated to adequacy of transportation access, should NOT be considered an instance of Transportation SDOH. However, if there is a case of explicit textual representation that a patient is absent for treatment *and* that absence is due to transportation issues, then this IS considered an instance of Transportation SDOH.

**Attributes:** {Distance, Resource, Other}

**Distance:** The patient has issues getting to and from the treatment site due to the distance from their residence.

*"Pt lives 30mi away from hospital and and complains about needing to transfer three times each way."*

**Resource:** The patient has issues (financial or material) to get to and from the treatment site.

*"Pt missed appointment because her sister couldn't drive her today."*

**Other:** Sentence represents a transportation adequacy issue that is not described by the other two attributes, including mentions of time.

*"Pt is worried about making appointments because the metro is under construction this month."*

*"Pt is worried about the two hour drive."*

*"She is having trouble lying flat for treatment, she thinks it is because her back hurts after the two hour car ride into clinic."*

### Notes and Edge Cases

Discussions of transportation that are not related to a patient's ability to get to treatment should NOT be annotated.

**EG:** *"Patient has to urinate 3-4 times during the car ride home."* [No annotation]

Annotators should consider sentences that have only implicit mentions of transportation, within reason. That is to say, an explicit textual mention of a form of transportation is not always necessary for a Transportation SDOH tag.

**EG:** *"Pt felt that coming to Los Angeles was hard for them and asked to be referred to Santa Cruz."* [Transportation\_Distance]

### 2.1.3 Relationship Status

Whether or not a patient is in a partnered relationship is an abundant SDOH in the clinical notes. A sentence represents relationship status if it **expresses evidence that a patient is married, in a partnership, divorced/separated, single, or widowed.**

**Attributes:** {Married, Partnered, Divorced, Widowed, Single}

**Married:** Patient is married.

*"Pt and her husband came into my office today."*

**Partnered:** In a non-marital relationship with at least one other person.

*"Pt and her fiancée came into my office today."*

*"He is here with his boyfriend."*

**Divorced:** Patient is divorced or separated.

*"Pt is meeting ex-wife at appointment."*

*"Pt is married but separated."*

**Widowed:** Patient is widowed.

*"Pt spouse passed away in October of last year."*

**Single:** Patient is not in a relationship and does not meet the above criteria.

*"Pt is single."*

### **Notes and Edge Cases**

For the purposes of this annotation task, we consider any sentence that describes that a patient is divorced or widowed at any point to be annotated as such. For example- if there is evidence, even within the same sentence, that a patient has been divorced, and they are also currently partnered, both annotations should be present.

EG: *"Pt arrived with his girlfriend, and his ex-wife will attend with him at next week's session."*

**[RELATIONSHIP\_PARTNERED] [RELATIONSHIP\_DIVORCED]**

In cases of ambiguity between divorced and widowed, with NO context to resolve the ambiguity, we will defer to the divorced attribute.

EG: *"Pt has 3 kids from former marriage"* **[RELATIONSHIP\_DIVORCED]**

Only label sentences if they describe relationship status. If a sentence refers to an individual who you know is a partner based on information elsewhere in the note, but the sentence is not describing relationship status, do not label it.

EG: Sentence 1: *"He is married to Sheila."* **[RELATIONSHIP\_MARRIED]**

Sentence 2: *"Sheila asked many questions."* [no annotation]

#### **2.1.4 Parental Status**

This SDOH should be used for descriptions of a patient being a parent to at least one child who is a minor (under the age of 18 years old). The only evidence necessary for this SDOH is the existence of a patient's child **under the age of 18**. For the purposes of this task, "teenage children" can be considered minors. This SDOH category is binary and has no attributes, so the full annotation will just be the SDOH.

*"Pt has 2 children ages 9 and 13."* **[PARENT]**

*"Pt has 2 teenage children."* **[PARENT]**

### 2.1.5 Employment Status

This SDOH pertains to expressions of a patient's employment status. A sentence should be annotated as an Employment Status SDOH if **it expresses if the patient is employed (a paid job), unemployed, retired, or a current student.**

**Attributes:** {Employed, Unemployed, Under-Employed, Disability, Retired, Student}

**Employed:** The patient is currently employed, and there is no evidence that the position is financially insecure. Financial security does not need to be explicitly stated.

*"Pt works as an electrician in Rockland."*

**Under-Employed:** The patient is employed in a financially insecure position.

*"Pt works part-time at Jim's Fish and is struggling to pay rent."*

**Unemployed:** The patient is not currently employed in a paid position.

*"Pt has been living off of unemployment for the past 2 months."*

**Disability:** The patient is currently on disability.

*"Used to be a car mechanic, but he has been on disability for the past 2 years since his diagnosis."*

**Retired:** The patient is retired and living off of income other than a paid position.

*"Pt is a 75yr old retiree."*

**Student:** The patient is a student as their primary occupation.

*"Pt is attending Cool University full time."*

### Notes and Edge Cases

For the purposes of this annotation task, we will NOT consider a sentence that describes a patient as a full-time homemaker or volunteer as a candidate for any Employment Status SDOH tag.

**EG:** *"Pt left his job a number of years ago to stay at home to take care of the house and family."*

**Employment Status:** [No annotation]

Sentences tagged as Disability should not by default be tagged as Retired, Employed, Under-employed, or Unemployed **unless** the patient independently fits into both categories. We do not expect this to occur often, but an example where two employment status labels would be used is:

**EG:** *"He is currently on disability and is also occasionally working as an Uber driver to help cover the bills."* **[Under-Employed, Disability]**

A patient can sometimes be described as "semi-retired" or something similar that indicates some level of employment. In these cases we ask annotators to consider whether or not the sentence describes any level of financial insecurity. If the sentence does NOT describe any financial insecurity, it should be tagged as Employed as opposed to Under-Employed.

**EG:** *"Pt is a semi-retired marketing consultant."* **[Employed]**

### 2.1.6 Social Support

A sentence may describe that a **patient is actively receiving care support (emotional, health, financial, etc.) from people that are NOT a part of the patient's healthcare team (doctor, nurse, therapist, social worker, case manager, etc.)**. The sentence **MUST** describe an act of care, participation in the patient's care, or an explicit statement that the person in the patient's life is "supportive", "caring for them", etc. In these cases, we wish to capture a patient's Social Support with this annotation. On the contrary, a sentence may describe a noticeable lack of Social Support. Importantly, the absence of evidence for social support should **NOT** be considered evidence for the lack of social support. Only if there is explicit textual mention of a lack of support, will that be annotated per this task. Please note that statements that the patient simply lives with or near someone, without any mention that they are receiving support, should not be annotated as +Support.

**Attributes:** {+Support, -Support}

**+ Support:** The patient is receiving social support from people they know personally.

*"Here today is Pt, her daughter, and supportive wife."*

*"Pt is living with his parents during treatment, while his neighbors watch his cat."*

*"Pt had to borrow money from her friend to catch the bus today."*

**- Support:** The patient is not receiving social support from people they know personally.

*"Pt lives alone."*

*"Pt is struggling to find someone to watch his cat on the days he has to come for treatment."*

### Notes and Edge Cases

Sentences that only mention someone who is in the life of the patient with no other information about support provided, should **NOT** be considered for a Social Support tag.

**EG:** *"Patient's daughter works as a firefighter."* [No Annotation]

Sentences that **JUST** describe that a patient is accompanied / unaccompanied to the appointment should not be considered for a Social Support tag at all. This may seem in contrast with the instruction of taking in the note writer's intent into account (e.g., "why would the physician write this if it wasn't important for the reader?"), but we have found physicians will often write this with no SDOH-type intent.

**EG:** *"Here today is Pt, her daughter, and wife."* [No Annotation]

**EG:** *"Pt is unaccompanied today."* [No Annotation]

More subtle context may show that the mention of a person in the patient's life is indicative of support for the patient. Take the first example below- the fact that the sentence states that the patient is now living with a relative **BECAUSE** of the treatment points to the relative providing support. By contrast, in the 2<sup>nd</sup> example below, sentences that simply describe living with or near family and friends should not be labeled.

**EG:** *"Pt is currently living with nephew while receiving treatment."* [ **+ Social Support**]

**EG:** *"Lives in an apartment a few doors down from her daughter."* [No annotation]

**EG:** *"Lives with her 90 year old father."* [No annotation]

There are descriptions of actions that may be suggestive of support, but this support is not explicit. For example, picking up a phone call could suggest that someone is providing support, but this is not explicit; instead, **one would need to infer that the phone call was about PROVIDING support to the patient**. These instances should NOT be annotated as + Support.

**EG:** *“Made contact / Spoke on the phone with pt’s daughter.”* [No Annotation]

**EG:** *“Unless otherwise noted, patient/family is in agreement with this plan”* [No Annotation]

Finally, a description of someone being a Health Care Proxy (HCP) should NOT be considered a Social Support annotation.

**EG:** *“Pt’s mother is his HCP and can be reached at 555-555-5555.”* [No Annotation]

### 3. Annotation Process

#### 3.1 Annotation Tool

*Multi-document Annotation Environment (MAE)*

<http://keighrim.github.io/mae-annotation/>

Please follow the documentation in the above link to download the latest version of MAE.

1. Download Java: <https://www.java.com/en/download/>

2. Run MAE:

- Either click the downloaded jar file icon or run “java -jar mae-VERSIONNUMBER-fatjar.jar” from the command line (changing VERSIONNUMBER to the appropriate version e.g., mae-2.2.11-fatjar.jar).

3. Load the DTD file

- File → New Task Definition (or CTRL/⌘+N) → Upload SDOH.DTD
- This is the label set for annotations. You will not be able to load or annotate documents without first uploading SDOH.dtd.

4. Load a document

- File → Open Document (or CTRL/⌘+O) → open txt document
- From your data directory, load the text file into MAE.
- MAE will automatically create an xml file based on the txt file you just uploaded. It is fine to have this xml in the same directory as the txt files.
  - NOTE: After all annotations are complete and you save the document (CTRL/⌘+S), MAE will overwrite the xml it created in the beginning with your annotations.
- If you need to load a document that you have already annotated, load the XML file instead of the TXT. If you have already annotated some of a document, when you open the TXT file again it may overwrite the XML that has your annotations with a blank one. Please ensure you are opening the TXT to BEGIN annotations and XML to CONTINUE annotations.

5. Annotate the document → **PLEASE READ THIS SECTION CAREFULLY**

- Documents will be formatted with each sentence on its own line, with the first characters being the index number surrounded by brackets.

[0] this is the first example sentence in the document.

[1] this is another example sentence.

[2] here is a third.

...

[n] this is the last sentence.

- Please **HIGHLIGHT ONLY THE NUMBER** that is surrounded in brackets, at the beginning of the sentence.
- Then right click the number you just highlighted, and select the appropriate SDOH category.
- Once the correct category is selected, you can choose the appropriate attribute on the bottom-right side of the screen under the “type” column.
  - Click the attribute and a drop-down list will appear showing the relevant attributes for that SDOH tag. Select the appropriate attribute.
  - The SDOH tag options are summarized in [Appendix B – SDOH Tag List](#)
  - **NOTE:** for convenience, each SDOH will have a default attribute (e.g., for Relationship, it will be “married”), which are **bolded** in [Appendix B – SDOH Tag List](#). Please ensure you are choosing the correct attribute.

6. Save the Document, close it, and repeat with the next one.

#### Example Document

The below note is a fake patient note. No PHI is included in this document. Please zoom in to see the annotation tool more clearly.

**[REDACTED]**

### 3.2. SDOH Selection and Challenging Situations

For most instances of SDOH annotations, the annotators will not need to bring in outside context to choose the appropriate SDOH tag. A reminder- annotations are completed at the sentence level, and a document (and even a single sentence) can have multiple of the same SDOH category with differing attributes.

Below we lay out some common instances that may be confusing to annotators and how to deal with them for this task.

#### 1: Partial phrases

If an SDOH is represented by a partial sentence, but is part of a larger complete sentence, the entire sentence should be annotated as an SDOH.

*“They could not make the appointment today because their ride canceled on them, and they will receive medication next Thursday. [TRANSPORTATION\_RESOURCE]”*

If multiple distinct SDOHs are captured in a single sentence, that sentence should be annotated with both SDOHs.

*"Patient is taking care of her husband who has chronic back pain.*

**[RELATIONSHIP\_MARRIED]"**

If a single sentence contains multiple instances of the same SDOH attribute, the sentence should only be annotated with the appropriate SDOH tag/attribute once.

*"Pt is divorced, and his ex-wife is driving him to the appt. [RELATIONSHIP\_DIVORCED]*

**[SUPPORT\_PLUS]"**

## 2: Two adjacent sentences

If two sentences are adjacent and they both describe the same SDOH, each sentence should be treated as its own instance of an SDOH.

*"Pt and her husband visited the office on Friday. [RELATIONSHIP\_MARRIED] Her husband will fill out the paperwork before they leave. [RELATIONSHIP\_MARRIED]"*

## 3: Ungrammatical / incomplete sentences

If a sentence is not well formed or is only a partial phrase, and that phrase contains an SDOH, then the phrase should be annotated as if it were a complete and grammatical sentence.

*"Pt not here no ride [TRANSPORTATION\_RESOURCE]"*

## 4. Typos (that potentially cause ambiguity)

There is no way to completely eliminate ambiguity and this in turn may cause disagreement between annotators. In truly ambiguous sentences like the example below, we ask annotators to use their best judgment to resolve the typo. Below "Attend" is ungrammatical in the sentences and could be resolved to "attends" or "attended". If the former then it should be given an Employment – Student tag, and if the latter it should not be annotated. In this case, we resolved the typo to "attends" as this is the closest word and easiest typo to make, and so tag the sentence as Employment – Student.

*"Attend Cool University to become a marketing specialist." [EMPLOYMENT\_STUDENT]*

## 5. The sentence does not refer to the patient.

There may be sentences that at first glance would be a candidate for an SDOH tag, but given the context of the note, we find that the sentence does not refer to a patient's SDOH. This will commonly occur when a proper name is used as opposed to a pronoun or the term "patient / pt". These should not be given any annotations, as in sentence [2] below.

[1] *Pt has undergone treatment and is accompanied by her mother Tilda...*

[2] *Tilda works as a contractor for the city, and pays most of her daughter's bills. [No Annotation]*

## APPENDIX A - Fully Annotated Examples

Note that these examples are fake notes and for the guidelines ONLY. Notes in the actual task will include PHI.

### **Physician Note**

[0] Patient: John Doe Date of Exam: 01/01/2023 Referring Provider: Dr. Jane Smith  
[1] Chief Complaint: Right-sided chest pain and difficulty breathing  
[2] History of Present Illness: Mr. Doe is a 65-year-old male with a past medical history significant for smoking and hypertension.  
[3] He presents today with a chief complaint of right-sided chest pain and difficulty breathing that began yesterday.  
[4] The pain is described as a sharp, stabbing sensation that radiates to his back and is rated 8/10 in severity.  
[5] He also reports shortness of breath and fatigue.  
[6] He has no history of previous cardiac issues.  
[7] Social History: Mr. Doe is married with two adult children. [RELATIONSHIP\_MARRIED]  
[8] He is currently employed as a truck driver and works long hours. [EMPLOYMENT\_EMPLOYED]  
[9] He reports limited physical activity due to the demands of his job. [EMPLOYMENT\_EMPLOYED]  
[10] He smokes 1 pack of cigarettes per day and has a 30-year history of smoking.  
[11] He denies any alcohol or drug use.  
[12] Physical Exam:  
[13] Vital signs: BP: 162/90 mmHg, HR: 90 bpm, T: 98.6 F, RR: 18 breaths/minute  
[14] General: Patient appears uncomfortable and in distress Cardiovascular: Regular rate and rhythm, no murmurs or gallops Lungs: Decreased breath sounds on the right side, mild wheezing  
[15] Diagnostic Studies:  
[16] Chest X-ray: Moderate right-sided pleural effusion CT scan of the chest: Mass in the right upper lobe, likely malignant  
Diagnosis: Right upper lobe mass, likely malignant  
[17] Plan: Referral to medical oncology for further evaluation and treatment.  
[18] Follow-up with radiation oncology in one week for further management.  
[19] Consideration for thoracentesis to alleviate symptoms.  
[20] Smoking cessation counseling to be provided.  
[21] Follow-up: Mr. Doe is to follow-up with radiation oncology in one week for further management.  
[22] Signed, Dr. John Smith Radiation Oncologist

### **Social Worker Note**

[REDACTED]

## APPENDIX B – SDOH Tag List

**Bolded** Attributes are the default value for each label in MAE

<i>Category</i>	<i>Attribute</i>
Transportation	Distance
	<b>Resource</b>
	Other
Housing	<b>Poor</b>
	Undomiciled
	Other
Relationship Status	<b>Married</b>
	Partnered
	Divorced
	Widowed
	Single
Parental Status	<b>N/A</b>
Employment Status	<b>Employed</b>
	Under-Employed
	Unemployed
	Disability
	Retired
	Student
Social Support	<b>+ Support</b>
	- Support