“Scare quotes”

# Coding Ontology

Following the identification of sample size, inappropriate examples will be removed after assessment of results, and then a random sample of the quote-containing chart segments (sentence-level) will be subsequently qualitatively coded as Probably Useful, Probably Harmful, or Potentially Misinterpreted labeling derived from Park et al.’s 2021 study.[1](https://www.zotero.org/google-docs/?FN1vOF)

## Quote\_use CODES:

### Probably Useful (CODE AS 0)

* **Clinical Info**, **Effect on Life**, **Values or Preferences**
* Provides important contextual clues for **clinical info**
  + Examples:
    - Chest pain that “feels like an elephant is on my chest”
    - Reported that “this is the worst headache I’ve had in my life”
* Conveys **effect of illness on patient’s life**
  + Examples:
    - Has a persistent low mood, endorsing “I don’t want to live like this.”
* Conveys Patient **Values or Preferences (PVP)**
  + Examples:
    - When discussing treatment goals, she said “if I cannot breathe without a tube, I don’t want to live. I do not want to suffer. I want to make sure that my family are with me at the end.”

### Probably Harmful (CODE AS 1)

* Cast **doubt on** integrity of patient to provide **reliable testimony**
  + Examples:
    - Reports she had a “reaction” to the medication.
    - Stated “migraine” was due to “stress”
* **Convey ridicule, contempt, or frustration** by highlighting **unsophisticated language** or **limited knowledge**
  + Examples:
    - Does not believe he has prostate cancer because “his bowels are working fine”

### Potentially Misinterpreted (CODE AS 2)

* Neutral phrases where quotes serve no clear purpose but could be read as scare quotes conveying doubt or judgment.
* On the fence use of quotes
  + Examples:
    - She reports she has been off of cigarettes for “a year”
    - She states her living situation is “less than ideal”

### Isn’t quoting the Patient (Code as 3)

* I.e. “Respiratory stated patient was ‘cooperative’ “
* These are times where quotes are used to describe words from other providers or family members, not the patient’s own words.
* Patient answers “Yes or No” questions
* Per PT: “patient is ambulatory”

## not\_applicable CODES:

### Unsure (code as 2)

* If not sure which of the 3 categories the quote falls under. Please provide comments as to why in the Third column.

**Is a repeat of another sentence (code as 3)**

* When the quote of a patient is the exact same or wording seems copied from another chart verbatim or near-verbatim. This may indicate multiple

**Nonsense: 4**

## THIRD COLUMN: Comments

### Free-text

* + If there are any issues where the text is difficult to understand or uninterpretable (like just a bunch of random characters) or if you cannot categorize based on the ontology. This is very helpful for me to know, but try to code this section only if it can absolutely not be coded one way or the other.

Notes for me:

* Make sure that annotation columns fit new ontology
* Make sure columns are formatted correctly
* Format as table
* Chop at number of occurrences

# First round Gold Standard coding notes

* There were several charts (~12) out of the x number found that were just pasted into the csv file where it seemed like each line of the note was its own row in the csv file. May be an error in the preprocessing or a weird character causing this
* I had to go manually and delete the ~12 notes i found that were chopping up the dataframe in this way.
  + Dataset: gold\_standard\_bias\_annotation\_doc.xlsx
  + I should create new excel column Rand() and sort by the random number