

# Supplementary Material for *A Corpus with Multi-Level Annotations of Patients, Interventions and Outcomes to Support Language Processing for Medical Literature*

Anonymous ACL Submission

## A Corpus Details

### A.1 PIO: Participants

In the Participants task, we sought primarily to capture the specific characterizations of the population that would affect treatment decisions or inclusion/exclusion criteria in systematic reviews. The starting spans for the participants were often a full sentence in length, describing the relevant demographic information of the subjects. The targeted information and hierarchy of labels are depicted in Figure A.1.

The available ranges for the Age label were derived from combining the age-specific MeSH terms into higher level categories. For example, the [YOUNG] category covers the MeSH terms for [INFANT, NEWBORN], [INFANT], [CHILD, PRESCHOOL] and [CHILD].

### A.2 PIO: Interventions

For the intervention task, the starting spans were almost always short and contained only the specific intervention used in the trial and therefore the task focused less on anchoring specific types of information in the text and more on assigning the intervention type and semantic labels. The intervention types chosen for this task are the subset of the the labels provided by a domain expert (an MD on our team) that remain after removing overly technical and rare categories.

To compliment the MeSH terms from the article, we provide two additional semantic labels: COMPLEX to indicate when the intervention was applied in conjunction with at least one other intervention, and DOSAGE CHANGE for cases where the same intervention is repeated but at different dosages.

### A.3 PIO: Outcomes

The labels for the outcomes task shown in Figure A.3 are derived from the vocabulary used by the Cochrane Collection, adjusted to remove rare categories and better capture the types of spans observed in the initial dataset.

A common occurrence in the discussion of outcomes is to provide a general description of the outcome (e.g. "average pain"), and then later go in to more detail about the way the measurement was taken (e.g. "pain score on a visual analog scale from 0-10"). To capture this relationship we provided the annotators with the ability to select if any particular outcome is GENERAL or SPECIFIC, and instructed them to mark any outcomes differing only in their level of specificity as co-referent.

A recurring issue with aggregation across noisy labels is the necessity of handling low quality workers. Due to the high variance of worker skill levels on mechanical turk, even filtering workers leaves some disparity in the quality of annotations. Using models that encode worker reliability is therefore highly desirable, and evaluation of results is best done with this aspect in mind as well.

One saving grace is that prolific workers tend to fall in to two categories: spammers submitting low quality work which is easily detected and disregarded, and those with an aptitude or interest for the work.

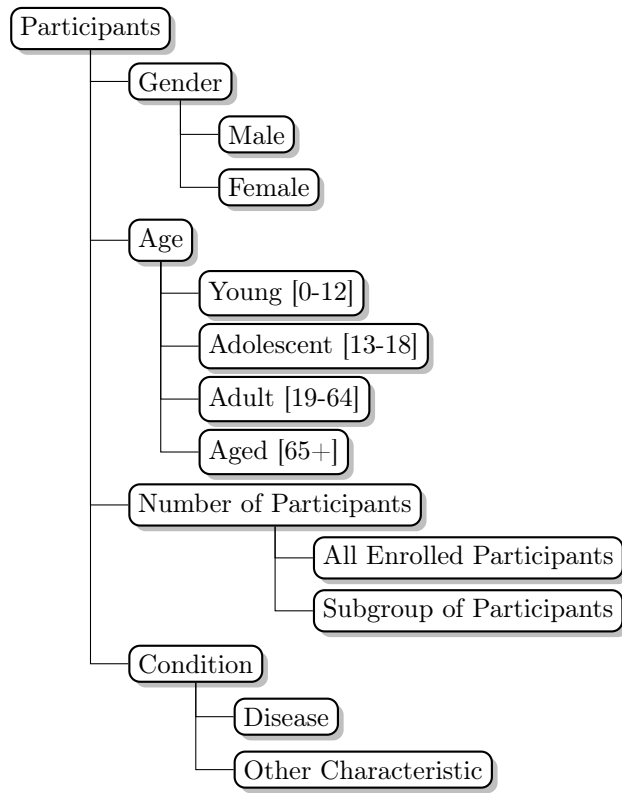


Figure 1: Participant task label hierarchy

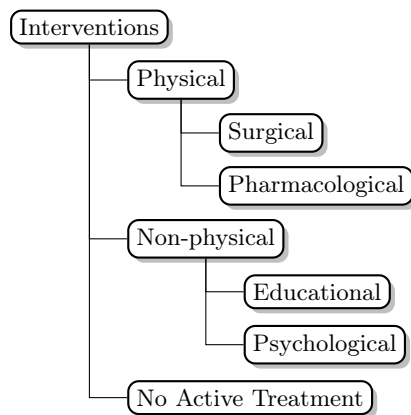


Figure 2: Intervention task label hierarchy



## A.4 MeSH terms in at least 1% of documents

Humans, Female, Male, Adult, Middle Aged, Aged, Double-Blind Method, Treatment Outcome, Adolescent, Prospective Studies, Child, Follow-Up Studies, Time Factors, Child Preschool, Aged 80 and over, Young Adult, Autistic Disorder, Cross-Over Studies, Dose-Response Relationship Drug, Drug Therapy Combination, Clinical Trials as Topic, Random Allocation, Drug Administration Schedule, Blood Pressure, Antineoplastic Combined Chemotherapy Protocols, Risk Factors, Analysis of Variance, Questionnaires, Combined Modality Therapy, Infant, Placebos, Administration Oral, Hypertension, Postoperative Complications, Pain Measurement, Single-Blind Method, Quality of Life, Pilot Projects, Pregnancy, Severity of Illness Index, Animals, Child Development Disorders Pervasive, Prognosis, Breast Neoplasms, Psychiatric Status Rating Scales, Heart Rate, Infusions Intravenous, Survival Rate, Antineoplastic Agents, Pain Postoperative, Infant Newborn, Survival Analysis, Acute Disease, Patient Satisfaction, Recurrence, Neoplasm Staging, Chronic Disease, Blood Glucose, Disease-Free Survival, Drug Combinations, Myocardial Infarction, Hemodynamics, Fluorouracil, Age Factors, Pain, Chi-Square Distribution, Injections Intravenous, Anti-Bacterial Agents, Patient Compliance, Exercise, Cyclophosphamide, Patient Education as Topic, Statistics Non-parametric, Antipsychotic Agents, Heart Failure, Recombinant Proteins, Antihypertensive Agents, Social Behavior, Biological Markers, Neoplasms, United States, Attention, Incidence, Neoplasm Recurrence Local, Neuropsychological Tests, Reproducibility of Results, Reference Values, Research Design, Doxorubicin, Dietary Supplements, Disease Progression, Insulin, Risk Assessment, Exercise Therapy, Body Mass Index, Sensitivity and Specificity, Lung Neoplasms, Randomized Controlled Trials as Topic, Retrospective Studies, Diabetes Mellitus Type 2, Anesthetics Local, Behavior Therapy, Electrocardiography, Body Weight, Asthma, Psychomotor Performance, Anti-Inflammatory Agents Non-Steroidal, Cognitive Therapy, Coronary Disease, Length of Stay, Exercise Test, Methotrexate, Risperidone, Obesity, Administration Topical, Parents, Reaction Time, Aspirin, Muscle Skeletal, Proportional Hazards Models, Cisplatin, Longitudinal Studies, Lipids, Prostatic Neoplasms, Chemotherapy Adjuvant, Equipment Design, Regression Analysis, Diet, Cost-Benefit Analysis, Adenocarcinoma, Communication, Neoplasm Metastasis, Area Under Curve, Cognition, Activities of Daily Living

## B Annotation Instructions

### B.1 Span Instructions

Below (beginning on the next page) we attach span the instructions provided to annotators for marking spans corresponding to populations, interventions/comparators and outcomes (the respective PICO elements) in order.

Following these, in Section [B.2](#) we attach the instructions provided for the subsequent pass, which entailed more granular annotations on the spans.

## Help doctors search the medical literature

Often, doctors personalize treatment depending on the characteristics of the patient. Searching the medical literature for personalized information is difficult. We are building interfaces to make their search more efficient. Here we mark descriptions of participants in medical studies, as described in the abstracts of medical articles.

### Annotation Instructions

The task is to find in a short text all phrases and sentences that describe the participants in a medical study. Most text will be abstracts of medical articles. The descriptions of participants can include details about how and where the participants were recruited in the medical study and what characteristics or requirements they needed to meet to be included in the study. Specific common population descriptors include: gender, how many people were in the study, medical diagnoses or conditions they may have, location where they were treated as well as many other details relevant to a particular study. Descriptions of what was done in the study or the findings of the study should not be marked, even when they describe subgroups of participants.

Read the text and highlight all phrases that describe the people who participated in the study. Mark the longest contiguous text that includes such a description.

It may be helpful to read the entire text once and only after that mark the text snippets describing the people in the study.

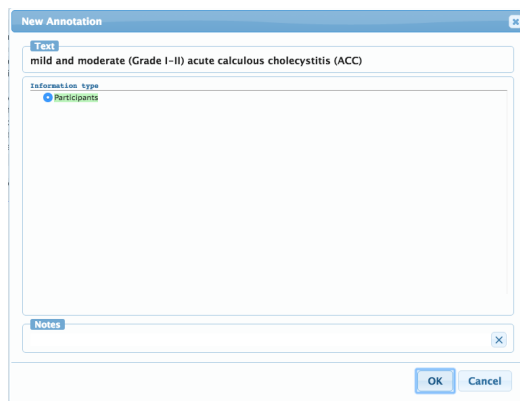


Figure 1: Markup dialog window. A "delete" button will appear for a span already marked.

To mark a portion of the text, highlight the text. A menu will appear. At the top, in the "Text" field you will see the selected text. The "Information type" section has the "Participants" tag selected by default. At the bottom, there is a "Notes" section where you can leave comments about this particular mark-up. Select "OK" at the bottom right to finalize the mark-up.

If you wish to delete a mark-up section, double click on the existing label, then select "Delete" at the bottom right.

The easiest cases to mark is in sentences that directly describe what people were selected to participate in a medical study. Below are some examples. The highlighted text snippets show the descriptions of the participants. Sometimes the descriptions can be just a phrase other times it can cover the entire sentence:

*Participants with mild-to-moderate AD (Mini-Mental State Examination score of 13-26) were recruited from December 1999 to November 2000 using clinic populations, referrals from community physicians, and local advertising.*

*A total of 270 patients with MCI were enrolled in a 24-week, multicenter, randomized, double-blind, placebo-controlled study.*

*Of 474 participants screened, 351 were enrolled.*

Some texts contain headings such as "SETTING", "PARTICIPANTS" and "RESULTS". Still read the text: some sentences in the "SETTING" or "PARTICIPANTS" text may not directly describe the participants and sometimes "RESULTS" texts contain participants descriptions.

*RESULTS: Seventy institutionalized subjects with AD (mean age 84) completed the study.*

### Special cases of what to mark:

Text in the titles **should** be marked.

Expressions that provide some information about the participants but not directly describing them **should** be marked. These include demographics information, for example, if an article reports that the study was conducted in "Finnish hospitals", this should be marked; in the following example, the sentence was used to elicit that the participants are from Taiwan:

*However, the laboratory data in these studies were not obtained among Taiwan population.*

Some sentences mention the type of participants in a study indirectly, for example while describing the goal of their study in sentences of the form "We wanted to see if treatment X will help a given group of patients." These descriptions should also be annotated.

*OBJECTIVE: To determine whether treatment with a selective cyclooxygenase (COX) -2 inhibitor (rofecoxib) or a traditional nonselective NSAID (naproxen) slows cognitive decline in patients with mild-to-moderate AD.*

Similarly sentences describing the result of the study can mention the type of people who participated in the study. Mark these descriptions. For example:

*The results of this study indicate that rofecoxib or low-dose naproxen does not slow cognitive decline in patients with mild-to-moderate AD.*

Some sentences describe behaviors or characteristics of the participants in the study, without mentioning the participants explicitly. If you understand from the context that these describe the participants, mark them. For example in the sentence below one can infer that it was allowed that the participants in the study use the listed medication, even though the word participant is not directly mentioned.

*Stable use of cholinesterase inhibitors, estrogen, low-dose aspirin, and vitamin E was allowed.*

Other sentences may describe who was excluded from the study, which allows one to indirectly infer information about the participants, either of whom will have the property which led to others being excluded.

*Participants with inflammatory diseases that might respond to the study medications were excluded.*

### Special cases of what NOT to mark:

Some sentences describe not what characteristics of the participants had but rather what they were asked to do in the study or when describing the results of the study. Do not mark these.

*The 1-year mean (SD) change in ADAS-Cog scores in participants treated with naproxen (5.8 [8.0]) or rofecoxib (7.6 [7.7]) was not significantly different from the change in participants treated with placebo (5.7 [8.2]).*

A complete annotated article should look like Figure 2.

1	Abstract
2	BACKGROUND:
3	Although the preoperative management of mild and moderate (Grade I-II) acute calculous cholecystitis (ACC) has been standardized, there is no consensus on the value of abdominal drainage after early
4	METHODS:
5	In a post hoc analysis of a randomized controlled trial (NCT01015417) focused on the value of postoperative antibiotic therapy in patients with ACC, we determined the value of abdominal drainage in
6	patients having undergone laparoscopic cholecystectomy for Grades I-II ACC.
7	All postoperative complications were analyzed after using a propensity score.
8	RESULTS:
9	Of the 414 enrolled patients, 178 did not have abdominal drainage (forming the no-drainage group) and 236 had drainage (the drainage group).
10	After matching on PS, the deep incisional site infection was 1.1 versus 0.8 %, p = 0.78.
11	This result is similar for the superficial incisional site infections; the distant infections; the overall morbidity, and the readmission rate.
12	Only the hospital length of stay was significantly longer in the drainage group (3.3 vs. 5.1 days, p = 0.003).
13	Neither abdominal drainage nor the absence of postoperative antibiotic therapy was found to be a risk factor for deep incisional site infections.
14	CONCLUSIONS:
15	The use of abdominal drainage depends on the surgeon's personal preferences but is often used in high-risk populations.
16	However, abdominal drainage does not appear to be of any benefit (in terms of postoperative outcomes) and may even compromise recovery in
	patients having undergone early laparoscopic cholecystectomy for mild or moderate ACC.

Figure 2: A completed annotation.

## Navigation Instructions



Figure 3: Navigation bar.

Each HIT consists of three short texts to annotate. For qualification you will mark one text. Use the left and right arrows at the top bar to navigate through the articles. When you are finished with all texts, press the "right" arrow again and a "submit" button will appear, shown in Figure 3. There you will see a code unique for the HIT you just completed. **Do NOT refresh or zoom the annotation page (this will cause your annotations to be lost). Make sure to leave the Mechanical Turk window open as you complete the annotation.** When you are finished, you will return to the Mechanical Turk page to paste the code into the box.

In the annotation page you can always view these instructions again by clicking on the "help" button at the upper right corner of the navigation bar.

## Help doctors search the medical literature: treatments

### Annotation Instructions

We are interested in finding all phrases and sentences that describe the **medical treatments** studied in a clinical trial. Medical treatments are often called **interventions** in the medical literature. A medical treatment can be a specific drug ("aspirin"), surgery ("inguinal hernia repair"), talking therapy ("cognitive behavioural therapy") or even a lifestyle modification like changing one's diet or the type of toothpaste the patients use.

Many studies will have a group who receive a "control" treatment, such as a "placebo", no treatment at all, or may who may receive whatever is currently standard practice ("usual care"). These are also regarded as *interventions*, and a description of these, too, should be marked.

Clinical trials are designed to find out whether a medical treatment works or not. People who take part in clinical trials receive one or more of the studied interventions to find which works the best. **The task is to highlight text snippets that describe all medical treatments studied in the short text you will read.**

We **do NOT** want to annotate the details about how they are given to participants. For example medications may be given by mouth (orally), or by injection ("intramuscular", "subcutaneous", or "intravenous"), and will normally have a dose described (e.g. "325 mg"). The frequency and duration of interventions is often described (an exercise session might last "30 minutes", talking therapies might be scheduled as "6 monthly sessions each lasting 1 hour", and a medication might be given "twice daily for 12 months"). All of this is extraneous information that does NOT need to be marked.

Read the text and highlight all phrases that name the interventions. Mark **the longest contiguous texts** that includes such descriptions. A given intervention may be referenced or described more than once in an abstract; **all such occurrences should be marked**. It may be helpful to read the entire text once and only after that mark the text snippets describing the interventions in the study.

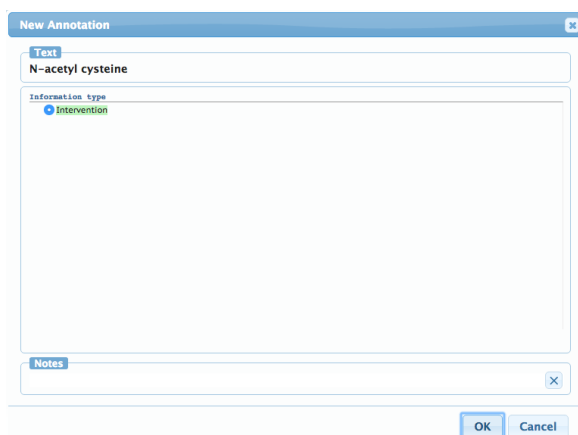


Figure 1: Markup dialog window. A "delete" button will appear for a span already marked.

To mark a portion of the text, highlight the text. A menu will appear. At the top, in "Text" you will see the selected text. The "Information type" section has the "Intervention" tag. At the bottom, there is a "Notes" section where you can leave comments about this particular mark-up. Select "OK" at the bottom right to finalize the mark-up.

If you wish to delete a mark-up section, double click on the existing label, then select "Delete" at the bottom right.

A highlighted span should look like Figure 2.

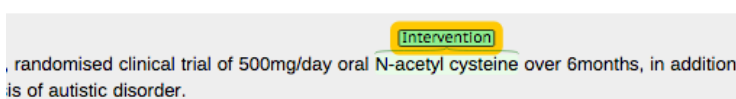


Figure 2: A highlighted span.

In some cases, interventions will be called out in article abstracts. In these the entire sentence may contain a description of the treatment and its administration. We want to mark **only the treatments involved, not the other details**. For example:

INTERVENTION: [Zinc lozenges](#), 10 mg, orally dissolved, 5 times a day (in grades 1-6) or 6 times a day (in grades 7-12).

INTERVENTION: The patients received either [azithromycin](#) (600 mg/d for 3 days during week 1, then 600 mg/wk during weeks 2-12; n = 3879) or [placebo](#) (n = 3868).

Some typical examples are shown below, where again descriptions of interventions are shown in blue:

We report findings of a pilot RCT for [a parent training intervention](#) with a focus on the development of joint attention skills and joint action routines. Twenty-four children meeting ICD-10 criteria for childhood autism (mean age = 23 months) were identified using the CHAT screen and randomised to [the parent training group](#) or to [local services](#) only.

We randomly assigned 6595 men, 45 to 64 years of age, with a mean (+/- SD) plasma cholesterol level of 272 +/- 23 mg per deciliter (7.0 +/- 0.6 mmol per liter) to receive [pravastatin](#) (40 mg each evening) or [placebo](#).

[Antihypertensive therapy](#) was started immediately after randomization in the active treatment group, but only after termination of the double-blind trial in the control patients. Treatment consisted of [nitrendipine](#) (10-40 mg/d), with the possible addition of [enalapril maleate](#) (5-20 mg/d), [hydrochlorothiazide](#) (12.5-25 mg/d), or [both add-on drugs](#).

METHOD: Seventy-two people residing in National Health Service (U.K.) care facilities who had clinically significant agitation in the context of severe dementia were randomly assigned to [aromatherapy with Melissa essential oil](#) (N = 36) or [placebo \(sunflower oil\)](#) (N = 36). The active treatment or [placebo oil](#) was combined with a base lotion and applied to patients' faces and arms twice a day by caregiving staff.

We next review special cases of what to mark.

Text in titles **should** be marked. For instance, consider the following title:

[Individualized Acupuncture](#) for Symptom Relief in Functional Dyspepsia: A Randomized Controlled Trial.

As shown, "Individualized Acupuncture" should be marked as the intervention here, even though this appears in title study title, rather than the abstract. Of course, descriptions of the acupuncture intervention should also be marked wherever they appear in the abstract text.

## Navigation Instructions



Figure 3: Navigation bar.

Each HIT consists of three short texts to annotate. For qualification you will mark one text. Use the left and right arrows at the top bar to navigate through the articles. When you are finished with all texts, press the "right" arrow again and a "submit" button will appear, shown in Figure 3. There you will see a code unique for the HIT you just completed. **Do NOT refresh or zoom the annotation page** (this will cause your annotations to be lost). **Make sure to leave the Mechanical Turk window open as you complete the annotation.** When you are finished, you will return to the Mechanical Turk page to paste the code into the box.

In the annotation page you can always view these instructions again by clicking on the "help" button at the upper right corner of the navigation bar.



## Help doctors search the medical literature: outcomes

### Annotation Instructions

We are interested in finding all phrases that describe **outcomes** in a clinical trial.

Outcomes describe what is measured in people taking part in a trial, to find out if their treatment has worked. In clinical trials, the researchers compare outcomes for two or more groups of patients, each of whom receive a different treatment.

Outcomes can be physical measurements of patients (*blood pressure, weight*), the score on a medical test or questionnaire (*Autism Diagnostic Observation Schedule assessment, Quality of Life Scales*), or positive or negative events in the patient groups (*quit smoking, improved social communication; deaths, stroke, number of caries, number of hospital readmissions*). Often there are multiple outcomes; you should mark the description of all of them.

Outcomes may be described in general terms (*quality of life*), or sometimes by naming the specific score or measurement tool which was used (*the Quality of Life Scale*). Some outcomes are only described generally. For example, a medical treatment may be described simply as *preventive*, without a description of how the researchers verified that claim. Similarly, a treatment may be described as *improving quality of life*, without specific details of how this was measured. Mark such general description, as well as any information that describe the specifics of what was measured.

Mark the description of what outcomes were measured. Do NOT annotate the text reporting numbers or results, or their interpretation. For instance, if the outcome is *quit smoking* this should be highlighted, but the number of people who quit smoking at the end of the trial in different groups should not be. Similarly abstracts often report if certain treatment improved, reduced or was beneficial for some outcome of interest. We do not want to mark these evaluations, only the description of what was measured and how it was measured.

Occasionally abstracts will report adverse reactions. Mark these as outcomes.

Text in titles **should** be marked if it contains description of an outcome, as in *Garlic lowers [blood pressure](#)*.

Some medical papers have structured abstracts that have an explicitly marked section called OUTCOMES. These usually describe the outcomes that were measured, followed by reports of the actual values of the measurements in the following section of the abstract. Again, you would annotate only the former.

Read the text and highlight all phrases that describe the outcomes. Mark **the longest contiguous texts** that include such descriptions. An outcome may be referenced or described more than once in an abstract; **all such occurrences should be marked**. It may be helpful to read the entire text once and only after that mark the text snippets describing the outcomes measured in the study.

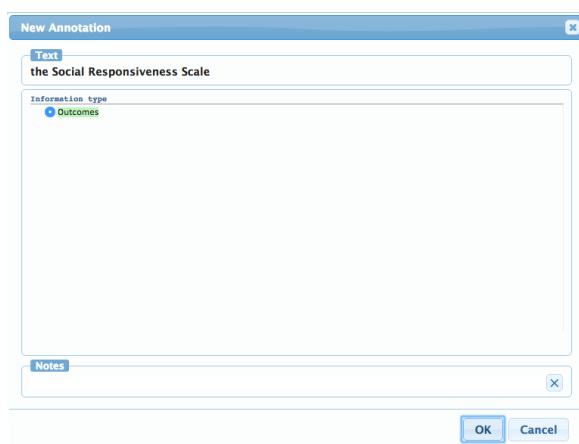


Figure 1: Markup dialog window. A "delete" button will appear for a span already marked.

To mark a portion of the text, highlight the text. A menu will appear. At the top, in "Text" you will see the selected text. The "Information type" section has the "Outcomes" tag. At the bottom, there is a "Notes" section where you can leave comments about this particular mark-up. Select "OK" at the bottom right to finalize the mark-up.

If you wish to delete a mark-up section, double click on the existing label, then select "Delete" at the bottom right.

A highlighted span should look like Figure 2.



Below are some example annotations:

***Effects of 12 weeks' treatment with a proton pump inhibitor on [insulin secretion](#), [glucose metabolism](#) and [markers of cardiovascular risk](#) in patients with type 2 diabetes: a randomised double-blind prospective placebo-controlled study.***

**AIMS/HYPOTHESIS:**

Recent studies suggest that proton pump inhibitor treatment may increase [insulin secretion](#) and improve [glucose metabolism](#) in type 2 diabetes. In a randomised double-blind prospective placebo-controlled 2 × 2 factorial study, we examined the effect of esomeprazole on [insulin secretion](#), [HbA\(1c\)](#) and [cardiovascular risk factors](#) in type 2 diabetes.

**METHODS:**

Forty-one patients with type 2 diabetes using dietary control or oral glucose-lowering treatment were randomised to receive add-on esomeprazole 40 mg (n = 20) or placebo (n = 21) for 12 weeks. Randomisation was carried out prior to inclusion on the basis of a computer-generated random-number list. The allocation sequence was concealed in sealed envelopes from the researcher enrolling and assessing participants. The study was undertaken at Steno Diabetes Center, Gentofte, Denmark. The primary outcome was change in [AUC for insulin levels during a meal test](#). Secondary outcomes were the [levels of HbA\(1c\)](#) and [biochemical markers of cardiovascular risk](#), including [lipids](#), [coagulation factors](#), [inflammation markers](#), [markers of endothelial function](#) and [24 h ambulatory BP measurements](#).

**RESULTS:**

Forty-one participants were analysed. In the esomeprazole-treated group the [AUC for insulin](#) did not change (before vs after treatment: 28,049 ± 17,659 vs 27,270 ± 32,004 pmol/l × min (p = 0.838)). In the placebo group [AUC for insulin](#) decreased from 27,392 ± 14,348 pmol/l × min to 22,938 ± 11,936 pmol/l × min (p = 0.002). Esomeprazole treatment (n = 20) caused a ninefold increase in the [AUC for gastrin](#). [HbA\(1c\)](#) increased from 7.0 ± 0.6% (53 ± 5 mmol/mol) to 7.3 ± 0.8% (56 ± 6 mmol/mol) in the esomeprazole-treated group and from 7.0 ± 0.6% (53 ± 5 mmol/mol) to 7.4 ± 0.8% (57 ± 6 mmol/mol) in the placebo group (n = 21) (p for difference in change >0.05). Except for [BP](#), there were no differences between the groups in the [markers of cardiovascular risk](#) (p > 0.05). Monitoring of [24 h ambulatory BP](#) showed a significant decrease in [daytime systolic BP](#), [daytime diastolic BP](#) and [24 h diastolic BP](#) in the placebo group (p < 0.05). No change in [BP](#) was seen in the patients treated with esomeprazole.

**CONCLUSIONS/INTERPRETATION:**

Treatment with esomeprazole over 12 weeks did not improve [insulin secretion](#), [glycaemic control](#) or [cardiovascular disease biomarkers](#) in patients with type 2 diabetes.

=====

***Safety, Feasibility, and Efficacy of Vagus Nerve Stimulation Paired With Upper-Limb Rehabilitation After Ischemic Stroke***

**BACKGROUND AND PURPOSE:**

Recent animal studies demonstrate that vagus nerve stimulation (VNS) paired with movement induces [movement-specific plasticity in motor cortex](#) and improves [forelimb function](#) after stroke. We conducted a randomized controlled clinical pilot study of VNS paired with rehabilitation on [upper-limb function](#) after ischemic stroke.

**METHODS:**

Twenty-one participants with ischemic stroke >6 months before and moderate to severe upper-limb impairment were randomized to VNS plus rehabilitation or rehabilitation alone. Rehabilitation consisted of three 2-hour sessions per week for 6 weeks, each involving >400 movement trials. In the VNS group, movements were paired with 0.5-second VNS. The primary objective was to assess [safety](#) and [feasibility](#). Secondary end points included change in [upper-limb measures](#) (including [the Fugl-Meyer Assessment-Upper Extremity](#)).

#### RESULTS:

Nine participants were randomized to VNS plus rehabilitation and 11 to rehabilitation alone. There were no **serious adverse device effects**. One patient had **transient vocal cord palsy** and **dysphagia** after implantation. Five had **minor adverse device effects** including **nausea and taste disturbance on the evening of therapy**. In the intention-to-treat analysis, the change in **Fugl-Meyer Assessment-Upper Extremity scores** was not significantly different (between-group difference, 5.7 points; 95% confidence interval, -0.4 to 11.8). In the per-protocol analysis, there was a significant difference in change in **Fugl-Meyer Assessment-Upper Extremity score** (between-group difference, 6.5 points; 95% confidence interval, 0.4 to 12.6).

#### CONCLUSIONS:

This study suggests that VNS paired with rehabilitation is **feasible** and has not raised **safety concerns**. Additional studies of VNS in adults with chronic stroke will now be performed.

=====

#### **Risperidone in children with autism and serious behavioral problems**

##### BACKGROUND:

Atypical antipsychotic agents, which block postsynaptic dopamine and serotonin receptors, have advantages over traditional antipsychotic medications in the treatment of adults with schizophrenia and may be beneficial in children with autistic disorder who have **serious behavioral disturbances**. However, data on the safety and efficacy of atypical antipsychotic agents in children are limited.

##### METHODS:

We conducted a multisite, randomized, double-blind trial of risperidone as compared with placebo for the treatment of autistic disorder accompanied by **severe tantrums, aggression, or self-injurious behavior** in children 5 to 17 years old. The primary outcome measures were the **score on the Irritability subscale of the Aberrant Behavior Checklist and the rating on the Clinical Global Impressions - Improvement (CGI-I) scale** at eight weeks.

##### RESULTS:

A total of 101 children (82 boys and 19 girls; mean [±SD] age, 8.8±2.7 years) were randomly assigned to receive risperidone (49 children) or placebo (52). Treatment with risperidone for eight weeks (dose range, 0.5 to 3.5 mg per day) resulted in a 56.9 percent reduction in the **Irritability score**, as compared with a 14.1 percent decrease in the placebo group ( $P<0.001$ ). The rate of a positive response, defined as at least a 25 percent decrease in the **Irritability score** and a rating of much improved or very much improved on the **CGI-I scale**, was 69 percent in the risperidone group (34 of 49 children had a positive response) and 12 percent in the placebo group (6 of 52,  $P<0.001$ ). Risperidone therapy was associated with an average **weight gain** of 2.7±2.9 kg, as compared with 0.8±2.2 kg with placebo ( $P<0.001$ ). **Increased appetite, fatigue, drowsiness, dizziness, and drooling** were more common in the risperidone group than in the placebo group ( $P<0.05$  for each comparison). In two thirds of the children with a positive response to risperidone at eight weeks (23 of 34), the benefit was maintained at six months.

##### CONCLUSIONS:

Risperidone was used for the treatment of **tantrums, aggression, or self-injurious behavior** in children with autistic disorder. The short period of this trial limits inferences about **adverse effects** such as **tardive dyskinesia**.

=====

#### **Video-feedback Intervention to promote Positive Parenting adapted to Autism (VIPP-AUTI): A randomized controlled trial**

In a randomized controlled trial, we evaluated the early intervention program Video-feedback Intervention to promote **Positive Parenting** adapted to Autism (VIPP-AUTI) with 78 primary caregivers and their child (16-61 months) with Autism Spectrum Disorder. VIPP-AUTI is a brief attachment-based intervention program, focusing on improving **parent-child interaction** and reducing the child's individual **Autism Spectrum Disorder-related symptomatology** in five home visits. VIPP-AUTI, as compared with usual care, demonstrated efficacy in reducing **parental intrusiveness**. Moreover, parents who received VIPP-AUTI showed increased **feelings of self-efficacy in child rearing**. No significant group differences were found on **other aspects of parent-child interaction** or on **child play behavior**. At 3-months follow-up, intervention effects were found on **child-initiated joint attention skills**, not mediated by intervention effects on parenting. Implementation of VIPP-AUTI in clinical practice is facilitated by the use of a detailed manual and a relatively brief training of interveners.

=====

### **Results of the Randomized Danish Lung Cancer Screening Trial with Focus on High-Risk Profiling**

#### **RATIONALE:**

As of April 2015, participants in the Danish Lung Cancer Screening Trial had been followed for at least 5 years since their last screening.

#### **OBJECTIVES:**

*Mortality, causes of death, and lung cancer findings* are reported to explore the effect of computed tomography (CT) screening.

#### **METHODS:**

A total of 4,104 participants aged 50-70 years at the time of inclusion and with a minimum 20 pack-years of smoking were randomized to have five annual low-dose CT scans (study group) or no screening (control group).

#### **MEASUREMENTS AND MAIN RESULTS:**

Follow-up information regarding *date and cause of death, lung cancer diagnosis, cancer stage, and histology* was obtained from national registries. No differences between the two groups in *lung cancer mortality* (hazard ratio, 1.03; 95% confidence interval, 0.66-1.6;  $P = 0.888$ ) or *all-cause mortality* (hazard ratio, 1.02; 95% confidence interval, 0.82-1.27;  $P = 0.867$ ) were observed. More *cancers* were found in the screening group than in the no-screening group (100 vs. 53, respectively;  $P < 0.001$ ), particularly *adenocarcinomas* (58 vs. 18, respectively;  $P < 0.001$ ). More *early-stage cancers* (stages I and II, 54 vs. 10, respectively;  $P < 0.001$ ) and *stage IIIa cancers* (15 vs. 3, respectively;  $P = 0.009$ ) were found in the screening group than in the control group. *Stage IV cancers* were nonsignificantly more frequent in the control group than in the screening group (32 vs. 23, respectively;  $P = 0.278$ ). For the *highest-stage cancers* (T4N3M1, 21 vs. 8, respectively;  $P = 0.025$ ), this difference was statistically significant, indicating an absolute stage shift. Older participants, those with chronic obstructive pulmonary disease, and those with more than 35 pack-years of smoking had a significantly increased risk of *death* due to lung cancer, with nonsignificantly fewer *deaths* in the screening group.

#### **CONCLUSIONS:**

No statistically significant effects of CT screening on *lung cancer mortality* were found, but the results of post hoc high-risk subgroup analyses showed nonsignificant trends that seem to be in good agreement with the results of the National Lung Screening Trial.

## **Navigation Instructions**



Figure 3: Navigation bar.

Each HIT consists of three short texts to annotate. For qualification you will mark one text. Use the left and right arrows at the top bar to navigate through the articles. When you are finished with all texts, press the "right" arrow again and a "submit" button will appear, shown in Figure 3. There you will see a code unique for the HIT you just completed. **Do NOT refresh or zoom the annotation page** (this will cause your annotations to be lost). **Make sure to leave the Mechanical Turk window open as you complete the annotation.** When you are finished, you will return to the Mechanical Turk page to paste the code into the box.

In the annotation page you can always view these instructions again by clicking on the "help" button at the upper right corner of the navigation bar.

## **B.2 Detailed Span Annotation Instructions**

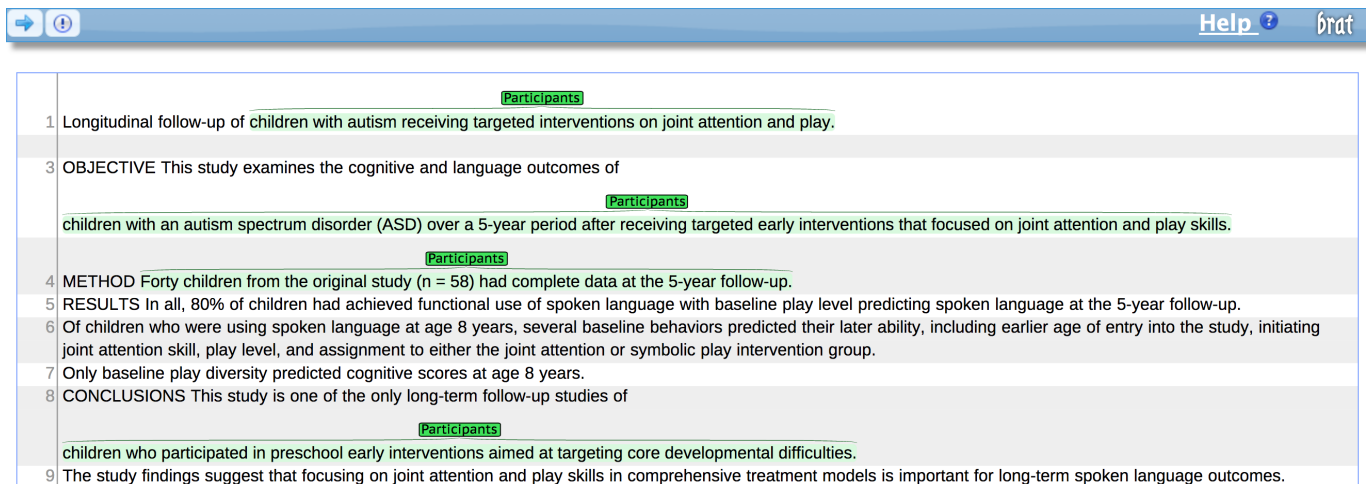
We attach span the instructions provided to annotators for marking data at a granular level (within spans) corresponding for populations, interventions and outcomes below, in that order.

## Instructions

In this task, we want to collect more detailed information about Participants in medical studies. Please read the instructions carefully.

Use your worker ID to log in to the tool when prompted.

We are interested in supplementing information about the participants in a clinical trial. This task involves adding and refining information about pre-selected snippets of text from medical journal articles. Each HIT contains three documents to update.



The screenshot shows the brat annotation tool interface. At the top, there is a blue header bar with navigation icons on the left and 'Help' and 'brat' on the right. Below the header is a document viewer with a line-numbered list of text segments. Several segments are highlighted in green, and each has a small green box labeled 'Participants' above it. The segments are:

- Line 1: Longitudinal follow-up of children with autism receiving targeted interventions on joint attention and play.
- Line 3: OBJECTIVE This study examines the cognitive and language outcomes of children with an autism spectrum disorder (ASD) over a 5-year period after receiving targeted early interventions that focused on joint attention and play skills.
- Line 4: METHOD Forty children from the original study (n = 58) had complete data at the 5-year follow-up.
- Line 5: RESULTS In all, 80% of children had achieved functional use of spoken language with baseline play level predicting spoken language at the 5-year follow-up.
- Line 6: Of children who were using spoken language at age 8 years, several baseline behaviors predicted their later ability, including earlier age of entry into the study, initiating joint attention skill, play level, and assignment to either the joint attention or symbolic play intervention group.
- Line 7: Only baseline play diversity predicted cognitive scores at age 8 years.
- Line 8: CONCLUSIONS This study is one of the only long-term follow-up studies of children who participated in preschool early interventions aimed at targeting core developmental difficulties.
- Line 9: The study findings suggest that focusing on joint attention and play skills in comprehensive treatment models is important for long-term spoken language outcomes.

Figure 1: Sample document with pre-selected snippets highlighted in green.

## Part 1: Marking snippets of the Participants labels

You be shown will a document that has already has text snippets that describe the participants of the study highlighted. Your task is to highlight pieces of these of these Participants snippets that convey more specific information. There are four types of information you should highlight:

- Gender - If it's clear that the participants are either males, females, or mixed genders
- Age - If any age or range of ages is given, either with specific numbers or with qualitative terms (such as "young")
- Condition - If an aspect of the participants' health (usually a disease) is named
- Number\_of\_participants - If the quantity of participants is given

When you find a snippet within one of the Participants text segments that matches one of these four categories, click and drag to select the shortest snippet that pertains to the tag. This will open the annotation interface where you can select the correct label for the new snippet.

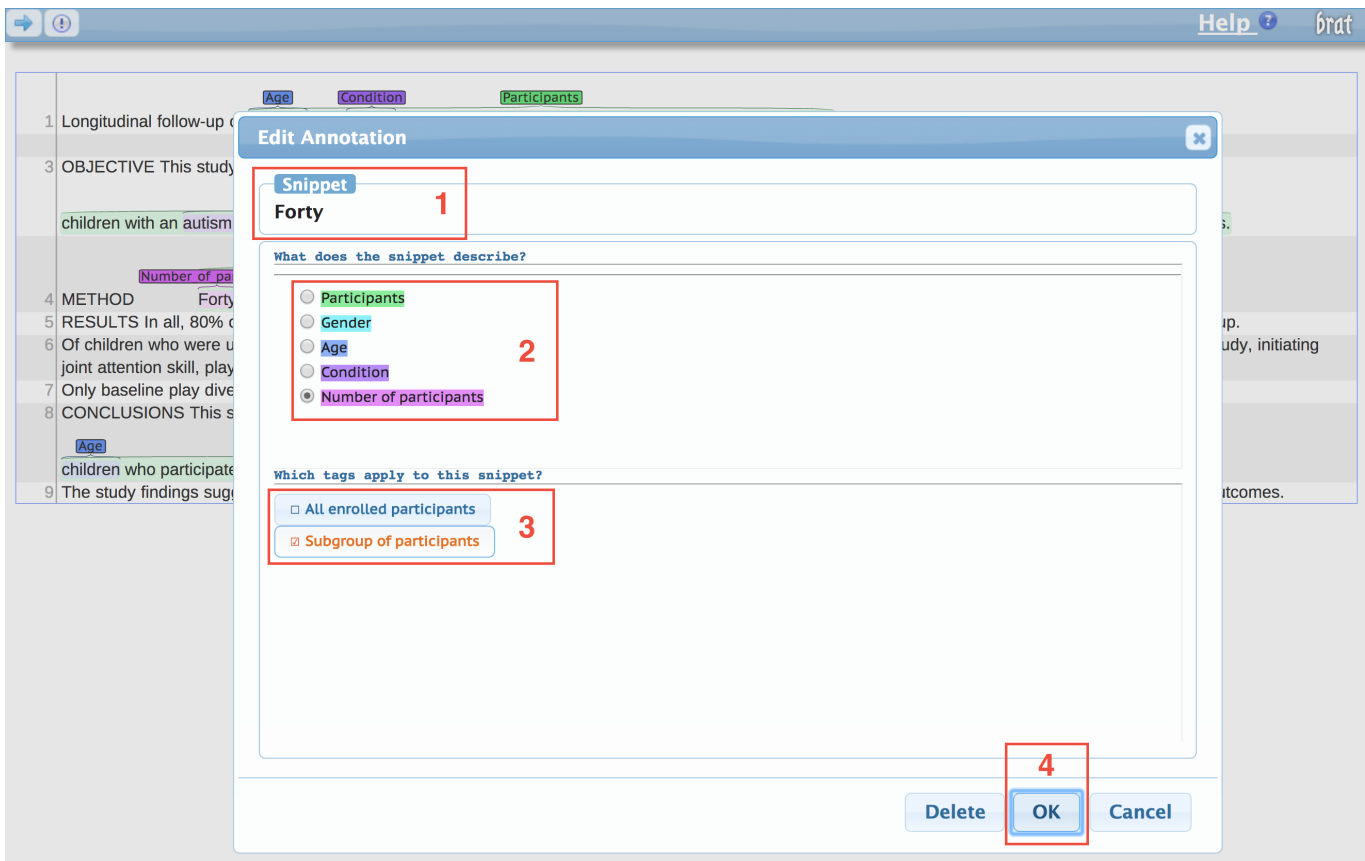


Figure 2: Annotation interface.

The editing interface has four main components:

1. This is to help keep track of which snippet you are currently editing.
2. Set the label by selecting one of the options from this list.
3. Click the tag(s) that apply to the current snippet.
4. Submit your changes by clicking OK.

If you need to delete or modify one of your annotations, you can double click its label to open an editing window.

## Example annotation

This document is fully annotated with new labels.

### Important notes:

- Highlighted snippets cannot overlap with each other
- Highlighted snippets must be contained within the Participants text
- Keep the snippets as short as possible



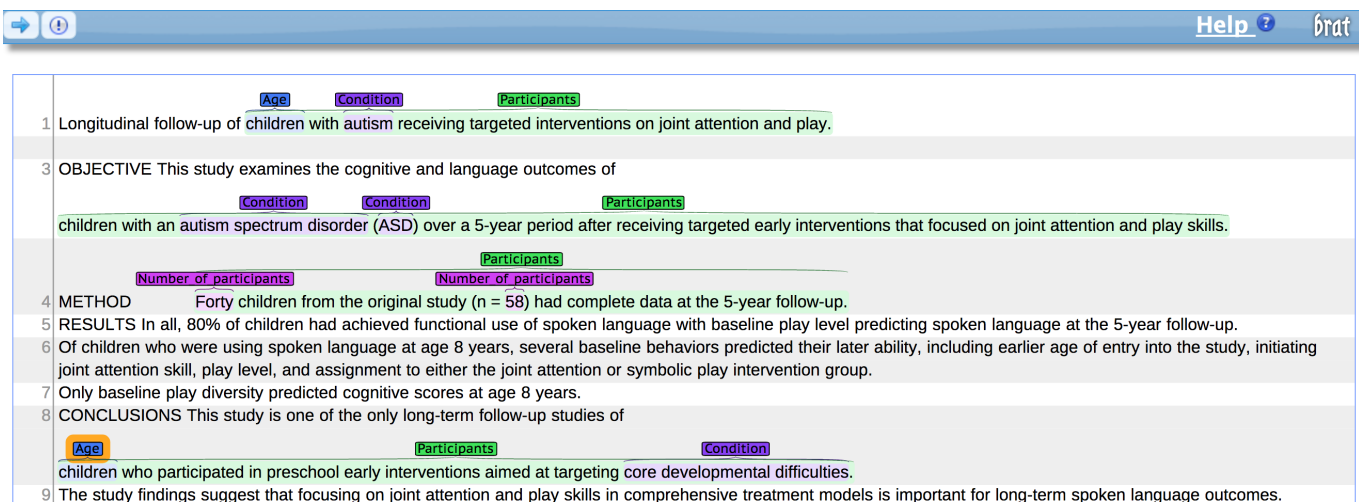


Figure 3: Annotated document.

## Part 2: Supplementing the snippets

Once you have finished marking snippets, you must complete the matching task before you are finished with the document. Click on the "!" button in the top left to open the snippet-matching window. The goal of this task is to match snippets to medical names and to indicate which of the snippets refer to the same information.

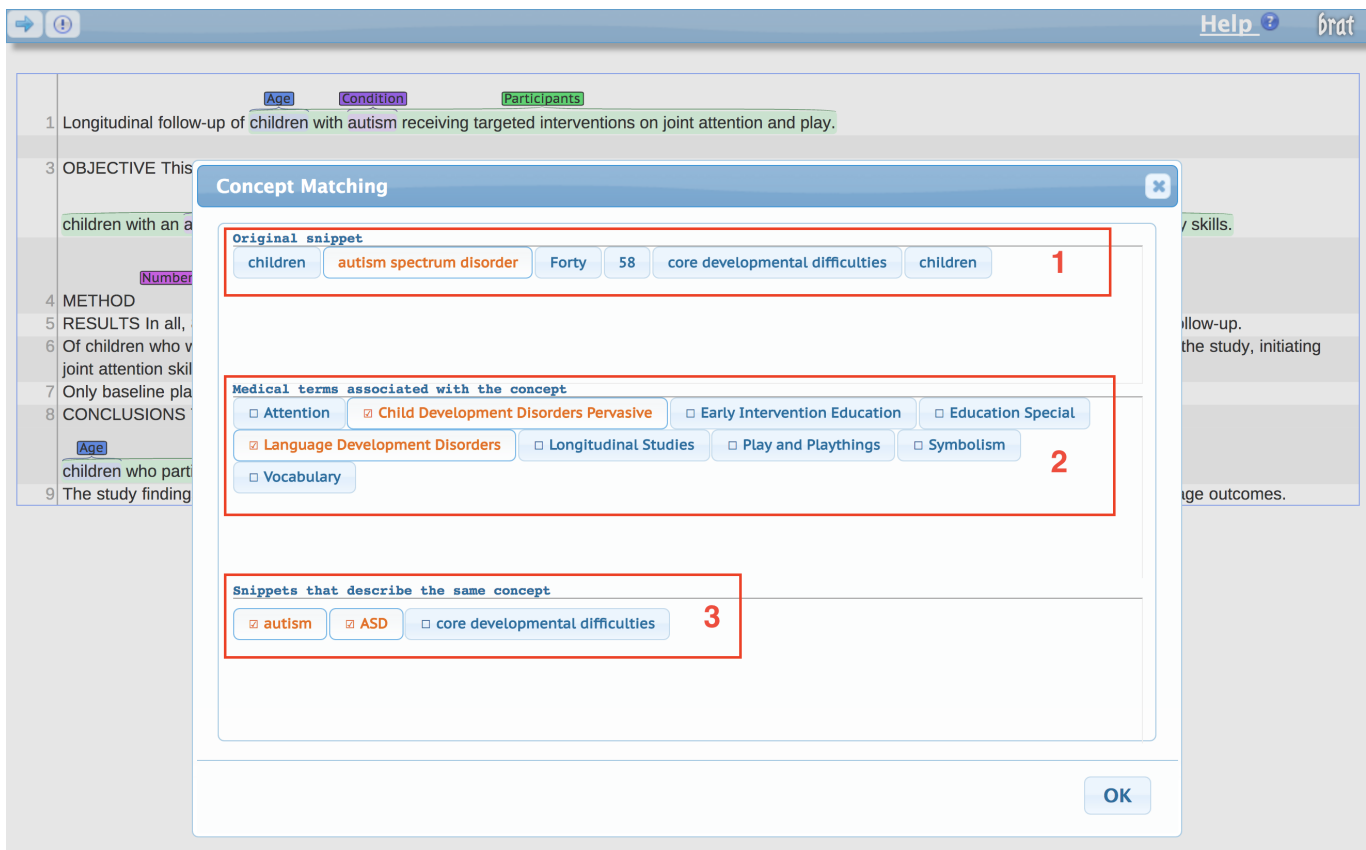


Figure 4: Snippet-matching interface.

The matching interface has three main components

1. All of the snippets you highlighted



2. The list of possible medical tags that might match the snippet
3. The list of other snippets that might mean the same thing as the current snippet

Click on one of the snippets in the top section, and then mark any of the medical terms in the middle section that you think match the currently selected snippet. Then click on each of the snippets in the bottom section that **refer to exactly the same information**. Once you indicate that a snippet refers to the same information as the current snippet, that matching snippet will be removed from the top panel so you don't need to enter redundant information. Repeat this process for each of the snippets displayed in the top section.

Notes:

- Each snippet can only match other snippets with the same type of label.
- Only submit this task once. After you click 'OK' move to the next document (or finish the HIT).

## Example matching

In the image above, the currently selected snippet is the Condition snippet "autism spectrum disorder." In the middle panel two of the tags seem directly related to the snippet so they have been checked. In the bottom panel all of the other Condition snippets are listed. Two of them, "autism" and "ASD" refer to exactly the same thing so they have been checked. Note that "autism" and "ASD" no longer appear in the top panel after being checked off. The remaining Condition snippet, "core developmental difficulties," is related to "autism spectrum disorder" but is not exactly the same thing so it is left unchecked.

## Completing the HIT

You will be shown three documents to annotate. After annotating and completing the matching task for each document, click the right arrow button in the top left corner. When you click this button after finishing the third document, you will see a pop-up window with your unique ID for the task. Enter this ID into the form below to complete the HIT.

In the annotation page you can always view these instructions again by clicking on the "help" button at the upper right corner of the navigation bar.

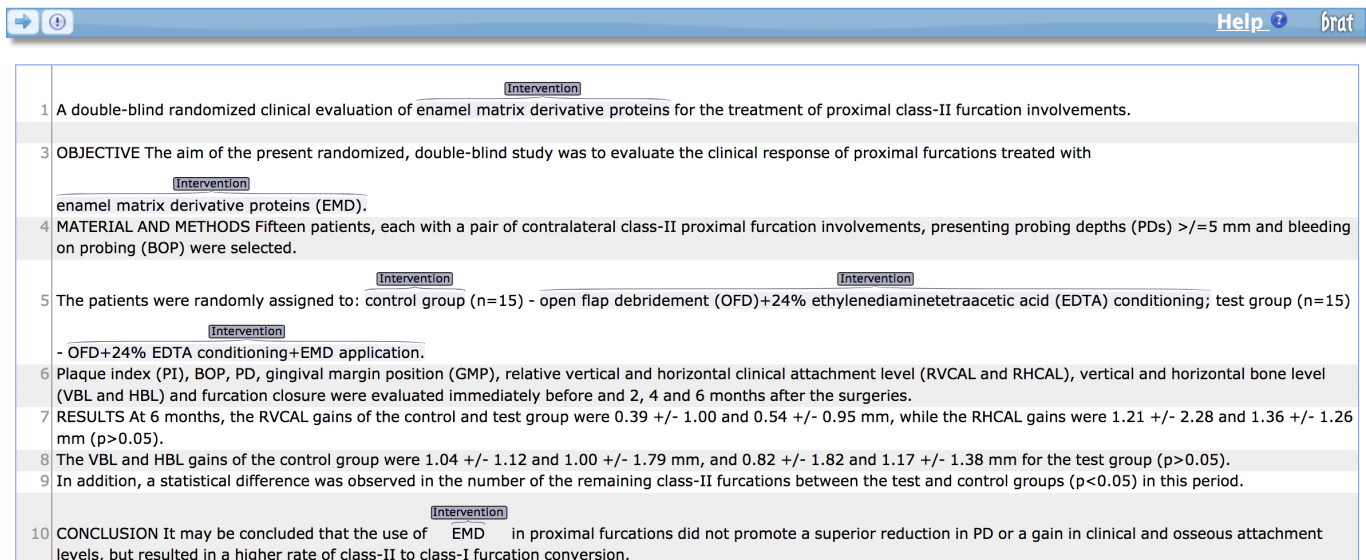
## Instructions

# This task is about medical study Interventions. Please read the instructions carefully.

This task involves adding and refining information about pre-selected snippets of text from medical journal articles. All of the snippets describe the interventions of a medical study. Interventions are the actions or treatments used to modify the outcomes of the trials.

There are three steps to the task for each document:

1. Highlight and label each individual intervention within the starting Interventions snippets
2. Group together repeated mentions of the same intervention
3. Select relevant tags for each group of interventions



The screenshot shows a web interface for a medical document. At the top, there is a blue header bar with navigation icons on the left, a 'Help' button with a question mark, and the 'brat' logo on the right. Below the header, the document text is displayed in a light gray background. Several snippets of text are highlighted in a light blue color, and each snippet is labeled with a small blue box containing the word 'Intervention'. The snippets are numbered 1 through 10. Snippet 1 is a title: 'A double-blind randomized clinical evaluation of enamel matrix derivative proteins for the treatment of proximal class-II furcation involvements.' Snippet 3 is the objective: 'The aim of the present randomized, double-blind study was to evaluate the clinical response of proximal furcations treated with enamel matrix derivative proteins (EMD).' Snippet 4 is the material and methods: 'Fifteen patients, each with a pair of contralateral class-II proximal furcation involvements, presenting probing depths (PDs) >=5 mm and bleeding on probing (BOP) were selected.' Snippet 5 is the intervention: 'The patients were randomly assigned to: control group (n=15) - open flap debridement (OFD)+24% ethylenediaminetetraacetic acid (EDTA) conditioning; test group (n=15) - OFD+24% EDTA conditioning+EMD application.' Snippet 6 is the plaque index and furcation closure evaluation: 'Plaque index (PI), BOP, PD, gingival margin position (GMP), relative vertical and horizontal clinical attachment level (RVCAL and RHCAL), vertical and horizontal bone level (VBL and HBL) and furcation closure were evaluated immediately before and 2, 4 and 6 months after the surgeries.' Snippet 7 is the results: 'At 6 months, the RVCAL gains of the control and test group were 0.39 +/- 1.00 and 0.54 +/- 0.95 mm, while the RHCAL gains were 1.21 +/- 2.28 and 1.36 +/- 1.26 mm (p>0.05).' Snippet 8 is the VBL and HBL gains: 'The VBL and HBL gains of the control group were 1.04 +/- 1.12 and 1.00 +/- 1.79 mm, and 0.82 +/- 1.82 and 1.17 +/- 1.38 mm for the test group (p>0.05).' Snippet 9 is a statistical difference: 'In addition, a statistical difference was observed in the number of the remaining class-II furcations between the test and control groups (p<0.05) in this period.' Snippet 10 is the conclusion: 'It may be concluded that the use of EMD in proximal furcations did not promote a superior reduction in PD or a gain in clinical and osseous attachment levels, but resulted in a higher rate of class-II to class-I furcation conversion.'

Figure 1: Sample document with pre-selected snippets highlighted.

## Part 1: Refining the Interventions snippets

You will receive a document that has already been annotated by other workers in order to identify the snippets of text that describe the interventions of the study. **These starting snippets are not always completely accurate.** Sometimes there will be interventions that haven't been highlighted, and sometimes the snippets that are highlighted won't contain an intervention. Please do your best with the provided snippets.

For each starting Interventions snippet, highlight any individual intervention contained in the snippet. Sometimes there will be multiple interventions reported in a single Interventions snippet. Highlighting the text will open the annotation interface, where you can select the best label and specificity for the intervention.

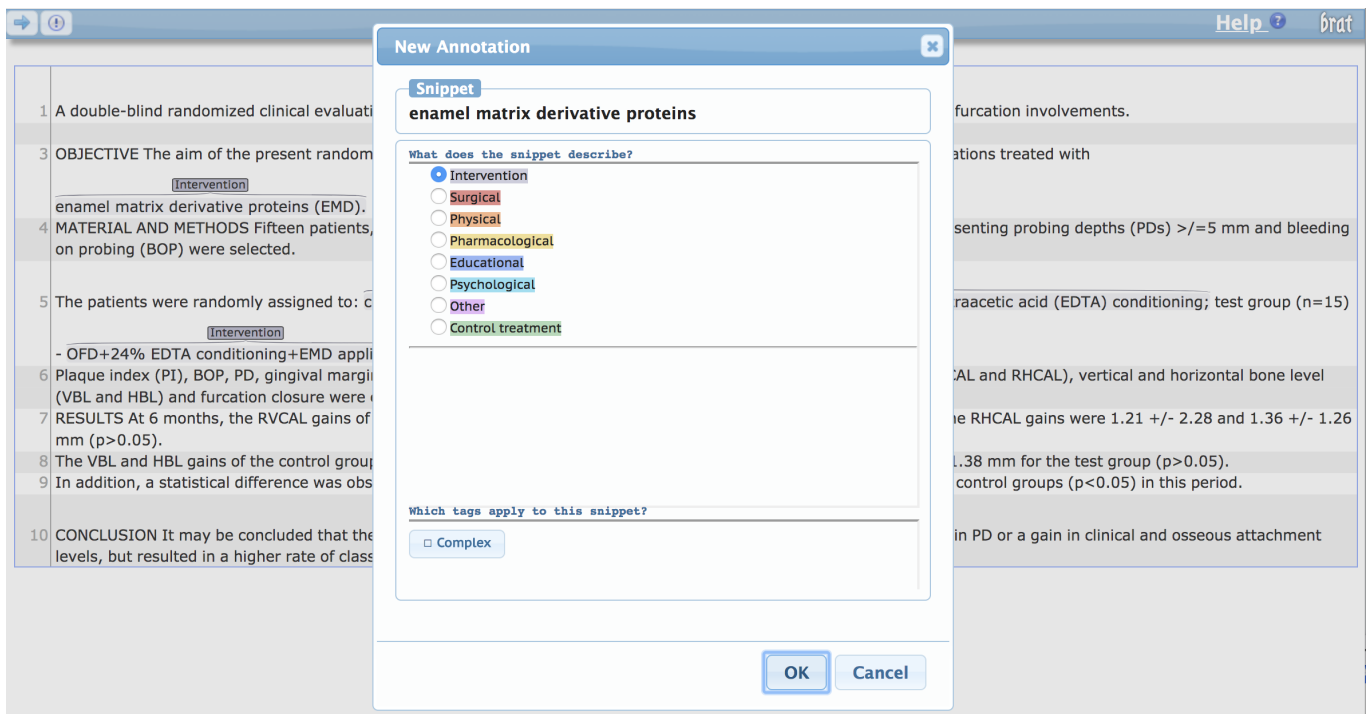


Figure 2: Annotation interface for marking a highlighted section as an individual intervention.

## Intervention label categories

There are three main intervention labels, each of which has more specific labels for common intervention types. When selecting the best label, use the specific labels if possible.

1. Surgical: Surgeries or operations.
2. Physical: Non-surgical interventions that are primarily physical.
3. Pharmacological: Drugs or other biochemical interventions.
4. Educational: Any intervention attempting to change patient behavior via education.
5. Psychological: Other mental interventions such as therapy.
6. Other: Valid interventions not captured by any of the above labels.
7. Control treatment: The comparator treatment. This is often but not always a placebo or non-active treatment.

When selecting the best label, there will be a check box in the bottom pane to indicate if this intervention is part of a complex intervention (e.g. used together with other interventions). If so, check this box.

## Example

1 PROCLAIM: pilot study to examine the effects of **Pharmacological Intervention** clopidogrel on inflammatory markers in patients with metabolic syndrome receiving low-dose **Intervention Pharmacological** aspirin.

3 Metabolic syndrome is associated with intravascular inflammation, as determined by increased levels of inflammatory biomarkers and an increased risk of ischemic atherothrombotic events.

4 Evidence suggests that atherothrombosis and intravascular inflammation share predictive biomarkers, including high-sensitivity C-reactive protein, CD40 ligand, P-selectin, and N-terminal pro-brain natriuretic peptide.

5 Patients who had metabolic syndrome were randomized to receive

**Pharmacological Pharmacological Control\_treatment Pharmacological**  
 clopidogrel 75 mg/day plus aspirin 81 mg/day (n = 89) or placebo plus aspirin 81 mg/day (n = 92) for 9 weeks to assess the efficacy of each treatment in suppression of inflammatory markers.

6 Change from baseline in the levels of high-sensitivity C-reactive protein, CD40 ligand, P-selectin, and N-terminal pro-brain natriuretic peptide at 6 weeks was assessed to evaluate each treatment.

7 There was a significant difference at Week 6 in model-adjusted CD40-ligand levels in favor of **Intervention Pharmacological Pharmacological** clopidogrel plus aspirin compared with **Intervention Pharmacological** placebo plus aspirin in both the intent-to-treat population (difference between least-squares means = -186.5; 95% confidence interval, -342.3 to -30.8; P = 0.02) and the per-protocol population (P = 0.05).

8 No significant differences were observed between the treatment arms for high-sensitivity C-reactive protein, P-selectin, and N-terminal pro-brain natriuretic peptide.

9 There were no deaths or serious adverse events in either treatment arm.

10 Data from this study suggest that **Pharmacological Intervention** clopidogrel can decrease the expression of the CD40-ligand biomarker.

Figure 3: Completed annotation for initial document.

Example annotation 1: note how some of the starting spans contain multiple interventions, each of which are selected individually.

1 Morphological onset and early diagnosis in apical hypertrophic cardiomyopathy: a long term analysis with **Intervention Physical** nuclear magnetic resonance imaging.

3 OBJECTIVES A long-term follow-up study with **Intervention Physical** nuclear magnetic resonance (NMR) imaging was undertaken to detect the morphological onset and to establish the early diagnosis in apical hypertrophic cardiomyopathy (HCM).

4 BACKGROUND A spadelike configuration on left ventriculogram (LVG) is regarded as a diagnostic criterion for the classical apical HCM.

5 There also exists a segmented hypertrophy at the apical level without indicating the spadelike features (a nonspade configuration).

6 To detect the hypertrophied myocardium of the nonspade configuration, circumferential scrutiny of the apex is required.

7 Although both configurations can be underlying causes of giant negative T waves, etiological relationship between the two is not clarified.

8 METHODS The criteria for the spadelike configuration defined on left ventricular short-axis **Intervention** NMR images were as follows: (apical maximal thickness > or = 15 mm), (apical anterior thickness over basal anterior thickness > or = 1.3) and (apical posterior thickness over basal posterior thickness > or = 1.3).

9 Thirteen patients who had predominant hypertrophy (> or = 15 mm) at the apical level without the spadelike configuration underwent **Physical Intervention** NMR imaging twice before and after 54+/-10 months' follow-up.

10 RESULTS Apical hypertrophy that had been confined to the lateral wall in four, the anterior-lateral wall in two, and the septal-anterior wall in one developed to become circumferential hypertrophy that fulfilled the criteria for the spadelike configuration after the follow-up period.

11 CONCLUSIONS The spadelike configuration can begin with the nonspade configuration and therefore, both can constitute a single disease entity of apical HCM.

12 The early diagnosis of apical HCM can be achieved by identifying the hypertrophy frequently confined to the lateral wall at the apical level.

Figure 4: Completed annotation for initial document.

Example annotation 2: not all of the starting spans contain a true intervention. The "NMR images" are an output from the imaging technique but not an intervention on their own.

## Recap:

1. Highlight each intervention contained in the snippet. This will open the annotation window.

2. Select the most closely matching label for the intervention you have highlighted.
3. Indicate if the intervention is part of a complex intervention.
4. Click "OK" to add your selection. Repeat this process for each intervention in the article.

**Notes:**

1. The intervention snippets cannot overlap with each other.
2. The intervention snippets must be contained completely within the original Interventions snippet.
3. Each different intervention should be highlighted as its own snippet.

## Part 2: Grouping repeated mentions of the same intervention

Once you have annotated all of the interventions, open the next task by clicking the arrow in the top-left of the interface (or by pressing the right arrow key). This will open the snippet-matching window. Your task here is to group together any snippets that refer to the same intervention. The top panel lists all of the snippets in the document, and the bottom panel lists all of the other snippets that are potentially the same.

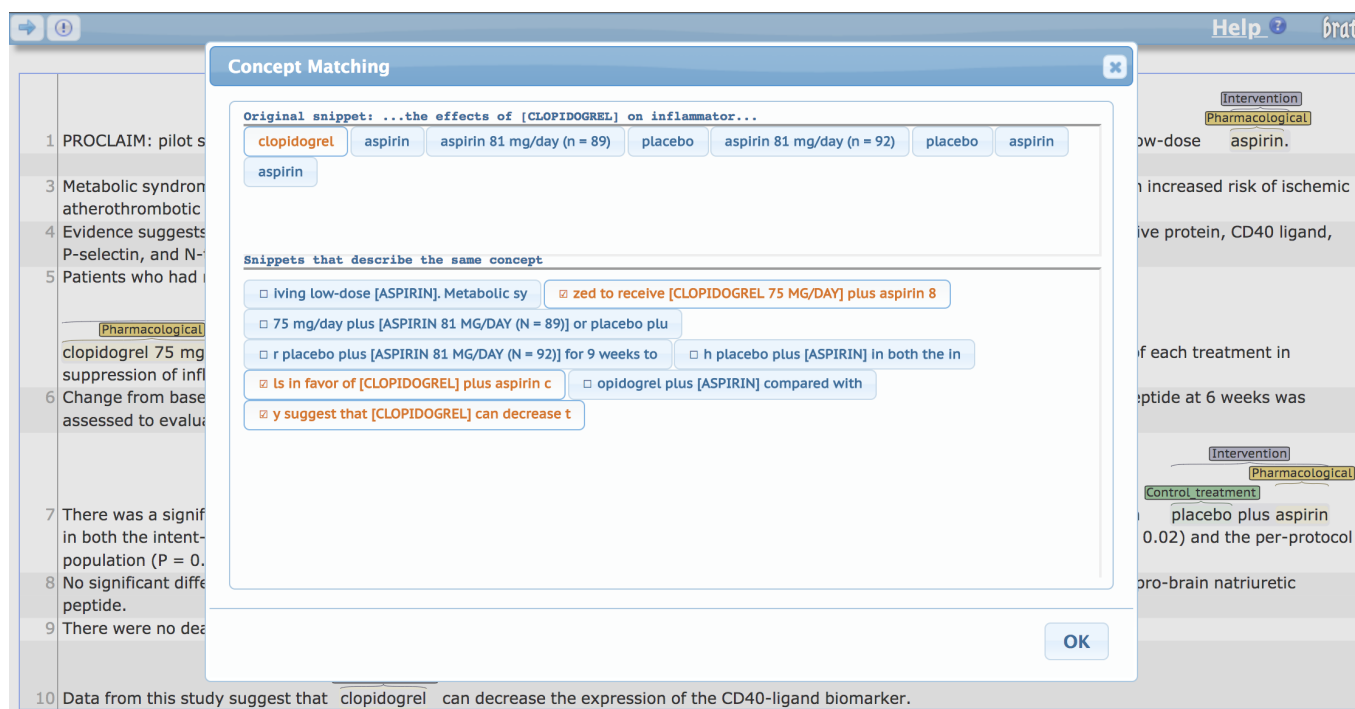


Figure 5: Snippet matching interface.

1. All of the annotated interventions. Note that the label at the top shows some context around the snippet to help you remember which snippet is which.
2. The potentially matching snippets (the other snippets with the same label type).

Start by clicking on one of the snippets in the top panel, and then clicking any snippets in the bottom panel that refer to the same intervention. **Repeat this process by clicking on each snippet in the top panel** and marking anything that matches it until you are satisfied that you have grouped together everything that matches. Once you have done so, click the "OK" button to proceed to part three of the task.

**Notes:**

- If two snippets refer to different same intervention but differ only by the dosage, mark them as matching.
- When you mark a snippet as matching another one, it will be removed from the top panel since you've already indicated which group it belongs to.

## Part 3: Adding information to the intervention groups

Once you have finished grouping the interventions, you will see the group-labeling window. Your task here is to assign any applicable labels to each group of interventions. The top-left panel lists all of the groups of interventions, the bottom-left panel contains the potential labels, and the right panel displays each intervention snippet included in the currently selected group.

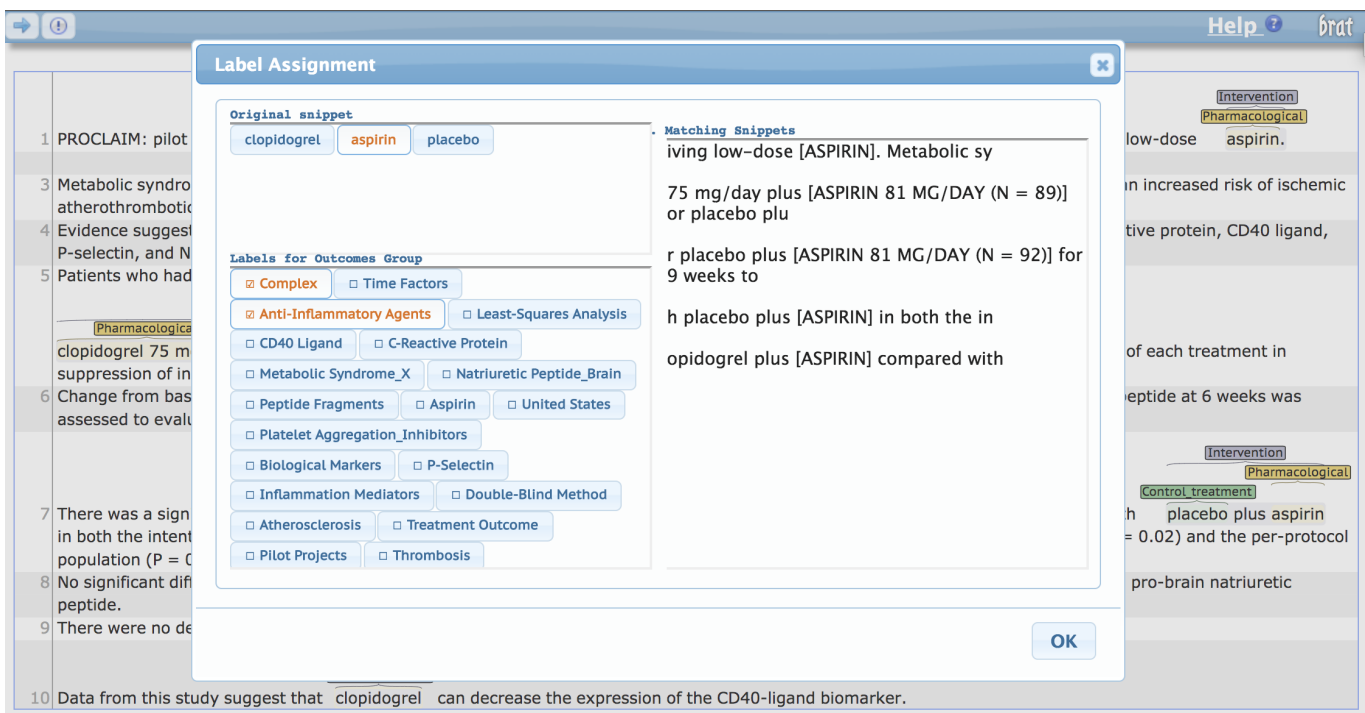


Figure 6: Snippet labeling interface.

1. The first snippet from each snippet group.
2. Potential labels for the snippet group.
3. Each other snippet included in the group.

Start by clicking on one of the snippets in the top-left panel, and then clicking any label in the bottom-left panel that applies to this group. **Repeat this process by clicking on each snippet in the top-left panel** and marking any labels that applies until you are satisfied that you have selected the relevant labels for each group. Once you click "OK" you will have completed the document.

## Completing the HIT

You will be given three documents to annotate. After annotating and completing the matching task for each document, click the right arrow button in the top left corner. When you click this button after finishing the third document, you will see a pop-up window with your unique ID for the task. Enter this ID into the form below to complete the HIT.

In the annotation page you can always view these instructions again by clicking on the "help" button at the upper right corner of the navigation bar.

**Link:** Link to annotation page (`${brat_link}`)

**Provide the unique code here:**

Submit

]]> 1000



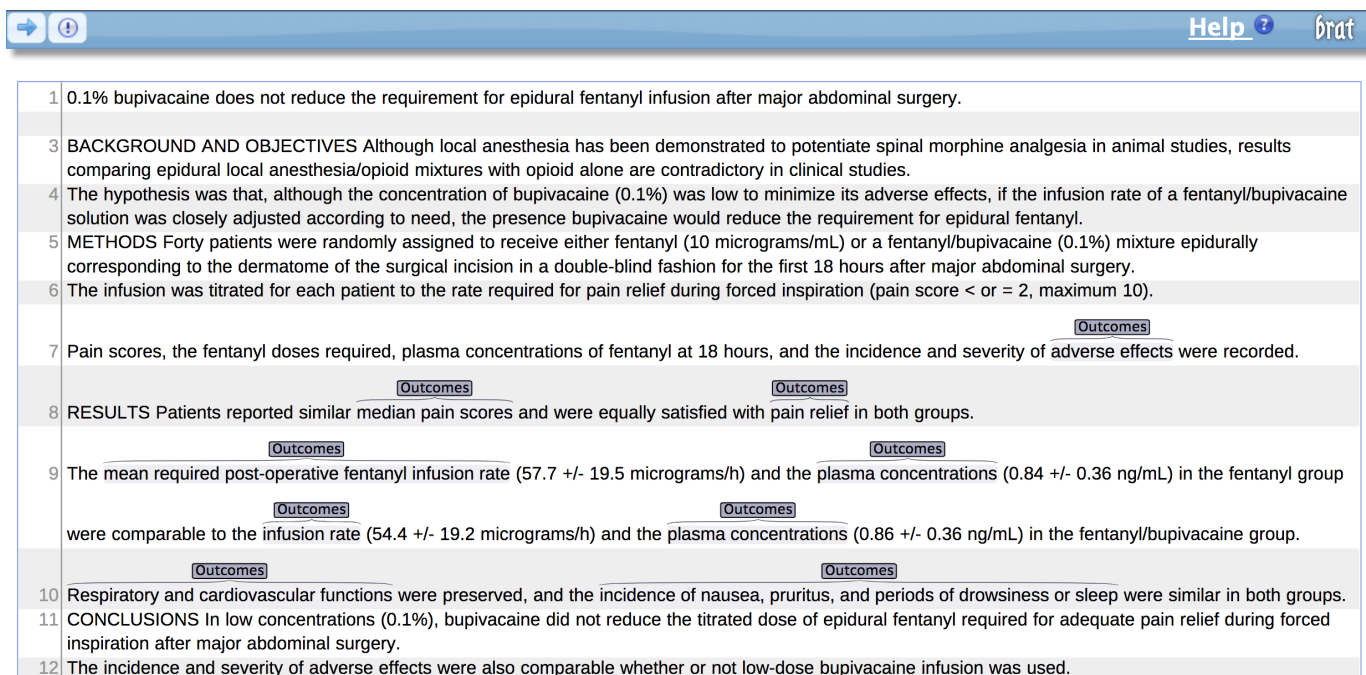
## Instructions

# This task is about medical study Outcomes. Please read the instructions carefully.

This task involves adding and refining information about pre-selected snippets of text from medical journal articles. All of the snippets describe the outcomes of a medical study. Outcomes are the measures reported to determine if a medical intervention is successful, along with any side effects experienced by the study participants.

There are three steps to the task for each document:

1. Highlight and label each individual outcome within the starting Outcomes snippets
2. Group together repeated mentions of the same outcome
3. Select relevant tags for each group of outcomes



The screenshot shows a document interface with a blue header bar containing navigation icons, a 'Help' button, and the 'brat' logo. The document text is displayed in a table-like format with line numbers on the left. Several snippets of text are highlighted in light blue, and each has a small 'Outcomes' label above it. The text includes sections for Background and Objectives, Methods, Results, and Conclusions. The highlighted snippets are: '0.1% bupivacaine does not reduce the requirement for epidural fentanyl infusion after major abdominal surgery.', 'Pain scores, the fentanyl doses required, plasma concentrations of fentanyl at 18 hours, and the incidence and severity of adverse effects were recorded.', 'RESULTS Patients reported similar median pain scores and were equally satisfied with pain relief in both groups.', 'The mean required post-operative fentanyl infusion rate (57.7 +/- 19.5 micrograms/h) and the plasma concentrations (0.84 +/- 0.36 ng/mL) in the fentanyl group were comparable to the infusion rate (54.4 +/- 19.2 micrograms/h) and the plasma concentrations (0.86 +/- 0.36 ng/mL) in the fentanyl/bupivacaine group.', 'Respiratory and cardiovascular functions were preserved, and the incidence of nausea, pruritus, and periods of drowsiness or sleep were similar in both groups.', and 'The incidence and severity of adverse effects were also comparable whether or not low-dose bupivacaine infusion was used.'

Figure 1: Sample document with pre-selected snippets highlighted.

## Part 1: Refining the Outcomes snippets

You will receive a document that has already been annotated by other workers in order to identify the snippets of text that describe the outcomes of the study. **These starting snippets are not always completely accurate.** Sometimes there will be outcomes that haven't been highlighted, and sometimes the snippets that are highlighted won't contain an outcome. Please do your best with the provided snippets.

For each starting Outcomes snippet, highlight any individual outcome contained in the snippet. Sometimes there will be multiple outcomes reported in a single Outcomes snippet. Highlighting the text will open the annotation interface, where you can select the best label and specificity for the outcome.



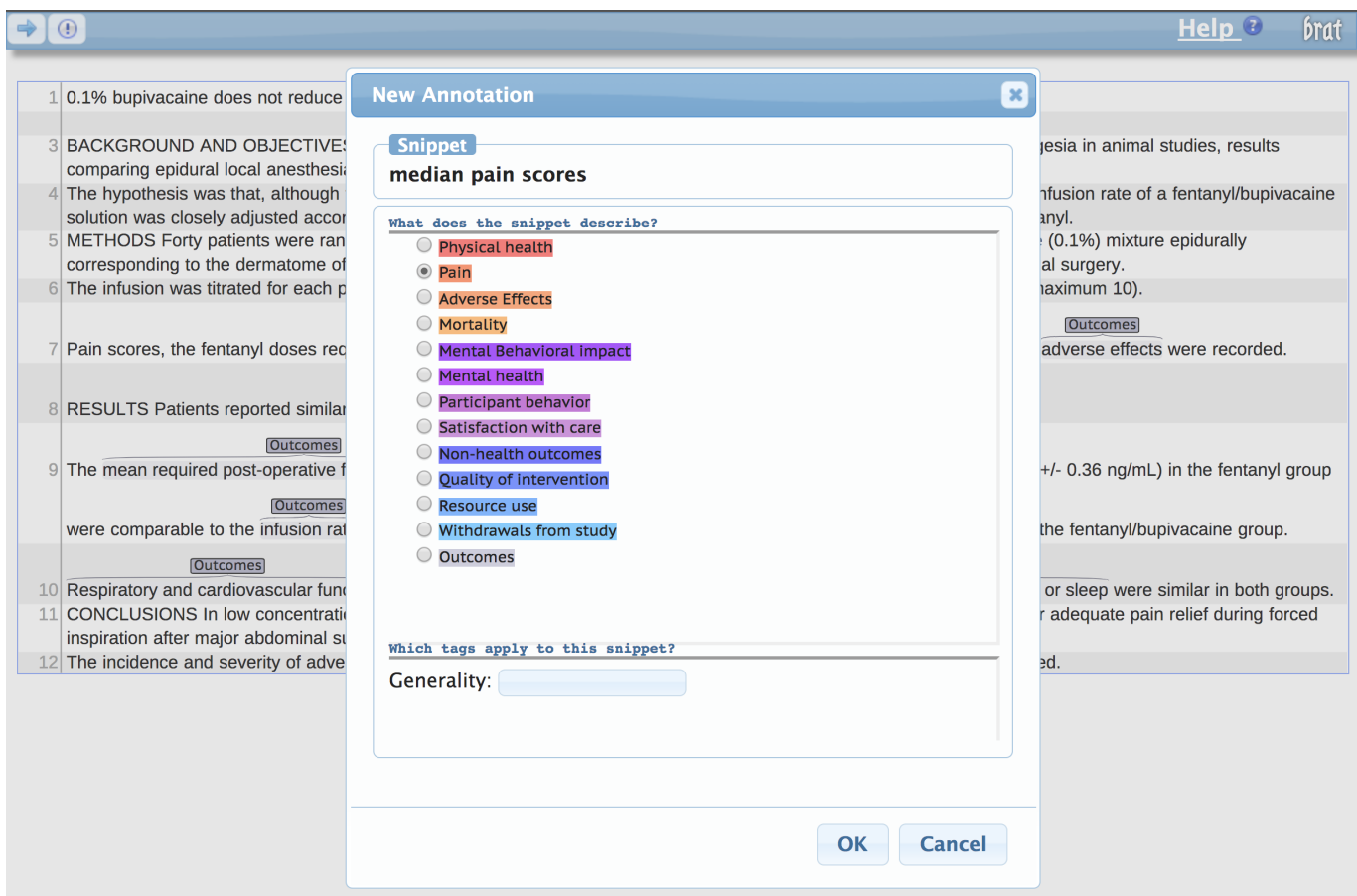


Figure 2: Annotation interface for marking a highlighted section as an individual outcome.

## Outcome label categories

There are three main outcome labels, each of which has more specific labels for common outcome types. When selecting the best label, use the specific labels if possible.

### 1. Physical health

Any measurement or property of the participant's physical health.

- Pain
  - Evaluation of participant's pain level.
    - Example: daily pain measurements
- Adverse Effects
  - Any physical side effects of the intervention.
    - Example: infection
- Mortality
  - Anything related to death or survival rates.

### 2. Mental/Behavioral impact

Any evaluation of the participant's mental state, capability, or behavior.

- Mental health
  - Measurements or analysis of brain function/health.
    - Example: anxiety
    - Example: executive brain function
    - Example: word classification accuracy

- Participant behavior  
Actions or behaviors of the participants.
  - Example: opioid usage
  - Example: higher social skills
  - Example: compliance with treatment
- Satisfaction with care  
How happy or content participants are.

### 3. Non-health outcomes

Any outcome type that isn't a direct evaluation of the participant's health/behavior.

- Quality of intervention  
Any measurement or qualification of how successful the intervention was.
  - Example: recovery time
  - Example: safety
  - Example: efficacy
- Resource use  
Specific mentions of the amount of resources (including money and medical resources) used.
- Withdrawals from study  
If participants dropped from the study.

## Specific vs. General

Often articles will sometimes provide a general description of the outcome that is being measured, and then give specific measurements for this outcome. Whenever this happens, select the appropriate "General" or "Specific" tag in the annotation interface.

## Example

1 0.1% bupivacaine does not reduce the requirement for epidural fentanyl infusion after major abdominal surgery.

3 BACKGROUND AND OBJECTIVES Although local anesthesia has been demonstrated to potentiate spinal morphine analgesia in animal studies, results comparing epidural local anesthesia/opioid mixtures with opioid alone are contradictory in clinical studies.

4 The hypothesis was that, although the concentration of bupivacaine (0.1%) was low to minimize its adverse effects, if the infusion rate of a fentanyl/bupivacaine solution was closely adjusted according to need, the presence bupivacaine would reduce the requirement for epidural fentanyl.

5 METHODS Forty patients were randomly assigned to receive either fentanyl (10 micrograms/mL) or a fentanyl/bupivacaine (0.1%) mixture epidurally corresponding to the dermatome of the surgical incision in a double-blind fashion for the first 18 hours after major abdominal surgery.

6 The infusion was titrated for each patient to the rate required for pain relief during forced inspiration (pain score < or = 2, maximum 10).

7 Pain scores, the fentanyl doses required, plasma concentrations of fentanyl at 18 hours, and the incidence and severity of adverse effects were recorded.

8 RESULTS Patients reported similar median pain scores and were equally satisfied with pain relief in both groups.

9 The mean required post-operative fentanyl infusion rate (57.7 +/- 19.5 micrograms/h) and the plasma concentrations (0.84 +/- 0.36 ng/mL) in the fentanyl group were comparable to the infusion rate (54.4 +/- 19.2 micrograms/h) and the plasma concentrations (0.86 +/- 0.36 ng/mL) in the fentanyl/bupivacaine group.

10 Respiratory and cardiovascular functions were preserved, and the incidence of nausea, pruritus, and periods of drowsiness or sleep were similar in both groups.

11 CONCLUSIONS In low concentrations (0.1%), bupivacaine did not reduce the titrated dose of epidural fentanyl required for adequate pain relief during forced inspiration after major abdominal surgery.

12 The incidence and severity of adverse effects were also comparable whether or not low-dose bupivacaine infusion was used.

Figure 3: Completed annotation for initial document.

In the document annotated in Figure 3, there are two examples of the general/specific relationship.

1. One observed outcome is mentioned as adverse effects, and the later in the document the specific effects are listed (nausea, pruritus, and drowsiness or sleep). Note that the specific effects are all listed in one Outcomes snippet but are all different specific outcomes so they have been labeled individually.
2. The first description of the mean required post-operative fentanyl infusion rate is much more specific than when it is referred to again as simply the infusion rate.

### Recap:

1. Highlight each outcome contained in the snippet. This will open the annotation window.
2. Select the most closely matching label for the outcome you have highlighted.
3. Indicate if the outcome is general or specific.
4. Click "OK" to add your selection. Repeat this process for each outcome in the article.

### Notes:

1. The outcome snippets cannot overlap with each other.
2. The outcome snippets must be contained completely within the original Outcomes snippet.
3. Each different outcome should be highlighted as its own snippet.

## Part 2: Grouping repeated mentions of the same outcome

Once you have annotated all of the outcomes, open the next task by clicking the arrow in the top-left of the interface (or by pressing the right arrow key). This will open the snippet-matching window. Your task here is to group together any snippets that refer to the same outcome. The top panel lists all of the snippets in the document, and the bottom panel lists all of the other snippets that are potentially the same.

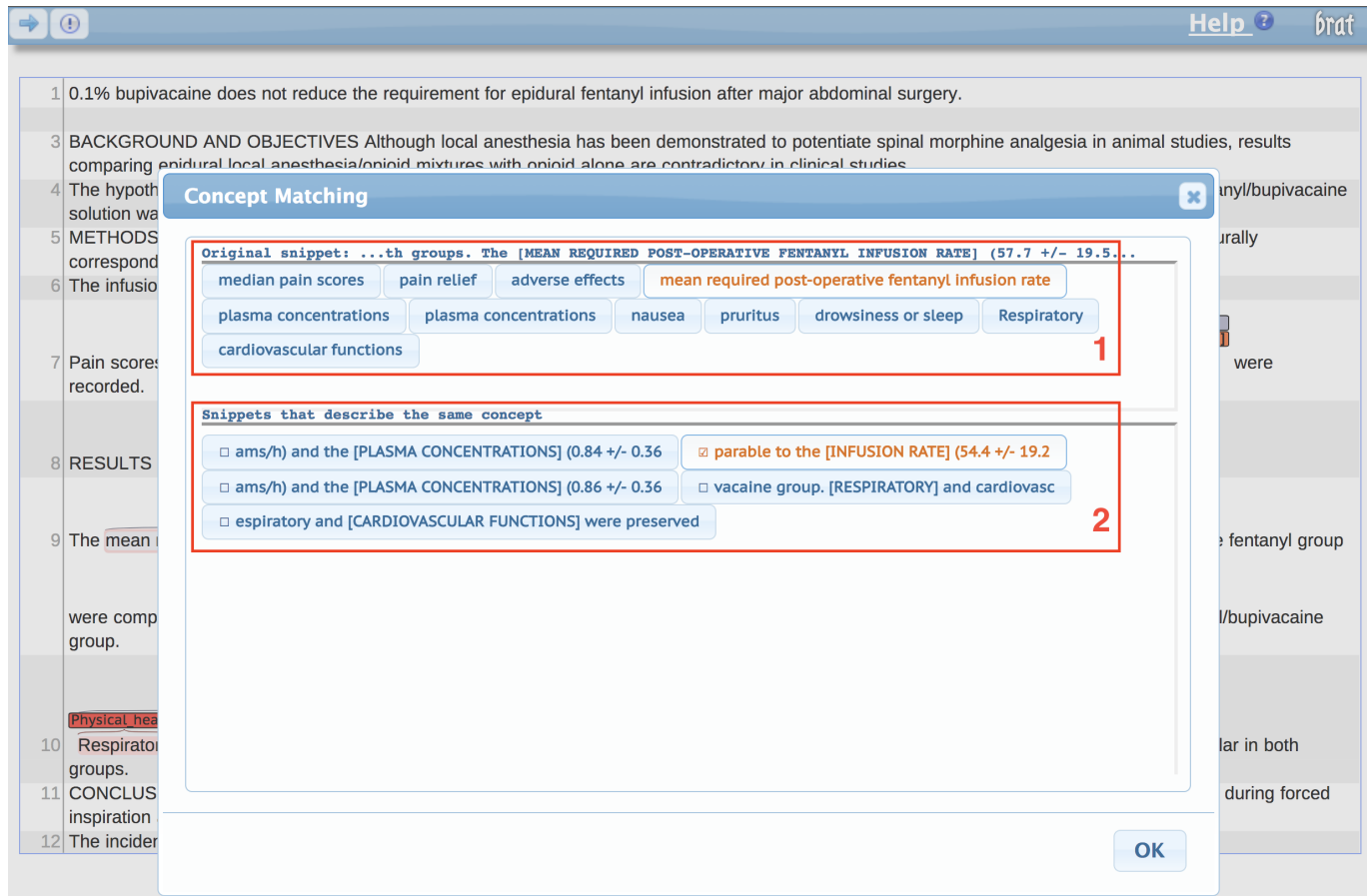


Figure 4: Snippet matching interface.

1. All of the annotated outcomes. Note that the label at the top shows some context around the snippet to help you remember which snippet is which.
2. The potentially matching snippets (the other snippets with the same label type).

Start by clicking on one of the snippets in the top panel, and then clicking any snippets in the bottom panel that refer to the same outcome. **Repeat this process by clicking on each snippet in the top panel** and marking anything that matches it until you are satisfied that you have grouped together everything that matches. Once you have done so, click the "OK" button to proceed to part three of the task.

#### Notes:

- If two snippets refer to different levels of generality for the same outcome, mark them as matching.
- When you mark a snippet as matching another one, it will be removed from the top panel since you've already indicated which group it belongs to.

## Part 3: Adding information to the outcome groups

Once you have finished grouping the outcomes, you will see the group-labeling window. Your task here is to assign any applicable labels to each group of outcomes. The top-left panel lists all of the groups of outcomes, the bottom-left panel contains the potential labels, and the right panel displays each outcome snippet included in the currently selected group.

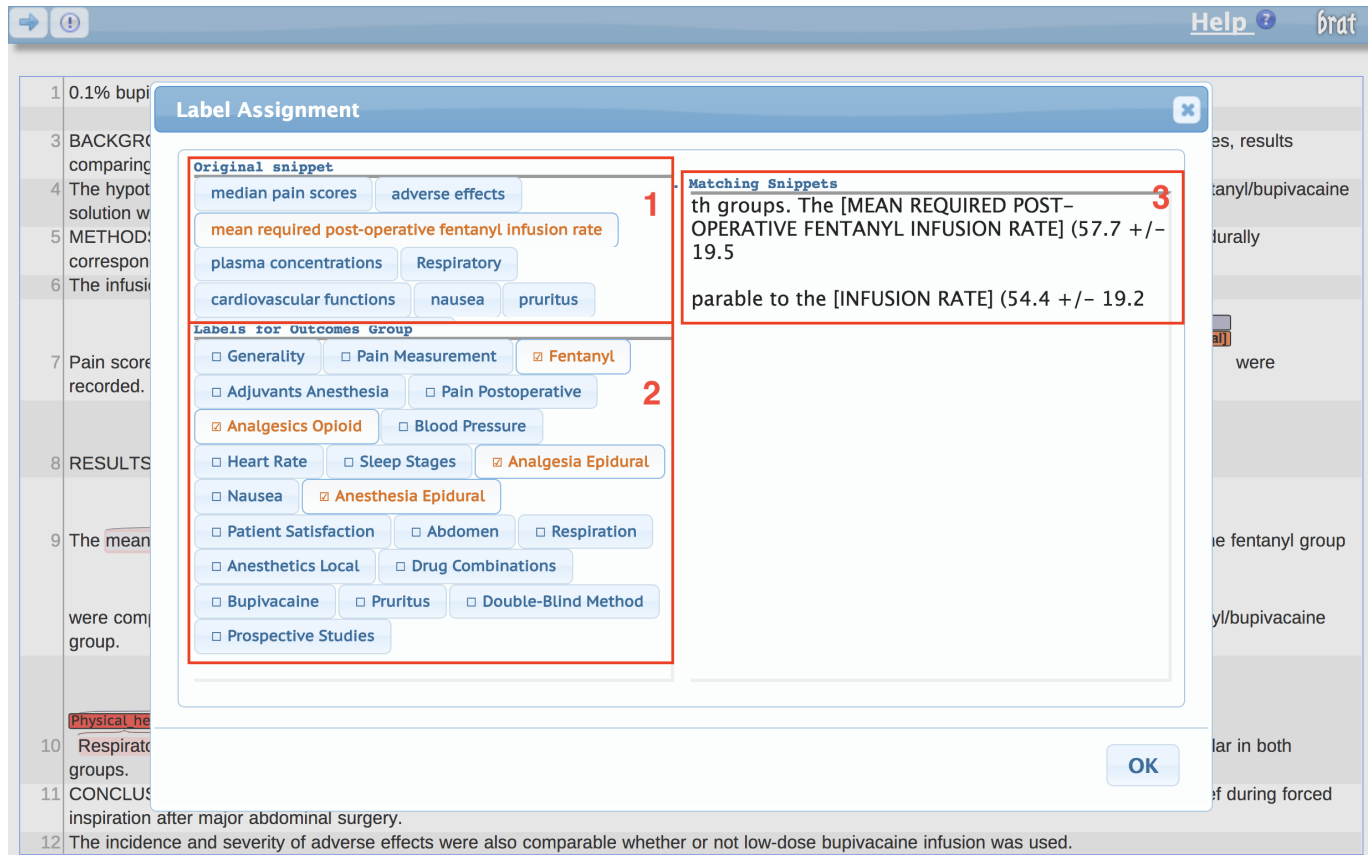


Figure 5: Snippet labeling interface.

1. The first snippet from each snippet group.
2. Potential labels for the snippet group.
3. Each other snippet included in the group.

Start by clicking on one of the snippets in the top-left panel, and then clicking any label in the bottom-left panel that applies to this group. **Repeat this process by clicking on each snippet in the top-left panel** and marking any labels that applies until you are satisfied that you have selected the relevant labels for each group. Once you click "OK" you will have completed the document.

## Completing the HIT

You will be given three documents to annotate. After annotating and completing the matching task for each document, click the right arrow button in the top left corner. When you click this button after finishing the third document, you will see a pop-up window with your unique ID for the task. Enter this ID into the form below to complete the HIT.

In the annotation page you can always view these instructions again by clicking on the "help" button at the upper right corner of the navigation bar.

**Link:**

Link to annotation page (`{brat_link}`)

**Provide the unique code here:**

e.g. 21EC2020-3AEA-4069-A2DD-08002B30309D

Submit

]]> 1000