Q-Pain Datasheet: Purpose & Documentation

Links

The Q-Pain dataset is hosted (and publicly available) on the PhysioNet platform, and licensed under the Creative Commons Attribution-ShareAlike 4.0 International Public License.
The DOI (currently being activated) and dataset citation are:

> Logé, C., Ross, E., Dadey, D. Y. A., Jain, S., Saporta, A., Ng, A., & Rajpurkar, P. (2021). Q-Pain: A Question Answering Dataset to Measure Social Bias in Pain Management (version 1.0.0). PhysioNet. https://doi.org/10.13026/61xq-mj56

The following Datasheet framework was modeled after Datasheets for Datasets.

Dataset Motivation

For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

> This dataset was created for the purpose of helping researchers characterize and quantify bias in medical QA systems (specifically the pain management context). It can also be used to assess bias in human medical decision-making. To our knowledge, no such datasets exist at present.

Who created this dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

> This dataset was created by a team of researchers at Stanford University within the Stanford Machine Learning Group.

Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.

> No funding or associated grant. No conflict of interest.

Dataset Composition

What do the instances that comprise the dataset represent? How many instances are there in total? Does the dataset contain all possible instances or is it a sample (not necessarily random)

of instances from a larger set? What data does each instance consist of? "Raw" data (e.g., unprocessed text or images) or features? In either case, please provide a description.

The dataset consists of 55 hypothetical medical "vignettes" (scenarios), in which a patient's pain symptom presentations are described. Associated with each vignette is also a question, an answer, and an explanation. The question asks whether or not the described patient should be prescribed the appropriate pain treatment, and if so, what dosage. Each answer ("Yes." or "No.") represents the medically-appropriate response, as validated by three physicians. Each dosage is "Low" by default—both the low and high dosages would be adequate for all scenarios depicted.

Dataset Files / Features	**Description**
The Q-Pain dataset is structured into **five csv files**, one for each medical contexts: 1. Acute Non-Cancer pain (*data_acute_non_cancer.csv*) , 2. Acute Cancer pain (*data_acute_cancer.csv*), 3. Chronic Non-Cancer pain (*data_chonic_non_cancer.csv*), 4. Chronic Cancer pain (*data_chronic_cancer.csv*), and 5. Post-Operative pain (*data_post_op.csv*)	Each csv file includes 10 "Yes." vignettes, and 1 "No." vignette. The fields (all *strings*) are: ● Vignette: the case presentation, ● Question: the question, ● Answer: the appropriate answer ● Dosage: the dosage ("Low" by default although both low and high would be adequate) ● Explanation: a brief explanation justifying the response
Across the Vignette, Question and Explanation fields are ***string* placeholders** used to allow for substitution of patient profiles.	Placeholders are used in the Yes vignettes: ● 'Patient D' (in the Yes vignettes) or 'Patient B' (in the No vignettes) to be kept as is or replaced by a patient's name / 'Patient A' / 'Patient C', ● [race] to be replaced by a patient's race (e.g. in our experiment we used: "Asian", "Black", "Hispanic", "White"), ● [gender] to be replaced by a patient's gender (e.g. in our experiment we used "man", "woman") ● [possessive] and [subject] to be replaced by pronouns (e.g. "his", "her" or "he", "she") always lowercase.
With the dataset is a **Python notebook**: ● Q-Pain_Experiments.ipynb	The notebook includes a walkthrough of the data as well as starter code to reproduce our Experimental Design framework.

Is the dataset self-contained, or does it link to or otherwise rely on external resources?

The dataset is self-contained.

Dataset Confidentiality

Does the dataset contain data that might be considered confidential? Does the dataset relate to people? Does the dataset identify any subpopulations (e.g., by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.

No. The data contains no real patient information, and is made up entirely of hypothetical medical scenarios.

Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why

No, we don't think that the data would be offensive, insulting or threatening to anyone.

Is it possible to identify individuals either directly or indirectly? Does the dataset contain data that might be considered sensitive in any way?

No.

Collection Process

How was the data associated with each instance acquired? What mechanisms or procedures were used to collect the data? Who was involved in the data collection process and how were they compensated?

Each vignette was written by a physician on the team, who then validated the dataset for medical appropriateness/relevance with two other physicians not involved in the data design or collection process.

Over what timeframe was the data collected / created?

The dataset was created between March 2021 and June 2021.

Were any ethical review processes conducted (e.g., by an institutional review board)?

No ethical review process was required.

Dataset Use Cases

Has the dataset been used for any tasks already? If so, please provide a description.

> The dataset was used to demonstrate a QA system bias assessment framework with GPT-2 and GPT-3 as proof of concept. Vignettes were presented to GPT-2 and GPT-3 multiple times, each with different race/gender profiles for the featured patient. We extracted the underlying probabilities that the reference QA systems would deny or prescribe treatment to the patient, along with the probabilities that the systems would prescribe a low vs high dosage. See Physionet Usage Notes for further details.

What (other) tasks could the dataset be used for?

> The dataset could also be used to measure bias within a population of physicians.

Are there tasks for which the dataset should not be used? If so, please provide a description

> This dataset should not be used to train or validate AI systems for medical decision support tasks of any kind.

Dataset Distribution

Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created? If so, please provide a description. How will the dataset be distributed (e.g., tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?

> The Q-Pain dataset is hosted on the PhysioNet platform, and licensed under the Creative Commons Attribution-ShareAlike 4.0 International Public License. Use of the dataset is free to all researchers on the PhysioNet platform. The DOI and citation are:
>
> Logé, C., Ross, E., Dadey, D. Y. A., Jain, S., Saporta, A., Ng, A., & Rajpurkar, P. (2021). Q-Pain: A Question Answering Dataset to Measure Social Bias in Pain Management (version 1.0.0). PhysioNet. https://doi.org/10.13026/61xq-mj56
>
> As of today, the dataset is also available to reviewers via the PhysioNet link and passphrase available at the top of this document.

Have any third parties imposed IP-based or other restrictions on the data associated with the instances?

No IP restrictions apply to the dataset.

Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?

No regulatory restrictions apply apart from the licensed use mentioned above.

Dataset Maintenance

Who is supporting/hosting/maintaining the dataset? How can the owner/curator/manager of the dataset be contacted?

The dataset is being hosted by PhysioNet. It is being maintained by the authors of the paper. The corresponding author can be contacted at : Cécile Logé (ceciloge@stanford.edu)

Is there an erratum? If so, please provide a link or other access point.

No, this is the first version of the dataset.

Will the dataset be updated? Will older versions of the dataset continue to be supported/hosted/maintained?

This will be posted on the dataset's webpage. If the dataset is updated in future, the older version will still be supported and kept around for consistency.

If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?

Others may do so and should contact the original authors about incorporating fixes/extensions.