**Application for Credentialed Access**

**Please use the form below to apply for PhysioNet credentialing. In order to apply:**

* **If you have an institutional/academic email address, please add it in your settings (**[**Emails settings**](https://physionet.org/settings/emails/)**).**
* **If you have an ORCID iD, please link to it in your settings (**[**ORCID settings**](https://physionet.org/settings/orcid/)**) as this may help us expedite your application by making it easier to verify your identity.**

**Recommendations for a timely review**

**The following steps will help to ensure a timely review of your application:**

* **You are required to provide a reference, whom we may contact with regard to your application.**
* **It is also helpful if your reference is a registered user and they have linked their ORCID iD to their profile.**
* **Check the form carefully before submission. Incomplete or erroneous applications will be rejected.**

**Required fields are indicated by a red asterisk (\*).**

**Personal Details**

First Name: Josha

Last Name: Thomas

Suffix, if applicable: Leaving this blank

Researcher category: Graduate Student

Organization name: Harvard School of Public Health

Job title or position: Student

City: Cambridge

State/Province (Required for Canada/U.S.): Massachusetts

ZIP/postal code: 02134

Country: United States

Webpage: I could put my linkedin here if we want to, though this is not a required field.

**Reference**

**Your reference's relationship to you. If you are a student or postdoc, this must be your supervisor. Otherwise, you may list a colleague. Do not list yourself or another student as reference. Remind your reference to respond promptly, as long response times will prevent approval of your application.**

Reference category: Supervisor

Reference name: Francesca Dominici

Reference email : fdominic@hsph.harvard.edu

Reference organization: Harvard School of Public Health

Reference job title or description: Clarence James Gamble Professor of Biostatistics, Population, and Data Science

**Research Area**

**Please provide a detailed description of how you plan to use the data, including the name of any specific dataset(s) you intend to use. If you will be using the data for a class, please also include the name and number of the course. (Minimum : 20 words)**

Research Topic: Our lab is developing and testing a pipeline for training large language models (LLMs) to predict the progression of diseases based on patient history and environmental effects. This dataset allows for experimentation with token-based structures, enabling us to refine the model’s capacity to analyze disease correlations. Testing this process with MIMIC-IV will help validate our approach and facilitate the integration of exposure data to further explore environmental impacts on respiratory diseases.