

Synthetic Dataset User Acknowledgement

Background

The Canadian Partnership for Tomorrow's Health (CanPath) has created a mock dataset, the *CanPath Synthetic Dataset*, designed to mimic the real data collected by CanPath. The Synthetic Dataset offers a valuable resource for teaching and learning about population health research and may be used for training and educational purposes.

Purpose: To provide students with hands-on experience working with population-level data without revealing any actual participant information.

Availability: The dataset is free for instructors at universities and colleges to use in academic courses.

Content: The synthetic dataset includes over 40,000 observations and 403 variables covering socio-demographic information, lifestyle behaviors, health perceptions, and certain self-reported diseases. The dataset also includes mock environmental variables based on data from the Canadian Urban Environmental Health Research Consortium.

Application Process: Applicants must submit an [application form](#), a brief CV, and a course syllabus to access the dataset. If using another dataset, additional documents, such as an REB approval or exemption, may be required.

Usage: The dataset is strictly for training and educational purposes and should not be used for research or research-related publications. Users can [apply to access](#) real CanPath data for research purposes.

This ***Synthetic Dataset User Acknowledgement*** sets out the terms and conditions governing the instructor's access to and use of the Synthetic Dataset for educational and training purposes. By signing this acknowledgment, the instructor agrees to adhere to the terms outlined in the *Synthetic Dataset Application Form* and the following terms and conditions:

1. **Ownership and Intellectual Property:** The Instructor shall not have intellectual property rights in any results arising from the Synthetic Dataset. Other intellectual property rights are in accordance with the CanPath [Intellectual Property Policy](#).
2. **Terms of Use:** The dataset is exclusively for the training and educational purposes outlined on the Synthetic Dataset application form approved by CanPath. Any changes to the use must be approved by CanPath. The course curriculum

must be updated to mention the “CanPath Synthetic Dataset” or “CanPath Student Dataset.”

3. **Access:** CanPath will provide the dataset to the instructor via secure file transfer. The instructor must limit access to the dataset to themselves and their students and notify CanPath when they no longer require access. The instructor shall remain responsible for their students’ actions and will ensure that students comply with the provisions of this acknowledgment. Instructors are encouraged to have their students review, sign and abide by the User Agreement (*Appendix A*). The dataset cannot be shared with third parties without CanPath’s written consent.
4. **Safeguards:** The instructor must use administrative, physical, and technological measures to protect the dataset. This includes using security software, encryption, firewalls, and secure servers. The instructor shall ensure that the dataset is only downloaded onto secure servers and not onto any personal devices. The instructor shall not attempt to combine or link the Synthetic Dataset to any other dataset to re-identify any person or for any other purpose.
5. **Publications:** The instructor shall ensure that no research-related publications are generated from the Synthetic Dataset. Requests for the review of pedagogical-related publications may be submitted to CanPath and assessed on a case-by-case basis. Please email apply@canpath.ca for further information.
6. **Reporting:** Upon completion of the course, the instructor shall complete the “[CanPath Synthetic Dataset Utilization Report](#),” which includes information on how the dataset was used in the course, the number of students who used it, and the title/abstract of any projects resulting from the dataset.
7. **Destruction:** Upon completion of the academic course and submission of the report or earlier, as directed by CanPath, the Instructor shall cease using the dataset and destroy any copies in their possession and shall confirm the destruction in writing to CanPath.
8. **Renewal:** The instructor must notify CanPath in writing to extend the use of the dataset for another academic term for the same course. If the instructor intends to use the dataset for a different course, a new Synthetic Dataset application must be submitted to CanPath.

If CanPath has concerns about the instructor’s compliance with the terms and conditions of this Acknowledgement, CanPath shall provide the instructor with written notice of such concerns and their reasons for them. The instructor shall, within five business days of receipt of the notice, investigate the matter and provide CanPath with a report stating the cause of the deficiency, if any, and the steps taken to prevent a recurrence, if required.

Appendix A

CanPath Synthetic Dataset User Agreement

As part of this course/project, you will have access to a dataset containing confidential information. By signing this agreement, you acknowledge and agree to the following terms:

- I will not share, distribute, or disclose the dataset or any portion of it to anyone outside of the authorized participants of this course.
- I will not create or store copies of the dataset on personal devices, laptops, or home computers.
- I will delete any local copies of the dataset after the course is completed or when directed by the instructor.
- I understand that I cannot publish the results of my analysis.
- I understand that I cannot present the results at conferences or other public forums.