Plan Overview

A Data Management Plan created using DMPonline

Title: Passenger Density on Stockholm Subway Platforms With and Without Blue-Light Exposure: A Quasi-experimental Observational Study

Creator: Jesper Alvarsson-Hjort

Principal Investigator: Gergö Hadlaczky

Data Manager: Jesper Alvarsson-Hjort

Project Administrator: Jesper Alvarsson-Hjort

Affiliation: Karolinska Institutet

Template: KIs Template for Swedish Research Council DMP

ORCID iD: 0000-0002-0556-6244

Project abstract:

This study is part of a larger project called BLISP and is a quasi-experimental observational field study examining how exposure to blue LED lighting (BLED) affects passenger density on subway platforms in the Stockholm metro system. The study focuses on four platforms at three stations, two equipped with BLED luminaires and two serving as non-exposed controls. On the exposed platforms, blue lighting is alternated on and off in two-week intervals over a total period between 15 September 2025 to 26 April 2026.

Passenger density is measured using automated video analytics applied to existing station security cameras. The system records minute-level counts of detected human presence within predefined spatial zones, including an exposure zone at the platform end where trains enter and an adjacent control zone under standard lighting. In addition, aggregated minute-level counts of entries and exits through ticket gates are collected to control for overall station usage, and environmental data on ambient daylight conditions are used as covariates.

The study aims to assess whether passenger density differs between exposed and non-exposed platform areas, whether any effects change over time due to habituation, and how observed patterns relate to overall crowding and outdoor light conditions. All data are aggregated and anonymised, with no personal or identifiable information collected.

ID: 203322

Start date: 30-11-2024

End date: 01-02-2027

Last modified: 04-05-2026

Grant number / URL: TRV 2024/29082, ID 2023-47

Copyright information:

Passenger Density on Stockholm Subway Platforms With and Without Blue-Light Exposure: A Quasi-experimental Observational Study

Description of data

How will data be collected, created or reused?

Data will be collected through automated recording systems already in use within the Stockholm metro system. Passenger density data will be generated using existing platform security cameras combined with video analytic software that automatically detects and counts entries into predefined spatial zones on subway platforms. Detections are aggregated at minute resolution and contain only anonymous count data indexed by time, station, platform, zone, and lighting condition.

In addition, aggregated minute-level data on passenger entries and exits will be obtained from station ticket gate systems managed by the public transport authority. These data serve as indicators of overall station usage and are used to contextualise and control for variations in passenger volume.

To account for ambient light conditions, environmental data on solar irradiance will be obtained from an external outdoor reference monitoring station. These measurements are time stamped and linked to the observational data via shared temporal identifiers.

No new personal data are created, and no identifiable information is collected or reused. All data are generated automatically without direct interaction with individuals and are provided to the researchers in pre-aggregated, anonymised tabular form. Data extraction and initial compilation are performed by the transport authority, after which the datasets are transferred securely to the research team for analysis.

What types of data will be created and/or collected, in terms of data format? Include version numbers if applicable.

The study will create and collect structured tabular data generated through automated systems operated by the Stockholm metro and an external environmental monitoring source. The core dataset consists of minute level count data derived from video analytic processing of existing platform security cameras. The tables contain numeric variables and textual identifiers for timestamp, station, platform and zone.

In addition, aggregated minute level station passage data from ticket gate systems are collected in tabular format, describing timestamped counts of passenger entries and exits per station.

Lastly solar irradiation data from SMHI open data will be obtained and are likewise provided as structured tabular files. These tables contain timestamped levels of, CIE UV irradiance, global horizontal irradiance, direct normal irradiance, direct horizontal irradiance, diffuse horizontal irradiance and photosynthetically active radiation (PAR)

All three data files will be stored in .csv format.

What volumes of data will be created and/or collected?

- < 100 GB

Data volumes are small likely below 50 MB

Documentation and data quality

How will the material be documented and described, with associated metadata relating to structure, standards and format for descriptions of the content, collection method, file naming-format-versioning, etc

The data will be stored long term in the Swedish National Data Service (SNDS) repository, where it will be preserved and made accessible in accordance with national standards for research data archiving and reuse. SNDS metadata records will provide high level descriptions of the dataset, including purpose, origin, temporal and geographic coverage, access conditions, and constraints.

The contents of the data files deposited at SNDS will be described in a dedicated README file accompanying the dataset. The README will document file structure, variable names and definitions, units, aggregation levels, collection methods, abbreviations, file naming conventions, and versioning. All analysis scripts necessary to reproduce the published results will be released together with the data, supporting transparency and reuse within the limits imposed by ethical approval and data access restrictions.

In parallel, the project will be continuously documented using the Karolinska Institutet Electronic Notebook (KI ELN), to ensure traceability, internal reproducibility, and compliance with institutional policies on research integrity and data management.

How will data quality be safeguarded and documented (for example repeated measurements, validation of data input, etc.)?

The data will be provided by AB Stockholms Lokaltrafik (SL). This includes aggregated passenger density registrations derived from automated analyses of existing platform camera systems, as well as aggregated counts of the number of passengers entering and exiting stations based on ticket gate systems. In both cases, the data are generated by SL as part of routine operations and delivered to the research project in pre-aggregated, anonymised tabular form. No raw video, images, or individual-level information are transferred to the research team.

Data quality will be ensured at Karolinska Institutet (KI) in accordance with the approved ethical application. Upon receipt, the data will be checked for completeness, internal consistency, and structural validity, including verification of timestamps, station and platform identifiers, spatial zones, and lighting condition labels. Implausible values, missing periods, or technical irregularities will be identified and documented before analysis. All quality control procedures, data exclusions, and processing steps will be documented in analysis scripts and project documentation, ensuring traceability and compliance with the ethical approval under which the study is conducted.

Storage and backup

How is storage and backup of data and metadata safeguarded during the research

process?

- ELN

During the research process, storage and backup of data and metadata will be safeguarded using Karolinska Institutet's approved IT infrastructure. Working datasets and associated metadata are stored on OneDrive, which is part of KI's centrally managed Microsoft 365 environment. This ensures automatic backup, version control, and controlled access in accordance with security policies.

Original data deliveries from AB Stockholms Lokaltrafik and fundamental updates to the datasets are documented in parallel in the KI Electronic Notebook (KI ELN). Data and metadata stored on OneDrive and in the KI ELN are subject to regular institutional backup routines managed by KI, ensuring protection against data loss and supporting traceability and reproducibility throughout the research lifecycle.

How is data security and controlled access to data safeguarded, in relation to the handling of sensitive data and personal data, for example?

No sensitive or personal data are collected in the project. Access to data prior to publication at SNDS is restricted to project members and designated KI superusers. Data and documentation are stored in KI managed systems such as OneDrive and the KI Electronic Notebook (KI ELN), with access controlled through individual authentication and role based permissions. All storage systems are covered by KI's standard security measures, including encrypted network traffic, regular backups, and audit trails, ensuring secure handling and traceability throughout the research process.

Legal and ethical aspects

How is data handling according to legal requirements safeguarded, e.g. in terms of handling of personal data, confidentiality and intellectual property rights?

The project does not involve personal data. However, if time stamped station passage data or minute level platform passenger density data are assessed by AB Stockholms Lokaltrafik as being overly detailed, which cannot be determined at present, these data will be aggregated to longer time intervals prior to publication. For reasons of traceability and reproducibility, the original unaggregated data will be retained and documented internally in the KI Electronic Notebook (ELN), but will not be made publicly available.

How is correct data handling according to ethical aspects safeguarded?

Ethical approval has been obtained (Dnr 2025-05468-01) and the project does not contain person data.

Accessibility and long-term storage

How, when and where will research data or information about data (metadata) be made

accessible? Are there any conditions, embargoes, licenses and limitations on the access to and reuse of data?

Research data and associated metadata will be made accessible through the Swedish National Data Service (SNDS). A preregistration will be published on SNDS prior to completion of data collection and analysis. This preregistration will include full metadata, detailed descriptions of the study design and planned analyses, and simulated data that mirror the structure of the final dataset, allowing assessment of analytic procedures without access to empirical data.

After publication of the main results in scientific journals, the final research data and complete metadata will be published as open access on SNDS. The data will be distributed with a standardized open license, allowing reuse for scientific purposes with appropriate attribution. Access conditions and any necessary limitations will be clearly stated in the SNDS metadata record.

In what way is long-term storage safeguarded, and by whom? How will the selection of data for long-term storage be made?

Long term storage of the data will be ensured through the KI Electronic Notebook (ELN), which will contain both raw and processed data. The data will be retained for at least ten years after publication. In addition, all scripts used for data processing and statistical analysis will be stored together with the corresponding datasets in both the ELN and the Swedish National Data Service (SNDS), ensuring traceability and reproducibility of the results.

Will specific systems, software, code or other types of services be necessary in order to open and use/analyse data in the long term?

No specific or proprietary systems are required to open or analyse the data in the long term. The data will be provided in open, non-proprietary CSV format, and all processing and analyses will be conducted using the open-source programming language R. Both the data files and the analysis scripts can therefore be accessed, used, and reanalysed with standard software tools that are freely available, ensuring long-term usability and reproducibility.

How will unique and persistent identifiers for the research data, such as a Digital Object Identifier (DOI), be obtained?

Unique and persistent identifiers for the research data will be provided by the Swedish National Data Service (SNDS). A DOI will first be assigned to the preregistered hypotheses and accompanying simulated datasets when these are published on SNDS. After publication of the scientific results, the empirical data will be added to the same DOI record, ensuring that the preregistration and the final dataset are linked within a single persistent identifier.

Responsibility and resources

Who is responsible for data management while the research project is in progress?

Data management is performed by the PI and a dedicated data manager in the research group with a PhD.

Who is responsible for data management, long-term storage after the research project has ended?

The Principal Investigator (PI) is responsible for overseeing data management throughout the project lifecycle and for ensuring that any necessary updates or amendments to the published datasets are made in coordination with SNDS. Responsibility for data management and long term storage is shared across roles and infrastructures. Karolinska Institutet (KI) is responsible for the long term storage of raw and processed data and project documentation maintained in the Electronic Notebook (ELN). The Swedish National Data Service (SNDS) is responsible for the long term preservation and access to the openly published research data and metadata.

What resources (costs, labour or other) will be required for data management (including storage, back-up, provision of access and processing for long-term storage)?

No additional resources, costs, or dedicated labour are required for data management. Storage, backup, access provision, and long term preservation are handled within centrally funded systems provided by Karolinska Institutet, including the KI Electronic Notebook, OneDrive, and the Swedish National Data Service. Data management activities are integrated into the regular research workflow and do not require project specific financial allocations.

What resources will be needed to ensure that data fulfil the FAIR principles?

Ensuring that the data fulfil the FAIR principles will primarily require support for data publication and metadata creation rather than additional technical or financial resources. Practical assistance will be sought from the Karolinska Institutet Library Data Access Unit to support preparation and upload of the datasets and metadata to the SND catalogue.

The datasets will be made findable by publication in a searchable data repository with metadata and DOI provided by SND. Accessibility will be ensured by linking the persistent identifier to the data and metadata. Interoperability will be supported through the use of open, non-proprietary file formats and established naming conventions and variables. Reusability will be ensured by openly providing, processing scripts, analytical scripts and data with a CC licence.