

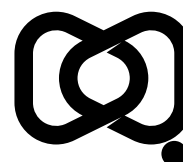


**Communicable
Diseases
Intelligence**

2026 • Volume 50 • cdc.gov.au/cdi • Epub date: 29.09.2026

Enhancing influenza surveillance in Australia using Google search data combined with machine learning: Supplementary information

Aminath Shausan, Adam G Dunn, Fiona J May, Satyamurthy Anuradha, Teresa M Wozniak, Amalie Dyda



**Australian
Centre for
Disease
Control**



Communicable Diseases Intelligence (CDI) is a peer-reviewed scientific journal published by the Australian Centre for Disease Control.

The journal aims to disseminate information on the epidemiology, surveillance, prevention and control of communicable diseases of relevance to Australia and the near region.

Editor

Dr Elise Firman

Deputy Editor

Simon Petrie

Design and Production

Lisa Thompson

Editorial Advisory Board

David Durrheim, Mark Ferson, Clare Huppatz, John Kaldor, Martyn Kirk and Meru Sheel

Submit an Article

Submit your next communicable disease related article to CDI for consideration.

Guidelines for authors and details on how to submit your publication is available on our website, or by email to the CDI Editor.

Contact us

Communicable Diseases Intelligence (CDI)
Australian Centre for Disease Control
PO Box 795, Canberra ACT 2601
Australia

Website: cdc.gov.au/cdi

Email: cdi.editor@cdc.gov.au

© 2026 Commonwealth of Australia as represented by the Australian Centre for Disease Control

ISSN: 2209-6051 Online

This journal is indexed by Index Medicus and Medline.

Creative Commons Licence

This publication is licensed under a Creative Commons Attribution-Non-Commercial-NoDerivatives 4.0 International Licence (Licence). You must read and understand the Licence before using any material from this publication.

Restrictions

The Licence does not cover, and there is no permission given for, use of any of the following material found in this publication (if any):

- the Commonwealth Coat of Arms (by way of information, the terms under which the Coat of Arms may be used can be found on the Department of Prime Minister and Cabinet website);
- any logos (including the Australian Centre for Disease Control's logo) and trademarks;
- any photographs and images;
- any signatures; and
- any material belonging to third parties.

Disclaimer

Opinions expressed in *Communicable Diseases Intelligence* are those of the authors and not necessarily those of the Australian Government or the Australian Centre for Disease Control. Data may be subject to revision.

Enquiries

Enquiries regarding any other use of this publication should be addressed to the CDI Editor.

Supplementary information: Summary of study design and execution



Version: 11-January-2024

Section/topic	Item	Development/ evaluation ^a	Checklist item	Reported on page
Title				
Title	1	D;E	Enhancing influenza surveillance in Australia using Google search data combined with machine learning	1
Abstract				
Abstract	2	D;E	See TRIPOD+AI for Abstracts checklist ^{S1,S2}	4
Introduction				
Background	3a	D;E	Explain the healthcare context (including whether diagnostic or prognostic) and rationale for developing or evaluating the prediction model, including references to existing models	5
	3b	D;E	Describe the target population and the intended purpose of the prediction model in the context of the care pathway, including its intended users (e.g. healthcare professionals, patients, public)	5
	3c	D;E	Describe any known health inequalities between sociodemographic groups	N/A
Objectives	4	D;E	Specify the study objectives, including whether the study describes the development or validation of a prediction model (or both)	5
Methods				
Data	5a	D;E	Describe the sources of data separately for the development and evaluation datasets (e.g., randomised trial, cohort, routine care or registry data), the rationale for using these data, and representativeness of the data	5
	5b	D;E	Specify the dates of the collected participant data, including start and end of participant accrual; and, if applicable, end of follow-up	5
Participants	6a	D;E	Specify key elements of the study setting (e.g. primary care, secondary care, general population) including the number and location of centres	5
	6b	D;E	Describe the eligibility criteria for study participants	5
	6c	D;E	Give details of any treatments received, and how they were handled during model development or evaluation, if relevant	N/A
Data preparation	7	D;E	Describe any data pre-processing and quality checking, including whether this was similar across relevant sociodemographic groups	6–8

Section/topic	Item	Development/ evaluation ^a	Checklist item	Reported on page
Outcome	8a	D;E	Clearly define the outcome that is being predicted and the time horizon, including how and when assessed, the rationale for choosing this outcome, and whether the method of outcome assessment is consistent across sociodemographic groups	6
	8b	D;E	If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of the outcome assessors	N/A
	8c	D;E	Report any actions to blind assessment of the outcome to be predicted	N/A
Predictors	9a	D	Describe the choice of initial predictors (e.g., literature, previous models, all available predictors), and any pre-selection of predictors before model building	5
	9b	D;E	Clearly define all predictors, including how and when they were measured (and any actions to blind assessment of predictors for the outcome and other predictors)	5–6
	9c	D;E	If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of the predictor assessors	N/A
Sample size	10	D;E	Explain how the study size was arrived at (separately for development and evaluation), and justify that the study size was sufficient to answer the research question. Include details of any sample size calculation	5
Missing data	11	D;E	Describe how missing data were handled. Provide reasons for omitting any data	6
Analytical methods	12a	D	Describe how the data were used (e.g., for development and evaluation of model performance) in the analysis, including whether the data were partitioned, considering any sample size requirements	6–8
	12b	D	Depending on the type of model, describe how predictors were handled in the analyses (functional form, rescaling, transformation, or any standardisation)	7
	12c	D	Specify the type of model, rationale ^b , all model-building steps, including any hyperparameter tuning, and method for internal validation	7
	12d	D;E	Describe if and how any heterogeneity in estimates of model parameter values and model performance was handled and quantified across clusters (e.g., hospitals, countries). See TRIPOD-Cluster for additional considerations ^c	N/A
	12e	D;E	Specify all measures and plots used (and their rationale) to evaluate model performance (e.g., discrimination, calibration, clinical utility) and, if relevant, to compare multiple models	8
	12f	E	Describe any model updating (e.g, recalibration) arising from the model evaluation, either overall or for particular sociodemographic groups or settings	N/A

Section/topic	Item	Development/ evaluation ^a	Checklist item	Reported on page
	12g	E	For model evaluation, describe how the model predictions were calculated (e.g., formula, code, object, application programming interface)	8
Class imbalance	13	D;E	If class imbalance methods were used, state why and how this was done, and any subsequent methods to recalibrate the model or the model predictions	N/A
Fairness	14	D;E	Describe any approaches that were used to address model fairness and their rationale	8
Model output	15	D	Specify the output of the prediction model (e.g., probabilities, classification). Provide details and rationale for any classification and how the thresholds were identified	8
Training versus evaluation	16	D;E	Identify any differences between the development and evaluation data in the healthcare setting, eligibility criteria, outcome, and predictors	12
Ethical approval	17	D;E	Name the institutional research board or ethics committee that approved the study and describe the participant-informed consent or the ethics committee waiver of informed consent	13
Open science				
Funding	18a	D;E	Give the source of funding and the role of the funders for the present study	13
Conflicts of interest	18b	D;E	Declare any conflicts of interest and financial disclosures for all authors	13
Protocol	18c	D;E	Indicate where the study protocol can be accessed or state that a protocol was not prepared	N/A
Registration	18d	D;E	Provide registration information for the study, including register name and registration number, or state that the study was not registered	N/A
Data sharing	18e	D;E	Provide details of the availability of the study data	13
Code sharing	18f	D;E	Provide details of the availability of the analytical code ^d	13
Patient and public involvement				
Patient and public involvement	19	D;E	Provide details of any patient and public involvement during the design, conduct, reporting, interpretation, or dissemination of the study or state no involvement	No involvement
Results				
Participants	20a	D;E	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful	N/A
	20b	D;E	Report the characteristics overall and, where applicable, for each data source or setting, including the key dates, key predictors (including demographics), treatments received, sample size, number of outcome events, follow-up time, and amount of missing data. A table may be helpful. Report any differences across key demographic groups	5

Section/topic	Item	Development/ evaluation ^a	Checklist item	Reported on page
	20c	E	For model evaluation, show a comparison with the development data of the distribution of important predictors (demographics, predictors, and outcome)	N/A
Model development	21	D;E	Specify the number of participants and outcome events in each analysis (e.g., for model development, hyperparameter tuning, model evaluation)	N/A
Model specification	22	D	Provide details of the full prediction model (e.g, formula, code, object, application programming interface) to allow predictions in new individuals and to enable third-party evaluation and implementation, including any restrictions to access or re-use (e.g., freely available, proprietary) ^e	7
Model performance	23a	D;E	Report model performance estimates with confidence intervals, including for any key subgroups (e.g., sociodemographic). Consider plots to aid presentation	10–11
	23b	D;E	If examined, report results of any heterogeneity in model performance across clusters. See TRIPOD Cluster for additional details ^c	N/A
Model updating	24	E	Report the results from any model updating, including the updated model and subsequent performance	N/A
Discussion				
Interpretation	25	D;E	Give an overall interpretation of the main results, including issues of fairness in the context of the objectives and previous studies	12–13
Limitations	26	D;E	Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, missing data) and their effects on any biases, statistical uncertainty, and generalisability	13
Usability of the model in the context of current care	27a	D	Describe how poor quality or unavailable input data (e.g., predictor values) should be assessed and handled when implementing the prediction model	13
	27b	D	Specify whether users will be required to interact in the handling of the input data or use of the model, and what level of expertise is required of users	N/A
	27c	D;E	Discuss any next steps for future research, with a specific view to applicability and generalisability of the model	13

a D: items relevant only to the development of a prediction model; E: items relating solely to the evaluation of a prediction model; D;E: items applicable to both the development and evaluation of a prediction model.

b Separately for all model building approaches.

c TRIPOD-Cluster is a checklist of reporting recommendations for studies developing or validating models that explicitly account for clustering or explore heterogeneity in model performance (e.g., at different hospitals or centres).^{S1}

d This relates to the analysis code, for example, any data cleaning, feature engineering, model building, evaluation.

e This relates to the code to implement the model to get estimates of risk for a new individual.

References

- S1. Debray TPA, Collins GS, Riley RD, Snell KIE, Van Calster B, Reitsma JB et al. Transparent reporting of multivariable prediction models developed or validated using clustered data: TRIPOD-Cluster checklist. *BMJ*. 2023;380:071018. doi: <https://doi.org/10.1136/bmj-2022-071018>.
- S2. Collins GS, Moons KGM, Dhiman P, Riley RD, Bearn AL, Van Calster B et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:078378. doi: <https://doi.org/10.1136/bmj-2023-078378>.