Title	Epidemiological alchemy in the AI era comes under close scrutiny—but is it entirely without meaning?
Author(s)	Ide, Kazuki; Nakayama, Takeo
Citation	Journal of Epidemiology. 2025, p. JE20250602
Version Type	AM
URL	https://hdl.handle.net/11094/103364
rights	This article is licensed under a Creative Commons Attribution 4.0 International License.
Note	

1   *Journal of Epidemiology*

2   *Letters to the Editor*

3

4   **Epidemiological alchemy in the AI era comes under close scrutiny—but is it entirely**

5   **without meaning?**

6

7   Kazuki Ide[1,2*], Takeo Nakayama[3]

8   [1]Division of Scientific Information and Public Policy, Center for Infectious Disease Education

9   and Research (CiDER), The University of Osaka, Osaka, Japan; [2]Research Center on Ethical,

10   Legal and Social Issues, The University of Osaka, Osaka, Japan; [3]Department of Health

11   Informatics, Graduate School of Medicine/School of Public Health, Kyoto University, Kyoto,

12   Japan

13

14   *Address for correspondence. Kazuki Ide, CiDER, The University of Osaka, Osaka, Japan (ide-*

15   *k@cider.osaka-u.ac.jp, ide.kazuki.2r@kyoto-u.jp; Tel +81-6-6879-4168)*

16

MAIN TEXT

17

18   Epidemiological open data obtained from the National Health and Nutrition

19   Examination Survey (NHANES), conducted in the United States, have provided

20   valuable opportunities for exploratory research [1]. What once embodied the public

21   good may now, in the artificial intelligence (AI) era, be revealing the seductive sheen of

22   a "forbidden fruit."

23   Suchak T et al. reported an increase in formulaic research articles using this AI-

24   ready dataset (four articles per year from 2014 to 2021, but 190 in 2024 alone as of

25   October 9), among which a considerable proportion were of questionable quality [2].

26   They suggested that this increase was partly driven by AI-assisted workflows and paper

27   mills.

28   After this problematic trend was reported in *Science*, two major publishers—

29   PLOS and Frontiers—announced that they would no longer consider studies based

30   solely on this dataset without additional research [3]. This means that authors are now

31   required to include experiments demonstrating robustness or to incorporate additional

32   data from their own institutions to show the unique value of the study. The Editor-in-

33   Chief of PLoS One stated that "*the rejection rate for such papers has increased from*

34   *40% to 94%*," although it remains difficult to judge whether this change is desirable. A

35 94% rejection rate seems high. We argue that frequent reuse of a dataset tends to signal

36 FAIR (Findable, Accessible, Interoperable, Reusable)-aligned quality. In this context, Dr

37 Mons has underscored the importance of "Fully AI-Ready" data [4]. We should

38 therefore remain mindful of the risk of becoming unduly restrictive or overly

39 exclusionary toward exploratory, early-stage work. Not limited to NHANES, the journal

40 *Globalization and Health* has also discussed issues in Global Burden of Disease (GBD)

41 studies, which may stem from comparable problems [5].

42   In Japan, collaborative studies such as the Japan COVID-19 and Society Internet

43 Survey (JACSIS) and the Japan Society and New Tobacco Internet Survey (JASTIS)

44 have conducted multiple waves of data collection [6, 7]. These groups regularly invite

45 collaborators to use their data, although the datasets are not openly available. Their

46 controlled-access framework represents one possible solution for preventing

47 inappropriate use of data, provided it safeguards a minimum level of quality while

48 remaining open to diverse research. A similar approach is used by the Japan Multi-

49 Institutional Collaborative Cohort (J-MICC) Study, which provides open calls for

50 research proposals to researchers awarded Grants-in-Aid for Scientific Research

51 (KAKENHI) [8]. The initiative is also supported by the Platform of Supporting Cohort

52 Study and Biospecimen Analysis (CoBiA), which includes additional resources [9].

53   Another approach is to require authors to post a preprint upon submission, as

54   previously implemented by *eLife*. Such a policy may function as a deterrent against

55   inappropriate use of data and the submission of low-quality manuscripts, since authors'

56   names remain publicly associated with their work even if it is ultimately rejected. That

57   said, its impact might be limited, given that the generalist server hosted on *OSF*

58   *Preprints* released a blog post on October 16, 2025, announcing its decision to

59   indefinitely suspend operations, citing concerns over the rising volumes of AI-generated

60   content and paper-mill activities as one reason for the suspension [10].

61   Furthermore, discerning whether insufficient quality stems from malicious intent

62   or a mere lack of rigor remains a challenge. Care must be taken not to conflate the level

63   of expertise or methodological rigor with malicious intent, as such assumptions risk

64   stifling emerging research and discouraging ambitious researchers. Importantly, health-

65   related data and the findings derived from them may influence both individual and

66   population-level behavior.

67   We are living in an age of openness, with this trend being advanced at both

68   national and international levels [11]. It may be time for us not merely to advance

69   openness, but to proactively rethink its meaning and form—an invitation posed by the

70   very phenomena we are now witnessing.

Ethical Approval

This commentary is based exclusively on information from public sources that contain

no personal information; therefore, ethical approval was not required.

Data Availability Statement

All data are incorporated into the article.

104 **References**

105 1. Centers for Disease Control and Prevention (CDC). National Health and Nutrition

106   Examination Survey (NHANES).

107   https://www.cdc.gov/nchs/nhanes/about/index.html; Accessed 11.05.2025.

108 2. Suchak T, Aliu AE, Harrison C et al. Explosion of formulaic research articles,

109   including inappropriate study designs and false discoveries, based on the NHANES

110   US national health database. *PLoS Biol*. 2025;23(5): e3003152. doi:

111   10.1371/journal.pbio.3003152

112 3. O'Grady C. Journals and publishers crack down on research from open health data

113   sets. *Science*. 2025, *in press*. doi: 10.1126/science.z6eukgo

114 4. Farewell address Barend Mons. https://www.gofair.foundation/fairwell; Accessed

115   11.11.2025.

116 5. Globalization and Health. Aims and Scope.

117   https://globalizationandhealth.biomedcentral.com/about; Accessed 11.11.2025.

118 6. Tabuchi T, Shinozaki T, Kunugita N et al. Study profile: The Japan "Society and

119   New Tobacco" Internet Survey (JASTIS): A longitudinal internet cohort study of

120   heat-not-burn tobacco products, electronic cigarettes, and conventional tobacco

121   products in Japan. *J Epidemiol*. 2019;29:444-50. doi: 10.2188/jea.JE20180116

122   7.   The Japan COVID-19 and Society Internet Survey (JACSIS Study). https://jacsis-

123   study.jp/index.html; Accessed 11.05.2025.

124   8.   J-MICC STUDY. https://jmicc.com/; Accessed 11.11.2025. [In Japanese]

125   9.   Platform of Supporting Cohort Study and Biospecimen Analysis.

126   https://square.umin.ac.jp/cohort/english/index.html; Accessed 11.11.2025.

127   10. Center for Open Science (COS). OSF update on the future of OSF Preprints.

128   https://www.cos.io/blog/update-on-future-of-osf-preprints; Accessed 10.29.2025.

129   11. Ide K. Japan can embrace open science — but flexible approaches are key. *Nature*.

130   2024;629:286. doi: 10.1038/d41586-024-01340-w