**Online Resource 1.** Flow chart.

| Study | Included in the study | Information on breastfeeding. Dataset for analysis. | Information on duration of breastfeeding. HLA available. Dataset for analysis. |
|-------|-----------------------|-----------------------------------------------------|------------------------------------------------------------------|
| EIMS  | Cases                 | 2880                                                | 2677                                                             | 1304                                                             |
|       | Controls              | 6122                                                | 5734                                                             | 1805                                                             |
| KPNC  | Cases                 | 1163                                                | 993                                                              | 659                                                              |
|       | Controls              | 1178                                                | 1003                                                             | 537                                                              |

EIMS, Epidemiologic Investigation of Multiple Sclerosis; KPNC, Kaiser Permanente Northern California.
Online Resource 2. Variable descriptions

EIMS; Age was categorized into the following 8 strata: <20, 20-24, 25-29, 30-34, 35-39, 40-45, 45-49, and >50 years of age. Assessment of ancestry was based on whether the subject was born in Sweden or not, and whether either of the subject’s parents had immigrated to Sweden. A subject who was born in Sweden and whose parents had not immigrated, was classified as Swedish. Family history of MS was defined as having a first or second degree relative with MS. Alcohol consumption at inclusion in the study was categorized into no consumption, low consumption (<50 g/week for women and <100 g/week for men), moderate consumption (50-112 g/week for women and 100-168 g/week for men) or high consumption (>112 g/week for women and >168 g/week for men). The cutoffs were the same as those used by Statistics Sweden (http://www.scb.se). Based on three questions regarding exposure to UVR where each answer alternative was given a number ranging from 1 (the lowest exposure) to 4 (the highest exposure), we constructed an index by adding the responses together and thus acquired a value between 3 and 12. Educational level was categorized into no secondary education, post secondary education without university degree, or university degree. The last occupation during the year before the index year was used as a marker for socioeconomic class which was categorized into the following strata: 1, workers in goods production; 2, workers in service production; 3, employees at lower and intermediate levels; 4, employees at higher levels, executives, university graduates, and 5; others such as pensioners, students, and unemployed.
EIMS and KPNC; For each case, the time at the initial appearance of symptoms indicative of MS was used as an estimate of the disease onset, and the year in which this occurred was defined as the index year. Smoking habits were only considered before the index year, and was adjusted for by using a continuous variable (number of pack years of smoking). One pack years is equivalent to smoking 20 cigarettes daily for one year. Passive smoking was dichotomized into ever or never exposed before the index year. A history of infectious mononucleosis was dichotomized into yes, no, or unknown. Body mass index (BMI) at age 20 years was calculated by dividing self-reported weight in kilograms by self-reported height in meters squared and adolescent BMI was categorized into underweight (<18.5 kg/m²), normal weight (18.5-25 kg/m²), overweight (25-30 kg/m²) and obese (>30 kg/m²).
Online Resource 3

Mediation analysis

We evaluated evidence that any observed non-null association between prolonged breastfeeding and MS was mediated by being overweight during adolescence. Cases and controls were classified as being overweight during adolescence if their BMI was >25 during adolescence. The natural direct (NDE) and indirect effects (NIE) of prolonged breastfeeding on MS risk were estimated using both Baron-Kenney and causal mediation methods among EIMS and KPNC cases and controls. Additionally, we tested for interaction between reduced breastfeeding and BMI>25 and association with MS.

Missingness Bias analysis

Inverse probability of missing (IPM) weights was used to investigate the effects of 1) exclusion of study participants without breastfeeding data, and 2) grouping participants who did not know duration into the unknown duration category of breastfeeding on observed associations. The IPM weights, \( w_i \), were defined as the inverse of the conditional probability of the \( i \)th participant’s breastfeeding status being observed, \( w_i = Pr(O_i | Z_j)^{-1} \), where \( O \) was an indicator of observing breastfeeding status and \( Z \) were covariates. Logistic regression and conditional logistic regression were used to estimate \( Pr(O_i | Z_j) \) and subsequent IPM weighted breastfeeding associations among EIMS and KPNC participants, respectively.

Misclassification Bias Analysis
Due to the retrospective nature of the study, and length of time between the exposure of interest and study enrollment, we also investigated the potential influence of differential misclassification of prolonged breastfeeding exposure among males. Record level probabilistic bias analysis was conducted to assess effect of possible misclassification of prolonged pregnancy exposure among males on observed protective associations. Beta distribution priors were used to model ranges of positive predictive value (PPV) and negative predictive value (NPV) proportions. The record level correction algorithm is presented below:

1. Bootstrap sample male subjects.

2. Randomly sample PPV and NPV proportions from Beta distributions.

3. Identify number of false positive (FP) and false negative (FN) observations by multiplying number prolonged breastfeeding exposed and unexposed male cases by sampled 1-PPV and 1-NPV proportions, respectively.

4. Randomly sample number of FP and FN cases identified in previous step and “correct” their exposure status to unexposed and exposed, respectively.

5. Estimate odds ratio of bias-corrected prolonged breastfeeding exposure among males using logistic regression (EIMS) and conditional regression (KPNC).

6. Iterate steps 1-5 10,000 times.
7. Estimate odds ratios and 95% confidence intervals from mean and 2.5% and 97.5% quantiles of empirical distribution of odds ratios estimated in each bootstrap iteration.

1. VanderWeele TJ, Vansteelandt S. Odds Ratios for Mediation Analysis for a Dichotomous Outcome. Am. J. Epidemiol. 2010;172(12):1339–1348.

2. Greenland S, Lash TL. Bias analysis. In: Rothman KJ, Greenland S, Lash TL, editors. Modern Epidemiology. 3. Philadelphia: Lippincott-Williams-Wilkins; 2008. pp. 345–380.
Online Resource 4. a) Interaction between breastfeeding <4 months and being overweight (BMI>25) among males in EIMS and KPNC datasets, and b) Total, natural direct, and natural indirect effects of breastfeeding >4 months mediated by being not being overweight or obese among males in EIMS and KPNC datasets.

a)

|                  | Breastfeeding <4 mo. & BMI<25 | Breastfeeding >4 mo. & BMI>25 | Breastfeeding <4 mo. & BMI>25 | RERI^c |
|------------------|-------------------------------|------------------------------|-------------------------------|--------|
|                  | OR_{10} 95% CI                 | OR_{01} 95% CI               | OR_{11} 95% CI               | OR 95% CI |
| **EIMS^a**       |                               |                              |                               |        |
| Unadj.           | 1.4 1.0, 1.9                   | 1.5 1.0, 2.1                 | 1.1 0.6, 2.1                 | 0.3 -0.9, 1.5 |
| Adj.             | 1.4 1.0, 1.9                   | 1.4 0.9, 2.0                 | 1.1 0.6, 2.1                 | 0.3 -0.9, 1.5 |
| **KPNC^b**       |                               |                              |                               |        |
| Unadj.           | 2.0 0.8, 4.8                   | 4.0 0.7, 22.6                | 0.3 0.1, 1.8                 | -1.4 -4.9, 2.0 |
| Adj.             | 2.5 0.8, 7.3                   | 3.8 0.6, 24.8                | 0.3 0.1, 2.4                 | -2.5 -9.9, 5.0 |

Abbreviations: BMI, body mass index; OR, odds ratio; CI, confidence interval, mo., months; RERI, relative excess risk due to interaction; EIMS, Epidemiologic Investigation of Multiple Sclerosis; KPNC, Kaiser Permanente Northern California. The reference group comprises those with BMI<25kg/m² who were breastfed for 4 months or longer.

^a Adjusted for Swedish ancestry, smoking pack years, and age

^b Adjusted for pack years, and inverse probability weights for BMI data missingness.

^c RERI estimated from ORs from logistic regressions with an interaction term for Breastfeeding <4 mo. and BMI>25, RERI = OR_{11} - OR_{10} - OR_{01} + 1 (Van Der Weele TJ, Knol MJ. A tutorial on interaction. Epidemiol Method. Walter de Gruyter GmbH; 2014;3:33–72).
b)

|                          | Breastfeeding >4 months<sup>a</sup> | Baron and Kenney          | Causal Mediation Analysis<sup>b</sup> |
|--------------------------|-------------------------------------|---------------------------|--------------------------------------|
|                          | OR  | 95% CI<sup>c</sup> | OR  | 95% CI<sup>c</sup> | OR  | 95% CI<sup>c</sup> |
| **EIMS**<sup>e</sup> (N=1,167) |     |                   |     |                   |     |                   |
| Total Effect             | 0.7 | 0.5, 0.9          | 0.7 | 0.5, 0.9          | 0.7 | 0.5, 0.9          |
| Natural Direct Effect    | 0.7 | 0.5, 0.9          | 0.7 | 0.5, 0.9          | 0.7 | 0.5, 0.9          |
| Natural Indirect Effect  | 0.99| 0.96, 1.01        | 0.98| 0.94, 1.00        |     |                   |
| **KPNC**<sup>d</sup> (N=166) |     |                   |     |                   |     |                   |
| Total Effect             | 0.5 | 0.2, 1.3          | 0.5 | 0.2, 1.4          | 0.5 | 0.2, 1.5          |
| Natural Direct Effect    | 0.5 | 0.2, 1.3          | 0.5 | 0.2, 1.5          | 0.5 | 0.2, 1.5          |
| Natural Indirect Effect  | 1.0 | 0.9, 1.2          | 0.99| 0.8, 1.1          |     |                   |

Abbreviations: OR, odds ratio; 95% CI, 95% confidence interval; EIMS, Epidemiologic Investigation of Multiple Sclerosis; KPNC, Kaiser Permanente Northern California.

<sup>a</sup>Breastfeeding >4 compared to <4 months among males.

<sup>b</sup>Assumed no interaction between not being overweight or obese and breastfeeding.

<sup>c</sup>Adjusted for Swedish ancestry, smoking packyears, and age.

<sup>d</sup>Adjusted for packyears.

<sup>e</sup>95% CIs obtained using nonparametric bootstrap.
Online Resource 5. Inverse probability of missing (IPM) weighted logistic regression results.

|        | EIMS |                   | KPNC\(^1\) |                   |
|--------|------|-------------------|------------|-------------------|
|        | OR (95% CI) | Excluded\(^2\) | Unknown Duration\(^3\) | OR (95% CI) | Excluded\(^2\) | Unknown Duration\(^3\) |
| All    |                   |                 |               |                   |                 |               |
| >4 months | 1.0 (0.9, 2.2) | 0.9 (0.8, 1.1) | 1.0 (0.9, 1.1) | 0.6 (0.4, 0.9) | 0.6 (0.5, 0.8) | 0.6 (0.4, 0.8) |
| Unknown duration | 0.8 (0.7, 0.9) | 0.8 (0.7, 1.0) | 1.3 (0.9, 2.8) | 0.9 (0.7, 1.2) |               |               |
| Women  |                   |                 |               |                   |                 |               |
| >4 months | 1.0 (0.9, 2.2) | 1.0 (0.9, 1.2) | 1.0 (0.9, 1.1) | 0.7 (0.5, 2.0) | 1.0 (0.7, 1.4) | 0.8 (0.5, 1.3) |
| Unknown duration | 0.8 (0.7, 2.0) | 0.9 (0.8, 1.0) | 1.5 (1.0, 2.1) | 0.8 (0.6, 1.2) |               |               |
| Men    |                   |                 |               |                   |                 |               |
| >4 months | 0.7 (0.6, 0.9) | 0.7 (0.6, 0.9) | 0.7 (0.5, 0.9) | 0.3 (0.1, 0.9) | 0.4 (0.1, 1.4) | 0.5 (0.1, 2.0) |
| Unknown duration | 0.7 (0.6, 0.9) | 0.7 (0.6, 0.9) | 0.8 (0.4, 2.6) | 1.0 (0.5, 2.0) |               |               |

Abbreviations: OR, odds ratio; 95% CI, 95% confidence interval; EIMS, Epidemiologic Investigation of Multiple Sclerosis; KPNC, Kaiser Permanente Northern California.

\(^1\) Conditional logistic regression was used for KPNC analyses.
2 Missing analysis of those who were excluded because of missing breastfeeding data.

3 Missing analysis of those who were missing breastfeeding duration.
**Online Resource 6.** Bias-corrected OR with 95% CI of developing MS among male subjects under different assumed rates of false positive and false negative prolonged breastfeeding exposure.

| Case PPV | Case NPV | Control PPV | Control NPV | EIMS (95% CI) | KPNC (95% CI) |
|----------|----------|-------------|-------------|---------------|---------------|
| Observed |          |             |             | 0.7 (0.6, 0.9) | 0.3 (0.1, 0.9) |
| 0.85 - 1 | 0.7 - 1  | 0           | 0           | 0.7 (0.6, 1.0) | 0.2 (0.1, 1.0) |
| 0.9 - 1  | 0.8 - 1  | 0           | 0           | 0.7 (0.6, 1.0) | 0.3 (0.1, 1.0) |
| 0.95 - 1 | 0.9 - 1  | 0           | 0           | 0.7 (0.6, 1.0) | 0.3 (0.1, 1.0) |
| 0.95 - 1 | 0.7 - 1  | 0           | 0           | 0.8 (0.6, 1.0) | 0.3 (0.1, 1.0) |
| 0.95 - 1 | 0.8 - 1  | 0           | 0           | 0.8 (0.6, 1.0) | 0.3 (0.1, 1.0) |
| 0.85 - 1 | 0.7 - 1  | 0.95 - 1    | 0.9 - 1     | 0.8 (0.6, 1.0) | 0.3 (0.1, 1.0) |
| 0.85 - 1 | 0.7 - 1  | 0.9 - 1     | 0.8 - 1     | 0.7 (0.6, 1.0) | 0.3 (0.1, 1.0) |
| 0.9 - 1  | 0.8 - 1  | 0.85 - 1    | 0.7 - 1     | 0.7 (0.5, 0.9) | 0.3 (0.1, 1.0) |
| 0.95 - 1 | 0.9 - 1  | 0.85 - 1    | 0.9 - 1     | 0.8 (0.6, 1.1) | 0.3 (0.1, 1.0) |
| 0.95 - 1 | 0.7 - 1  | 0.85 - 1    | 0.8 - 1     | 0.7 (0.6, 1.0) | 0.3 (0.1, 1.0) |

Abbreviations: OR, odds ratio; 95% CI, 95% confidence interval; EIMS, Epidemiologic Investigation of Multiple Sclerosis; KPNC, Kaiser Permanente Northern California.

1 Bias-corrected ORs and CIs estimated from mean and 2.5% and 97.5% quantiles of the bootstrap distributions.

2 Conditional logistic regression was used for KPNC analyses.