Article

Milk intake and risk of mortality risk in the Japan Collaborative Cohort Study - a Bayesian survival analysis

Chaochen Wang[1,*], Hiroshi Yatsuya[2], Yingsong Lin[1], Tae Sasakabe[1], Sayo Kawai[1], Shogo Kikuchi[1], Hiroyasu Iso[3], Akiko Tamakoshi[4]

1 Department of Public Health, Aichi Medical University School of Medicine, Nagakute, Japan;
2 Departmet of Public Health, Fujita Health University School of Medicine, Toyoake, Japan;
3 Public Health, Department of Social Medicine, Osaka University Graduate School of Medicine, Osaka, Japan;
4 Department of Public Health, Faculty of Medicine, Hokkaido University, Sapporo, Japan;
* Correspondence: Email.: chaochen@wangcc.me; Tel.: +81-561-62-3311. Department of Public Health, Aichi Medical University School of Medicine, 1-1 Yazakokarimata, Nagakute, Aichi, 480-1195, Japan (C.W.)

Version July 7, 2020 submitted to Nutrients

Abstract: A single paragraph of about 200 words maximum. For research articles, abstracts should give a pertinent overview of the work. We strongly encourage authors to use the following style of structured abstracts, but without headings: 1) Background: Place the question addressed in a broad context and highlight the purpose of the study; 2) Methods: Describe briefly the main methods or treatments applied; 3) Results: Summarize the article's main findings; and 4) Conclusion: Indicate the main conclusions or interpretations. The abstract should be an objective representation of the article, it must not contain results which are not presented and substantiated in the main text and should not exaggerate the main conclusions.

Keywords: keyword 1; keyword 2; keyword 3 (list three to ten pertinent keywords specific to the article, yet reasonably common within the subject discipline.).

1. Introduction

2. Materials and Methods

2.1. The database

We used data from the Japan Collaborative Cohort (JACC) study, which was sponsored by the Ministry of Education, Sports, Science, and Technology of Japan. Sampling methods and details about the JACC study have been described extensively in the literature [1–3]. Participants of the JACC study completed self-administered questionnaires about their lifestyles, food intake (food frequency questionnaire, FFQ), and medical histories of cardiovascular disease or cancer. In the final follow-up of the JACC study, data from a total of 110585 individuals (46395 men and 64190 women) were successfully retained for the current analysis. We further excluded samples if they meet one of the following criteria: 1) with any disease history of stroke, cancer, myocardial infarction, ischemic heart disease, or other types heart disease (n = 6655, 2931 men and 3724 women); 2) did not answer the question regarding their milk consumption in the baseline FFQ survey (n = 9545, 3593 men and 5952 women). Finally, 94385 (39386 men and 54999 women) are left in the database. The study design and informed consent procedure were approved by the Ethics Review Committee of Nagoya University School of Medicine.

2.2. Exposure and the outcome of interest

Frequency of milk intake during the preceding year of the baseline was assessed by FFQ from "never", "1-2 times/month", "1-2 times/week", "3-4 times/week", and "Almost daily". The exact amount of milk consumption was difficult to assess here. However, good reproducibility and validity were confirmed previously (Spearman rank correlation coefficient between milk intake frequency and weighed dietary record for 12 days was 0.65) [4].

The causes and date of death were obtained from death certificates and were systematically reviewed. The follow-up period was defined as from the time of the baseline survey was completed, which was between 1988-1990, until the end of 2009 (administrative censor), or the date when move-out of study area, or the date of death from stroke recorded, whichever occurred first. Other causes of death were treated as censored and assumed not informative. The causes of death were coded by the 10th Revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10), therefore stroke was defined as I60-I69. We further classified these deaths into hemorrhagic stroke (I60, I61 and I62) or cerebral infarction (I63) when subtypes of stroke in their death certificates were available.

2.3. Statistical approach

We calculated sex-specific means (standard deviation, sd) and proportion of selected baseline characteristics according to the frequency of milk intake. Overall difference across the milk intake groups were tested by either analysis of variance for continuous variables or χ^2 test for categorical variables.

Full parametric proportional hazard models under Bayesian framework with Weibull distribution were fitted using Just Another Gibbs Sampler (JAGS) program [5] version 4.3.0 in R version 4.0.1 [6]. JAGS program is similar to the OpenBUGS [7] project that uses a Gibbs sampling engine for Markov Chain Monte Carlo (MCMC) simulation. In the current analysis, we specified non-informative prior distributions for each of the parameters in our models ($\beta_n \sim N(0, 1000)$, and $\kappa_{\text{shape}} \sim \Gamma(0.001, 0.001)$). The Brooks-Gelman-Rubin diagnostic [8] was used to refine the approximate point of convergence, the point when the ratio of the chains is stable around 1 and the within and between chain variability start to reach stability was visually checked. The auto-correlation tool further identified if convergence has been achieved or if a high degree of auto-correlation exists in the sample. Then, the number of iterations discarded as 'burn-in' was chosen. All models had a posterior sample size of 100000 from three separated chains with a "burn-in" of 2500 iterations. Posterior means (sd) and 95% Credible Intervals (CrI) of the estimated HR were presented for each category of milk intake frequency taking the "never" category as the reference. Posterior probability that the estimated hazard of dying from stroke for the milk intake for frequency that higher or equal to "1-2 times/month" is smaller compared with those who chose "never" to their milk intake frequency were calculated as $P(\text{HR} < 1)$.

The parametric forms of the models fitted in the Bayesian survival analyses include three models: 1) the crude model, 2) the age-centered adjusted model, 3) and a model further adjusted for potential confounders which includes: age (centered, continuous), smoking habit (never, current, former), alcohol intake (never or past, $<$ 4 times/week, Daily), body mass index ($<$ 18.5, $\geq$ 18.5 and $<$ 25, $\geq$ 25 and $<$ 30, $\geq$ 30 kg/m^2), history of hypertension, diabetes, kidney/liver diseases (yes/no), exercise (more than 1 hour/week, yes/no), sleep duration ($<$ 7, $\geq$ 7 and $<$ 8, $\geq$ 8 and $<$ 9, $\geq$ 9, hours), coffee intake (never, $<$ 3-4 times/week, almost daily), education level (attended school till age 18, yes/no)

3. Results

The total follow-up was 1555073 person-years (median = 19.3 years), during which 2675 death from stroke was confirmed (1352 men and 1323 women). Among these stroke mortality, 952 were hemorrhagic stroke (432 men and 520 women), and 957 were cerebral infarction (520 men and 437 women).

Table 1. Sex-specific baseline characteristics according to the frequency of milk intake (JACC study, 1988-2009).

	Never	Drinker	Milk drinkers				*P* value
			1-2 times/ Month	1-2 times/ Week	3-4 times/ Week	Almost Daily	
Men (n = 39386)							
number of subjects	8508	30878	3522	5928	5563	15865	
Age, year (mean (SD))	56.8 (9.9)	56.8 (10.2)	55.2 (10.1)	55.4 (10.1)	55.4 (9.9)	58.1 (10.1)	<0.001
Current smoker, %	58.7	49.8	57.4	55.9	51.1	45.4	<0.001
Daily alcohol drinker, %	51.9	47.8	50.9	48.4	48.6	46.5	<0.001
BMI, kg/m^2 (mean (SD))	22.6 (3.4)	22.7 (3.4)	22.8 (2.8)	22.8 (2.8)	22.9 (5.4)	22.6 (2.8)	<0.001
Exercise (> 1h/week), %	19.0	27.6	26.5	25.0	25.5	29.5	<0.001
Sleep duration, 8-9 hours, %	35.6	35.9	34.6	36.2	35.1	36.3	<0.001
Vegetable intake, daily, %	21.3	25.4	20.1	20.4	20.8	30.1	<0.001
Fruit intake, daily, %	14.8	22.4	15.4	16.3	17.3	28.1	<0.001
Green tea intake, daily, %	76.5	79.2	79.9	78.3	77.9	79.8	<0.001
Coffee intake, daily, %	43.8	50.7	50.5	48.0	47.5	52.9	<0.001
Educated over 18 years old, %	25.5	34.7	33.8	33.3	31.0	36.6	<0.001
History of diabetes, %	5.5	6.3	4.5	4.2	5.5	7.7	<0.001
History of hypertension, %	18.4	17.9	17.5	17.1	16.8	18.7	0.039
History of kidney diseases, %	3.0	3.4	3.8	3.0	3.0	3.5	<0.001
History of liver diseases, %	5.8	6.5	6.3	6.0	5.4	7.2	<0.001
Women (n = 545999)							
number of subjects	10407	44592	3640	7590	8108	25254	
Age, year (mean (SD))	58.0 (10.2)	56.9 (9.9)	56.5 (10.2)	55.6 (10.1)	55.6 (9.9)	57.9 (9.9)	<0.001
Current smoker, %	6.9	4.2	6.1	5.5	4.3	3.5	<0.001
Daily alcohol drinker, %	4.3	4.5	5.5	4.3	4.2	4.6	<0.001
BMI, kg/m2 (mean (SD))	23.0 (3.4)	22.9 (3.7)	23.0 (3.8)	23.1 (4.4)	23.1 (3.1)	22.8 (3.6)	<0.001
Exercise (> 1h/week), %	13.6	20.8	17.1	18.5	18.8	22.6	<0.001
Sleep duration, 8-9 hours, %	27.7	25.6	25.1	25.9	25.4	25.7	<0.001
Vegetable intake, daily, %	24.7	30.4	25.0	24.6	24.2	34.8	<0.001
Fruit intake, daily, %	25.0	35.7	26.6	29.2	29.2	41.1	<0.001
Green tea intake, daily, %	73.8	76.8	77.0	76.4	75.8	77.3	<0.001
Coffee intake, daily, %	39.6	48.2	46.2	46.4	44.4	50.2	<0.001
Educated over 18 years old, %	19.9	31.6	27.9	29.8	27.4	34.0	<0.001
History of diabetes, %	2.6	3.7	3.2	2.7	2.7	4.4	<0.001
History of hypertension, %	21.5	19.7	20.5	19.1	18.9	20.0	<0.001
History of kidney diseases, %	3.6	4.1	3.9	3.7	3.7	4.4	<0.001
History of liver diseases, %	3.5	4.6	4.9	3.9	3.9	5.0	<0.001

As listed in **Table 1**, compared with those who chose "never" as their milk intake frequency at baseline, milk drinkers were less likely to be a current smoker or a daily alcohol consumer in both men and women. Furthermore, people consumed milk more than 1-2 times/month were more likely to be a daily consumers of vegetable, fruit as well as coffee, and more likely to join exercise more than 1 hour/week among either sex.

Detailed results from the Bayesian survival models (crude, age-adjusted and multivariable adjusted) according to the frequency of milk intake separated by sex are listed in **Table 2** (men) and **Table 3** (women). Compared to those who never had milk, both men and women had lower hazard of dying from total stroke in crude models. Chances that the posterior HRs were estimated to be lower than 1 for those who had at least 1-2 times/month was higher than 86.5% in men and more than 94.6% in women. However, lower hazard was observed to remain after age or multivariable adjustment only in men. Specifically, the mean (sd; 95% CrI) of posterior multivariable-adjusted HRs for daily male consumers of milk was 0.80 (sd = 0.07; 95% CrI: 0.69, 0.93) with a probability of 99.0% to be smaller than the null value (=1). Daily female milk consumers had posterior HRs that was distributed with mean (sd) of 0.95 (sd = 0.12; 95% CrI: 0.80, 1.17) which had 78.0% of chance that their HRs could be lower than 1.

Posterior distributions of HRs for mortality from hemorrhagic stroke were found to contain the null value for either men or women among all fitted models. In contrast, men who had milk intake frequency higher than 1-2 times/week were found to be associated with averagely 26%-37% lower hazard of dying from cerebral infarction compared with men who never drank milk (Model 2 in **Table** 2). We were more than 95.7% sure that these posterior HRs could be lower than 1. No associations were observed between milk intake and hazard of cerebral infarction among women.

4. Discussion

In the JACC study cohort, we observed that

5. Conclusion

Acknowledgments: All sources of funding of the study should be disclosed. Please clearly indicate grants that you have received in support of your research work. Clearly state if you received funds for covering the costs to publish in open access.

Author Contributions: "X.X. and Y.Y. conceive and designed the experiments; X.X. performed the experiments; X.X. and Y.Y. analyzed the data; W.W. contributed reagents/materials/analysis tools; Y.Y. wrote the paper.' '

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, an in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

JACC	Japan Collaborative Cohort
FFQ	Food Frequency Questionnaire
MCMC	Markov Chain Monte Carlo
JAGS	Just Another Gibbs Samplers
HR	hazard ratio
sd	standard deviation
CrI	credible interval

References

1. Ohno, Y.; Tamakoshi, A.; Group, J.S.; others. Japan collaborative cohort study for evaluation of cancer risk sponsored by monbusho (JACC study). *Journal of epidemiology* **2001**, *11*, 144–150.
2. Tamakoshi, A.; Yoshimura, T.; Inaba, Y.; Ito, Y.; Watanabe, Y.; Fukuda, K.; Iso, H. Profile of the JACC study. *Journal of epidemiology* **2005**, *15*, S4–S8.
3. Tamakoshi, A.; Ozasa, K.; Fujino, Y.; Suzuki, K.; Sakata, K.; Mori, M.; Kikuchi, S.; Iso, H. Cohort profile of the Japan Collaborative Cohort Study at final follow-up. *Journal of epidemiology* **2013**, p. JE20120161.
4. Date, C.; Fukui, M.; Yamamoto, A.; Wakai, K.; Ozeki, A.; Motohashi, Y.; Adachi, C.; Okamoto, N.; Kurosawa, M.; Tokudome, Y.; others. Reproducibility and validity of a self-administered food frequency questionnaire used in the JACC study. *Journal of epidemiology* **2005**, *15*, S9–S23.
5. Plummer, M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling, 2003.
6. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.
7. Lunn, D.; Spiegelhalter, D.; Thomas, A.; Best, N. The BUGS project: Evolution, critique and future directions. *Statistics in medicine* **2009**, *28*, 3049–3067.
8. Brooks, S.P.; Gelman, A. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics* **1998**, *7*, 434–455.

Table 2. Summary of posterior Hazard Ratios (HR) of mortality from total stroke, different stroke type according to the frequency of milk intake in men (JACC study, 1988-2009).

	Never	1-2 times/Month	1-2 times/Week	3-4 times/Week	Almost Daily
Person-year	135704	56551	97098	92153	252364
N	8508	3522	5928	5563	15865
Total Stroke	326	122	181	177	546
Model 0					
MeanHR (SD)	1	0.89 (0.09)	0.77 (0.07)	0.79 (0.07)	0.90 (0.06)
95% CrI	-	(0.73, 1.08)	(0.63, 0.91)	(0.66, 0.94)	(0.79, 1.03)
MCSE	-	0.0022	0.0019	0.0022	0.0018
Pr(HR < 1)	-	86.5%	99.9%	99.7%	93.5%
Model 1					
MeanHR (SD)	1	0.98 (0.11)	0.84 (0.08)	0.86 (0.08)	0.76 (0.05)
95% CrI	-	(0.79, 1.19)	(0.70, 1.00)	(0.71, 1.02)	(0.66, 0.87)
MCSE	-	0.0027	0.0022	0.0021	0.0016
Pr(HR < 1)	-	58.7%	97.3%	96.1%	100.0%
Model 2					
MeanHR (SD)	1	1.01 (0.12)	0.87 (0.09)	0.90 (0.09)	0.80 (0.07)
95% CrI	-	(0.81, 1.24)	(0.72, 1.05)	(0.74, 1.08)	(0.69, 0.93)
MCSE	-	0.0041	0.0036	0.0038	0.0031
Pr(HR < 1)	-	50.6%	93.7%	89.6%	99.0%
Hemorrhagic stroke	100	42	58	56	176
Model 0					
MeanHR (SD)	1	1.03 (0.19)	0.82 (0.14)	0.84 (0.15)	0.97 (0.13)
95% CrI	-	(0.70, 1.46)	(0.56, 1.14)	(0.60, 1.17)	(0.75, 1.26)
MCSE	-	0.0066	0.0057	0.0057	0.0062
Pr(HR < 1)	-	47.2%	88.4%	86.3%	63.1%
Model 1					
MeanHR (SD)	1	1.11 (0.21)	0.88 (0.16)	0.90 (0.16)	0.88 (0.12)
95% CrI	-	(0.75, 1.58)	(0.63, 1.25)	(0.63, 1.24)	(0.67, 1.14)
MCSE	-	0.0091	0.0067	0.0073	0.0076
Pr(HR < 1)	-	31.6%	79.7%	76.6%	87.6%
Model 2					
MeanHR (SD)	1	1.14 (0.22)	0.92 (0.17)	0.95 (0.18)	0.95 (0.14)
95% CrI	-	(0.75, 1.61)	(0.63, 1.29)	(0.65, 1.37)	(0.71, 1.27)
MCSE	-	0.0093	0.0075	0.0098	0.0082
Pr(HR < 1)	-	28.8%	72.4%	64.4%	69.3%
Cerebral infarction	151	41	64	66	198
Model 0					
MeanHR (SD)	1	0.65 (0.12)	0.59 (0.09)	0.64 (0.09)	0.71 (0.09)
95% CrI	-	(0.46, 0.92)	(0.43, 0.79)	(0.47, 0.85)	(0.56, 0.89)
MCSE	-	0.0042	0.0037	0.0036	0.0038
Pr(HR < 1)	-	99.1%	99.9%	99.7%	99.5%
Model 1					
MeanHR (SD)	1	0.73 (0.13)	0.65 (0.10)	0.70 (0.11)	0.58 (0.07)
95% CrI	-	(0.49, 1.02)	(0.48, 0.88)	(0.51, 0.94)	(0.46, 0.72)
MCSE	-	0.0045	0.0035	0.0041	0.0029
Pr(HR < 1)	-	96.9%	99.8%	98.9%	100.0%
Model 2					
MeanHR (SD)	1	0.73 (0.14)	0.67 (0.11)	0.72 (0.12)	0.61 (0.08)
95% CrI	-	(0.50, 1.04)	(0.48, 0.91)	(0.52, 0.99)	(0.48, 0.79)
MCSE	-	0.0049	0.0047	0.0061	0.0052
Pr(HR < 1)	-	96.1%	99.1%	97.5%	99.8%

Note:
Abbreviations: SD, standard deviation; CrI, credible interval; MCSE, Monte Carlo Standard Error; Pr(HR < 1) indicates the prabability for posterior HR to be smaller than 1.
Model 0 = Crude model; Model 1 = age-adjusted model; Model 2 = multivariable adjusted model.
Covariates included in Model 2: age, smoking habit, alcohol intake, body mass index, history of hypertension, diabetes, kidney/liver diseases, exercise, sleep duration, coffee intake, education level.

Table 3. Summary of posterior Hazard Ratios (HR) of mortality from total stroke, different stroke type according to the frequency of milk intake in women (JACC study, 1988-2009).

	Never	1-2 times/Month	1-2 times/Week	3-4 times/Week	Almost Daily
Person-year	173222	59904	129233	139919	418925
N	10407	3640	7590	8108	25254
Total Stroke	300	84	182	172	585
Model 0					
MeanHR (SD)	1	0.83 (0.10)	0.81 (0.08)	0.70 (0.07)	0.81 (0.07)
95% CrI	-	(0.64, 1.05)	(0.68, 0.97)	(0.58, 0.85)	(0.71, 0.93)
MCSE	-	0.0029	0.0022	0.0021	0.0023
Pr(HR < 1)	-	94.6%	98.7%	99.9%	99.6%
Model 1					
MeanHR (SD)	1	1.00 (0.14)	1.18 (0.14)	1.03 (0.12)	0.92 (0.09)
95% CrI	-	(0.76, 1.31)	(0.95, 1.47)	(0.82, 1.28)	(0.78, 1.09)
MCSE	-	0.0038	0.0045	0.0042	0.0034
Pr(HR < 1)	-	52.3%	6.3%	42.0%	86.8%
Model 2					
MeanHR (SD)	1	1.01 (0.17)	1.19 (0.15)	1.03 (0.15)	0.95 (0.12)
95% CrI	-	(0.75, 1.36)	(0.96, 1.52)	(0.81, 1.31)	(0.80, 1.17)
MCSE	-	0.0083	0.0096	0.0079	0.0075
Pr(HR < 1)	-	52.8%	6.4%	44.4%	78.0%
Hemorrhagic stroke	108	27	78	76	231
Model 0					
MeanHR (SD)	1	0.73 (0.16)	0.98 (0.15)	0.87 (0.14)	0.89 (0.11)
95% CrI	-	(0.47, 1.08)	(0.71, 1.31)	(0.64, 1.16)	(0.71, 1.15)
MCSE	-	0.0058	0.0071	0.0061	0.0065
Pr(HR < 1)	-	94.7%	58.1%	83.1%	83.0%
Model 1					
MeanHR (SD)	1	0.84 (0.18)	1.17 (0.18)	1.06 (0.17)	0.93 (0.12)
95% CrI	-	(0.54, 1.24)	(0.86, 1.58)	(0.76, 1.45)	(0.73, 1.19)
MCSE	-	0.0065	0.0089	0.0084	0.0069
Pr(HR < 1)	-	81.6%	16.9%	38.9%	74.6%
Model 2					
MeanHR (SD)	1	0.85 (0.18)	1.21 (0.18)	1.13 (0.17)	1.03 (0.12)
95% CrI	-	(0.56, 1.30)	(0.90, 1.62)	(0.84, 1.52)	(0.81, 1.29)
MCSE	-	0.0051	0.0067	0.0039	0.0047
Pr(HR < 1)	-	79.6%	12.7%	33.4%	71.4%
Cerebral infarction	102	35	63	50	187
Model 0					
MeanHR (SD)	1	1.01 (0.13)	0.82 (0.14)	0.59 (0.14)	0.76 (0.16)
95% CrI	-	(0.69, 1.48)	(0.60, 1.13)	(0.43, 0.84)	(0.59, 0.99)
MCSE	-	0.0087	0.0079	0.0064	0.0071
Pr(HR < 1)	-	51.9%	75.6%	97.6%	96.1%
Model 1					
MeanHR (SD)	1	1.29 (0.14)	1.15 (0.13)	0.86 (0.14)	0.86 (0.15)
95% CrI	-	(0.88, 1.89)	(0.84, 1.58)	(0.61, 1.20)	(0.67, 1.09)
MCSE	-	0.0071	0.0072	0.0095	0.0054
Pr(HR < 1)	-	36.7%	42.1%	78.7%	88.9%
Model 2					
MeanHR (SD)	1	1.30 (0.11)	1.18 (0.09)	0.87 (0.10)	0.93 (0.07)
95% CrI	-	(0.88, 1.92)	(0.86, 1.63)	(0.62, 1.23)	(0.72, 1.19)
MCSE	-	0.0055	0.0057	0.0067	0.0047
Pr(HR < 1)	-	35.1%	39.9%	75.6%	80.1%

Note:
Abbreviations: SD, standard deviation; CrI, credible interval; MCSE, Monte Carlo Standard Error; Pr(HR < 1) indicates the prabability for posterior HR to be smaller than 1.
Model 0 = Crude model; Model 1 = age-adjusted model; Model 2 = multivariable adjusted model.
Covariates included in Model 2: age, smoking habit, alcohol intake, body mass index, history of hypertension, diabetes, kidney/liver diseases, exercise, sleep duration, coffee intake, education level.