



Prevalence of *H. Pylori* Infection by birth year

A meta regression analysis

Chaochen Wang 王超辰

2017/7/14 in NCC

METHODS

Data extraction (1)

- 43 papers
 - + JPHC Cohort II
 - + JPHC-NEXT Cohort Studyincluded for meta-regression analysis ([Table 1](#))(←click here).
- Prevalence of *H. pylori* infection reported by birth year group of participants^[1~7].
 1. Ueda J. *Helicobacter*. 2014;
 2. Watanabe M. *Cancer Sci*. 2015;
 3. Reploge M.L. *Int J Epidemiol*. 1996;
 4. Shimoyama T. *Gastric Cancer*. 2012;
 5. Shimatani T. *J Gastroenterol Hepatol*. 2005;
 6. JPHC Cohort II;
 7. JPHC-NEXT

Data extraction (2)

- 38 studies reported prevalence with age groups.
 - 34 studies reported data collection period:
 - Example:
 - 60-70 years old group;
 - data collection done in 1990;
 - birth year should be 1920~1930.
 - 4 studies reported age groups with **data collection year unavailable**, year of publication was used instead of data collection period.

Data extraction example

(rules are decided by researchers)

Sasazuki, S. et al. *Cancer Epidemiol Biomarkers Prev* 15, 1341–1347 (2006).
(Study No. 12)

Table 2. Adjusted ORs (95% CIs) of developing gastric cancer for all subjects with *H. pylori* IgG seropositivity and with stratification by several factors

	No. <i>H. pylori</i> –positive cases/controls	Adjusted OR* (95% CI)
All subjects (511 pairs)	478/383	5.1 (3.2-8.0)
Gender		
Men (342 pairs)	327/259	6.8 (3.6-12.6)
Women (169 pairs)	151/124	4.6 (2.1-9.9)
Age at baseline (y)		
40-49 (77 pairs)	72/52	7.0 (1.8-27.4)
50-59 (235 pairs)	220/180	4.4 (2.2-8.7)
60-69 (199 pairs)	186/151	5.4 (2.5-11.6)
Duration between blood donation and cancer diagnosis (y)		
0-4 (205 pairs)	190/144	7.4 (3.4-16.2)
4-8 (181 pairs)	170/141	4.9 (2.2-10.9)
≥8 (125 pairs)	118/98	4.5 (1.8-11.4)
Tumor location		
Upper third of stomach, including cardia (39 pairs)	37/33	3.7 (0.2-68.4)
Distal portion of stomach (368 pairs)	344/274	5.1 (3.0-8.6)
Histologic type		
Differentiated type (299 pairs)	281/223	5.8 (3.1-10.8)
Undifferentiated type (159 pairs)	149/122	5.1 (2.1-12.3)

*Matched for age, gender, resident area, blood donation date, and fasting times at blood donation. Adjusted for smoking status, consumption of fish gut, green, yellow vegetables, other vegetables, fruit, green tea, body mass index, and

- Control samples of this study (JPHC I/II) were collected during 1990-1993.

⇒ 1990 (the earlier one) used as the year of research

- Birth year groups:
 - 40-49 ⇒ 1941-1950
 - 50-59 ⇒ 1931-1940
 - 60-69 ⇒ 1921-1930
 - Many studies reported no upper limit of age (e.g. 70 or older) ⇒ 1920

273 Data Points from 45 Studies were available for meta-regression

Show entries Search:

No	Author	adultdults.or.childchildren	Source.population	S
1	Kikuchi, 1998	adult	General	s
1	Kikuchi, 1998	adult	General	s
1	Kikuchi, 1998	adult	General	s
1	Kikuchi, 1998	adult	General	s
2	Fujisawa, 1999	adult	General	s
2	Fujisawa, 1999	adult	General	s

STATISTICAL ANALYSIS

Step 1: Data Point Weight Calculation

```
library(meta)

library(metafor)

library(mgcv)

meta <- metaprop( event = Number_of_Positive,
                  n      = Number_of_Subjects_in_the_corresponding_group,
                  byvar  = Birth_Year,
                  sm      = "PLOGIT",      # Logit transformation
                  method.tau = "REML")    # Restricted Maximum-likelihood
                                          # estimator to estimate the
                                          # between-study variance

weight<-meta$w.random # Weight of each data point extracted
```


Step 2: Meta-regression (Generalized Additive Mixed Model, GAMM)

```
##### MODEL 1 #####  
res1 <- gam(cbind(event,n) ~ s(Birth_Year, bs="cr") +  
              # Cubic Spline Regression 三次スプライン曲線  
s(Study_ID, bs="re") + # Study ID as random effect  
  
Source_of_population + # Community OR Clinical based  
  
Specimen_type          + # Serum OR Others (urinary, salivary, stool,  
  
Kit.from                + # Antigen derived from domestic or foreign  
  
early,                  # Data collection period, cutoff = 2000  
  
data = data, weights=weight,  
  
family="binomial"(link=logit), method="REML")
```

Summary from Model 1 comparable to [Table 3](#) ← click here

```
##
## Family: binomial
## Link function: logit
##
## Formula:
## cbind(mp, n_total - mp) ~ s(birth.year_high, bs = "cr") + s(No,
##      bs = "re") + Source.population + Specimen.type + kit.from +
##      early
##
## Parametric coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.17064    0.14022  -1.217    0.2236
## Source.populationPatient  0.28630    0.20234   1.415    0.1571
## Specimen.typeother    -0.41271    0.19256  -2.143    0.0321 *
## kit.fromforeign      0.01836    0.16847   0.109    0.9132
## kit.fromunknown     -0.11517    0.25923  -0.444    0.6569
## earlylate          -0.25471    0.15813  -1.611    0.1072
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##              edf Ref.df Chi.sq p-value
## s(birth.year_high)  7.372  8.158   4255  <2e-16 ***
## s(No)              34.754 38.000   1910  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Step 3: Model Modification

```
##### MODEL 2 #####
```

```
res2 <- gam(cbind(event,n) ~ s(Birth_Year, bs="cr") +  
           s(Study_ID, bs="re") +  
           Specimen_type,  
           data = data, weights=weight,  
           family="binomial"(link=logit), method="REML")
```

```
##### MODEL 3 #####
```

```
res3 <- gam(cbind(event,n) ~ s(Birth_Year, bs="cr") +  
           s(Study_ID, bs="re"),  
           data = data, weights=weight,  
           family="binomial"(link=logit), method="REML")
```

TABLE 2. Informations for tested models.			
	AIC	BIC	LogLik
Model 1: $\text{Logit}(P) = s(\text{birth year}) + r(\text{study ID}) + f(\text{source of population}) + f(\text{diagnostic test}) + f(\text{ELIZA kits}) + f(\text{research year})$	1716.444	1895.216	-808.6935 (df=49.53)
Model 2: $\text{Logit}(P) = s(\text{birth year}) + r(\text{study ID}) + f(\text{diagnostic test})$	1730.349	1904.178	-817.0157 (df=48.16)
Model 3: $\text{Logit}(P) = s(\text{birth year}) + r(\text{study ID})$	1731.451	1906.366	-817.2658 (df=48.46)
Abbreviations and definitions: AIC: Akaike's information criterion; BIC: Bayesian information criterion; LogLik: Log-likelihood; P: prevalence; s: penalized cubic spline; r: random effect; f: fixed effect; df: degree of freedom.			

Step 4: Point Estimation

```
library(visreg) # Visualization of Regression Models
library(plyr)
use(com_data)
plotdata_res0 <- visreg(res0, type = "contrast", plot = F)
smooths <- ldply(plotdata_res0, function(part)
  data.frame(Variable = part$meta$x,
             x=part$fit[[part$meta$x]],
             smooth=part$fit$visregFit,
             lower=part$fit$visregLwr,
             upper=part$fit$visregUpr))

lg=smooths$smooth
p=exp(lg)/(1+exp(lg))
p_low <- exp(smooths$lower)/(1+exp(smooths$lower))
p_high <- exp(smooths$upper)/(1+exp(smooths$upper))

dt_plot <- data.frame(smooths$x[1:101],p[1:101], p_low[1:101], p_high[1:101])

names(dt_plot) <- c("Birthyear", "Prevalence", "95%CI_low", "95%CI_high")
print(dt_plot)
```

Table 4. Predicted Prevalence of *H. pylori* infection from 1908 to 2003

Show entries

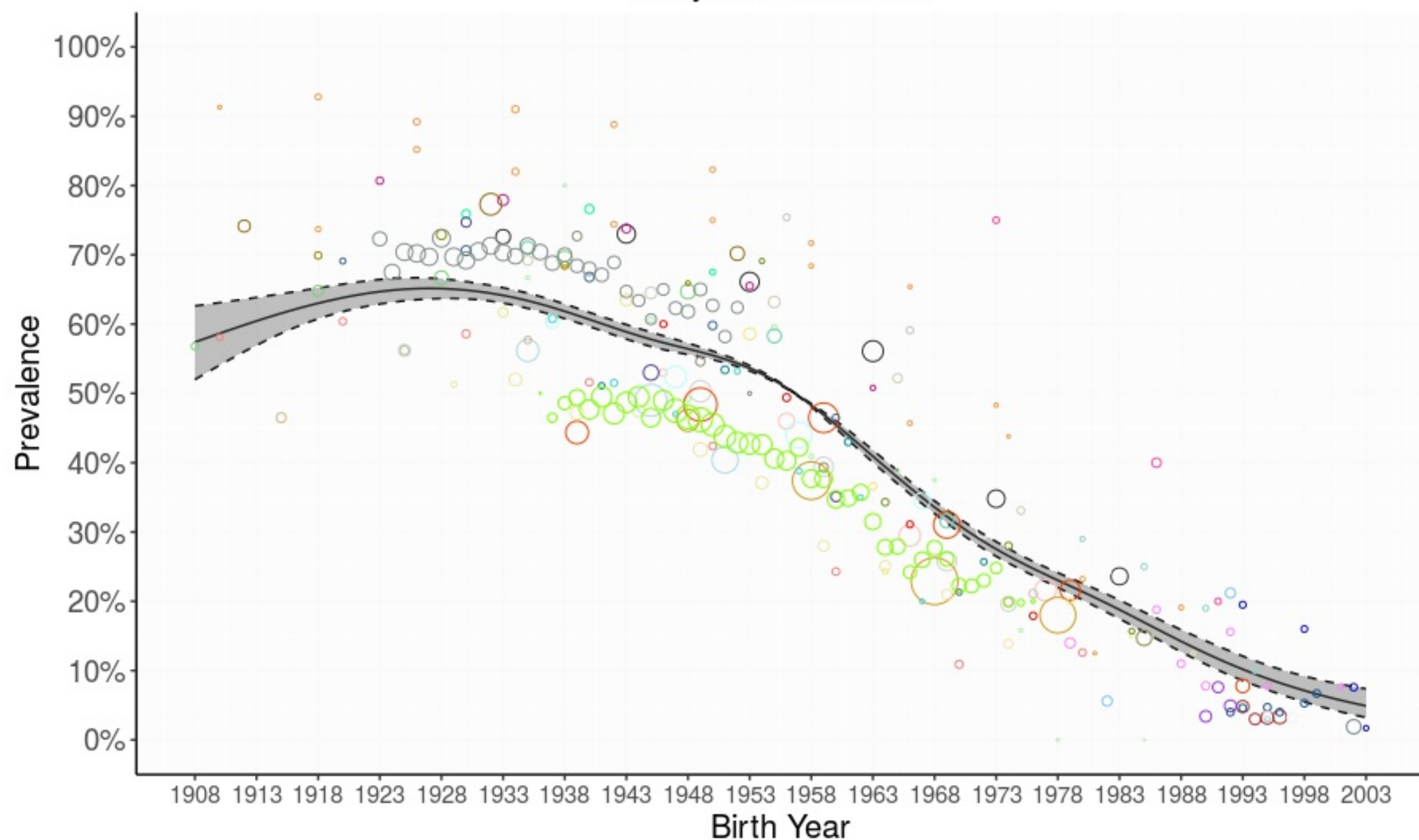
Search:

Birthyear	Prevalence	95%CI_low	95%CI_high
1908	0.574	0.520	0.626
1909	0.580	0.530	0.628
1910	0.586	0.540	0.631
1911	0.592	0.549	0.633
1912	0.597	0.559	0.635
1913	0.603	0.567	0.638
1914	0.609	0.576	0.640
1915	0.614	0.584	0.643
1916	0.619	0.591	0.645

PLOT

by package **ggplot2**

Figure 2. Multivariable adjusted prevalence of *H. pylori* infection in Japanese by birth year
from year of 1908 to 2003



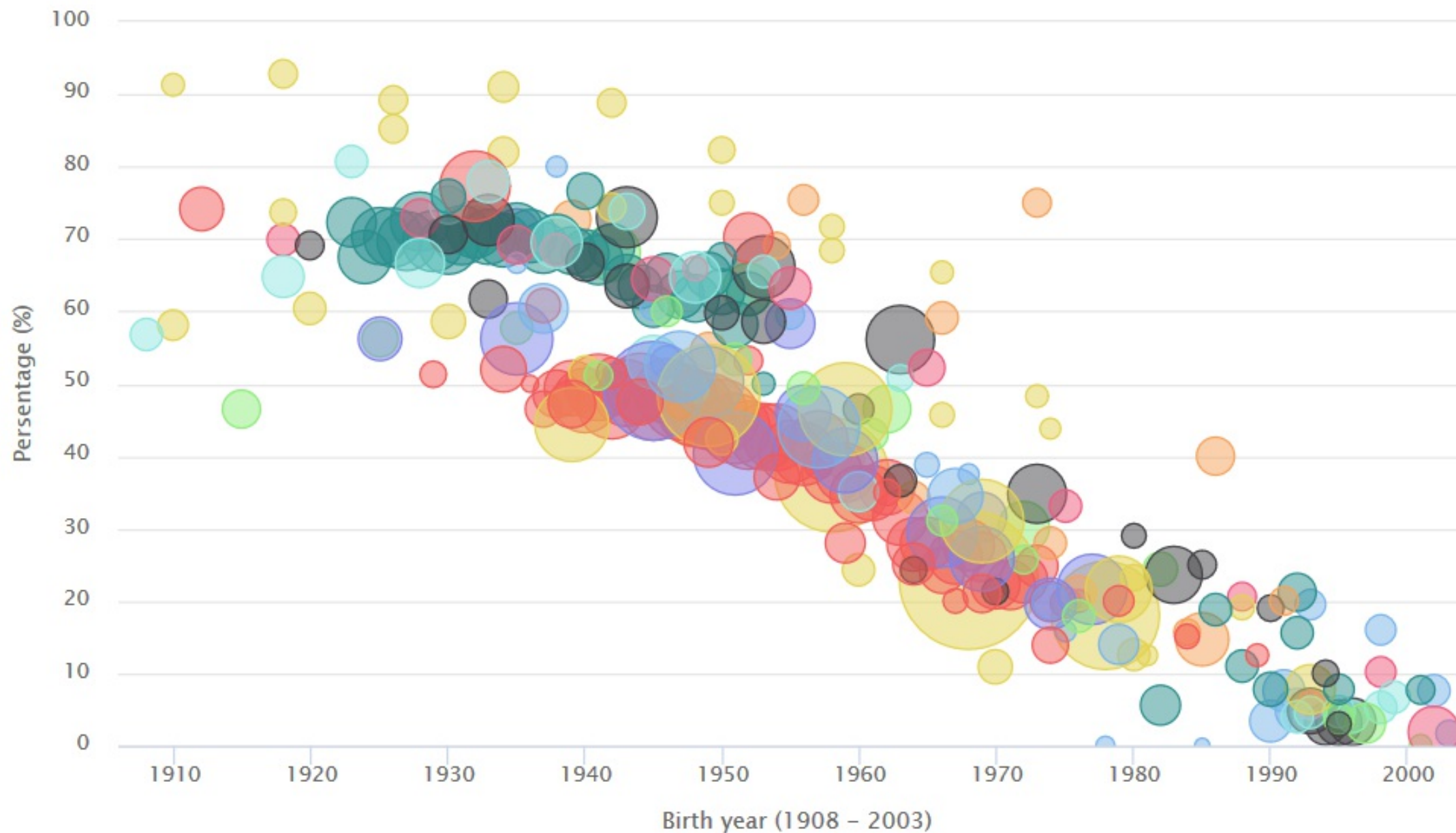
Sample Size ○ 500 ○ 1000 ○ 2000 ○ 4000 ○ 6000

Author, Publish year	○ Akamatsu, 2011	○ Kato, 2003	○ Naito, 2008	○ Shibata, 2000	○ Urita, 2013
	○ Akamatsu, 2015	○ Kato, 2004	○ Nakajima, 2010	○ Shibata, 2002	○ Watanabe, 2015
	○ Fujimoto, 2007	○ Kawade, 2005	○ Nakao, 2011	○ Shimatani, 2005	○ Yamagata, 2000
	○ Fujisawa, 1999	○ Kawai, 2010	○ Nakayama, 2016	○ Shimoyama, 2012	○ Yamaji, 2001
	○ Fukao, 1993	○ Kikuchi, 1998	○ Nobuta, 2004	○ Shimoyama, 2014	○ Yamashita, 2001
	○ Hirayama, 2014	○ Kikuchi, 2005	○ Ogihara, 2000	○ Shiotani, 2008	○ Yang, 1999
	○ JPHC, CohortII	○ Kumagai, 1998	○ Okuda, 2014	○ Tamura, 2012	○ Youn, 1998
	○ JPHCnext	○ Kurosawa, 2000	○ Reploge, 1996	○ Toyoda, 2012	
	○ Kamada, 2015	○ Mizuno, 2010	○ Sasazuki, 2006	○ Ueda, 2014	

Interactive PLOT

Click above ↑

Prevalence of *H.pylori* in Japanese by birth year



Author, publish year

- | | | | | |
|------------------|------------------|------------------|-------------------|------------------|
| ● Akamatsu, 2011 | ● Akamatsu, 2015 | ● Fujimoto 2007 | ● Fujisawa, 1999 | ● Fukao, 1993 |
| ● Fukuda, 2003 | ● Hirayama, 2014 | ● JPHC, CohortII | ● JPHCnext | ● Kamada, 2015 |
| ● Kato, 2004 | ● Kawade, 2005 | ● Kawai 2010 | ● Kikuchi, 1998 | ● Kikuchi, 2005 |
| ● Kurosawa, 2000 | ● Mizuno, 2010 | ● Naito, 2008 | ● Nakajima, 2010 | ● Nakao 2011 |
| ● Nakayama, 2016 | ● Nobuta, 2004 | ● Ogihara, 2000 | ● Okuda, 2014 | ● Reploge, 1996 |
| ● Sasazuki, 2006 | ● Shibata, 2000 | ● Shibata, 2002 | ● Shimatani, 2005 | ● Shimoyama 2012 |
| ● Shimoyama 2014 | ● Shiotani, 2008 | ● Tamura, 2012 | ● Toyoda, 2012 | ● Ueda, 2014 |
| ● Watanabe, 2015 | ● Yamagata, 2000 | ● Yamaji, 2001 | ● Yamashita, 2001 | ● Yang, 1999 |
| | | | | ● Youn, 1998 |