Differences in COVID-19 Incidence by Health Check-up Questionnaire Responses

Takao Yabuki
Nippon Life Insurance Company

© Takao Yabuki

**This presentation has been prepared for the 33rd International Congress of Actuaries.
The Institute of Actuaries of Japan (IAJ) and the International Actuarial Association (IAA) assume no responsibility for the opinions or content expressed in these presentation slides. Furthermore, the opinions and content contained herein do not necessarily reflect the views of the IAJ or the IAA.**

Research question

Do lifestyle habits affect COVID-19 incidence?

Exploratory analysis using check-up questionnaire responses

Question

Does COVID-19 incidence differ by lifestyle responses on the check-up questionnaire? If so, for which items, in which direction, and by how much?

Lifestyle items

- Smoking
- Alcohol (frequency, amount)
- Sleep
- Exercise, walking, walking speed
- Eating habits (speed, late dinner, snacking, breakfast)

Approach

Stratify by sex and age; compare incidence across response groups.
Compute sex- and age-adjusted rate ratios (aRR) vs a reference group.
Exploratory and descriptive only.

Notes:
This is a statistical exploration and does not address medical causation
The views expressed are the presenter's own and do not represent the organization

Context | Public health insurance

Japan's universal health insurance consists of three schemes

Employees' insurance

≈77 million

Employees & dependents

Health insurance societies, JHIA, mutual aid

Employee premium (share of income)
≈6–7%
(Employer pays an additional share)

National Health Insurance

≈26 million

Self-employed & others

Prefectures, municipalities, NHI associations

Premium (share of income)
≈9%
(Municipal NHI average shown)

Late-stage elderly care (75+)

≈20 million

Aged 75 and older

Prefecture-wide regional unions

Premium (share of income)
≈8%
(Average across the scheme)

This study uses Wellness-Star data held by Nippon Life (provided by health insurance societies)

Premiums are indicative FY2023 averages
Source: MHLW (https://www.mhlw.go.jp/content/kiso_r05.pdf)

Context | Health check-ups (1)

What is a health check-up?

A health check-up is an examination that insured persons receive to confirm their own health status and to prevent lifestyle-related and other diseases

Specific health check-up

Covers: insured persons and dependents aged 40–74

Other health check-ups

Covers: everyone else (e.g., periodic check-ups provided by employers)

* The legal basis differs by scheme; details are omitted here

In either case, employees receive a health check-up once a year without exception

Source: MHLW (https://www8.cao.go.jp/kisei-kaikaku/kisei/meeting/wg/2210_03medical/230424/medical11_0103.pdf)

Context | Health check-ups (2)

Of the 22 standard questionnaire items, the 11 lifestyle-related items were analyzed

No.	Question (summary)	Used
1	Blood-pressure medication	
2	Blood-sugar medication / insulin	
3	Cholesterol medication	
4	History of stroke	
5	History of heart disease	
6	Kidney disease / renal failure	
7	Told you have anemia	
8	Currently smoke habitually	●
9	Weight gain ≥10 kg since age 20	
10	Exercise ≥30 min ×2/week, ≥1 year	●
11	Walking ≥1 hour/day	●

No.	Question (summary)	Used
12	Walks faster than peers	●
13	Chewing ability	
14	Eats faster than others	●
15	Dinner within 2 h of bedtime	●
16	Snacks / sweet drinks	●
17	Skips breakfast ≥3 times/week	●
18	Alcohol frequency	●
19	Alcohol amount per day	●
20	Enough rest through sleep	●
21	Intends to improve lifestyle	
22	Received specific health guidance	

Source: MHLW (https://www.mhlw.go.jp/content/10900000/001153030.pdf)

Context | Wellness-Star data

The data consist of three parts

Data type	What it records	When created / updated
Enrollment	Member attributes, enrollment, withdrawal	When enrollment / withdrawal / changes are filed Updated as eligibility or attributes change
Claims	Diagnoses, procedures, drugs, medical costs	Each month with treatment or dispensing Billed by providers/pharmacies next month
Check-up	Height, weight, blood pressure, blood tests, etc.	Each time a check-up is taken Usually once a year; recorded once results are final

Link enrollment, claims and check-up data to follow each individual

Context | Enrollment data (1)

Enrollment data records who was enrolled, and when

Basic member information needed to read claims and check-up records

Main information	What it tells us / how it is used
Individual identifier	Links the same person's claims and check-up data
Member attributes	Sex, date of birth, etc. to describe the study population
Eligibility start / end dates	Enrollment, withdrawal and the observable period
Insurer / employer information	The group a member belongs to and linkable industry information

The basis for capturing all members, not only those who visited a hospital

Context | Enrollment data (2)

Both members and dependents have enrollment records

Person	Birth	Sex	Type	Enrolled	Left	Data from
P-A	1972/01	Male	M	2023/03	—	2017/04
P-B	1998/02	Male	M	2019/04	2023/10	2017/04
P-C	1956/10	Female	M	1999/04	2019/05	2017/04
P-D	1952/01	Female	D	2022/06	2024/06	2017/04
P-E	2017/08	Female	D	2017/08	—	2017/04

A: enrolled after data provision began; claim in the month of enrollment
B: enrolled after provision began; since withdrawn
C: enrolled before provision began
D: enrolled as a spouse
E: child enrolled in the month of birth

Type: M = member, D = dependent
Data from: start of data provision by the member's health insurance society
Only part of the data is shown

Context | Enrollment data (3)

Distinguish enrollment from the point where claims become visible



Grey: enrollment history / Green: period in which the individual's claims are visible
Observation starts at the later of enrollment and start of data provision, and ends at the earlier of withdrawal and end of provision

Context | Claims data

Example of claims data

Person ID	Month	Diagnosis	Suspected flag	ICD-10	Diagnosis code
P-A	2023/01	Influenza A	0	J10.1	8842080
P-A	2023/01	COVID-19	1	U07.1	8850104

Linked to enrollment data by individual ID and month of treatment
Only part of the data is shown

Context | Check-up data

Example of check-up data

Person	Month	Age	Ht (cm)	Wt (kg)	DBP (mmHg)	Smoker	Exercise
P-A	2022/07	50	170	65	80	1	0
P-A	2023/08	51	171	64	85	0	1

DBP: diastolic blood pressure

Linked to enrollment data by individual ID and month of check-up
Only part of the data is shown

Context | Computing incidence

Link the three datasets; compute incidence by sex, age and item

1 Select the population

Members observable throughout 2024 with a check-up (questionnaire) in 2024. Those who enrolled or withdrew during the year are excluded.

2 Define a case

"Case" = at least one confirmed COVID-19 diagnosis (ICD-10) in 2024 claims; suspected diagnoses are excluded.

3 Compute stratified incidence

For each of 12 strata (sex × 6 age groups): cases ÷ population, by response category.

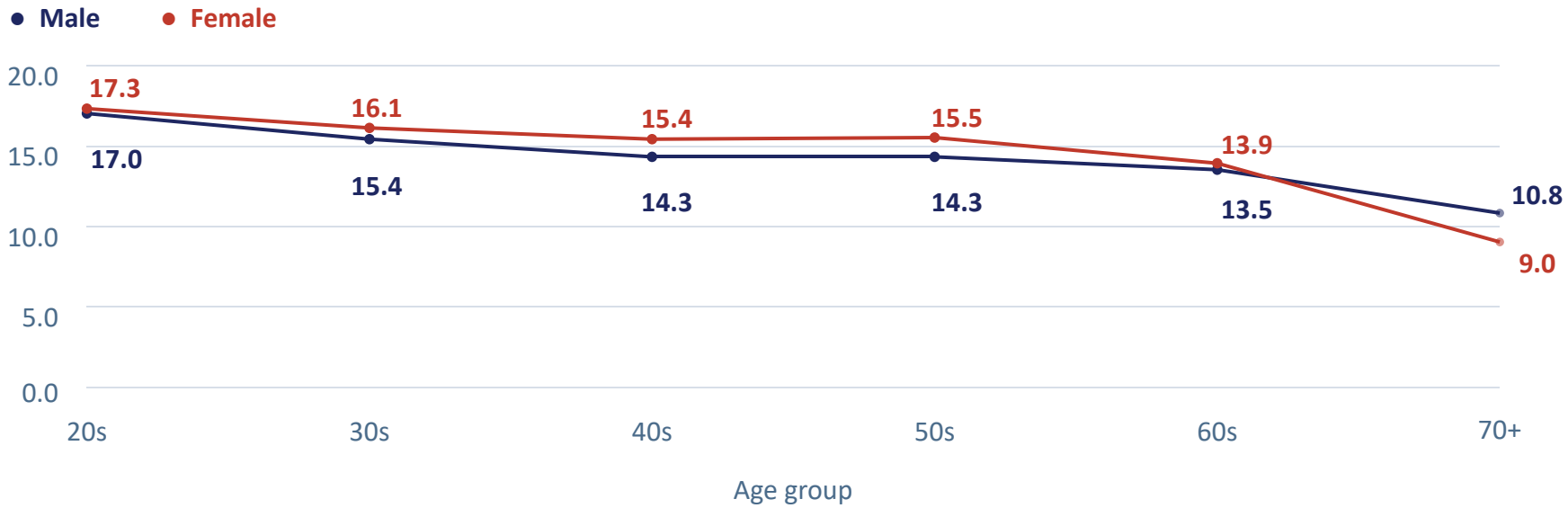
4 Compare response groups

Pool the 12 stratum rates; compute the sex- and age-adjusted rate ratio (aRR, 95% CI) vs the reference (healthier) response.

The RESULTS that follow focus on this sex- and age-adjusted rate ratio (aRR)

RESULT 1: Age pattern

COVID-19 incidence declines with age in both sexes

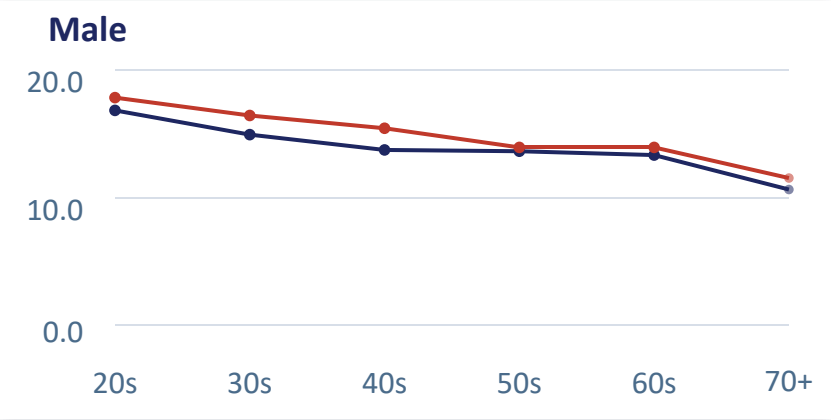


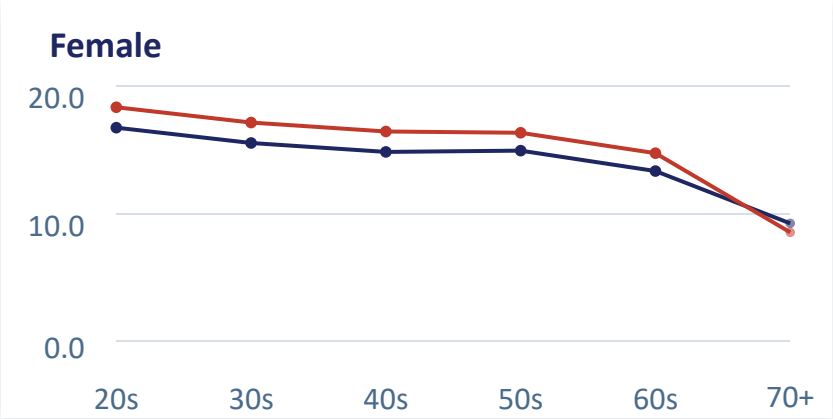
Declines from the 20s to 70+ in both sexes (men 17.0% → 10.8%, women 17.3% → 9.0%).
Note: the 70+ stratum is small, and its values fluctuate in the age-specific comparisons that follow.

RESULT 2: Differences by questionnaire item (1/2)

Insufficient sleep: a broadly consistent gap

● Enough sleep ● Insufficient sleep





The insufficient-sleep group is higher in 11 of 12 strata (exception: women 70+).

RESULT 2: Differences by questionnaire item (2/2)

Late dinner: little consistent difference



The ≥3 times/week group is higher in only 4 of 12 strata; among women the direction flips with age.

RESULT 3: Items where incidence differs

Higher incidence is intuitive; some items are lower, against intuition

▲ Higher incidence (aRR > 1)

Item	aRR	95% CI
Snacking: daily	**1.13**	1.10 – 1.15
Exercise habit: none	**1.12**	1.10 – 1.13
Sleep: insufficient	**1.11**	1.09 – 1.12
Daily walking: none	**1.10**	1.08 – 1.11
Snacking: sometimes	**1.08**	1.07 – 1.10

▼ Lower incidence (aRR < 1)

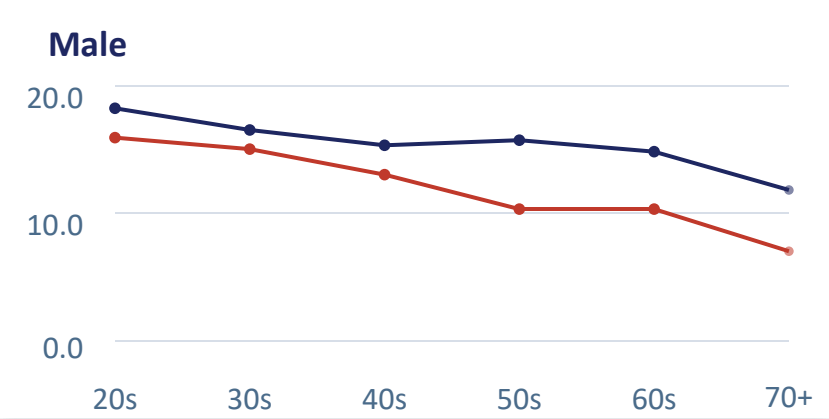
Item	aRR	95% CI
Smoking: current smoker	**0.81**	0.80 – 0.83
Alcohol amount: heavy	**0.84**	0.81 – 0.87
Breakfast: skipped ≥3/week	**0.87**	0.86 – 0.88
Alcohol amount: moderate	**0.90**	0.88 – 0.92
Drinking: every day	**0.92**	0.90 – 0.93

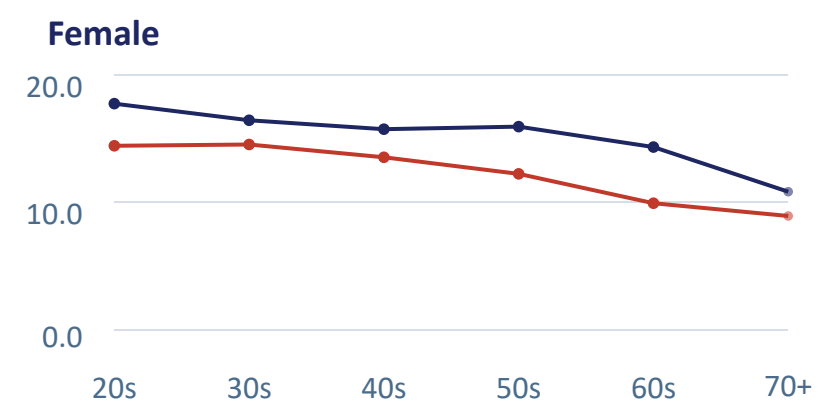
Sex- and age-adjusted rate ratios. The reference group for each item is the response considered healthier (e.g., "enough sleep").

RESULT 4: Items with lower incidence (1/3)

Smoking: smokers show lower incidence — in both men and women

● Non-smokers ● Smokers



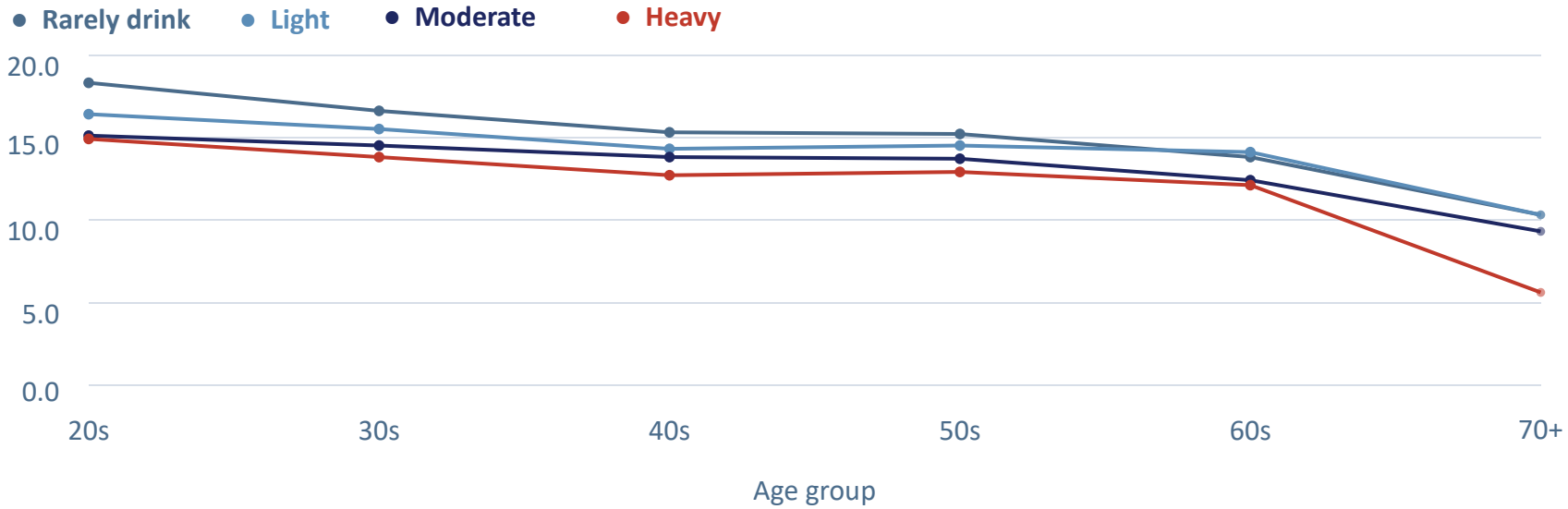


Smokers are lower in 11 of 12 strata (exception: women 70+).
The ratio falls with age: 0.86 in the 20s → 0.70 in the 60s (see Result 6).

Smokers vs non-smokers, sex- and age-adjusted rate ratio: 0.81 (95% CI 0.80–0.83)

RESULT 4: Items with lower incidence (2/3)

The more alcohol consumed, the consistently lower the incidence



Incidence falls as alcohol intake rises (light 0.95, moderate 0.90, heavy 0.84).
The heavy group exceeds the "rarely drink" group in 0 of 12 strata.

RESULT 4: Items with lower incidence (3/3)

No difference by sex

Men only

0.815

Age-adjusted rate ratio, smokers vs non-smokers

≈

Women only

0.814

Age-adjusted rate ratio, smokers vs non-smokers

Age-adjusted separately by sex, the ratios are almost identical (0.815 / 0.814). Sex composition does not explain the gap.

RESULT 5: Why incidence appears lower

Three candidate hypotheses — none can be proven with this data

1

Care-seeking bias

Incidence counts confirmed diagnoses in claims, not symptoms. If smokers and heavy drinkers tend not to see a doctor for mild illness, no claim is recorded.

2

Survivor bias

People whose health has deteriorated are more likely to quit smoking or drinking, so the "current smoker" group retains relatively healthy people — reverse causation.

3

Response bias

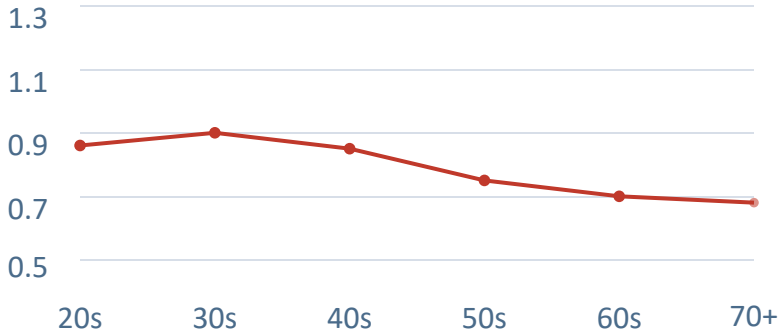
Smoking is self-reported at the check-up. People who smoke but want to avoid health-guidance interviews may answer "No".

None of these can be confirmed or ruled out with the present data — a topic for future research

RESULT 6: Stability across age groups

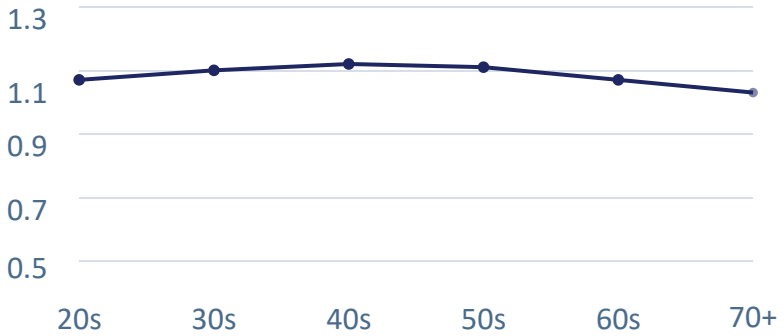
Smoking varies widely

Smoking — smokers vs non-smokers



Range (20s–60s): 0.21 — largest of all items

Sleep — insufficient vs enough



Range (20s–60s): 0.05 — stable

Sex-adjusted ratios within each age group (20s–60s).
Sleep stays within 1.07–1.12; smoking falls with age from 0.86 to 0.70.
70+ excluded from the range because the stratum is small.

RESULT 7: Comparison with other diseases (1/2)

Most items point the same way for all three diseases. Smoking and alcohol reverse for influenza.

Item (vs reference)	Common cold	COVID-19	Influenza
Insufficient sleep	1.09	1.11	1.06
No exercise habit	1.06	1.12	1.08
No daily walking	1.06	1.10	1.07
Daily snacking	1.06	1.13	1.06
Heavy drinking	0.98	0.84	1.08
Current smoker	0.87	0.81	1.09

Example: smoking

Smokers vs non-smokers

0.87
Common cold

0.81
COVID-19

1.09
Influenza

Sleep, exercise, walking and snacking have aRR > 1 for all three diseases. Smoking and heavy drinking are lower or no different for COVID-19 and the common cold, but higher for influenza.

RESULT 7: Comparison with other diseases (2/2)

Strong correlation with the common cold; almost none with influenza

Correlation of sex- and age-adjusted rate ratios across the 11 lifestyle items (16 categories), Pearson's r

COVID-19 vs Common cold

r = 0.88

Very similar pattern

COVID-19 vs Influenza

r = 0.00

Almost unrelated

Several explanations are conceivable (e.g., diagnostic pathway, timing of epidemics), but they cannot be tested with this data.

Summary

Limitations of this analysis

Representativeness

Data come from members of selected health insurance societies (mainly employees and their families) and do not represent the general Japanese population.

Definition of "incidence"

Based on confirmed diagnoses in claims, so infected people who did not see a doctor are excluded. What is measured is "visited and diagnosed", which differs from true incidence.

Temporal ordering

The order of questionnaire response and infection is not tracked (cross-sectional comparison within the same year).

Single year; univariate only

Responses are from 2024 only; results may change over several years. Each item is adjusted for sex and age only; correlations between items (e.g., smoking and drinking) are not adjusted.

Summary

Implications from this analysis

1 **COVID-19 incidence differs by lifestyle item**

Most items point the expected way: insufficient sleep, no exercise habit and no daily walking are all associated with higher incidence. The effect sizes, however, are not large.

2 **Some items point the opposite way — no conclusion from this data alone**

Smoking and heavy drinking show "lower" incidence. Claims-based measurement, survivor bias and response bias remain candidates; none has been demonstrated.