#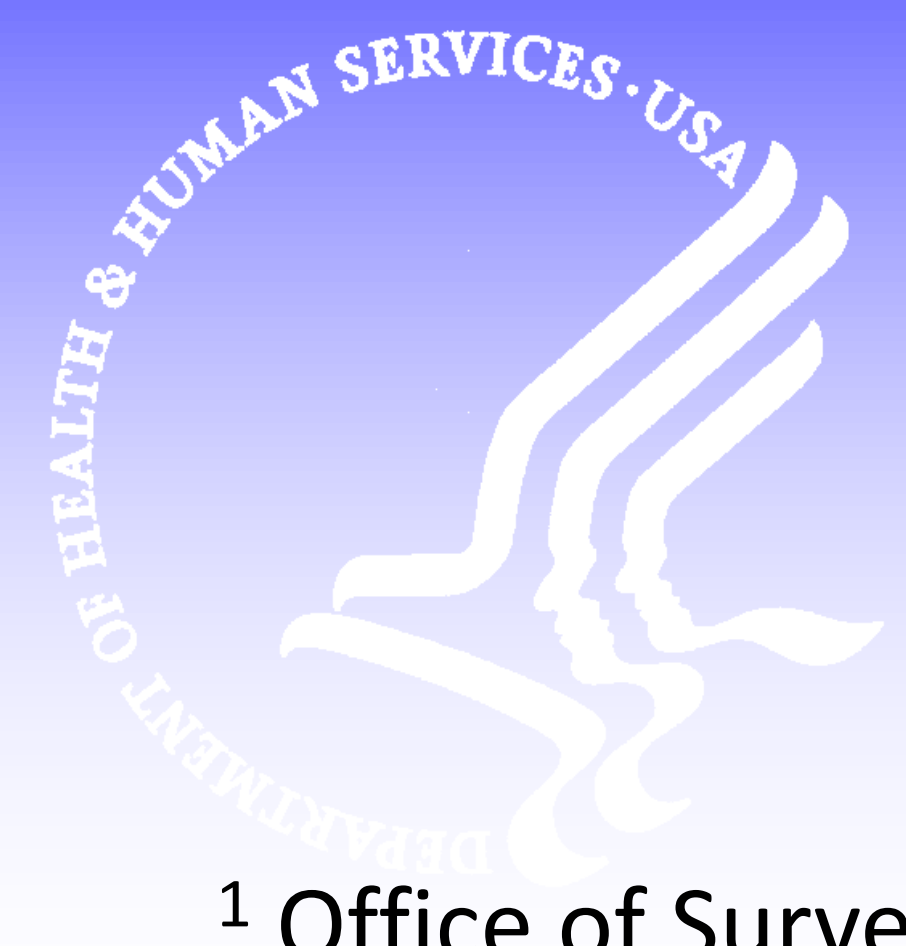 Confounding variable capture in large healthcare administrative claims database: a trend analysis in the Sentinel system

Wei Liu[1], Austin Cosgrove[2], Sarah Dutcher[1], and Justin Bohn[2]

[1] Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, FDA, Silver Spring, MD; [2] Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA

INTRODUCTION

- Use of large healthcare administrative claims databases for observational studies has become an increasingly popular source for comparative effectiveness and safety research.
- Several variables including smoking and obesity/overweight are traditionally considered important confounders in observational studies, yet are reported inconsistently and rarely in administrative claims databases because they generally do not impact health plan's decision for reimbursement.
- There have been anecdotal reports suggesting that the capture of these confounding variables in claims is improving over years.[1,2]

Objective

To conduct a trend analysis in the Sentinel Distributed Database (SDD) to determine how well several confounding variables are captured.

METHODS

- We analyzed healthcare administrative data from 17 Data Partners contributing to the SDD, from 1/1/2006 to 6/30/2018. Study end date varied by data partners from 5/31/2015 to 6/30/2018.
- Descriptive analyses were conducted to examine the prevalence of five confounding conditions, including smoking, obesity, overweight, alcohol abuse or dependence, drug abuse or dependence, which were identified by diagnosis, procedure, and dispensing codes.
- Eligible subjects were required to be continuously enrolled in health plans with medical and pharmacy coverage for at least 365 days before the date of first recording of a confounder of interest during the study time.
- We estimated the prevalence of confounder (e.g., number of enrollees with at least 1 claim for each confounder per 10,000 eligible members), stratified by calendar year, age, and sex.

RESULTS

- The number of eligible beneficiaries in the SDD increased from 15 million in 2007 to the highest of 67 million in 2016. Partial year data were reported in 2018 and the number of data partners contributing to the analyses was small in 2006. Thus, prevalence estimates in both years were deemed as imprecise.
- During 2007-2017, the estimated prevalence of obesity increased 1.9-fold from 69 to 197 per 10K members, paralleled by a 2.5-fold increase in prevalence of overweight from 22 to 78 per 10K members. Prevalence of smoking, alcohol abuse or dependence, drug abuse or dependence increased by 82% (from 65 to 118 per 10K members), 50% (from 14 to 21 per 10K members), and 67% (from 6 to 10 per 10K members), respectively (Fig. 1).
- Similar increasing trend in prevalence of obesity, overweight, and smoking was seen in men and women. In women, the prevalence of obesity was greater than prevalence of smoking after 2010; whereas prevalence of smoking was higher than prevalence of obesity in men. (Fig. 2).
- The transition of the ICD 9 to 10 coding system in Oct/2015 does not seem to change the increasing trend of prevalence estimates, except for drug/alcohol abuse. The spike in and fall in prevalence of drug/ alcohol abuse around the end of 2016 may be due to a change in policies for substance abuse reporting.
- For smoking, obesity, and overweight, their prevalence estimates increased as a function of age. For alcohol and drug abuse, the highest prevalence were observed among subjects aged 44-64 and 18-43 years.
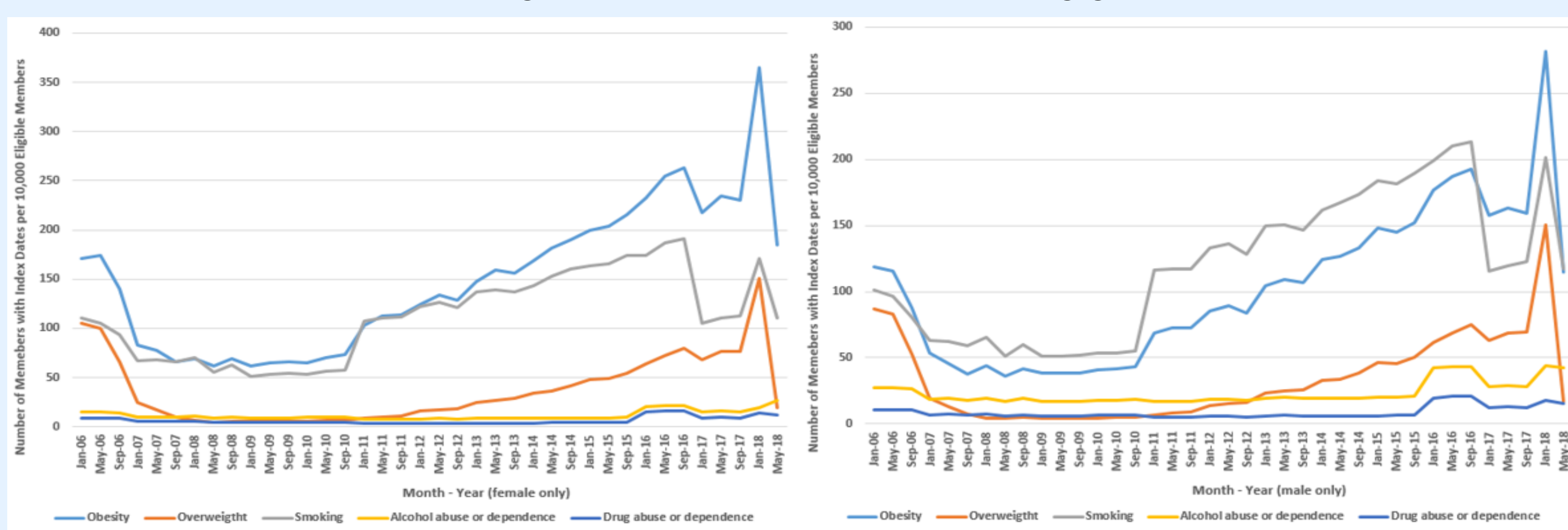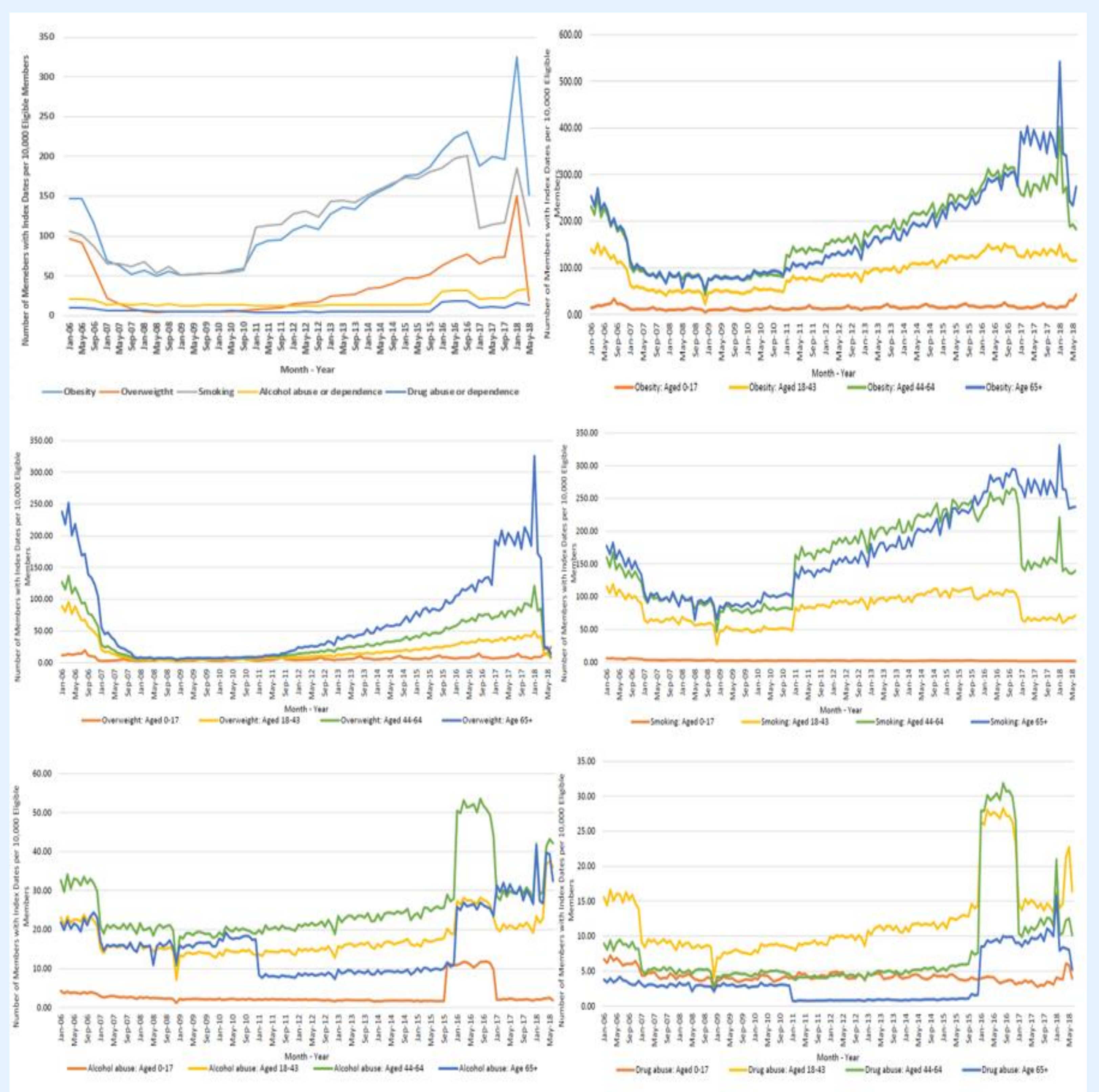
Figure 1. Prevalent capture of confounding conditions in the Sentinel Distributed Database between January 1 2006 and June 30 2018, overall and by age groups

Figure 2. Prevalent capture of confounding conditions in the Sentinel Distributed Database between January 1 2006 and June 30 2018, by year, month and sex

DISCUSSION

- The results presented here underline a continuous increase in the prevalence of identified confounding conditions in administrative claims databases.
- However, these estimates were significantly lower than the estimates obtained from survey based studies. For example, in 2015-2016, the prevalence of obesity was 39.8% in adults,[3] 1 in 3 adults were considered to be overweight in the US.[4] Estimates from the 2017 National Health Interview Survey found the prevalence of current smokers in adults was about 19.3%.[5]

	Sentinel data (2017)	National estimates
Obesity	9.3%	39.8%[3]
Overweight	4.0%	1 in 3 adults[4]
Smoking	6.3%	19.3%[5]
Drug abuse or dependence	0.59%	11.2%[6]
Alcohol abuse or dependence	0.96%	7%[7]

- Unlike in some electronic health record (EHR) systems where sometimes there are government incentives to promote documenting of some of these confounding variables,[8] the absence of firm guidance and financial incentive for recording these confounding variables in claims data may result in their not being measured routinely.[9]
- The Sentinel analysis has several limitations. First, the claims data reflects only diagnosis, procedure, or prescription drugs dispensing that were submitted for reimbursement (e.g., patients received smoking cessation treatment); it does not capture patients who did not seek medical care or medications from healthcare system. Second, the SDD represents primarily commercially insured patients and thus prevalence estimates cannot be generalized to low income population or uninsured individuals.
- In this context, future analysis or interpretation of administrative claims data should consider these possible limitations.

CONCLUSION

A continuous increase in the recording of confounding conditions in claims data was observed; though low prevalence suggests these conditions remain inadequately documented in US claims. Future studies should evaluate whether this represents an actual improved recording of these conditions or an increase in the underlying prevalence of these conditions or a combination of both factors.

References:
1. Huo J, Yang M, Tina Shih YC. Sensitivity of claims-based algorithms to ascertain smoking status more than doubled with meaningful use. Value Health 2018; 21(3): 334-340.
2. George Jaiben, Newman JM, Ramanathan D, et al. Administrative databases can yield false conclusions – an example of obesity in total joint arthroplasty. The J of Arthroplasty 2017; 32: S86-S90. 3. www.cdc.gov/nchs/data/databriefs/db288.pdf
4. https://www.niddk.nih.gov/health-information/health-statistics/overweight-obesity 5. https://www.cdc.gov/mmwr/volumes/67/wr/mm6744a2.htm?s_cid=mm6744a2_w 6. https://www.samhsa.gov/data/report/2017-nsduh-annual-national-report
7. https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/alcohol-facts-and-statistics 8. https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/downloads/9_Record_Smoking_Status.pdf
9. Al Kazzi ES, Lau B, Li T, et al. Differences in the prevalence of obesity, smoking and alcohol in the United States Nationwide Inpatient Sample and Behavioral Risk Factor Surveillance System. PloS One 2015; 10(11): e0140165.