Machine Learning in Healthcare

#L05-Odds and odds ratio

Technion-IIT, Haifa, Israel

Asst. Prof. Joachim Behar
Biomedical Engineering Faculty, Technion-IIT
Artificial intelligence in medicine laboratory (AIMLab.)
https://aim-lab.github.io/
Twitter: @lab_aim

Odds and odds ratio

Interpreting LR

- Let’s take the former example of AF diagnosis:
  - The positive class as AF and negative class as non-AF.
- We learned the corresponding LR model:
  - $h_w(x) = \sigma(w^T x) = \frac{1}{1+e^{-w^T x}}$
  - This means that we have learned the weights w from the dataset.
- How do we interpret the weights value?

Interpreting LR

- Let's write $p = \frac{1}{1+e^{-w^T x}}$
- Then we have:
  - $\frac{p}{1-p} = \frac{1}{e^{-w^T x}}$
  - $\log\left(\frac{p}{1-p}\right) = w^T x$
- Let's define the following quantity $\frac{p}{1-p}$ that we will call the **odds**.
  - p is the probability of AF and $1 - \mathrm{p}$ is the probability of non-AF.
  - $odds = \frac{p}{1-p} = \frac{p(occurding)}{p(not\ occuring)}$ is the likelihood of some event to happen.

Probability versus odds

- If we roll an even dice and look for the chance of obtaining a 4 then we can say that the probability of 4 is 1/6=17% or equivalently that the odds of a 4 is (1/6)/(5/6)=0.2 or 1:5.
  - Probabilities: “The probability of rolling a four is 17%”
  - Odds: “For one roll of a 4 you will have 5 non-4.” The odds is 1:5.
- On the AF example, say $p = 20\%$ for an adult then:
  - Probabilities: “The probability of a patient being AF is 20%”.
  - Odds: “For 1 patient having AF 4 will be non-AF.” The odds is 1:4

Odds and odds ratio

- On the AF example, say $p = 20\%$ for an adult then:
  - Probabilities: "The probability of a patient being AF is 20%".
  - Odds: "For 1 patient having AF 3 will be non-AF." The odds is 1:4
- For an elderly the odds are 1:3. then:
  - Odds ratio = 1/3/(1/4)=1.33
  - Odds ratio: "The odds of getting AF for an elderly person are 1.33 greater than an adult one."
  - **Odds ratio:** ratio of two odds!

Back to LR

- If we assume a single feature, say body mass index (BMI) then:
  - $\log\left(\frac{p}{1-p}\right) = \log(odds) = w_0 + w_1 \cdot BMI$

Interpreting LR

- If we write this equation for a given value of BMI and one increment i.e. $BMI\ +1$.
  - (1) $\log(odds_{BMI}) = w_0 + w_1 \cdot BMI$,
  - (2) $\log(odds_{BMI+1}) = w_0 + w_1 \cdot (BMI + 1)$.
  - $(2) - (1) = \log(odds_{\mathrm{BMI+1}}) - \log(odds_{BMI}) = w_1$
    - Thus w_1 corresponds to the difference between the log odds for one unit increase in BMI.
  - $e^{w_1} = \frac{odds_{BMI+1}}{odds_{BMI}} = \textbf{odds ratio}$
    - **Odds ratio**: ratio of two odds! The odds ratio for a variable in LR represents how the odds change with one unit increase in that variable holding all other variables constant.

BMI	Weight status
Below 18.5	Underweight
18.5-24.9	Normal weight
25.0-29.9	Overweight
30.0-34.9	Obesity class I
35.0-39.9	Obesity class II
Above 40	Obesity class III

Interpreting LR

- Say $e^{w_1} = 1.15$, how do we read that?
  - "For one unit increase in BMI the odds of having AF will increase by 15%."
  - By how much will the odds increase for 5 units of BMI?
  - Is that true at any point in the weight spectrum?

Interpreting LR

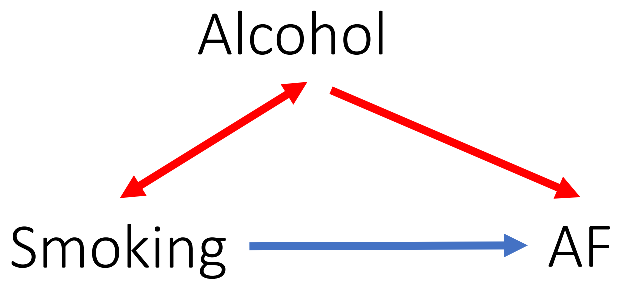
- Why do we use odds and odds ratio?
  - We use the concept of odds and odds ratio here to provide some interpretation to the weights we have learned in the LR model.
- What if we standardize features?
  - In this case the “one unit increase” will read as “one standard deviation increase”.

Interpreting LR

- What about if you have many features?
  - $log(odds) = w_0 + w_1 \cdot BMI + w_2 \cdot Age + w_3 \cdot Smoking + w_4 \cdot Alcohol$
  - How the odds change with one unit increase in one variable while holding all other variables constant.

Interpreting LR

- We may include many features to learn about the association between them and the outcome but also to correct for **confounding** variables. E.g.



Interpreting LR

Contents lists available at ScienceDirect

Clinical Nutrition

journal homepage: http://www.elsevier.com/locate/clnu

Original Article

Chocolate consumption is inversely associated with prevalent coronary heart disease: The National Heart, Lung, and Blood Institute Family Heart Study

Luc Djoussé [a,b,*], Paul N. Hopkins [c], Kari E. North [d], James S. Pankow [e], Donna K. Arnett [f], R. Curtis Ellison [g]

[a] Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[b] Massachusetts Veterans Epidemiology and Research Information Center and Geriatric Research, Education and Clinical Center, Boston Veterans Affairs Healthcare System, Boston, MA, USA
[c] Cardiovascular Genetics, University of Utah, Salt Lake, UT, USA
[d] Department of Epidemiology, School of Public Health, University of North Carolina, Chapel Hill, NC, USA
[e] Division of Epidemiology and Community, University of Minnesota, Minneapolis, MN, USA
[f] Department of Epidemiology, University of Alabama, Birmingham, AL, USA
[g] Section of Preventive Medicine & Epidemiology, Evans Department of Medicine, Boston University School of Medicine, Boston, MA, USA

Table 2
Prevalence odds ratios (95% confidence intervals) of coronary heart disease according to chocolate consumption in 4970 participants in the NHLBI Family Heart Study[a].

Frequency of chocolate intake	Cases/N	Crude	Model 1[b]	Model 2[c]
0	168/1093	1.0	1.0	1.0
1–3 per month	147/1167	0.79 (0.62–1.01)	1.01 (0.76–1.37)	1.05 (0.77–1.43)
1–4 per week	182/1931	0.57 (0.46–0.72)	0.74 (0.56–0.98)	0.75 (0.56–1.01)
5+ per week	43/779	0.32 (0.23–0.45)	0.43 (0.28–0.67)	0.43 (0.27–0.68)
P for linear trend		<0.0001	<0.0001	0.0002

[a] Coronary heart disease was defined as history of myocardial infarction, PTCA, or CABG.
[b] Adjusted for age, sex, and risk group (random vs. high risk) using generalized estimating equations (GEE).
[c] Variables in Model 1 plus additional adjustment for dietary linolenic acid, education, exercise (min/d), smoking (yes/no), alcohol intake (yes/no), fruit and vegetables, energy intake, and non-chocolate candy (4 groups) consumption.

AIMLab.

Interpreting LR

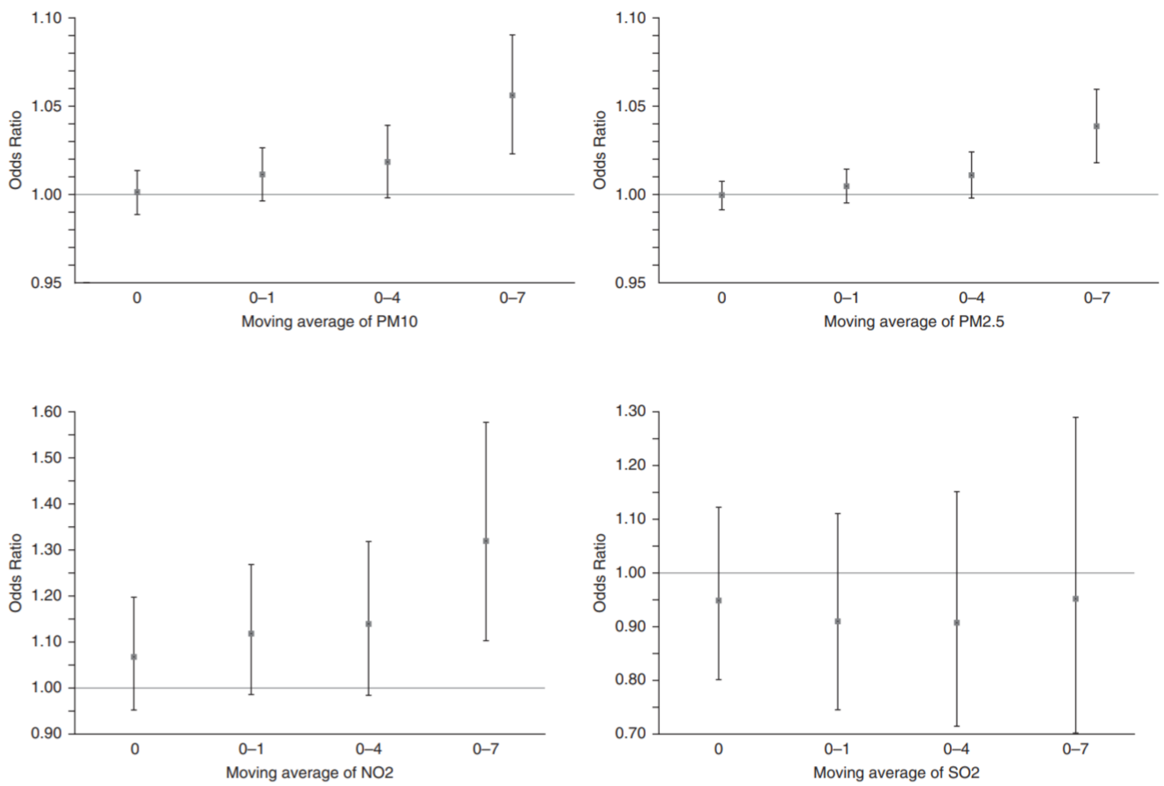


Figure 1. The association between air pollution and bronchiolitis: case–cross-over data showing the results of single-pollutant models of the association between bronchiolitis and moving average concentrations of the pollutants 0, 0–1, 0–4, and 0–7 days before the event. Each exposure was modeled separately and was adjusted for the average temperature that corresponds to the period of the exposure tested. Results are presented as odds ratios and 95% confidence intervals for interquartile range increase in particulate matter (PM) less than 10 μm (PM_{10}; 25 μg/m^3), PM less than 2.5 μm ($PM_{2.5}$; 5 μg/m^3), nitrogen dioxide (NO_2; 13 μg/m^3), or sulfur dioxide (SO_2; 3.14 μg/m^3).

ORIGINAL RESEARCH

Air Pollution and Hospitalization for Bronchiolitis among Young Children

Maayan Yitshak-Sade[1,2]*, Dror Yudovitch[1]*, Victor Novack[1,2], Asher Tal[3], Itai Kloog[4], and Aviv Goldbart[3]

[1]Clinical Research Center, Soroka University Medical Center, Beer Sheva, Israel; [2]Department of Medicine, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer Sheva, Israel; [3]Department of Pediatrics, Soroka University Medical Center, Beer Sheva, Israel; and [4]Department of Geography and Environmental Development, Faculty of Humanities and Social Sciences, Ben-Gurion University of the Negev, Beer Sheva, Israel

Abstract

Rational: Several studies have found higher risks for childhood respiratory illness, associated with exposure to particulate matter (PM) less than 10 μm in diameter (PM_{10}) and $PM_{2.5}$ and gaseous pollution.

Objectives: We analyzed the association between air pollution and hospitalizations due to bronchiolitis, an obstructive pulmonary disorder, commonly caused by respiratory syncytial virus infant infection.

Methods: Data were obtained from a local tertiary medical center providing services for a population of 700,000 comprising two ethnic groups: predominantly urban Jews and rural Bedouin Arabs. The latter

Results: We identified 4,069 bronchiolitis hospitalizations (3,889 children), with 55.3% being Bedouin Arabs, of whom 16.8% resided in temporary dwellings. An increase in interquartile range of average weekly air pollutants was associated with an increased odds of bronchiolitis (odds ratio [95% confidence interval]): PM_{10} (1.06 [1.02–1.09]), $PM_{2.5}$ (1.04 [1.02–1.06]) and nitrogen dioxide (1.36 [1.12–1.65]). Higher effect-estimates for PM were observed among Bedouin Arabs residing in temporary dwellings (1.14 [1.01–1.30] and 1.07 [1.01–1.15]) compared with Jewish individuals (1.05 [0.99–1.11] and 1.03 [1.01–1.07]) and other Bedouin Arabs (1.05 [1.01–1.10] and 1.03 [1.01–1.07]), and among males (1.11 [1.06–1.16] and 1.06 [1.03–1.09]) compared with females (0.99 [0.94–1.05] and 1.01 [0.97–1.04]).

Yitshak-Sade, Maayan, et al. "Air pollution and hospitalization for bronchiolitis among young children." Annals of the American Thoracic Society 14.12 (2017): 1796-1802.

Take home

- Probability versus odds.
  - Odds: likelihood of some event to happen.
  - Odds ratio: ratio of the odds.
- Interpretation of LR coefficients.
  - The odds ratio for a variable in LR represents how the odds change with one unit increase in that variable holding all other variables constant.

References

[1] Szumilas, Magdalena. "Explaining odds ratios." Journal of the Canadian academy of child and adolescent psychiatry 19.3 (2010): 227.