Stat 27850/30850: real data analysis critique

This is a group assignment—please work in groups of size 2, 3, or 4.
Only one member from each group should submit the assignment to Canvas.

Select one of the following papers to critique:

1. "Optimism is associated with exceptional longevity in 2 epidemiologic cohorts of men and women", Lee et al, PNAS 2019
   `https://www.pnas.org/content/116/37/18357`
2. "Lung cancer incidence decreases with elevation: evidence for oxygen as an inhaled carcinogen", Simeonov & Himmelstein, PeerJ 2015
   `https://peerj.com/articles/705/`
3. "Adolescents in Peer Groups Make More Prudent Decisions When a Slightly Older Adult Is Present", Silva et al, Psychological Science 2015
   `http://pss.sagepub.com/content/early/2016/01/14/0956797615620379.full.pdf+html`

Please note that you may need to be on campus to access the full text of the articles (or log in through UChicago Library).

These papers are chosen as recent research across a range of health & psychology disciplines; they are chosen to represent typical studies, not necessarily as examples of good or bad statistical methodology. Your job is to critique the statistical analyses performed in the paper you choose from the point of view of multiple testing and replicability. You should also assess the validity of the conclusions from a statistician's point of view: think about possible confounding variables, whether the data collected is representative of the claim being made, etc.

Some questions you might think about:

- Are there explicit issues of multiple comparisons—e.g. looking for effects of multiple variables on the response? If so, were these accounted for in the analysis?
- Are there implicit issues of multiple comparisons, where the analyses that were run, were chosen by examining the data? Can you think of a hypothesis which was not tested, which is equally plausible to the one that was tested?
- Are there many "degrees of freedom" in the choices made for the analysis, such as which subjects to exclude, how to partition variables into categories, which variables to control for, etc? Were these chosen in advance or as part of the analysis?
- Statistical issues not related to multiple comparisons, e.g. possible confounding variables, representativeness of the population, issues of study design, dropping some individuals or observations from the data set, etc.
- And any other relevant issues that come to mind.

What you should hand in: a short report (e.g. 2–4 pages, but this is flexible) that summarizes the study (including methods, who was in the study, main questions asked, etc), then discusses any relevant points or concerns regarding multiple comparisons or validity of the statistical analyses in the paper. You are encouraged to consider designing a simulation to illustrate any possible issues that you identify.