Diabetes and Prediabetes Prevalence Estimation

By:

Alex Peterson [petersonale@mail.missouri.edu](mailto:petersonale@mail.missouri.edu)

Andrew Brumitt [acb7w6@mail.missouri.edu](mailto:acb7w6@mail.missouri.edu)

Justin Gallina [jag9t5@mail.missouri.edu](mailto:jag9t5@mail.missouri.edu)

Sampling Techniques, STAT 4310, Chong He

*Abstract --* The National Center for Health Statistics (NCHS), part of the Center for Disease Control (CDC), performed a study titled “Diabetes and Prediabetes Among Adults in Los Angeles County and the United States, 1996-2004 and 2007-2014,” which examined the prevalence of diagnosed diabetes, undiagnosed diabetes, prediabetes, and total diabetes in hopes of discovering an increase in persons over the age of twenty from the years 1999-2006 to 2007-2014. Their findings were published on April 19, 2019. Data for this study is from in-home interviews and laboratory testing conducted as a part of the National Health and Nutrition Examination Survey (NHANES). This study was performed in Los Angeles County with constructed sampling weights so that estimates accurately represent the county. Prevalence and 95% confidence intervals were created for diagnosed, undiagnosed, total diabetes, and prediabetes. These were then compared with sample data collected from the entire U.S.. The results of this survey show that Los Angeles County is a good estimator for the U.S. as a whole. It is also found that total diabetes increased by a statistically significant amount between 1999-2006 and 2007-2014. The difference between age groups is also proven to be statistically significant. Monitoring these trends in diabetes of adults in Los Angeles County may provide guidance for future actions taken towards diabetes prevention and control.

**Keywords:** diabetes -- plasma fasting glucose -- Los Angeles County -- National Health and Nutrition Examination Survey (NHANES)

**Introduction**

*Motivation --* Diabetes is a major cause of mortality and disability in the United States. The NCHS is involved with finding trends than can provide insights to improve the health of the citizens of the United States. The NCHS investigated a hypothesis that diabetes and prediabetes are becoming more prevalent among Americans. Their goal within this study is to find if there is significant data to support the theory that diabetes is becoming more common among citizens of the country.

*Research Problem --* There are two things which must be proven for this survey. One: does the sample survey significantly represent the target population (United States)? and Two: does the sample survey show that the prevalence of diabetes and prediabetes is significantly greater in 2007-2014 compared to 1999-2006?

*Proposed Approach --* The sample survey is conducted by cluster sampling inside Los Angeles County with sampling weights based on age group (20-39;40-59;60+). Furthermore, the survey has three stages of sampling: segments within LA County, dwelling units within segments, persons within dwelling units. The measurement methods consist of participants being randomly assigned to a morning or afternoon session. Those that were sampled in the morning were asked to fast for between 8 and 24 hours. Participants that were sampled in the afternoon were not asked to fast, and they were not included in the overall survey data due to this inconsistency. Participants were classified as having diabetes if they answered “yes” to the question “Other than during pregnancy, have you ever been told by a doctor or health professional that you have diabetes or sugar diabetes?”. Diabetes and prediabetes were defined by the Current American Diabetes Association. It was also reported that the study oversampled Hispanic persons due to the large Hispanic population in Los Angeles County and the sampling weights they applied to the sampled population.[[1]](#footnote-0)

*Analysis --* The analysis methods used was a two sample t-test v.s. U.S. sample data in order to determine if the differences between Los Angeles County and the U.S. data were significant. This test result was that the sample is significant to predict the U.S.. Then, a univariate 2-sided t-test was performed on the differences in prevalence for all adults in 1999-2006 compared with 2007-2014, as well as differences in prevalence by sex and race in 2007-2014. Linear trends across age categories were tested using logistic regression. Reliability of the estimates was tested by the NCHS standards for effective sample size, degrees of freedom, confidence interval width, and relative confidence interval width. The oversampling of Hispanics in this study results in a wide 95% confidence interval. Because of the low statistical power in subgroups, differences between race, age, and sex in Los Angeles County vs U.S. was not tested statistically.[[2]](#footnote-1) [[3]](#footnote-2) [[4]](#footnote-3)

*Results* --

* The age-adjusted prevalence of **diagnosed diabetes** in Los Angeles County was 7.5% in 1999-2006 and increased to 10.2% in 2007-2014. In the U.S. diagnosed diabetes was 7.2% in 1999-2006 and increased to 8.6% in 2007-2014.
* The age-adjusted prevalence of **undiagnosed diabetes** in Los Angeles County between 1999–2006 (3.0%) and 2007–2014 (4.3%). In the United States, the age-adjusted prevalence of undiagnosed diabetes was similar in 1999–2006 (3.5%) and 2007–2014 (3.8%).
* Among **adults with diabetes** in 2007-2014: **Undiagnosed diabetes** was recorded at 29.9% in Los Angeles County and 30.6% in the U.S.
* In Los Angeles County, the age-adjusted prevalence of **total diabetes** increased from 10.5% in 1999–2006 to 14.4% in 2007–2014. In the United States, it went from 10.7% to 12.4%
* In Los Angeles County, the age-adjusted prevalence of **prediabetes** was 37.3% in 1999–2006 and 40.2% in 2007–2014, and in the United States it went from 37.2% to 42.4%
* The difference between the prevalence of diagnosed diabetes, undiagnosed diabetes, total diabetes, or prediabetes in Los Angeles County vs U.S. was **NOT statistically significant**.
* In Los Angeles County, the prevalence of diagnosed diabetes increased with age from 0.3% among adults aged 20–39 to 14.9% among those aged 40–59 and 19.4% among adults aged 60 and over.
* The difference between Men and Women was **NOT** deemed statistically significant.

**Related Works**

A 2015 study by Menke et al. titled “Prevalence of and Trends in Adult Obesity in the United States” is a similar study that aimed to determine if there was a significant increase in obesity as a whole and among varying subgroups within the United States over different periods of time ranging from 1988-2012. This study came to similar conclusions that the survey by Hales et al. determined in terms of the overall increase being seen in diabetes rates in each progressive period of time that was studied. They found that the age-adjusted prevalence for diabetes increased from 9.8% in 1988-1994, to 10.8% in 2001-2002, and was finally recorded at 12.4% in 2011-2012.[[5]](#footnote-4)

The publication that much of the study by Hales et al. was compared with was the 2017 National Diabetes Statistics Report, published by the Center for Disease Control (CDC), which was a large, highly-funded examination of the prevalence of diabetes in America, which estimated that 9.4% of the American population, or more than 30.3 million people, suffer from diabetes. Further, the NDSR shows that there are differences in diabetes prevalence among subgroups. Non-Hispanic Asians had an overall diabetes prevalence of 16.0%, non-Hispanic blacks were sampled at 17.7%, Hispanic populations at 16.4%, with non-Hispanic whites having only a 9.3% diabetes rate, based on the data from 2011-2014.[[6]](#footnote-5)

**Our Approach**

Our approach would include slightly changing the methods used in the survey by Hales et al. in the National Health Statistics Report to only sample those that have been fasting for 8-24 hours so that we are not hindering the power of our sample to be able make conclusions about important subgroups. Half of the participants sampled by the survey had their examination and testing done in the afternoon or evening, which resulted in their blood test being unfit for use. As a result, we are unable to gather their samples’ plasma fasting glucose level (in mg/mL) in the blood of the patients, which is how we determine the patients’ status as having undiagnosed diabetes, prediabetes, or not having diabetes. As the report states, this resulted in the NHSR not being able to make conclusions about the significance of diabetes prevalence in various subgroups. Had such conclusions been made, the CDC would be able to more effectively target at-risk groups based on age, race, and sex with prevention and control efforts to combat diabetes. In addition, the sampling weights for Hispanic Americans and Mexican-Americans was admittedly high, despite the fact that Los Angeles County already contains large populations of these oversampled groups. By adjusting the sampling weights so that they more accurately reflect the demographic proportions of all Americans would make our confidence intervals for other groups, such as non-Hispanic whites and blacks, more narrow and thus more precise. As for additional analysis, we will be evaluating two sample t-tests on sex, race, and age. Similarly to the original study, we will also be comparing these results with data collected from the whole U.S.. If these findings are conclusive, it would result in future spending being cut by a large amount since the entire U.S. would not need to be sampled but rather estimated using Los Angeles County. The findings would also aid researchers even further than the original study on who to specifically target as high risk for diabetes and prediabetes so that measures of health concern may be taken.

**Conclusion**

In conclusion, the study that we chose to analyze seems to provide results that are fairly reliable. However, there were a few factors that provided skewed results. The primary factor being that the races of people were not sampled in equal weight so the results ultimately show very reliable data for Hispanics and diabetes but not necessarily the population of the United States as a whole. Therefore, we would conclude that using sampling weights and balancing the sampling groups would be the best way to provide more accurate results. In the future this is what we would do should we choose to conduct a study of the same kind.

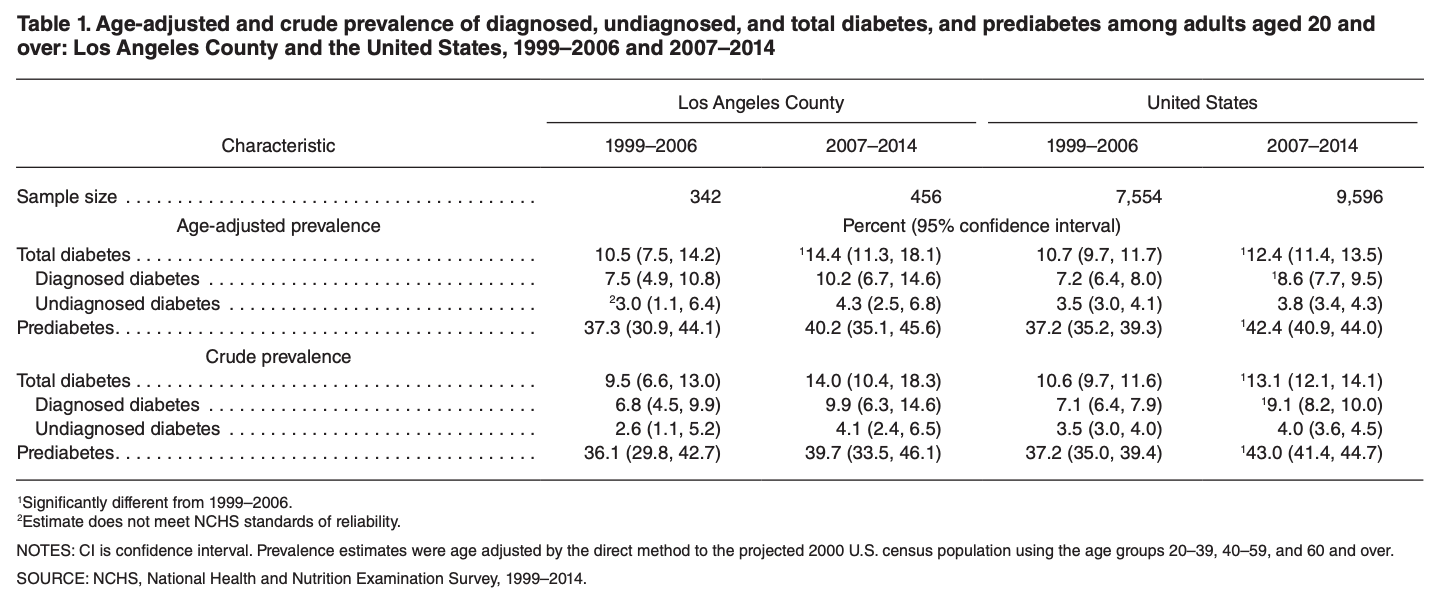
**References**

1. Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2017.   
    Atlanta, GA: Centers for Disease Control and Prevention, U.S. Dept of Health and   
    Human Services; 2017.
2. Johnson CL, Dohrmann SM, Burt VL, Mohadjer LK. National Health and Nutrition   
    Examination Survey: Sample design, 2011–2014. National Center for Health   
    Statistics. Vital Health Stat 2(162). 2014
3. Klein RJ, Schoenborn CA. Age adjustment using the 2000 projected U.S. population.   
    Healthy People 2010 Stat Notes no 20. 2001.
4. Kish L. Chapter 12, Special selection techniques. In: Survey sampling. New York: John   
    Wiley & Sons, Inc., 440–508. 1995.
5. Korn EL, Graubard BI. Confidence intervals for proportions with small expected number   
    of positive counts estimated from survey data. Surv Methodol 24(2):193–201.   
    1998.
6. Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and Trends in Diabetes   
    Among Adults in the United States, 1988-2012. *JAMA.* 2015;314(10):1021–1029.

**Reference to Full Report**

1. Hales, Craig M, et al. “Diabetes and Prediabetes Among Adults in Los Angeles County and the United States, 1999–2006 and 2007–2014 .” San Diego, California, 19 Apr. 2019.

**Appendix**



(From the National Health Statistics Report, 2019)

1. Johnson CL, Dohrmann SM, Burt VL, Mohadjer LK. National Health and Nutrition Examination Survey [↑](#footnote-ref-0)
2. Klein RJ, Schoenborn CA. Age adjustment using the 2000 projected U.S. population. Healthy People 2010 Stat Notes no 20. 2001. [↑](#footnote-ref-1)
3. Kish L. Chapter 12, Special selection techniques. In: Survey sampling. New York: John Wiley & Sons, Inc., 440–508. 1995. [↑](#footnote-ref-2)
4. Korn EL, Graubard BI. Confidence intervals for proportions with small expected number of positive counts estimated from survey data. Surv Methodol 24(2):193–201. 1998. [↑](#footnote-ref-3)
5. Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and Trends in Diabetes Among Adults in the United States, 1988-2012. *JAMA.* 2015;314(10):1021–1029. [↑](#footnote-ref-4)
6. Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2017. Atlanta, GA: Centers for Disease Control and Prevention, U.S. Dept of Health and Human Services; 2017. [↑](#footnote-ref-5)