Comparison of cardiometabolic risk biomarkers from a national clinical laboratory with the US adult population

Elliott Wolin, PhD, Justin White, MS, James V. Pottala, PhD,
Maciek Sasinowski, PhD, MD, Tara Dall, MD, Thomas D. Dayspring, MD,
Joseph P. McConnell, PhD, Daniel M. Hoefner, PhD, Stephen A. Varvel, PhD,
Dawn L. Thiselton, PhD, G. Russell Warnick, MS, William S. Harris, PhD*

Department of Research and Development, Health Diagnostic Laboratory, Inc., Richmond, VA, USA (Drs Wolin, White, Pottala, Sasinowski, Dall, Dayspring, McConnell, Hoefner, Varvel, Thiselton, Warnick, Harris); and Department of Internal Medicine, Sanford School of Medicine, University of South Dakota, Sioux Falls, SD, USA (Drs Pottala, Harris)

KEYWORDS:
US population;
NHANES;
Lipids;
Clinical laboratory;
Biomarkers;
Cardiovascular risk

BACKGROUND: Clinical laboratory patient databases are an untapped source of valuable diagnostic and prognostic information. However, the lack of associated clinical and/or demographic information and questionable generalizability to nonpatient populations often limit utility of these data.

OBJECTIVES: This study compared levels of cardiometabolic biomarkers between a national clinical laboratory patient cohort (Health Diagnostic Laboratory [HD Lab]) and the US population as inferred from the National Health and Nutrition Examination Survey (NHANES, 2011–2012).

METHODS: Sample sizes for HD Lab ranged from 199,000 to 739,000 and for NHANES from 2200 to 5300. The latter were weighted to represent the adult US population ($\sim$220 million). Descriptive statistics were compared for body mass index, 5 lipid biomarkers, and 3 glycemic biomarkers.

RESULTS: Using age- and sex-matched data, mean biomarker values (mg/dL unless noted) and percent differences (%) for HD Lab vs NHANES were body mass index (kg/m^2), 29.1 vs 28.6 (1.7%); total cholesterol, 185 vs 193 (–4.1%); apolipoprotein B, 92 vs 90 (2.2%); low-density lipoprotein cholesterol, 107 vs 115 (–7%); high-density lipoprotein cholesterol, 53 vs 53 (0%); triglycerides, 128 vs 127 (0.8%); glucose, 99 vs 108 (–8.3%); insulin (uU/mL), 13.7 vs 13.4 (2.2%); and hemoglobin A1c (%), 5.6 vs 5.8 (–3.4%). Although all differences were statistically significant, only low-density lipoprotein cholesterol and glucose differed by more than 5%. These may reflect a greater use of medications among HD Lab patients and/or preanalytical factors.

CONCLUSIONS: Cardiometabolic risk markers from a national clinical laboratory were broadly similar to those of the US population; thus, with certain caveats, data from the former may be generalizable to the latter.
© 2015 National Lipid Association. All rights reserved.

* Corresponding author. Health Diagnostic Laboratory, 737 N. 5th St, Suite 103, Richmond, VA 23219.

E-mail address: bharris@hdlabinc.com

Introduction

The use of "big data" in health care has surged in recent years, allowing consolidation and use of new types of data in nontraditional ways. Big data is a general term that describes data sets that are so large, complex, and/or unstructured that they challenge traditional analytic approaches. Fueled in part by the proliferation of electronic medical record systems, electronic medical claim submission systems, and an increasingly large amount of health information managed online, health care organizations are experimenting with new methods and techniques to extract value from their data. One such nontraditional source of data with great potential value can be found in clinical laboratories (CLs) across the United States. CLs receive millions of blood samples each year and generate billions of data elements on known and emerging biomarkers. The rationale for using very large data sets (hundreds of thousands to millions of samples) from CLs for discovery of biomarker associations and/or correlations and the identification of rare diseases has been previously described by others.[1]

Although such data sets have obvious limitations (eg, limited demographic and clinical information, such as medical history and current medication status) and thus are inappropriate for many purposes, several ways of extracting clinically meaningful value have been demonstrated–particularly useful when data are not yet available for large "healthy" populations. For example, defining age- and sex-based population norms for novel biomarkers such as red blood cell fatty acids[2] or noncholesterol sterols,[3] describing the extent of discordance between different measures of atherogenic lipoprotein particles (and of the association of that discordance with insulin resistance)[4,5] and reporting sex-based differences in lipoprotein subclass distributions.[6] Another recent "big data" study demonstrated that red blood cell omega-3 fatty acid status and lipid parameters did not vary by apolipoprotein E (*APOE*) genotype[7] suggesting that the use of fish oil supplements may not need to be restricted in patients carrying the *APOE* ε4 allele as had previously been recommended.[8]

Most CL data–based studies to date have assumed, either implicitly or explicitly, that these data were reflective of the population as a whole, but that assumption has not been directly examined. There is a potential selection bias problem, that is, patients being seen by health care providers, presumably for diagnosis and treatment of disease, may not reflect "typical" Americans. Hence, even if more patient-specific data were available from CLs, whether the information deduced from it would be generalizable to the wider population is unclear. Indeed, although very basic information is typically available (eg, age, sex, and occasionally body mass index [BMI]), data on disease comorbidities, medical history, physical activity, socioeconomic status, drug usage, diet, family history, and race are typically not. The purpose of this study was to address this potential selection bias limitation by comparing 8 typical blood lipid and glycemic control biomarkers from samples submitted to a national CL with those collected as part of the National Health and Nutrition Examination Survey (NHANES).

Methods

Subjects

NHANES cohort

The NHANES Survey is performed biannually by the US Center for Disease Control. Survey data are made available to the public in 2-year releases, the latest being for 2011–2012, in which a nationally representative sample of US adults and children (n = 9756) underwent a thorough physical examination, interview, and laboratory testing. NHANES uses a complex survey design to allow inference from the sample of subjects to the full civilian, noninstitutionalized US population. Subjects are selected based on gender, age, ethnicity, and geographic location, with subpopulations being undersampled or oversampled and carefully weighted. For each individual biomarker of interest (glucose, insulin, triglycerides [TGs], low-density lipoprotein cholesterol [LDL-C], high-density lipoprotein cholesterol [HDL-C], apolipoprotein B [apoB], hemoglobin A1c [HbA1c], and total cholesterol), the following exclusion criteria were applied to arrive at a final analytical data set: value missing, sex missing, not fasting (applied to the first 4 biomarkers listed previously), weighting value missing, or age <20 years. It was not mandatory that each NHANES subject provide data for all lipid and glycemic biomarkers of interest. Accordingly, the sample sizes ranged from about 5300 for demographics to 2200 for some blood markers. The NHANES program is further described here: http://www.cdc.gov/nchs/nhanes.htm.

National clinical laboratory data set

Data from Health Diagnostic Laboratory, Inc (HD Lab, Richmond, VA) were used in this study. Deidentified patient data from the first blood sample analyzed during 2011–2012 were used. All samples tested for glucose, insulin, LDL-C, and TG were from subjects who had fasted for a minimum of 8 hours. The original data set included about 739,000 unique patients. As was the situation for the NHANES survey, not all data were available for all patients; hence, the sample sizes ranged from approximately 739,000 for demographics, down to 424,000 for some lipids, and 199,000 for some glycemic markers.

Laboratory methods

Compatible measurement methods were used for lipid and glycemic markers measured by NHANES and HD Lab. The test methodologies used by NHANES are described on their Web site (http://wwwn.cdc.gov/nchs/nhanes/search/nhanes11_12.aspx). For the HD Lab cohort, all biomarkers