**Letters**

**Comments and Responses**

**Drawing Conclusions about Short-Term Variability in Liver Function Test Results**

TO THE EDITOR: We read with great interest the recent study by Lazo and colleagues (1) on the short-term variability of various biochemical liver tests and want to share some concerns. First, this study is based on a nonrandom, convenience sample that represents only 9.5% (1864 out of 19618) of the original NHANES (National Health and Nutrition Examination Survey) study population, raising questions about the generalizability of the findings, given the strong likelihood for selection bias. Second, the very small percentage of persons from racial or ethnic groups other than white, black, and Hispanic restricts the applicability of the results, particularly for Asians.

The cutoff values of “normal” levels of serum alanine aminotransferase (ALT) that were used in this study—40 IU/L for men and 31 IU/L for women—are probably overestimates, as has been shown in 2 large studies from Italy and Korea (2, 3). These studies recommend using cutoff values of 30 IU/L for men and 19 IU/L for women. On the basis of these recommended cutoff values, a substantial percentage of patients in Lazo and colleagues’ study classified as having elevated ALT levels at examination 1 (median, 43 IU/L) who returned to normal in examination 2 (median, 27 IU/L) would probably still have abnormal levels of ALT. These distinctions are not trivial, as evidenced by the increased risk for death in individuals with ALT levels greater than 20 IU/L compared with those with ALT levels less than 20 IU/L (relative risk for patients with ALT levels of 20 to 29 IU/L, 2.9; relative risk for patients with ALT levels of 30 to 39 IU/L, 9.5) (3). The American Association for the Study of Liver Diseases has also called for recalibration of the normal range for ALT level (4).

Finally, the Gilbert syndrome, a genetic disease with a prevalent homozygosity of 9% in the Western population, can result in values outside the normal range of total bilirubin level, varying with fasting status or stress. This may affect the determination of normal versus abnormal values in Lazo and colleagues’ study.

Gaurav Arora, MD
George Triadafilopoulos, MD
Stanford University School of Medicine
Stanford, CA 94304

**Potential Financial Conflicts of Interest:** None disclosed.

**References**

1. Lazo M, Selvin E, Clark JM. Brief communication: clinical implications of short-term variability in liver function test results. Ann Intern Med. 2008;148:348-52. [PMID: 18316753]
2. Prati D, Taioli E, Zanella A, Della Torre E, Butelli S, Del Vecchio E, et al. Updated definitions of healthy ranges for serum alanine aminotransferase levels. Ann Intern Med. 2002;137:1-10. [PMID: 12093239]
3. Kim HC, Nam CM, Jee SH, Han KH, Oh DK, Suh I. Normal serum aminotransferase concentration and risk of mortality from liver diseases: prospective cohort study. BMJ. 2004;328:983. [PMID: 15028636]
4. American Association for the Study of Liver Diseases. Know your ALT [press release]. 30 October 2006. Accessed at https://www.aasl.org/eweb/docs/meetings/06annualmeeting/pressadvisoryALT.pdf on 6 May 2008.
Discordant test results have the same long-term outcome as those with abnormal liver test results on a first determination be routinely retested before undergoing further evaluation.

In our opinion, caution should be used in drawing practice recommendations from epidemiologic studies. We are particularly concerned that the proposed strategy could be highly misleading in a clinical setting. Liver tests results, particularly aminotransferases and \( \gamma \)-glutamyltransferase, typically fluctuate in patients with chronic liver disease (2). When evaluating a patient, even an asymptomatic one, with abnormal liver biochemistries, clinicians should interpret results according to the clinical context and consider an adequate work-up (2, 3). A repeated value in the normal range does not ensure that the initial value was truly erroneous.

In addition, substantial evidence indicates that high aminotransferase values are statistically significantly correlated with increased future mortality, suggesting that these blood tests are valuable indicators of long-term prognosis (4, 5). How should one differentiate between a clinically insignificant fluctuation of normality and a predictor of mortality in a single patient? To support their recommendations, Lazo and colleagues should have noted that patients with 2 discordant test results have the same long-term outcome as those with 2 concordant normal test results. Otherwise, normalization cannot be defined as proof of normality.

Lazo M, Selvin E, Clark JM. Brief communication: clinical implications of short-term variability in liver function test results. Ann Intern Med. 2008;148:348-52. [PMID: 18316753]

TO THE EDITOR: The recent article by Lazo and colleagues (1) shows that in the context of an epidemiologic survey, such as NHANES, abnormal liver test results in more than one third of participants (levels of aspartate aminotransferase, ALT, alkaline phosphatase, \( \gamma \)-glutamyltransferase, and bilirubin) would be reclassified as normal if retested 17 days apart. On the basis of their findings, Lazo and colleagues recommend that, to avoid unnecessary testing, individuals with abnormal liver test results on a first determination be routinely retested before undergoing further evaluation.

In our opinion, caution should be used in drawing practice recommendations from epidemiologic studies. We are particularly concerned that the proposed strategy could be highly misleading in a clinical setting. Liver tests results, particularly aminotransferases and \( \gamma \)-glutamyltransferase, typically fluctuate in patients with chronic liver disease (2). When evaluating a patient, even an asymptomatic one, with abnormal liver biochemistries, clinicians should interpret results according to the clinical context and consider an adequate work-up (2, 3). A repeated value in the normal range does not ensure that the initial value was truly erroneous.

In addition, substantial evidence indicates that high aminotransferase values are statistically significantly correlated with increased future mortality, suggesting that these blood tests are valuable indicators of long-term prognosis (4, 5). How should one differentiate between a clinically insignificant fluctuation of normality and a predictor of mortality in a single patient? To support their recommendations, Lazo and colleagues should have noted that patients with 2 discordant test results have the same long-term outcome as those with 2 concordant normal test results. Otherwise, normalization cannot be defined as proof of normality.

Lazo M, Selvin E, Clark JM. Brief communication: clinical implications of short-term variability in liver function test results. Ann Intern Med. 2008;148:348-52. [PMID: 18316753]

TO THE EDITOR: The recent article by Lazo and colleagues (1) shows that in the context of an epidemiologic survey, such as NHANES, abnormal liver test results in more than one third of participants (levels of aspartate aminotransferase, ALT, alkaline phosphatase, \( \gamma \)-glutamyltransferase, and bilirubin) would be reclassified as normal if retested 17 days apart. On the basis of their findings, Lazo and colleagues recommend that, to avoid unnecessary testing, individuals with abnormal liver test results on a first determination be routinely retested before undergoing further evaluation.

In our opinion, caution should be used in drawing practice recommendations from epidemiologic studies. We are particularly concerned that the proposed strategy could be highly misleading in a clinical setting. Liver tests results, particularly aminotransferases and \( \gamma \)-glutamyltransferase, typically fluctuate in patients with chronic liver disease (2). When evaluating a patient, even an asymptomatic one, with abnormal liver biochemistries, clinicians should interpret results according to the clinical context and consider an adequate work-up (2, 3). A repeated value in the normal range does not ensure that the initial value was truly erroneous.

In addition, substantial evidence indicates that high aminotransferase values are statistically significantly correlated with increased future mortality, suggesting that these blood tests are valuable indicators of long-term prognosis (4, 5). How should one differentiate between a clinically insignificant fluctuation of normality and a predictor of mortality in a single patient? To support their recommendations, Lazo and colleagues should have noted that patients with 2 discordant test results have the same long-term outcome as those with 2 concordant normal test results. Otherwise, normalization cannot be defined as proof of normality.

Lazo M, Selvin E, Clark JM. Brief communication: clinical implications of short-term variability in liver function test results. Ann Intern Med. 2008;148:348-52. [PMID: 18316753]

TO THE EDITOR: The recent article by Lazo and colleagues (1) shows that in the context of an epidemiologic survey, such as NHANES, abnormal liver test results in more than one third of participants (levels of aspartate aminotransferase, ALT, alkaline phosphatase, \( \gamma \)-glutamyltransferase, and bilirubin) would be reclassified as normal if retested 17 days apart. On the basis of their findings, Lazo and colleagues recommend that, to avoid unnecessary testing, individuals with abnormal liver test results on a first determination be routinely retested before undergoing further evaluation.

In our opinion, caution should be used in drawing practice recommendations from epidemiologic studies. We are particularly concerned that the proposed strategy could be highly misleading in a clinical setting. Liver tests results, particularly aminotransferases and \( \gamma \)-glutamyltransferase, typically fluctuate in patients with chronic liver disease (2). When evaluating a patient, even an asymptomatic one, with abnormal liver biochemistries, clinicians should interpret results according to the clinical context and consider an adequate work-up (2, 3). A repeated value in the normal range does not ensure that the initial value was truly erroneous.

In addition, substantial evidence indicates that high aminotransferase values are statistically significantly correlated with increased future mortality, suggesting that these blood tests are valuable indicators of long-term prognosis (4, 5). How should one differentiate between a clinically insignificant fluctuation of normality and a predictor of mortality in a single patient? To support their recommendations, Lazo and colleagues should have noted that patients with 2 discordant test results have the same long-term outcome as those with 2 concordant normal test results. Otherwise, normalization cannot be defined as proof of normality.

Lazo M, Selvin E, Clark JM. Brief communication: clinical implications of short-term variability in liver function test results. Ann Intern Med. 2008;148:348-52. [PMID: 18316753]
Not Sold on Performance Measures

TO THE EDITOR: I’ve practiced internal medicine with 3 colleagues for 28 years and have no hope to impede the improvement of patient safety and care quality. However, in their recent article, Landon and Normand (1) evoke more questions than answers, leaving me unconvinced of the value of performance measurements and unenthusiastic about embracing them.

Where is the evidence that public reporting of performance measures and pay-for-performance programs improve (rather than worsen) outcomes and reduce (rather than increase) health care costs?

Are these programs the best place to invest limited resources in repairing today’s pressing health care problems, such as caring for the uninsured and addressing the decreased affordability and access of care?

Do these programs encourage career choices in primary care and address shortages of new primary care physicians or accelerate the exodus of currently practicing primary care physicians to other specialties or retirement? Do they boost dwindling physician morale or—more likely—further sap it?

Health plan reports to physicians are currently reported in aggregate, deidentified format. Without referencing an identifiable episode of care, are generic process recommendations (prescribe more generic drugs, order more mammograms) useful in changing physician behavior?

In regard to profit measures, are these programs ultimately designed to enhance insurer profits and investor returns? Are these zero-sum programs? Are payments shifted from poorly measured to good ones and does this hold the line on spending? Or do insurers profit by reducing payments to underperforming practices?

Do these programs assuage the nagging suspicion that the major benefit of physician-purchased electronic medical records to insurers is easy access to, acquisition of, and tracking of physician care, process, and prescribing measures? Why should small practices pay to prescribe electronically, when pharmaceutical benefit managers, pharmacy monitoring programs, and the American Medical Association sell the data for profit?

I’m swamped with dreaded requests to complete forms and copy charts for insurers. If patients benefit from these chores, shouldn’t they foot part of the cost, perhaps a premium surcharge returned to physicians to recoup costs?

Should physicians be penalized for patients who choose to be nonadherent, patients whose care and prescribing are shared by multiple physicians, and patients enrolled in self-directed or high-deductible plans who deliberately choose not to spend money on proven preventive measures included in performance assessment?

The United Kingdom experience of rewarding physicians for complying with performance measures resulted in substantial financial gain for physicians at great government cost. Given the current difficult economic conditions in the United States and negative attitudes toward physician compensation, can a similar program be developed here?

Please convince me!

Michael E. Miller, MD
Medical Care Affiliates
Boston, MA 02199

Potential Financial Conflicts of Interest: None disclosed.

References
1. Landon BE, Normand SL. Performance measurement in the small office practice: challenges and potential solutions. Ann Intern Med. 2008;148:353-7. [PMID: 18316754]

IN RESPONSE: We sympathize with the sentiments expressed by Dr. Miller. One of us is a practicing physician and is aware of the extra burden imposed by many performance measurement programs with unclear patient benefits. Evidence suggests, however, that the proliferation, measurement, and dissemination of quality information have a substantial impact on measured areas of quality. Indeed, one measure (β-blocker use after a myocardial infarction) has been retired because performance has approached perfection (1). It is highly unlikely that performance on this measure and others would be so high if a spotlight had not been aimed at them. Although pay-for-performance and other programs have been shown to have a generally small impact over short periods, their cumulative effects over time remain unknown. The hope is that better use of population health management techniques and electronic resources, such as electronic health records and decision support, will improve the capacity of physician organizations to achieve higher-quality care.

Although space constraints prohibit us from addressing each of Dr. Miller’s questions, we comment on a few key points. First, given recent evidence that the quality of care produced by the U.S. health care system is suboptimal, we believe not only that limited resources should be directed toward improving care but that this investment should be much more substantial (2–4). Second, we disagree that these programs are at the root of the current primary care crisis. In fact, the United Kingdom has instituted a broad pay-for-performance program that includes substantial additional resources directed toward general practitioners, in part to stabilize the primary care workforce. Like Dr. Miller, we hope that onerous utilization

Potential Financial Conflicts of Interest: None disclosed.

References
1. Prati D, Taioli E, Zanella A, Della Torre E, Butelli S, Del Vecchio E, et al. Updated definitions of healthy ranges for serum alanine aminotransferase levels. Ann Intern Med. 2002;137:1-10. [PMID: 12093239]
2. Neuschwander-Tetri BA, Unalp A, Creer MH. Nonalcoholic Steatohepatitis Clinical Research Network. Influence of local reference populations on upper limits of normal for serum alanine aminotransferase levels [Letter]. Arch Intern Med. 2008;168:663-6. [PMID: 18362260]
3. Kaplan MM. Alanine aminotransferase levels: what’s normal? [Editorial]. Ann Intern Med. 2002;137:49-51. [PMID: 12093245]
4. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. STROBE initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Ann Intern Med. 2007;147:W163-9. [PMID: 17938389]

Mariana Lazo, MD, ScM
Elizabeth Selvin, PhD, MPH
Jeanne M. Clark, MD, MPH
Johns Hopkins University
Baltimore, MD 21205

Potential Financial Conflicts of Interest: None disclosed.
management tools and requests used by health plans will diminish with time as the interconnectedness of the health care system is improved. Nonetheless, it is unlikely that these programs will disappear while there is still substantial evidence of overuse and variations in use that cannot be explained by clinical need. Some of these other issues have been discussed in other papers (5). Despite these problems, we believe that increased measurement and transparency are required for improving health systems.

Bruce E. Landon, MD, MBA
Sharon-Lise T. Normand, PhD
Harvard Medical School
Boston, MA 02115

Potential Financial Conflicts of Interest: None disclosed.

References
1. Lee TH. Eulogy for a quality measure. N Engl J Med. 2007;357:1175-7. [PMID: 17881749]
2. Kohn KT, Corrigan JM, Donaldson MS, eds. Institute of Medicine. To Err Is Human: Building a Safer Health System. Washington, DC: National Academies Pr; 1999.
3. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Pr; 2001.
4. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. N Engl J Med. 2003;348:2635-45. [PMID: 12826639]
5. Landon BE, Normand SL, Blumenthal D, Daley J. Physician clinical performance assessment: prospects and barriers. JAMA. 2003;290:1183-9. [PMID: 12953001]

Corrections

Correction: Screening for Type 2 Diabetes Mellitus in Adults
In the recent U.S. Preventive Services Task Force recommendation statement on screening for type 2 diabetes mellitus in adults (1), the Figure contained an error. In the suggestions for practice regarding insufficient evidence, the first sentence should have read: “When BP [blood pressure] is = 135/80 mm Hg [not = 135/80 mm Hg], screening may be considered on an individual basis when knowledge of diabetes status would help inform decisions about coronary heart disease (CHD) preventive strategies, including consideration of lipid-lowering agents or aspirin.” This correction has been applied to the online version of the article.

Reference
1. U.S. Preventive Services Task Force. Screening for type 2 diabetes mellitus in adults: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2008;148:846-54.

Correction: A Leading Medical School Seriously Damaged
In the article by Ernst (1), there is a typographical error in the historical background section: “Duke Gibenau, a Frenchman, postulated in the 1850s that the purity of races would be a determining factor in history.” The name should be Arthur de Gobineau, and his preferred title was “le comte de Gobineau,” translated as either Count Gobineau or Count de Gobineau.

Reference
1. Ernst E. A leading medical school seriously damaged: Vienna 1938. Ann Intern Med. 1995;122:789-92. [PMID: 7717602]