LaMonte et al\(^1\) examined the associations of physical activity with fracture risk in 77,206 postmenopausal women with a mean 14 years of follow-up. With detailed information on physical activity and fracture and large numbers of events, the authors\(^3\) were uniquely positioned to address this question. Inactivity was associated with a higher risk of fracture, with the highest tertile of total recreational physical activity associated with a 6% (95% CI, 2%-10%) lower hazard of total fracture and 18% (95% CI, 5%-28%) lower hazard of hip fracture. Associations between other summaries of physical activity and site-specific fractures were also significant for some reported comparisons, depending on activity measure, adjuster variables, and statistical tests used.

The simplicity of this summary belies the complexity of the underlying issues that affect the interpretation of these results. Because of the varied nature of the types of activity and because activity level tends to vary over time, physical activity is a complex exposure. The study by LaMonte et al\(^1\) relied on self-reported physical activity, a measure generally recognized to contain a sizeable amount of measurement error.\(^2\) How measurement error affects study results, causing attenuated or inflated risk estimates or increased uncertainty, can be difficult to estimate when variables associated with systematic error are also associated with the outcome being studied. The direction of bias is further complicated when a continuous measure is categorized into an ordinal variable, as LaMonte et al\(^1\) have done by reporting associations with tertiles of activity.\(^3\) Finally, it is unknown which aspect of physical activity is the most important for any given health outcome, so a natural tendency is to examine many components of physical activity. LaMonte et al\(^1\) chose to look at 14 different fracture outcomes, associating some with as many as 6 different activity summaries and sedentary time. Repeated statistical tests without adjustment for multiplicity increases the chance of finding spurious associations. The authors acknowledge each of these limitations.\(^1\) But is acknowledgment enough?

In many settings we proceed with some risk. Driving runs the risk of a collision, but we continue to drive. We wear seatbelts. Driving without is considered too high risk. The seatbelt is a simple, cost-effective solution shown to reduce the risk of the serious morbidity and mortality of motor vehicle collisions. When there are potential biases in a statistical analysis, not performing an analysis of how these biases could have affected outcomes is a high-risk activity. Like fastening a seatbelt, there are things we can do to prevent untoward outcomes. What can be done?

### Addressing Measurement Error

Data from internal validation studies, which compare an error-prone instrument with a reference with little error or at least unbiased (ie, random) error, can be used to develop an error-corrected exposure. These calibrated exposure values replace the unadjusted self-report in the target regression, using a method called regression calibration, which typically does a good job of reducing bias from exposure measurement error.\(^4\) In 450 women from the Women's Health Initiative observational study, Neuhouser et al\(^5\) compared self-reported activity-related energy expenditure with an estimate derived from 2 objective biomarkers, doubly labeled water\(^6\) and indirect calorimetry. Neuhouser et al\(^5\) developed calibration equations and found differences between self-reported and biomarker values that depended on participant characteristics. These calibration

---

**Open Access.** This is an open access article distributed under the terms of the CC-BY License.
equations are useful for adjusting associations for the error in total activity-related energy expenditure. They do not directly assess errors in other measures of activity, such as walking or mild physical activity.

When no or partial information regarding measurement error is available, sensitivity or bias analysis is an applicable and well-established method. In bias analyses, researchers consider a plausible range of mathematical models for the error, apply a method to adjust for this error, and examine the extent to which conclusions are robust to variability in these assumptions. Lash et al describe best practices for applying a bias analysis framework. In the Women’s Health Initiative observational study, a test-retest study was done that provided detailed descriptions of the reliability of different physical activity measures and assessed whether the reliability depended on covariates. Although the overall reliability was described as good, there were examples of covariate-dependent error. For mild physical activity, the intraclass correlation coefficient was 0.53 (95% CI, 0.46 to 0.60) for 390 white participants and 0.07 (95% CI, –0.19 to 0.31) for 60 African American participants. This reliability study provides a good starting point for investigating what levels of reporting error and systematic differences are plausible and whether the plausible levels are enough to nullify or reverse study results.

Categorizing an Error-Prone Exposure

Addressing bias from a categorized error-prone exposure is more complex. The level and direction of bias introduced can vary by category. Here, analyses using the original continuous measure may be more informative regarding whether there is a consistent trend of increasing or decreasing risk with increased activity. Bias analyses that model the effect of error on a continuous exposure and its categorized value may also provide insights into the effects of categorization on the direction and magnitude of the bias for different levels of physical activity.

Multiplicity Adjustment

Much has been written on when it is necessary and how to adjust for multiplicity. One view is that if a study is truly exploratory, then adjusting for multiplicity is impractical and unnecessary given the ad hoc nature of study analyses. However, conclusions contain the caveat that any found associations must be confirmed by another study because of the elevated false-positive (ie, type I) error rate from multiple comparisons. Are we still at the point of doing exploratory analyses? LaMonte et al cite several prior studies on physical activity and fracture. Had the authors chosen a priori to focus on a small set of previously found associations or those thought most plausible as a primary hypothesis and considered the rest as exploratory, then a small adjustment for multiple comparisons would have been necessary. Confirmed, a priori hypotheses adjusted for multiple comparisons provide a stronger level of evidence than significant results from unadjusted exploratory analyses involving many comparisons. There are other approaches, such as applying omnibus tests or multivariate tests, that can reduce multiplicity and efficiently test for exposure effects.

What Do We Conclude?

LaMonte et al contributed a comprehensive look at physical activity and fracture risk. The effects of their study could have been stronger had they taken full advantage of the available information to assess this association. In fact, the approach taken is prevalent. In a 2018 review of 40 cohort studies examining physical activity and health that mentioned measurement error, only 2 (5%) applied any error-correction method despite some also mentioning validation and/or calibration data. Why is this the dominant approach? When naive analyses produce an association in the
direction consistent with prior expectations, it can be tempting to report the result without adjustment. Unlike a motor vehicle collisions, there can be no overt signs that study analyses produced an adverse outcome, ie, misleading results. Conversely, there are no overt signs that this has not occurred. Results may be subject to overstated criticism, in that sensitivity analysis or a well-structured analysis handling multiple comparisions may reveal that the found conclusions were robust, thereby strengthening results. This seems a worthy trade-off for the extra analysis effort, which in many cases involves applying standard methods in existing software. It’s time to put on the seatbelt.

ARTICLE INFORMATION
Open Access: This is an open access article distributed under the terms of the CC-BY License. © 2019 Shaw PA. JAMA Network Open.

Corresponding Author: Pamela A. Shaw, PhD, Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, 423 Guardian Dr, Room 606, Philadelphia, PA 19104 (shawp@upenn.edu)

Author Affiliation: Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia.

Conflict of Interest Disclosures: None reported.

REFERENCES