

**James P. Scanlan
Attorney at Law
1529 Wisconsin Avenue, NW
Washington, D.C. 20007
(202) 338-9224
jps@jpscanlan.com**

October 26, 2012

Drew Gilpin Faust, Ph.D., President
Harvard University
Cambridge, MA 02138

Laura Miller, Interim President and CEO
National Quality Forum
601 Thirteenth Street, NW, Suite 500 North
Washington, DC 20005

Risa Lavizzo-Mourey, M.D., M.B.A., President and CEO
Robert Wood Johnson Foundation
Route 1 and College Road East, P.O. Box 2316
Princeton, NJ 08543

Peter L. Slavin, M.D., President
Massachusetts General Hospital
55 Fruit Street
Boston, MA 02114

Jeffrey S. Flier, M.D., Dean
Harvard Medical School
651 Huntington Ave.
Boston, MA 02115

Re: Withdrawal of *Commissioned Paper: Healthcare Disparities Measurement*ⁱ

Dear President Faust, President Miller, President Lavizzo-Mourey, President Slavin, and Dean Flier:

This is a follow-up to an October 9, 2012 letter to Harvard President Drew Gilpin Faust ([Harvard Letter](#)¹) and, in some respects, a follow-up to an October 22, 2009 letter to National Quality Forum (NQF) President and Chief Executive Officer Janet M. Corrigan ([NQF Letter](#)) and an

¹ To facilitate consideration of the issues raised in letters such as this I make available electronic copies of the letters on the Institutional Correspondence subpage of the Measuring Health Disparities page of jpscanlan.com. Underlinings in this letter reflect links to the underlined material in such a copy of the letter. When an online copy is corrected such fact is typically noted on the last page.

April 8, 2010 letter to Robert Wood Johnson Foundation (RWJF) President and Chief Executive Officer Risa Lavizzo-Mourey ([RWJF Letter](#)).

The Harvard Letter, the timing of which was prompted by the scheduling of an October 17, 2012 Applied Statistics Workshop at Harvard's Institute for Quantitative Social Science where I was to present a paper titled "[The Mismeasure of Group Differences in the Law and the Social and Medical Sciences](#)," discussed the subject of that workshop as it bore on research and teaching at Harvard about health disparities and other demographic differences. The letter described patterns by which standard measures of differences between outcome rates tend to be systematically affected by the prevalence of an outcome and explained that the failure to understand those patterns undermined efforts at Harvard and elsewhere to appraise demographic differences in outcome rates with respect to various issues in the law and the social and medical sciences. The letter urged Harvard to generally review its teaching and research concerning matters where the patterns are implicated. The letter also urged Harvard (after consultation with Massachusetts General Hospital, NQF, and RWJF) to withdraw the October 4, 2011 [Commissioned Paper: Healthcare Disparities Measurement](#) (*Commissioned Paper*), a joint project of Harvard Medical School and Massachusetts General Hospital sponsored by NQF and RWJF, because the document fails to consider the above-referenced patterns by which standard measures of differences in outcome rates tend to be systematically affected by the prevalence of an outcome. The letter explained that, while the *Commissioned Paper* is no more flawed than other measurement guidance documents created in the past, the paper will mislead the public, policy makers and other researchers, while causing the waste of substantial resources.

The 2009 NQF Letter had described the same patterns and urged NQF to consider the patterns in providing guidance on reviewing or measuring health and healthcare disparities. The 2010 RWJF Letter also described those patterns and urged RWJF to consider the patterns in conducting its activities relating to health and healthcare disparities issues and in encouraging young researchers to devote their time and talents to studying such issues. Both letters emphasized the importance of consideration of the patterns with respect to pay-for-performance (P4P) programs, referencing the [Pay for Performance](#) subpage of the [Measuring Health Disparities](#) page of [jpscanlan.com](#). That subpage explained (a) that failure to recognize the patterns by which absolute differences between outcome rates tend to change as the prevalence of an outcome changes had led to the mistaken perception in the United States that P4P would tend to increase healthcare disparities and (b) that failure to recognize the patterns by which each standard measure of differences between outcome rates was affected by the prevalence of an outcome would undermine any effort to tie P4P to measures of disparities.

This letter summarizes points made in the Harvard Letter regarding the *Commissioned Paper* and reasons for withdrawing it. It also augments those points with reference to points previously made in the 2009 NQF Letter and 2010 RWJF Letter, as well to the fact that the *Commissioned Paper* has already served as the foundation for NQF's September 2012 [Healthcare Disparities and Cultural Competency Consensus Standards Technical Report](#) (*Consensus Standards Technical Report*) in a way that undermines the latter document.

The patterns by which standard measures of differences between outcome rates tend to be affected by the prevalence of an outcome, and the problems with efforts to appraise the difference between the circumstance of two groups reflected by a pair of outcome rates without taking these patterns into account, are discussed in approximately 170 references collected on the [Measuring Health Disparities](#) page of [jpscanlan.com](#), as well as in the several score other pages and subpages of the site, especially the [Scanlan's Rule](#) and [Mortality and Survival](#) pages. But the best description of the patterns in a single item may be that found in the Harvard Letter.

Section A of the letter (at 3-18) explains the pattern by which the rarer an outcome, the greater tends to be the relative difference in experiencing it and the smaller tends to be the relative difference in avoiding it. Thus, for example, as mortality generally declines, relative differences in mortality tend to increase while relative differences in survival tend to decrease; as rates of receipt of beneficial healthcare procedures generally increase, relative differences in receipt of such procedures tend to decrease while relative differences in failing to receive such procedures tend to increase.

Section A also explains that in every instance where the *Commissioned Paper* makes a point based on the sizes of relative differences in favorable outcomes or relative differences in adverse outcomes the unmentioned relative difference for the opposite outcome would support an opposite point. Such instances in the *Commissioned Paper* generally occurred in the context of the paper's pointing out that a relative difference and the absolute difference yielded different conclusions as to the directions of changes over time. As explained in Section B of the Harvard Letter (at 19), anytime an observer notes that a relative difference and the absolute difference have changed in different directions over time, the unmentioned relative difference will have changed in the opposite direction of the mentioned relative difference and in the same direction as the absolute difference.

Section A (at 8) also explains that data in Figures 4 and 8 of the *Commissioned Paper* show that one would reach opposite conclusions as to which subpopulation has the largest racial disparity depending on whether one examined the relative difference in the favorable outcome or the relative difference in the adverse outcome. This is a particularly important matter with respect to the recommendation in the *Commissioned Paper* that healthcare disparities research examine interactions between race and other factors (a recommendation adopted in the *Cultural Competency Technical Report*), since one would commonly reach opposite conclusions as to the direction of an interactive effect depending on which outcome is examined.²

Section B of the letter (at 18-21) explains that absolute differences between rates, while yielding the same interpretation whether one examines the favorable or the adverse outcome, tend also to be affected by the prevalence of an outcome, though in a more complicated way than the two

² Figures 2 and 2a of the October 17, 2012 [Applied Statistics Workshop](#) illustrate the point with respect to the data in Figure 8 of the *Commissioned Paper*.

relative differences. Roughly, as uncommon outcomes (less than 50% for all groups being compared) become more common, absolute differences between rates tend to increase; as common outcomes (greater than 50% for all groups being compared) become even more common, absolute differences tend to decrease. In addition, as the prevalence of an outcome changes, the absolute difference will tend to change in the same direction as the smaller relative difference and the opposite direction of the larger relative difference. Thus, for example, as uncommon procedures become more common, absolute differences tend to increase; as already common procedures become even more common, absolute differences tend to decrease. Further, when outcome rates tend generally to be well above 50%, places or entities with comparatively high rates for such outcomes will tend to show smaller absolute differences between the rates of advantaged and disadvantaged groups than places or entities with comparatively low rates for such outcomes.

Section C of the letter (at 21-23) explains that the failure to understand the way absolute differences tend to increase when uncommon outcomes increase in overall prevalence led to the mistaken perception in the United States that P4P would tend to increase healthcare disparities, which in turn led the Commonwealth of Massachusetts to unwisely include a healthcare disparities measure in its Medicaid P4P program. The section also explains that the failure to understand that where outcome rates are generally above 50% higher rates tend to be associated with lower absolute differences resulted in a measurement approach in the Massachusetts program that is more likely to increase healthcare disparities than to reduce them. That is, the program employed a disparities measure that is a function of absolute differences between rates with respect to outcomes where overall rates averaged above 80%. Consequently, higher performing hospitals would be perceived as having smaller disparities simply because in the rate ranges at issue higher rates tend to be associated with smaller absolute differences between rates than lower rates. And because minorities comprise a lower proportion of patients at higher performing hospitals than at lower performing hospitals, the program, by diverting resources away from hospitals where minorities comprise a comparatively large proportion of patients, will tend to increase rather than decrease healthcare disparities.

Section D of the letter (at 24-28) also explains the fallacy of notions that opposite conclusions as to such things as the direction of change in healthcare disparities over time can both be valid or that choice of measure involves a value judgment. In a sense, this point is highlighted by the *Commission Paper* itself. For, while it stresses the importance of presenting different measures that yield different conclusions in order to “allow[] the readers to make their own interpretations” or “allow[] the readers to judge the importance by taking the context of the report into consideration” (at 56), the paper provides no guidance on how readers should make such judgments. As explained in Section D, any guidance for choosing between opposite interpretations as to whether and how the forces causing a disparity have changed over time, or are otherwise larger in one setting than another, without consideration of the way the measures are affected by overall prevalence, would be fundamentally unsound.

As explained in Section E.2.c of the Harvard Letter (at 41-43), the initial draft of the *Commissioned Paper* showed no recognition of the patterns by which standard measures of differences between outcome rates tended to be affected by the prevalence of an outcome. But in

response to my raising the issue in [comments](#) submitted to the authors, the authors made the following change, which they indicated was intended to address issues raised in my comment designated No. 113 on the online [collection](#) of comments) .

Originally, the first sentence of Section 4.c, which section is styled “Absolute Versus Relative Differences and Favorable Versus Adverse Outcomes,” read as follows (at 35):

While calculations of disparities can be straightforward, comparisons of disparities among entities or over time can be sensitive to the calculations chosen.

In the final version the sentence, which included eight additional words, read:

While calculations of disparities can be straightforward, comparisons of disparities among entities or over time can be sensitive to the calculations chosen, especially when the prevalence of an outcome changes.

It is not clear what the reader would likely make of this language since almost never will disparities change other than when the prevalence of an outcome changes.³ But the additional language obviously does not alert the reader to the fact that the key measures discussed in the paper tend to change in certain ways when prevalence changes, much less to the issue of the ways such fact calls into question the validity of each measure commonly used.

The Harvard Letter satisfactorily explains why the *Commissioned Paper* should be withdrawn and withdrawn immediately, before it might lead researchers to undertake studies using unsound statistical methods. But I add here several points relating to matters addressed in the 2009 NQF Letter and the 2010 RWJF Letter, as well as to the fact that the *Commissioned Paper* is serving as the foundation for future health and healthcare disparities measurement by NQF.

First, both letters discuss the study appearing in *Pediatrics* in 2008 by Morita et al.⁴ that won a Robert Wood Johnson Foundation award for addressing health disparities. Relying on relative differences in vaccination rates as a measure of disparity, the authors found that a school-entry Hepatitis B vaccination requirement that dramatically increased overall vaccination rates also dramatically reduced racial and ethnic vaccination disparities. By contrast, the National Center for Health Statistics, which always relies on relative differences in adverse outcomes, would have found dramatic increases in disparities.

³ The authors also made the following cryptic reference to prevalence in seeming agreement to my comment designated No. 115: "Commissioned paper authors considered the comment and agree prevalence is an important indicator of disparities. However, they believe the detail specified is not needed within the paper. "

⁴ Morita JY, Ramirez E, Trick WE. Effect of school-entry vaccination requirements on racial and ethnic disparities in Hepatitis B immunization coverage among public high school students. *Pediatrics* 2008;121:e547.

Table 1 below, which is an abbreviated version of Table A to my [Comment on Morita Pediatrics 2008](#) and Table 4 of the [Harvard Applied Statistics Workshop](#), shows the conflicting interpretations as to change over time according to whether one relies on relative differences in the favorable outcome (receipt of vaccination) or relative differences in the adverse outcome (failure to be vaccinated).

Table 1. White and Black Hepatitis B Vaccination Rates Before and After Implementation of School-Entry Vaccination Requirement with Rate Ratios of Receipt and Non-Receipt of Vaccination, Grades 5 and 9, from Morita et al., *Pediatrics* 2008

Grade	Year	Program	WhVacRt	BlVacRt	FavRatio	AdvRatio
5	1996	Pre	8%	3%	2.67	1.05
5	1997	Post	46%	33%	1.39	1.24
9	1996	Pre	46%	32%	1.44	1.26
9	1997	Post	89%	84%	1.06	1.45

The *Commissioned Paper* gives the reader no basis for believing that situations like this exist, much less a basis for understanding why such a pattern would be commonly found whenever the prevalence of an outcome changes substantially. And certainly it gives the reader no guidance on how to deal with the matter.

Indeed, while the *Commissioned Paper* discusses both relative differences in favorable outcome and relative differences in adverse outcomes, the main body of the document never suggests even the possibility that the two approaches could yield different conclusions as to the direction of changes over time. While a Summary Table (at 56) does indicate that relative differences in favorable and relative differences in adverse outcomes could yield different conclusions as to whether healthcare disparities are increasing or decreasing, the failure to address the matter in the body of the document caused the issue to be entirely ignored in the *Cultural Competency Technical Report*, which based its own measurement recommendations on the *Commissioned Paper*.

The report (at 4) listed “favorable and adverse measures” as one of the aspects of the *Commissioned Paper* that the NQF Cultural Competency Steering Committee accepted. The other mention of favorable and adverse outcomes appears in the heading of the following guidance paragraph (*id.*):

Absolute versus Relative Disparities and Favorable versus Adverse Measures: The absolute and relative changes in disparities can reveal different conclusions about whether gaps are actually closing and often can lead to different interpretations when making these comparisons. Both absolute and relative statistics should be calculated, and if they lead to conflicting conclusions, then both statistics should be reported, allowing users to reach their own conclusion. In addition, trends should be calculated and specific rates provided along with a narrative for explanation.

Though the *Cultural Competency Technical Report* retains the reference to “favorable versus adverse measures” that appeared in the title of Section 4.c of the *Commissioned Paper*, the *Cultural Competency Technical Report* says nothing whatever about choosing between favorable and adverse outcomes.⁵ Further, while the guidance notes the importance of reporting both relative differences and absolute differences when the two measures yield different conclusions, it says nothing about which relative difference should be examined. As noted, in every situation where the absolute difference and the mentioned relative difference yield different conclusions as to whether a gap is closing, the unmentioned relative difference will yield a conclusion that is the opposite of that yielded by the mentioned relative difference and that is the same as that yielded by the absolute difference. And commonly, as the prevalence of an outcome changes, the absolute difference will change in the same direction as the smaller relative difference and the opposite direction of the larger relative difference. In any case, whereas in point of fact relative differences in favorable outcomes and relative differences in adverse outcomes will tend to yield different conclusions as to directions of changes over time, not only does the *Cultural Competency Technical Report* fail to give any indication that such a thing is possible, but, apart from the reference in the heading and the bullet point, the document gives no indication that whether one examines the favorable or the adverse outcome is a matter of any consequence at all. Thus, researchers relying on the guidance in the document to conduct studies of the type that Morita *et al.* published in *Pediatrics* might well emphasize relative differences in the favorable outcome or relative differences in the adverse outcome without having any reason to believe that the relative difference not chosen would tend to show dramatic changes that are the opposite of the dramatic changes yielded by the chosen relative difference.

Second, both the NQF and RWJF letters discussed the importance of understanding measurement issues with respect to P4P. As mentioned above, Section C of the Harvard Letter discusses the way that failure to understand the way that absolute differences tend to be affected by the overall prevalence of an outcome led to an unfounded perception in the United States that P4P would tend to increase healthcare disparities, leading Massachusetts to unwisely include a healthcare disparities measure in its Medicaid P4P program and to do so in a manner that is more likely to result in increased healthcare disparities than in reduced healthcare disparities. Although the *Commissioned Paper* (at 37-38) discusses the study that led to the perception that P4P would tend to increase healthcare disparities (pointing out that the relative difference in the

⁵ As suggested in my comment designated No. 118, it appears that Section 4.c. of the *Commissioned Paper* intended to address the way relative differences in favorable and adverse outcomes could yield different conclusions as to change of time, but then instead addressed only the differences in conclusions yielded by a relative difference and the absolute difference. But contrary to my comments (which, as I noted, were written without having yet read the entire document), a summary table at page 55 of the draft listed an issue as: “Measuring rates of adverse and positive events can yield different conclusions about whether or not gaps are closing.” It recommended: “As above with respect to absolute and relative comparisons, public reporting of disparities should calculate statistics using both favorable and adverse events. If the results are notably different, both statistics should be reported, allowing the reader to judge the importance by taking the context of the report into consideration.” The same material then appeared at page 56 of the final document. But just as the material escaped my notice, it appears to have been overlooked by the NQF Steering Committee drafting the *Cultural Competency Technical Report*.

favorable outcome decreased while ignoring that the relative difference in the adverse outcome increased), it shows no recognition of the reasons why the increase in the absolute difference observed in that study would typically be found when an uncommon outcome increases in overall prevalence (or of the reasons why the relative difference in the favorable outcome would typically decrease and the relative difference in the adverse outcome would typically increase when the favorable outcome generally increases).⁶ And although the *Commissioned Paper* (at 40) specifically discusses P4P and the Massachusetts Medicaid P4P program and the measure it employs, it shows no recognition of the reasons why that measure would tend to cause higher performing hospitals to show smaller disparity values than lower performing hospitals.

The *Cultural Competency Technical Report* (at 7), in apparent reliance on the *Commissioned Paper*, includes two bullet points about tying P4P to measures of healthcare disparities. But it does so without discussion, or apparent recognition, of the way various measures would tend to yield different results.⁷ To the extent that such a recommendation is followed, if the Massachusetts program's measurement approach is used as a model, there is reason to believe that failure to understand measurement issues will result in the implementation of programs that, as in Massachusetts, are more likely to increase healthcare disparities than to decrease them.⁸ Other measures will tend also to yield certain broadly predictable results. In particular, reliance on relative differences in favorable outcomes will tend to favor higher performing hospitals while reliance on relative differences in adverse outcomes will tend to favor lower performing hospitals. In no case, however, will an appraisal of disparity have anything to do with actual differences in the comparative equity with which various hospital treat their minority and white patients unless the chosen measure is examined with a recognition of the way that it tends to be affected by the prevalence of the outcome at issue.

Third, the RWJF Letter, by reference to [Section E.7](#) of the [Measuring Health Disparities](#) page (as was also done in the Harvard letter), discussed the emerging recognition by other authors of the patterns by which standard measures of differences between rates tend to be affected by the prevalence of an outcome and of the need to consider these patterns in health disparities research, and note 4 of the NQF Letter specifically cited the works of such authors, including some of Europe's leading experts on health disparities measurement. The *Commissioned Paper* shows no

⁶ As discussed in the designated No. 118, one page later, the *Commissioned Paper* does exactly the opposite. In that instance, the paper contradicts the finding of an absolute difference on the basis of the relative difference in the adverse outcome, though again without mention that the relative difference in the opposite outcome would yield the same conclusion as the absolute difference.

⁷ In response to my emphasizing, in comment designated No. 123, the importance of measurement issues with respect to P4P, the authors indicated that "NQF will consider the suggestions within the context of the Healthcare Disparities consensus standards project." Whether or not the points were communicated to NQF with respect to P4P, the *Cultural Competency Technical Report* shows no understanding of the implications of the measurement issues as to P4P or any other matter.

⁸ As discussed in the Harvard Letter (at 23), reliance on absolute differences will tend to favor higher performing hospitals because the types of outcome that are likely to be examined as part of disparities elements of P4P programs commonly will involve rates that are well above 50% for all groups. To the extent that outcomes are examined where overall rates are quite low, reliance on absolute differences would tend to favor lower performing hospitals.

recognition whatever of this body of work. And although it cites a 2005 National Center for Health Statistics (NCHS) paper⁹ concerning several matters, the *Commissioned Paper* fails to show any recognitions that in that paper and three others (one official and two unofficial), NCHS statisticians specifically addressed the pattern by which relative differences in favorable outcomes and relative differences in adverse outcomes tend to yield different conclusions about whether health and healthcare disparities are increasing or decreasing over time. In any case, the existence of the work discussed in Section E.7 by itself invalidates any document that purports to provide guidance on the measurement of health or healthcare disparities without any consideration of the way the measures it discusses tend to be systematically affected by the prevalence of an outcome.

By email dated October 11, 2012, I advised the authors of the *Commissioned Paper* of my letter to Harvard suggesting that it withdraw the paper. In the email, I suggested that if the authors believed that the paper could be defended against the points I raised, they should make the case to Harvard, and, if not, they should consider themselves seeking withdrawal of the paper. By email of October 14, 2012, Joseph Betancourt responded on behalf of the authors declining to seek to have the paper withdrawn. I will not here attempt to summarize the response. But I will note that, to my mind, it did not address the substantive, entirely statistical issues. Nor do I believe that there exists a response that would cause the *Commissioned Paper* to be regarded as other than a fundamentally flawed document that will mislead the public, policy makers and researchers, in a way that is inimical to the goals that Harvard University, Harvard Medical School, Massachusetts General Hospital, the National Quality Forum, and the Robert Wood Johnson Foundation seek to promote through their various activities relating to health and healthcare disparities research.

Thus, I suggest, the responsible course for these entities is to withdraw the document. Further, as stressed in the Harvard Letter, if a document as potentially influential as the *Commissioned Paper* ought to be withdrawn, that should be done as soon as possible, before others rely upon it. Such point is underscored by the fact that, in consequence of employing the *Commissioned Paper* as a foundational document, the *Cultural Competency Technical Report*, a document that stresses the importance of “accurate and meaningful metrics to measure healthcare disparities,” is now itself contributing to confusion concerning the measurement of those disparities.

Sincerely,

/s/ **James P. Scanlan**

James P. Scanlan

⁹ Reference 60 of the *Commissioned Paper*, which it cites at page 35, 36, and 42, is the 2005 NCHS document “[Methodological Issues in Measuring Health Disparities](#)” discussed at page 30 of the Harvard Letter.

Minor corrections were made to this letter on December 29, 2012. None is of a nature to warrant specific description.

ⁱ By [letter](#) of December 12, 2012, from Gretchen Brodnicki, Dean for Faculty and Research Integrity of Harvard Medical School, and F. Richard Bringhurst, Research Integrity Office of Massachusetts General Hospital responded to this letter stating that they regarded issues I raised concerning the *Commissioned Paper* to involve “differences of scientific opinion” and not matters of research misconduct. The letter also advised that, while the institutions would bring the matters raised in my letter to the attention of the authors of the *Commissioned Paper*, the institutions do not assess the merits of papers of faculty members.