

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

March 15, 2019

ELECTRONICALLY TRANSMITTED

Shantanu Agrawal, MD, MPhil
President and CEO
Members of the Board of Directors
National Quality Forum
1030 15th Street, NW
Suite 800
Washington, DC 20005

Re: Innumeracy Problem at the National Quality Forum and Misleading Attendees of the 2019 Conference Regarding the Measurement of Health and Healthcare Disparities

Dear President Agrawal and Members of the Board of Directors:

This letter pertains to a longstanding innumeracy problem at the National Quality Forum (NQF) that has caused entities that rely on NQF guidance to devote substantial resources to unsound and misleading research into health and healthcare disparities. The letter also pertains to the probable further promotion of such research at NQF's 20th Anniversary Annual Conference to be held on March 22-24, 2019.

The [web page](#)¹ for the conference introduces the conference with the statement: "Join us to discuss Health Equity, National Quality and Improvement Priorities, and more, at the NQF's Annual Conference and Pre-Conference Event." The web page also states that the first day of the conference will focus on ensuring health equity and highlights among the topics to be addressed "Social Determinants and of Health and the Role of Data." Attendees will be charged up to \$1,509 for the conference itself and up to \$308 for pre-conference events

Some persons presumably will attend the conference to about measuring health and healthcare disparities in order that, consistent with the theme of the NQF September 14, 2017 document [A Roadmap for Promoting Health Equity and Eliminating Disparities: The Four I's](#)

¹ To facilitate consideration of issues raised in documents such as this I include links to referenced materials in electronic copies of the documents, in some cases, for the reader's convenience, providing the links more than once. Such copies are available by means of the [Measurement Letters](#) page of jpscanlan.com. If the online version of the letter is amended, such fact will be noted on its first page.

[for Health Equity \(Roadmap\)](#), they can devote substantial resources to identifying programs that reduce health and healthcare disparities and to implementing and incentivizing such programs. The great majority of persons familiar with that document will assume that there exist no complex or unresolved issues regarding how to determine whether a particular disparity has increased or decreased over time or how to determine the role of particular programs in causing or contributing to increases or decreases. Most such persons will also assume that it is impossible that different measures of health and healthcare disparities could yield opposite conclusions regarding such determinations. The assumptions would be supported by the reasonable belief on the part of readers of the *Roadmap* that, if there existed important issues as to such determinations – or even that it was possible for different measures to yield opposite conclusions about such determinations – a document like the *Roadmap* would mention such matters.

Actions of NQF officials and staff subsequent to issuance of the *Roadmap* would tend to bolster that assumption. For example, immediately following issuance of the *Roadmap*, an article titled “[NQF lays out tactics to measure healthcare disparities](#),” *Modern Healthcare* (Sept. 29, 2017), reported the following:

“There is a lot of work still to be done on the measurement side,” said Dr. Shantanu Agrawal, president and CEO of the NQF. “We know that data sources in this area are really lacking and it's not clear to healthcare stakeholders what the right data sources are.”

Such description of the measurement work still to be done, with an emphasis on data availability and uncertainty about sources, would strongly suggest to readers of the article that – as in fact seems to be the case – NQF is unaware of any measurement issues arising from (a) the fact that various commonly-used measures of differences between rates at which advantaged and disadvantaged groups experience some adverse or corresponding favorable health or healthcare outcome tend to yield opposite conclusions about such things as whether a disparity is increasing or decreasing over time or (b) the fact that all commonly-used measures of such differences tend to change solely because the overall prevalence of an outcome changes.

The same may be said of an article in the March 2019 issue of *Health Affairs*, co-authored by directors of NQF,² that purports to “demonstrate how the road map can be applied [by presenting] an example of how measurement and value-based payment can be used to reduce racial disparities in hypertension among African Americans.” Even though the aforementioned measurement issues are highly relevant to the subject of the article (see pages 4-5 of the August 29, 2017 letter and note 10 *infra*), the article gives no indication that the authors are aware of such issues.

² Anderson AC, O'Rourke E, Chin MH, *et al.* Promoting health equity and eliminating disparities through performance measurement. *Health Affairs* 2018 Mar;37(3):371-377. doi: 10.1377/hlthaff.2017.1301

The misleading nature of the *Roadmap* can be well illustrated by reference to the following facts. In four official and unofficial papers between 2004 and 2010,³ National Center for Health Statistics (NCHS) statisticians recognized that as health and healthcare generally improved, relative differences in the increasing, favorable outcomes (*e.g.*, infant survival, receipt of appropriate care) tended to decrease, while relative differences in the corresponding decreasing, adverse health and healthcare outcomes (*e.g.*, infant mortality, nonreceipt of appropriate care) tended to increase. See my [Comments for Commission on Evidence-Based Policymaking](#) (Nov. 14, 2016) (CEP Comments), "[The Mismeasure of Health Disparities](#)," *Journal of Public Health Management and Practice* (July/Aug. 2016), "[Race and Mortality Revisited](#)," *Society* (July/Aug. 2014), "[Measuring Health and Healthcare Disparities](#)," Proceedings of the Federal Committee on Statistical Methodology 2013 Research Conference (Mar. 2014), and "[Can We Actually Measure Health Disparities?](#)," *Chance* (Spring 2006). See also my letters to NQF leadership dated [August 29, 2017](#), [October 26, 2012](#), and [October 22, 2009](#) (which are attached hereto as Attachments A to C), as well as my comments at page 113 of the *Roadmap*.

Despite the fact that NCHS recognized this pattern long ago, the great majority of persons involved in the analyses of health and healthcare disparities are unaware that it is even possible for the relative difference in a favorable health or healthcare outcome and the relative difference in the corresponding adverse outcome to change in opposite directions as the prevalence of an outcome changes, much less that NCHS found that such pattern tends to occur systematically or that the pattern in fact is usually observed whenever there occur substantial changes in the overall prevalence of an outcome.

As discussed in the letters of August 29, 2017, and October 26, 2012 (and in "Race and Mortality Revisited" at 343-344), the NQF-sponsored [Commissioned Paper: Healthcare Disparities Measurement](#) (2012), which gave substantial attention to the fact that a relative difference and the absolute could yield opposite conclusions about directions of changes in disparities, at least recognized that it was possible for relative differences in a favorable health or healthcare outcome and relative differences in the corresponding adverse outcome to change in opposite directions. And, though unrecognized or acknowledged in the document, all examples the *Commissioned Paper* used to show that a relative difference in an outcome changed in the opposite direction of the absolute difference (as in Table 6 at page 37 and Table 7 at 39, as well as the discussion at page 37-38 regarding changes in coronary artery bypass grafting disparities), the unmentioned relative difference changed in the opposite direction of the mentioned relative difference and the same direction as the absolute difference.⁴ Further, in the instances where

³ The most significant of these papers is the monograph by Keppel Kenneth G., Pamuk Elsie, Lynch John, et al. titled [Methodological Issue in Measuring Health Disparities](#), *Vital Health Stat* 2005;2 (141), which was co-authored by the lead author of the NQF sponsored *Commissioned Paper: Healthcare Disparities Measurement* discussed *infra*.

⁴ The patterns by which measures tend to change as the prevalence of an outcome changes merely reflect tendencies, and, because other factors will be at work, those patterns may not always be observed even when the tendencies are

data were presented that allowed one to compare the size of relative differences for favorable and adverse outcomes within subgroups (as in Figure 4 at 33 and Figure 8 at page 53), relative differences in favorable outcomes and relative differences in adverse outcomes showed opposite patterns as to the comparative size of disparities.⁵

The situation presented in Table 6 of the *Commissioned Paper* is especially relevant to issues unaddressed in the *Roadmap*. In contrasting an increase in the relative racial difference in failure to be tested for prostate cancer with a decrease in the absolute difference between black and white rates (which is the same regardless of whether one examines the favorable or adverse outcome), the *Commissioned Paper* relied on authors who also authored the study underlying Table 5 (at 21) of the August 29, 2017 letter. In the case of the study underlying Table 5 of the letter, those authors had relied on an NCHS recommendation to measure all healthcare disparities in terms of relative differences in adverse healthcare outcome. They thus found that a large overall increase in mammography was accompanied by a very large increase in socioeconomic disparity based on the increase in the relative difference in nonreceipt of mammography, though they would have found a very large decrease in disparity based on the decrease in the relative

playing an important role. By contrast, it is mathematical fact that whenever a relative difference and the absolute difference change in opposite directions, the unmentioned relative difference will have changed in the opposite direction of the mentioned relative difference and the same direction as the absolute difference. See "Race and Mortality Revisited" (at 335) and the CEP Comments (at 15 note 26).

⁵ A corollary to the pattern recognized by NCHS is that in settings or within subgroups where adverse outcomes are comparatively uncommon, relative demographic differences in the adverse outcomes tend to be larger, while relative differences in the corresponding favorable outcomes tend to be smaller, than in settings where the adverse outcomes are comparatively common. Thus, among other examples that can be derived from Figure 4 of the *Commissioned Paper*, relative racial differences in nonreferral for kidney transplant are larger, while relative racial differences in referral for kidney transplant are smaller, among appropriate candidates than inappropriate candidates. Figure 8 of the *Commissioned Paper* shows that relative racial differences in self-rated health (SRH) less than good are larger, while relative racial differences in SRH good or better are smaller, within the highest income group than the lowest income group. See Figure 5 (slide 66) of my "[The Mismeasure of Health Disparities in Massachusetts and Less Affluent Places](#)," Quantitative Methods Seminar, Department of Quantitative Health Sciences, University of Massachusetts Medical School (Nov. 18, 2015). See also discussion at pages 9-10 of August 29, 2017 letter regarding confusion in the January 15, 2017 Final Report [Disparities in Healthcare and Health Outcomes in Selected Conditions](#) as to whether gender differences in advanced chronic kidney disease were greater among older persons or younger persons.

The pattern described in the preceding paragraph is not present in the illustration in the *Commissioned Paper's* discussion of Interaction Effects at page 43-44. That discussion and the illustration involve an unusual study that found identical cardiac catheterization referral rates for black men, white men, and white women, but lower referral rates for black women. See my [Comment on Schulman](#), *Journal Review* (June 2, 2007) regarding the many problems in that study. The *Commissioned Paper* then treated the concept of interaction as solely involving the situation where a disparity exists only as to a subgroup, something that is extremely rare if it in fact exists at all other than in that study. Perceptions of interaction typically involve situations where a factor shows different proportional effects on rates of different subgroups for the outcome (favorable or adverse outcome) one happens to be examining. Practically nothing said about such perceived interactions is sound, however, as a result of the failure to understand the patterns described in the preceding paragraph. See "Race and Mortality Revisited" at 340-341.

difference in receipt of mammography. As discussed with respect to Table 5 of the August 29, 2017 letter and in “The Mismeasure of Health Disparities,” in 2015 NCHS reversed that recommendation such that healthcare disparities would now be measured in terms of relative differences in receipt of screening.

Beginning with the examples in Table 1 at page 10, the *Roadmap* mentions “screening” regarding cancer and other adverse conditions 111 times as things that should be examined in an effort to reduce disparities. Yet the document gives no indication that relative differences in screening and relative differences in non-screening (the approach one might consider to be taken for granted by the *Commissioned Paper*) could yield opposite conclusions about directions of changes in the size of disparities (and the role of policies in causing such changes) and whether by screening the *Roadmap* actually means non-screening. But the same point could be made with regard to the great majority of things that the *Roadmap* suggests should be examined in order to quantify and monitor health and healthcare disparities,⁶ while failing to suggest even the possibility that different measures could yield opposite conclusions about the directions of changes in disparities or the effects of programs on the disparities, thus bolstering the view among the great majority of users of the document that there is no such possibility.⁷

The fundamental scientific problem with the *Commissioned Paper* was the failure to reflect any understanding of patterns by which measures tend to be affected by the prevalence of an outcome or to offer guidance for taking those patterns into account in appraising the effects of policies on differences in the circumstances of advantaged and disadvantaged groups. National Quality Forum’s more recent actions regarding health and healthcare disparities, while suffering from the same scientific problem, also suffer from the ethical problem of failing to alert entities that rely on NQF guidance of the existence of measurement issues, even as NQF encourages

⁶ The *Roadman* terms the things to be examined “measures.” While the usage appears to be consistent with NQF practice, it gives readers the impression that the *Roadmap* is actually providing guidance on how to measure disparities, as suggested by the *Modern Healthcare* article mentioned on page 2 *supra*. The *Roadmap* provides a great deal of information about the subjects to be examined with respect to disparities issue but nothing whatever about how to measure disparities in those subjects. It does not even indicate whether one should look at advantaged and disadvantaged groups’ rates of experiencing favorable outcomes or their rates of experiencing the corresponding adverse outcomes.

⁷ As discussed in the August 29, 2017 letter (at10), in finding that a general increases in cervical cancer screening rates was associated with a decrease in the disparity between screening rates of American Indian/Alaska Native women and white women, the NQF March 20, 2017 Final Report [*Effective Interventions in Reducing Disparities in Healthcare and Health Outcomes in Selected Conditions*](#), relied on absolute differences between rates to measure disparities. In doing so, however, it showed no recognition that general increases in screening rates would commonly be accompanied by an increase in relative differences in non-screening rates. Nor did it show an awareness that (a) increases in screening rates where rates are generally low would tend to increase absolute differences between rates (b) while increases in screening rates where rates are generally high will tend to reduce absolute differences between rates.

those entities to devote substantially greater resources to the study of health and healthcare disparities and to base policies on the results of such studies.

As discussed in the letters of October 26, 2012 (at 5), and August 29, 2017 (at 7), the *Commissioned Paper's* cryptic reference to changes in the prevalence of an outcome added in response to my comments did nothing to alert readers to the importance of considering the way various measures tend to be systematically affected by changes in the prevalence of an outcome, and nothing was done to follow through with the authors' suggestion that the matter would be addressed in the September 2012 Technical Report [Healthcare Disparities and Cultural Competency Consensus Standards](#) (*Consensus Standards* document). While saying nothing on that matter, the *Consensus Standards* document did at least recognize that a relative difference and the absolute difference could yield opposite conclusions as to the directions of changes in disparities. But it failed even to indicate an awareness that relative differences in favorable health and healthcare outcome and relative differences in the corresponding adverse health and healthcare outcomes could (or commonly would) change in opposite directions as the prevalence of an outcome changes or that this would always be the case in any situation where a relative and absolute difference showed opposite directions of changes in disparities. See note 4 *supra*.⁸

In August 2014, NQF issued [Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors](#) (*Risk Adjustment* document). The document, which principally addressed adjustment for patient characteristics in appraising hospital performance, gave considerable attention to healthcare disparities because of concerns that pay-for-performance programs could increase healthcare disparities. The concern rested in the fact that disadvantaged group representation tended to be greater at lower-performing hospitals than at higher-performing hospitals, and it was believed that adjustment for patient characteristics might at least address the role of patient characteristics in hospital performance differences. While it addressed some comparatively sophisticated adjustment issues, however, nothing in the *Risk Adjustment* document suggested an awareness that different measures might yield opposite conclusions about whether a disparity was larger in one hospital than another. Though various members of the Expert Panel producing the document and various references it cited had promoted the inclusion of a disparities element in pay-for-performance programs in order to counter the ways such programs might increase disparities, to my knowledge neither those members nor the references had indicated an awareness that different measures could yield opposite conclusions about the directions of changes in disparities or whether a disparity was larger in one hospital than another. Nothing in the document suggested an awareness of the issues addressed in the October 22, 2009 and October 26, 2012 letters regarding ways failure to address measurement issues would undermine the use of a disparities element in pay-for-performance programs or an understanding of the way that, as discussed in the October 26, 2012 letter (at 7-8) and "Race and

⁸ See pages 6-7 of the October 26, 2012 letter regarding the way the *Consensus Standards* document adopted language from the *Commissioned Paper* that suggested the *Consensus Standards* document was going to explain that relative differences in favorable outcomes and relative differences in the corresponding adverse outcomes could yield opposite conclusions about directions of changes in disparities but ultimately did not do so.

Mortality Revisited" (at 337-339), reliance on absolute differences between rates in the disparities element of the Massachusetts Medicaid pay-for-performance program would itself operate to increase health and healthcare disparities. See note 10 *infra*.

To my knowledge, the only reference cited in the *Risk Adjustment* document that showed an awareness that different measures could yield different conclusions about the comparative size of a disparity over time or in different hospitals was the *Consensus Standards* document. The *Risk Adjustment* document did not mention the *Commissioned Paper* at all. The *Risk Adjustment* document, however, did recommend that NQF appoint a standing Disparities Committee.

NQF's following of that recommendation then presumably led to the *Roadmap* and the other disparities documents issued in 2017.⁹ The production of these documents represented expenditure of substantial resources and the documents placed great emphasis on identifying and incentivizing the policies that were shown to reduce disparities. That is something that it is impossible to do without consideration of the ways the measures employed to quantify disparities tend to be affected by the overall prevalence of an outcome. Thus, one might expect that those documents would finally reflect an effort by NQF actually to address the crucial measurement issues ignored in the *Commissioned Paper* and the *Consensus Standards* and *Risk Adjustment* documents.

The recent documents, however, compound the problems created by earlier documents by failure even to recognize that different measures can yield opposite conclusions about directions of changes in disparities while recommending the expenditure of vastly greater resources to monitor disparities. That is, while failing to offer any indication that different measures commonly yield opposite conclusions about direction of changes in disparities regarding something like breast cancer screening, much less offering any guidance as to how to evaluate the policies involved in such situations, the *Roadmap* recommends that efforts be made to determine whether inequity is increasing or decreasing with respect to hundreds of such matters. In sum, in circumstances where it would be a manifest waste of resources to attempt to determine whether the inequity is increasing or decreasing or the role of policies in such increases or decreases even as to a single subject without consideration of the way the measure employed to quantify the disparity tends to be affected by the prevalence of the outcome (see discussion of Table 2 in "Race and Mortality Revisited" at 329-330, 343), the *Roadmap* recommends that such exercise be expanded exponentially.

As discussed above, NQF actions subsequent to issuance of the *Roadmap* further the misunderstanding created by the *Roadmap* itself and the other NQF 2017 document by creating

⁹ Such guidance includes (1) the January 15, 2017 Final Report [Disparities in Healthcare and Health Outcomes in Selected Conditions](#), (2) the March 20, 2017 Final Report [Effective Interventions in Reducing Disparities in Healthcare and Health Outcomes in Selected Conditions](#), (3) the June 15, 2017 Final Report [An Environmental Scan of Health Equity Measure Development](#).

the impression there exist no problems in determining whether a policy increased or decreased a disparity.¹⁰

There is reason to believe that, whatever the precise attention the 2019 NQF conference gives to health and healthcare equity issues, the currently planned presentations will merely further promulgate the misunderstandings promoted by the *Roadmap*. Such attention will likely also compound the damage caused by the *Roadmap* and related NQF activities by further encouraging the expenditure of additional resources in undertakings that so far have produced very little of value but a great deal that it is misleading about the effects of policies on health and healthcare disparities. Because the matter is entirely unaddressed in the *Roadmap* and the other 2017 documents, it is improbable that the pre-conference event “Evolving Measurement – A Learning Symposium” will even touch upon the possibility for different measures of health and healthcare disparities to yield different conclusions as to whether policies have improved or worsened such disparities.

Consider, however, a situation where an attendee at the conference asks panelists the simple question of whether – for purposes of determining if a program like cultural competency training actually achieved its desired result or was worth its cost or determining which programs aimed at reducing disparities should be incentivized and which should be discouraged – one should measure disparities in terms of relative differences in favorable outcomes like control of hypertension or screening for cancer or relative differences in the corresponding adverse outcomes like uncontrolled hypertension or failure to receive cancer screening. It is unlikely that panelists would be aware that it could make a difference which outcome one examined. Confronted with the similar questions with reference to the hypothetical situations illustrated in

¹⁰ The 2018 *Health Affairs* article co-authored by NQF directors referenced in note 2 provides a useful case study in the failure of understanding. The article relies on relative differences in adverse outcomes as an indicator of the problems arising from differences in hypertension rates of whites and racial minorities. But in discussing the failure of the disparities element in the Massachusetts Medicaid pay-for-performance program to achieve desired results, it shows no understanding of the way that (as discussed in "Race and Mortality Revisited" at 337-33, the October 26, 2012 letter at 3, and the August 29, 2017 letter at 5), reliance on a measure that was a function of absolute differences between rates to measure healthcare disparities, while examining subjects where favorable outcome rates were generally high, the disparities element of the program systematically favored hospitals that served comparatively minorities few for reasons that were unrelated to the degree of inequity within each hospital. And, even though National Health and Nutrition Survey data make it abundantly clear that general reductions in systolic blood pressure will tend to increase relative racial differences in hypertension while reducing relative differences in avoiding hypertension and that general improvements in control of hypertension will tend to increase relative differences in uncontrolled hypertension while reducing relative differences in rates of controlling hypertension (see the [NHANES Illustrations](#) subpage of the [Scanlan's Rule](#) page of [jpscanlan.com](#)), the article suggests no awareness even of a possibility for relative differences in favorable outcomes and relative differences in the corresponding adverse outcomes to change in opposite directions. Similarly, generally reducing rates of non-adherence to medication regimens will tend to increase relative racial differences in non-adherence rates, while reducing relative differences in adherence rates. It is of course true that actions particularly focused on racial minorities groups can reduce all measures of inequity related to hypertension. But it is impossible to identify such effects without understanding the patterns by which general improvements in control of hypertension and general reductions (or increases) in the causes of hypertension will tend to affect various measures of racial disparity.

Tables 1 and 2 of the August 29, 2017 letter, or the actual situations presented in the subsequent tables in the letter, panelists would likely be unaware that such situations could exist much less that they are commonplace. Ultimately, it would be unlikely that panelists could provide any guidance on such subjects beyond referring the questioners to the *Commissioned Paper*. As explained, however, that document provides no guidance whatever for distinguishing the effects of policies from effects on measures that are functions of changes in the prevalence of an outcome.

More likely, however, attendees will raise no such issues because, in part due to NQF actions, attendees will be unaware of the existence of the issues. Thus, to the extent that the conference addresses health equity, the probable effect of its doing so is to further persuade attendees of the importance of devoting resources to the study of how policies affect health equity, while leaving the attendees not only entirely in the dark as to how to identify the effects of policies on health equity, but entirely in the dark as to the existence of issues that complicate that process.

Given the nature of NQF efforts in this area so far, it is doubtful that NQF could effectively address these issues at the upcoming conference. But NQF can at least alert attendees to the existence of issues arising from the fact that (a) different measures commonly yield opposite conclusions about the comparative size of health and healthcare disparities at different points in time or in settings differentiated other than temporally and (b) that it is not possible to determine the effect of a policy on a health or healthcare disparity without consideration of the way the measures employed to quantify the disparity tend to be affected by the prevalence of an outcome. NQF should also withdraw the *Roadmap*, *Commissioned Paper*,¹¹ and *Consensus Standards* documents, as well as any other documents that purport to be provide guidance on quantifying health and healthcare disparities or that stress the importance of studying such issues without alerting readers as to the measurement issues that must be considered for such study to be of any value.

Thereafter, assuming that NQF believes it has or can secure the necessary expertise to do so, NQF should endeavor to study how actually to quantify health and healthcare disparities and how to appraise the effects of policies on those disparities. But researchers who are unaware either of the ways measures tend to change as the prevalence of an outcome changes, or of the need to distinguish the extent to which observed patterns are functions of changes in the

¹¹ See pages 343-344 or "Race and Mortality Revisited" regarding the reasons proffered by the research integrity officers of Massachusetts General Hospital and Harvard Medical School for refusing to withdraw the *Commissioned Paper* (in the letter from those officers attached as Attachment D). Whatever the validity of the reasons proffered by those institutions, the situation is different for NQF. Massachusetts General Hospital and Harvard Medical School merely took the position that they did not independently review the work of faculty. They were not standing behind the *Commissioned Paper*. NQF, however, is standing behind the *Commissioned Paper*. And if the failure of the *Commissioned Paper* to address the effects of the overall prevalence of an outcome on the measures it discussed could be excused by an intention to address the issue in subsequent NQF documents, the subsequent failure to follow through with that intention (and subsequent obscuring of the issues) vitiates that excuse.

Shantanu Agrawal, MD, MPhil, President and CEO
Members of the Board of Directors
National Quality Forum
March 15, 2019
Page 10

prevalence of an outcome and the extent to which the patterns reflect the effects of policies, can provide nothing of value in such undertaking.

In the meantime, however, NQF should discourage, rather than encourage, the expansion of areas of study regarding health and healthcare disparities without effective tools for quantifying the disparities or for determining the effects of policies on those disparities.

Inasmuch as NQF has now embarked on a joint [project](#) with Aetna Foundation to identify quality and payment innovations that can reduce health disparities, NQF should immediately address these issues with Aetna Foundation. For, given the impression created that *Roadmap* and other documents that NQF has substantial expertise in determining what practices reduce disparities, Aetna Foundation is likely to heavily rely on NQF guidance.

More broadly, at least with respect to NQF guidance regarding health and healthcare disparities, rather than celebrating its twenty years of contributions to sound healthcare policy as suggested by the conference web page, NQF should engage in a process of circumspections about whether its activities have done more harm than good. For, as reflected in the October 22, 2009 letter, for more than half of those twenty years NQF has promoted unsound health and healthcare disparities research as a result of failure to understand the ways measures of disparities tend to change solely because the prevalence of an outcome changes. And for almost ten years after these issues were specifically brought to the organization's attention, NQF has continued to ignore and has even obscured crucial issues at the same time that it has devoted increasing resources, and has encouraged numerous other entities to devote increasing resources, to disparities research that cannot be worthwhile without careful consideration of those issues.

Sincerely,

/s/ James P. Scanlan

James P. Scanlan

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

August 29, 2017

ELECTRONICALLY TRANSMITTED

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
1030 15th Street, NW
Suite 800
Washington, DC 20005

Re: Unsoundness of National Quality Forum Guidance on the Measurement
of Health and Healthcare Disparities

Dear President Agrawal:

This letter, which follows on a October 22, 2009 [letter](#)¹ to National Quality Forum (NQF) President and CEO Janet M. Corrigan and an October 26, 2012 [letter](#) to NQF Interim President and CEO Laura Miller (and a comment I submitted regarding the Draft Report *A Roadmap to Reduce Health and Healthcare Disparities through Measurement*), addresses the essential unsoundness of NQF guidance relating to health and healthcare disparities research and practices aimed at reducing those disparities. That unsoundness is a result of the guidance's failure to recognize both (a) that different measures of disparities commonly yield opposite conclusions about directions of changes in disparities and (b) that it is not possible to usefully analyze the effects of policies on disparities without consideration of the ways the measures employed in analyses tend to be affected by the prevalence of an outcome.

NQF guidance issued in 2011 and 2012, while recognizing, to a degree, that different measures of health and healthcare disparities can yield opposite conclusions about changes in directions of disparities over time, failed to recognize that measures commonly employed in health and healthcare disparities research tend to be systematically affected by the prevalence of an outcome. NQF guidance issued in final or draft form in 2017 fails to reflect an understanding

¹ To facilitate consideration of issues raised in documents such as this I include links to referenced materials in electronic copies of the documents, in some cases, for the reader's convenience, providing the links more than once. Such copies are available by means of the [Measurement Letters](#) page of jpscanlan.com. If the online version of the letter is amended, such fact will be noted on its first page.

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 2

that it is even possible for different measures to yield opposite conclusions about directions of changes in disparities.

Because one of the important documents discussed here is an NQF-sponsored guide that was funded with a grant from the Robert Wood Johnson Foundation (RWJF), I am sending the letter to RWJF President and CEO Richard Besser. I am also bringing the letter to the attention of persons and entities affiliated with NQF to encourage their participation in correcting unsound practices of the organization. I urge you to circulate the letter widely among staff and members of the NQF, both to educate them on the subject of the letter and to elicit their assistance in identifying other NQF (or NQF member) activities to which the issues raised in the letter pertain.

Further, the principal subject of the letter involves a substantial federal government contract (HHSM-500-2012-000091) awarded by the Center for Medicare & Medicaid Services (CMS), an agency within the Department of Health and Human Services (HHS). In a July 17, 2017 [letter](#) to the Secretary of HHS, I discussed (at 4) that HHS-funded activities involving analyses of demographic differences that failed to consider the ways measure employed in such analyses tend to be affected by the prevalence of an outcome had yielded very little of value and much that was misleading. I therefore suggested that the agency institute a moratorium on grants and contracts (and activities pursuant to grants and contracts already awarded) where implications of the failure to consider ways measures tend to be affected by the prevalence of an outcome are pertinent. As shown below, NQF activities pursuant to the referenced HHS contract that address reduction in health and healthcare disparities without reflecting an understanding of the pertinent measurement issues provide a compelling example of situations where expenditure of federal funds cannot be justified. Therefore, I am also sending the letter to the federal officials overseeing the contract. I may also make further reference to NQF activities involving this contract for exemplary purposes in seeking to prevent wasteful expenditure of federal funds.²

² The July 17, 2017 letter, which I discuss in "[Innumeracy at the Department of Education and the Congressional Committees Overseeing It](#)," Federalist Society Blog (Aug. 24, 2017), and "[The Government's Uncertain Path to Numeracy](#)," Federalist Society Blog (July 21, 2017), was sent also the heads of the Departments of Education and Justice and principally addresses the government's mistaken belief that generally reducing public school discipline rates will tend to decrease, rather than increase, relative differences in discipline rates and the proportions more susceptible groups make up of persons disciplined. The mistaken belief that generally reducing an outcome would be expected to reduce relative differences in rates of experiencing the outcome is as pervasive in the social and medical science research communities as it is in the government's civil rights establishment and has for decades undermined the interpretation of demographic differences in health, healthcare, and other outcomes. See my "[Race and Mortality](#)," *Society* (Jan./Feb. 2000). The persistence of a belief that is the exact opposite of reality, even among highly regarded experts in the analyses of demographic differences, illustrates the need for dramatically enhanced circumspection in the award of federal research funds. See my [Comments](#) for Commission on Evidence-Based Policymaking (Nov. 14, 2016). See discussion *infra* of the mistaken belief reflected in the NQF January 21, 2017 Final Report [Disparities in Healthcare and Health Outcomes in Selected Conditions](#) that a program aimed at generally reducing death from sudden infant death syndrome should tend to reduce, rather than increase, relative racial differences in SIDS deaths.

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 3

The principal purpose of this letter is to explain fundamental problems in the guidance NQF has so far provided (and intends to provide) on the measurement of health and healthcare disparities. As a result of the failure to recognize the ways measures employed in such research tend to be systematically affected by the prevalence of an outcome, the guidance will tend to promote wasteful and misleading health and healthcare disparities research.

The following is a simple summary of the pertinent statistical principles that may facilitate the reader's understanding of the points that follow. For reasons related to shapes of distributions of factors associations with likelihood of experiencing and failing to experience an outcome, as health and healthcare generally improve, relative demographic differences in favorable health and healthcare outcomes (*e.g.*, survival, receipt of appropriate care) tend to decrease, while relative differences in the corresponding adverse outcomes (*e.g.*, mortality, nonreceipt of appropriate care) tend to increase. This pattern was recognized by the National Center for Health Statistics (NCHS) more than a decade ago.

Absolute (percentage point) differences and differences measured by odds ratios tend also to be affected by the prevalence of an outcome, though in a more complicated way than the two relative differences. Roughly, as uncommon health and healthcare outcomes (less than 50% for both groups being compared) generally increase, absolute differences between rates tend to increase; as common health and healthcare outcomes (greater than 50% for both groups being compared) generally increase, absolute differences tend to decrease. The prevalence related patterns of changes in absolute differences is less predictable when any group's rate crosses the 50% point during the period examined. As the prevalence of an outcome changes, differences measured by odds ratios tend to change in the opposite direction of the absolute difference.

All of the measures may change in the same direction as the prevalence of an outcome changes, in which case one may infer that the observed pattern reflects something other than the consequence of a general change in the prevalence of an outcome. But anytime a relative difference and the absolute difference have changed in opposite directions, the other relative difference will necessarily have changed in the opposite direction of the first relative difference and the same direction as the absolute difference.

Tables 1 and 2 illustrate some of the implications of these patterns with regard to matters like interpretation of the effects of incentive programs on healthcare disparities. The two tables are variations on Table of 1 of "[Measuring Health and Healthcare Disparities](#)," Proceedings of Federal Committee on Statistical Methodology 2013 Research Conference (March 2014) (FCSM Paper), which explains its specifications. Versions of the table are also used in my [Comments for Commission on Evidence-Based Policymaking \(CEBP\)](#) (Nov. 14, 2016), "[The Mismeasure of Health Disparities](#)," *Journal of Public Health Management and Practice* (July/Aug. 2016), and "[Race and Mortality Revisited](#)," *Society* (July/Aug. 2014).³

³ See discussion of Table 5 in "Race and Mortality Revisited" (at 335-336) refuting arguments that a value judgment is involved in choosing between a relative difference and the absolute difference when the two yield opposite conclusions about directions of changes in (or the comparative size) of demographic differences.

Shantanu Agrawal, MD, MPhil
 President and CEO
 National Quality Forum
 August 29, 2017
 Page 4

The two pairs of rows in Table 1 show the effects of general increase in an uncommon procedure (first two rows) and a common procedure (second two rows) on various measures of differences between procedures rates of an advantaged group (AG) and a disadvantaged group (DG). The table shows that for both types of procedures, as rates increase, relative differences in the favorable outcome (receipt of the procedure) increase while relative differences in the adverse outcome (failure to receive the procedure) decrease.⁴ The table also shows that as procedure rates increase for the uncommon procedure, the absolute differences increases, and as procedure rates increase for the common procedure, the absolute difference decreases. As to each type of procedure, the difference measured by the odds ratio changes in the opposite direction of the absolute difference.⁵

Table 1. Hypothetical Rates of Receipt of Uncommon and Common Procedures of an Advantaged Group (AG) and a Disadvantaged Group (DG) at Two Points in Time, with Disparity Measures

Proc Type	Time	AG Rate	DG Rate	AG/DG Fav Ratio	DG/AG Adverse Ratio	Abs Diff (perc pts)	AG/DG Fav Odds Ratio
Uncommon	Year 1	20.0%	9.0%	2.22	1.14	11.0	2.53
Uncommon	Year 2	40.0%	22.6%	1.77	1.29	17.4	2.28
Common	Year 1	70.0%	51.0%	1.37	1.63	19.0	2.24
Common	Year 2	80.0%	63.4%	1.26	1.83	16.6	2.31

Table 2 uses the same data to show patterns of differences at lower-performing and higher-performing hospitals, assuming the latter have generally higher favorable outcome rates.

⁴ While I commonly refer to patterns of relative differences in discussions of these issues, the table actually presents rate ratios. The relative difference is the rate ratio minus 1 where the rate ratio is above 1 and 1 minus the rate ratio where the rate ratio is below one. In the former case, the larger the rate ratio, the larger the relative difference; in the latter case, the smaller the rate ratio, the larger the relative difference. One should be careful not to mistakenly refer to the rate ratio as the relative difference. But the distinction between the two terms is not pertinent to the discussion here of patterns by which the two relative differences tend to be affected by the prevalence of an outcome. In recent years I commonly present the rate ratios for both outcomes with the larger figure in the numerator, in which case, as to both outcomes, the larger the rate ratio, the larger the relative difference. Choice of numerator in the ratio, however, has no bearing on the patterns by which the two relative differences tend to be affected by the prevalence of the outcome.

⁵ There are four possible odds ratios depending on which outcome is used as the numerator in calculating each group's odds and which group's odds is used as the numerator of the odds ratio. Two are the same as each other and two are the reciprocals of the first two. The table shows the odds ratio in terms of the ratio of the favorable odds of AG to the favorable odds of DG, which is the same as the ratio of the adverse odds of DG to the adverse odds of AG. In discussion of the relationship of the odds ratios to the absolute difference, I try to be careful to say the difference by the odds ratio tends to change in the opposite direction of the odds ratio, since whether a change in the size of the odds ratio reflects an increase in the difference or a decrease in the difference depends on whether the odds ratio is above or below 1.

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 5

Table 2. Hypothetical Rates of Receipt of Uncommon and Common Procedures of an Advantaged Group (AG) and a Disadvantaged Group (DG) at Lower-Performing and Higher-Performing Hospitals, with Disparity Measures

Proc Type	Hospital	AG Rate	DG Rate	AG/DG Fav Ratio	DG/AG Adverse Ratio	Abs Diff (perc pts)	AG/DG Fav Odds Ratio
Uncommon	Higher	20.0%	9.0%	2.22	1.14	11.0	2.53
Uncommon	Lower	40.0%	22.6%	1.77	1.29	17.4	2.28
Common	Lower	70.0%	51.0%	1.37	1.63	19.0	2.24
Common	Higher	80.0%	63.4%	1.26	1.83	16.6	2.31

As to both types of procedures, higher-performing hospitals will tend to show smaller relative differences in favorable outcomes but larger relative differences in adverse outcomes. But for the uncommon procedures, absolute differences will tend to be larger at higher-performing hospitals; for the common procedures, absolute differences will tend to smaller at higher-performing hospitals. As to both types of outcomes, the difference measured by the odds ratio will tend to show a pattern that is the opposite of that shown by the absolute difference.

See "Race and Mortality Revisited" (at 337-339) regarding the way that the Massachusetts Medicaid pay-for-performance program included a disparities element that employed a measure that was a function of the absolute difference⁶ and examined types of care where rates were quite high (as in the second two rows of Table 2). Thus, the disparities element tended to favor higher-performing hospitals for reasons having nothing to do with the degree of within hospital equity. Since minorities tend to make up smaller proportions of patients at higher-performing than lower-performing hospitals, the disparities element of the program tended to divert resources away from hospitals where minorities comprise a comparatively high proportion of patients. Thus, the disparities element of the pay-for-performance would itself tend to worsen the comparative healthcare situation of minorities.

These patterns, of course, will not be observed in every situation since other factors are at work. Those other factors are in fact the principal concern of disparities research, especially with regard to things like understanding the effects of incentive programs on health and healthcare disparities. But without understanding these patterns, it is not possible to divine anything useful regarding whether changes in any measure reflect something other than a general change in the prevalence of the outcome. In addition to the aforementioned letter to the Secretary of HHS, see my "Race and Mortality Revisited" (especially with regarding to the discussion of its Table 2 (at

⁶ The particular measure in the Massachusetts program (the between group variance which is based on the absolute difference) has problems beyond those associated with absolute difference itself. See the [Between Group Variance](#) sub-page of [Measuring Health Disparities](#) page of [jpscanlan.com](#)

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 6

329-330, 337)) and my November 14, 2016 CEBP comments (especially the discussion regarding the fourth recommendation (at 46-47)).

By way of example of the described patterns with respect to a general increase in an uncommon outcome, a 2004 article co-authored by NQF Disparities Standing Committee member Jose Escarce,⁷ which I have discussed in numerous places and which I will discuss further in several places below, found that between 1986 and 1997 angiogram rates for Medicare patients increased from 8.6% to 22.8% for whites and from 4.3% to 16.10% for blacks. Thus, as commonly occurs in the circumstances (and as illustrated in the first two rows of Table 1), the relative difference in receipt of angiogram decreased, while both the relative difference in failure to receive angiogram and the absolute difference between rates increased. See Table 4 of my 2008 [Nordic Demographic Symposium](#) presentation.⁸

The failure of NQF experts (and NQF contractors) to understand patterns like this even after they were brought to the attention of NQF leadership has undermined all NQF guidance regarding health and healthcare disparities where quantification of disparities is pertinent. That is so even when the guidance has recognized that different measures can yield opposite conclusions about changes in directions (or the comparative size) of certain health and healthcare disparities, as in (a) the October 2011 [Commissioned Paper: Healthcare Disparities Measurement](#) (*Commissioned Paper*), which was commissioned by NQF based on an RWJF grant and produced by researchers at Harvard Medical School and Massachusetts General Hospital, and (b) the September 2012 Technical Report [Healthcare Disparities and Cultural Competency Consensus Standards](#) (*Consensus Standards Technical Report*), which was produced by NQF itself.

The *Commissioned Paper* recognized that the relative difference the observer happened to be examining (that is, in a favorable health and healthcare outcome or in the corresponding adverse outcome) and the absolute difference could yield opposite conclusions about directions of changes in disparities. The *Commissioned Paper* also recognized that the relative difference in a favorable outcome and the relative difference in the corresponding adverse outcome could yield opposite conclusions about changes in the directions of disparities. But the document showed no recognition that measures tend to be affected by the prevalence of an outcome or even that the NCHS had specifically recognized that the two relative differences would tend to change systematically in opposite directions as the prevalence of an outcome changes. Nor did the document show any recognition that anytime a relative difference and the absolute difference

⁷ Escarce JJ, McGuire TG. Changes in racial differences in use of medical procedures and diagnostic tests among elderly persons: 1986-1997. *Am J Public Health* 2004;94:1795-1799. My online comment regarding the article, titled "Perceptions of changes in healthcare disparities among the elderly dependant on choice of measure," which was originally posted on Journal Review, may be found [here](#).

⁸ The rates for the subjects examined in the Escarce study are rather lower than those in the first two rows of Table 1 (as is also the case for the rates at issue in the Werner study discussed *infra*). But the patterns shown in the first two rows of Table 1 hold for those lower rates as well.

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 7

have changed in opposite directions, the other relative difference will necessarily have changed in the opposite direction of the first relative difference and the same direction as the absolute difference.

In response to my [criticisms](#) of the failure to address the effects of the prevalence of an outcome on the measure employed in healthcare disparities analyses, the draft *Commissioned Paper* was amended such that in the final document (at 36) the italicized language was included in the following statement:

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 the outcome changes.*

But the added language hardly alerted readers (or NQF itself) to the implications of the facts that measures tend to change, and change in contrasting ways, solely because the prevalence of an outcome changes.

The authors of the *Commissioned Paper* suggested that implications of the prevalence of an outcome would be more fully addressed in the *Consensus Standards Technical Report*.⁹ That document recognized that the relative difference the observer happened to be examining and the absolute difference could yield opposite conclusions about directions of changes in disparities. But it showed no recognitions that it was even possible for the relative difference in a favorable outcome and the relative difference in the corresponding adverse outcome to yield opposite conclusions about changes in the directions of disparities over time, much less that the *Commissioned Paper* had specifically recognized that possibility or that the NCHS had recognized that such pattern would tend to occur systematically.

The problematic aspects of the above documents, while discussed somewhat below, is also addressed in the aforementioned October 26, 2012 [letter](#) to NQF Interim President and CEO Laura Miller and leadership of other entities responsible the *Commissioned Paper* seeking withdrawal of the paper, and in my “[Race and Mortality Revisited](#),” *Society* (July/Aug. 2014). That article (at 343-344) also discusses the soundness of reasons proffered by research integrity officers of Harvard Medical School and Massachusetts General Hospital for declining to withdraw the *Commissioned Paper* (a matter addressed further below).

That article also discusses (at 333) as a particularly egregious waste of resources a study by the Institute for Medicine and Public Health of the Vanderbilt University Medical Center that sought to evaluate the effectiveness of quality improvement in reducing healthcare disparities, while showing no recognitions that different measures could (or systematically tend to) yield opposite conclusions about whether quality improvements increase or decrease disparities.¹⁰ The study is

⁹ See item 113 of the online [collection](#) of comments and responses.

¹⁰ See also my [AHRO's Vanderbilt Study](#) webpage.

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 8

cited approvingly in (and may have been used as a model for similar work in) the second NQF item listed in the paragraph below. And the recent work of NQF regarding health and healthcare disparities may be compared to the Vanderbilt study with respect to the like failure to recognize that improvements in healthcare will tend to increase certain measures of disparities and reduce other measures of disparities.

Despite emphasis on incentivizing practices proven to reduce health and healthcare disparities, more recent NQF guidance involving health and healthcare disparities fails to recognize even that different measures can yield different conclusions about whether such disparities are increasing or decreasing. Such guidance includes (1) the January 15, 2017 Final Report [*Disparities in Healthcare and Health Outcomes in Selected Conditions*](#), (2) the March 20, 2017 Final Report [*Effective Interventions in Reducing Disparities in Healthcare and Health Outcomes in Selected Conditions*](#), (3) the June 15, 2017 Final Report [*An Environmental Scan of Health Equity Measure Development*](#), and (4) the July 21, 2017 Draft Report [*A Roadmap to Reduce Health and Healthcare Disparities through Measurement*](#).

While items 3 (at 12-13) and 4 (at 44) make reference to the *Commissioned Paper*,¹¹ none of the four documents reflects any awareness that *Commissioned Paper* (or the *Consensus Standards Technical Report*) specifically recognized that different measures could yield different conclusions about whether particular practices increase or decrease health or healthcare disparities.¹²

In fact, none of the documents says anything whatever about the measurement of disparities and only two of the items even mention the size of any disparities. Item 1 refers to the size of several disparities in terms of relative differences or relative odds and item 2 describes a change in absolute differences with regard to the effect of a program on disparities (though characterizing a change in percentage point difference as a change in percent difference). In addition to characterization problems in discussions of the size of disparities in both documents,¹³ each

¹¹ In those mentions of the *Commissioned Paper*, the documents describe it as produced by the Disparities Solution Center of Massachusetts. Such usage is also found on the NQF website. Given the information on the cover of the document, that would seem a reasonable way to describe the documents. But, as discussed in "Race and Mortality Revisited" (at 344), Massachusetts General Hospital was unwilling to take responsibility for the document save to state that the failure to address the ways measures it discussed tend to be affected by the prevalence of an outcome did not violate ethical guidelines.

¹² This failure to recognize that different measures can yield opposite conclusions about directions of changes in disparities occurs even though the NQF Disparities Standing Committee includes one co-author of the *Commissioned Paper* and several members of the Steering Committee for the *Cultural Competency Technical Report*.

¹³ When discussing a greater likelihood, item 1 typically uses terms like "times higher" or "times more likely" when it means times "as high" or "as likely." That usage, while predominating even in the most prestigious scientific journals (apart from the *New England Journal of Medicine*), can lead readers to believe a relative difference (or increased odds) is a 100 percentage points higher than it actually is, and has even led the Institute of Medicine to

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 9

document's discussion of the comparative size of disparities, or of a change in the size of disparities, involves an error of some pertinence to the issue addressed in this letter.

Item 1, the January 15, 2017 Final Report, citing Yu et al.,¹⁴ discusses gender differences in chronic kidney disease (CKD) as follows: "Although women had greater prevalence of advanced CKD, they had decreased odds of having CKD compared to men; these disparities were most prominent amongst the elderly." "Most prominent amongst the elderly" presumably referred to gender difference in advanced CDK, which accords with the Yu study's finding.¹⁵

That there is a difference adverse to men for CKD generally but a difference adverse to women for advanced CKD might be something worth studying, since this pattern is contrary to the usual pattern where group that is more likely to experience an outcome is also more likely to experience an advanced form or the outcome. But my focus here involves the seemingly larger gender disparity in advanced CKD (measured in terms of odds ratio) among the older group than the younger group.

Whenever some adverse outcome is substantially more prevalent among an older group than a younger young (as in the case of advanced CKD), the relative difference in the adverse outcome is almost always larger among the younger group than the older group (while the relative difference in the corresponding favorable is almost always larger among the older group than the younger group). In fact, data on patterns of relative differences in adverse outcomes and corresponding favorable outcomes among younger and older persons are among the more useful illustrations of the pattern whereby the rarer an outcome the greater tends to be the relative difference in experiencing it and the smaller tends to be the relative differences in avoiding it. See my [Life Tables Illustrations](#) webpage and Table 6 (slide 14) of my [2008 Nordic Demographic Symposium](#) presentation. See also Table 11 (at 22) of my FCSM Paper, and Table 1 (at 4) of my [Mortality and Survival](#) webpage, regarding the way relative differences in cancer mortality are generally greater among younger groups while relative differences in cancer survival are generally greater among older groups.

read a 20 percent lower odds as an 80 percent greater odds. See my [Times Higher](#) webpage. The usage of "percent" in describing a percentage point difference, as in item 2, is responsible for a great deal of confusion in the discussion of health and healthcare disparities, possibly including the instances discussed below where (a) where the 2010 National Healthcare Disparities Report reported as among the largest reductions in healthcare disparities over a particular period situations where the report would also regard the disparities to be much larger at the end of the period than the beginning of the period (see *infra* at 23) and (b) where researchers described two studies that showed the same results as showing opposite results (see *infra* at 24-25). See my [Percentage Points](#) webpage. An organization providing guidance on measurement should be meticulous in its usage.

¹⁴ Yu MK, Lyles CR, Bent-Shaw LA, et al. Risk factor, age and sex differences in chronic kidney disease prevalence in a diabetic cohort: The Pathways Study. *Am J Nephrol.* 2012;36:245-251.

¹⁵ With regard to CKD generally as distinguished from advanced CKD, the study found a lower female to male odds among the younger group (.66) than the older group (.75), which means a larger difference in odds among the younger group.

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 10

The general prevalence of advanced CKD is in a range where the odds ratio for experiencing the outcome tends to approximate the relative risk for that outcome. Thus, a larger gender odds ratio among the older group than the younger group, being contrary to the usual pattern, might be something warranting attention.

In fact, however, in accordance with what one typically observes in the circumstances, the odds ratio was larger among the younger group than the older group. As shown in the article's figure 1 (at 249), the female/male advanced CKD odds ratio was 2.66 among the under 60 group and 1.63 among the 60 and older group. But the former odds ratio was not statistically significant, which is hardly surprising given that there were only 6 female and 2 male cases in the under 60 group.¹⁶ The study, however, treated the fact that the difference was not statistically significant as indicating that there was no gender difference within the under 60 group. Such treatment then presumably underlay the study's statement that the gender disparity was larger among the older group than the younger group.

The fact that the female/male odds ratio for advanced CKD within the younger group was not statistically significant may be reasons not to conclude with a strong degree of confidence that the disparity (as measured by the odds ratio) was larger for the younger group than the older group. But such fact provides no basis whatever for concluding that the disparity was larger among the older group than the younger group.

But only with an understanding of the patterns described here, and in prior communications to NQF, will one recognize that a finding that the relative difference for an adverse outcome was greater within an older group than a younger group is something that may warrant scrutiny.

Item 2, the March 20, 2017 Final Report, in the context of discussing the favorable effects on healthcare disparities of an Oregon coordinated care organization (CCO) program and citing Irvin *et al.*,¹⁷ states (at 14) that "the introduction of CCOs improved the cervical cancer screening rates for American Indian/Alaska Native women relative to white women, reducing the difference between the groups from 7 to 8 percent in 2012 to 5 percent in 2013." Actually, the Irvin study (at 81) found that the differences dropped from a range of 7 to 8 percentage points before the program to 6 percentage points in 2012 and then dropped from 6 to 5 percentage points between 2012 and 2013. This is a minor error given that the Irvin study attributes the decline in 2012 as well as the declines between 2012 and 2013 to the CCO program. But the associated figure (Figure III.20) in the Irvin study shows a pattern whereby general increases in cervical cancer screening from a time when all rates were well below 50% were associated with

¹⁶ In fact there were so few cases among the younger group compared with the older group that the excess odds figure for the overall group was the same as that for the older group.

¹⁷ Irvin CV, Bigby J, Byrd V, et al. [Midpoint Evaluation of Oregon's Medicaid Section 1115 Demonstration: Mid-2012 through mid-2014](#). Salem, OR: Oregon Health Authority (OHA); 2015.

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 11

increasing absolute differences between rates, but that rates had increased to a point where further increases tend to reduce absolute differences. The decline in the absolute difference between 2012 and 2013 was a function of continuing increase for American Indian/Alaska Native women and a slight decline for white women. The latter, probably a single year anomaly or result of sampling variation, can hardly be attributed to program. But only with an understanding of the ways absolute differences tend to change as the prevalence of an outcome changes – including understanding that when rates are well under 50% for both groups general improvements will tend to increase absolute differences – will one be in a position to question whether a program aimed at generally increasing a favorable outcome can be expected to reduce or increase absolute differences between rates (or, more important, whether observed patterns are anything other than the consequence of a general change in the prevalence of an outcome).¹⁸

Most important, none of the documents informs the reader of a need to understand how to measure disparities in order to determine whether a particular intervention increases or decreases a disparity. For example, as suggested above and discussed further below, increases in screening and vaccination rates will tend to reduce relative differences in rates of receipt of these procedures rates and increase relative differences in rates of failing to receiving the procedures, and, depending on whether the particular type of screening or vaccination is uncommon or common, will tend either to increase or decrease the absolute difference between rates. But the four documents provide no basis for determining in such a case whether the disparity should be deemed to be increasing or decreasing. The same holds for, among numerous other common situations, the situation where improvements in cancer care reduce relative differences in rates of surviving cancer and increase relative differences in rates of failing to survive cancer, and, depending on the type of cancer, either increase or decrease absolute differences between rates.¹⁹ Even more important, the materials provide no guidance for considering the extent to which

¹⁸ See Table 4 (at 18) of my FCSM Paper regarding a situation where substantial increases in cervical cancer screening rates in the United Kingdom were accompanied by (a) substantial decrease in relative differences between screening rates of most and least deprived groups; (b) substantial increases in relative differences between in rates of not receiving screening for those groups; and (c) substantial decreases in absolute differences between rates; (d) substantial increases in the ratio of the screening odds of the least deprived group to the screening odds of the most deprived group. The final column shows that, to the extent the disparity can be measured, it increased to a small degree. Thus, it would have been a mistake to read the substantial decrease in the absolute difference as reflecting an improvement in the comparative situation of the disadvantaged group with regard to screening.

¹⁹ The Draft Report (at 61) lists cancer survival among the subjects to as to which disparities are to be measured. Presumably, some observers will measure cancer outcome disparities in terms of relative differences in cancer survival rates and others will measure them in terms of relative differences in cancer mortality rates (though sometimes describing their studies, especially in article titles, as examining differences in survival). There is little reason to expect that persons doing so will be aware that it is possible that (much less that typically) whether the disparity is increasing or decreasing will turn on whether one in fact examines relative differences in survival or relative differences in mortality. See Section A my [Comments for the Commission on Evidence-Based Policymaking](#) (Nov. 28, 2016). To my knowledge, no study of demographic differences in cancer outcomes has yet indicated an awareness of the possibility that patterns of changes, of the comparative size of, relative differences in mortality will be the opposite of the patterns for relative differences in survival.

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 12

observed patterns of changes are anything other than a consequence of general changes in the prevalence of the outcome and hence whether resources specifically devoted to addressing disparities had some role in observed patterns.

To my knowledge, no work of members of the Disparities Standing Committee (other than that reflected in the *Commissioned Paper* or the *Cultural Competency Technical Report*) has reflected awareness that it is possible for different measures to yield different conclusions about changes in the directions of disparities. This is not unusual. Apart from a passing mention in the 2005 National Healthcare Disparities Report, until recently, no arm of HHS other than NCHS has recognized it is possible for the relative difference the observer happened to be examining and the absolute difference to change in opposite directions as the prevalence of an outcome changes.²⁰ All arms of HHS other than NCHS may still be unaware that relative differences in a favorable health or healthcare outcome and relative differences in the corresponding adverse health or healthcare outcome can (and in fact tend to) change in opposite directions as the prevalence of an outcome changes.

Despite the substantial amount of health and healthcare disparities research conducted by arms of Harvard University, apart from the *Commissioned Paper*, nothing produced by those arms has recognized that different measures may yield different conclusions about directions of changes in disparities. Most institutions conducting health and healthcare disparities research or providing guidance on the measurement of such disparities, however, have never show any such recognition. And observers commonly discuss changes in disparities using their preferred measure while seemingly unaware, or in any event not mentioning, that a different measure would yield an opposite conclusion. That occurs even when the measure that yields an opposite conclusion is one more commonly employed in the circumstances. And never do they explore the crucial question of the extent to which an observed pattern reflects something other than a change in the prevalence of an outcome.

As discussed in "Race and Mortality Revisited" (at 344), Harvard University is in a better position than other institutions to understand the problematic nature of research that does not consider the ways measures employed tend to be affected the prevalence of an outcome. Hence, its production of research that ignores these issues is less excusable than in the case of other institutions. The same may be said of NQF, which, rather than build on the *Commissioned Paper* by addressing things it failed to address, ignores the paper's recognition that different measures in fact sometimes yield opposite conclusions about directions of changes in disparities.

²⁰ See my commentary "[The Mismeasure of Health Disparities](#)," *Journal of Public Health Management and Practice* (July/Aug. 2016), regarding recent recognitions by CDC personnel not part of NCHS that a relative difference and the absolute differences can change in opposite directions as the prevalence of an outcome changes. Presumably, "The Mismeasure of Health Disparities" alerted such persons also to the fact that the two relative differences tend to change in opposite directions as the prevalence of an outcome changes. The CDC *Health Inequalities & Disparities Reports* of 2011 and 2013, however, by presenting both relative and absolute differences for some subjects studied, have shown instances whether the two measures changed in opposite directions.

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 13

In any case, the fact that the failure of understanding reflected in the NQF's recent reports is commonplace, or even almost universal, is not a justification for supporting such work with funds of the federal government or any other entity or justification for publishing (or the continued publication of) such reports. Those documents will further promote unsound research and lead to the misallocation of resources aimed at reducing disparities. In the latter regard, the emphasis on incentive programs to reduce disparities, while ignoring measurement issues, may well promote anomalies such as observed in Massachusetts Medicaid pay-for-performance program whereby the disparities element in the program itself tends to increase healthcare disparities. I therefore urge NQF to withdraw the three recent Final Reports and to take no further action on the Draft Report without first considering the implications of the points raised in this letter and its references. Such consideration should be undertaken in consultation with the CMS officials overseeing the funding on this project. As a recipient of CMS funds, NQF's primary responsibility is to ensure that agency benefits from the expenditure of those funds.

See my August 24, 2017 [letter](#) to the American Institutes for Research (AIR) regarding that organization's responsibilities to alert the Department of Education and other agencies of the ways federally-funded AIR research has been undermined by a failure to recognize the ways the measures employed tend to be affected by the prevalence of an outcome. NQF bears a similar responsibility toward the entities that fund its activities to which issues addressed in this letter pertain.

And under no circumstances should NQF finalize the report until it is modified to both (a) reflect an understanding that different measures commonly yield opposite conclusions about whether health and healthcare disparities are increasing or decreasing and (b) specifically address the need for health and healthcare disparities research to consider the ways the measures employed tend to be affected by the prevalence of an outcome. For failure to address such issues will lead those relying on the document to assume no such issues exist.

I also urge NQF to withdraw the aforementioned *Commissioned Paper* and *Consensus Standards Technical Report*. Despite some recognition in each document that different measures can yield different conclusions about changes in health and healthcare disparities, both documents continue to provide fundamentally flawed guidance on the measurement of disparities.

In the two sections below, I describe my prior communications with the NQF on this subject and provide some the illustrations of the points made above. These illustrations include some striking examples of confused and misguided discussions of health and healthcare disparities issues resulting from the failure to understand the ways measures of differences between outcome rates tend to be affected by the prevalence of an outcome.

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 14

Prior Communications with National Quality Forum Regarding Health and Healthcare Disparities Measurement

My prior formal communications to NQF on this subject may be found in the aforementioned October 22, 2009 [letter](#) to NQF President and CEO Janet M. Corrigan (urging NQF to examine the failure of its guidance on health and healthcare disparities measurement to consider the ways the measures employed tend to be affected by the prevalence of an outcome) and an October 26, 2012 [letter](#) to NQF Forum Interim President and CEO Laura Miller and leadership of other entities responsible for the *Commissioned Paper* (urging withdrawal of the *Commissioned Paper* as a result of the document's failure to consider the ways the measures it discussed tend to be affected by the prevalence of an outcome).²¹

The 2009 letter to President Corrigan was focused on the failure of the March 2008 NQF document [National Voluntary Consensus Standards for Ambulatory Care—Measuring Healthcare Disparities](#) to recognize the ways measures of differences in outcome rates tend to be affected by the prevalence of an outcome. The letter briefly explained the patterns by which measures tend to be affected by the prevalence of an outcome and provided references explaining those patterns more fully. The letter gave particular attention to a study by Morita et al.,²² which examined the effects of school-entry Hepatitis-B vaccination requirement had dramatically increased vaccination rates. As commonly happens in the circumstances, relative differences in receipt of vaccination decreased while relative differences in nonreceipt of vaccination increased. The authors, who measured racial/ethnic disparities in terms of relative differences in vaccination rates, found that the requirement substantially reduced disparities. But NCHS, which in 2004-2005 had recognized that relative differences in receipt of care and nonreceipt of care would tend to change in opposite directions as care rates increased, had determined that, for purposes of Healthy People 2010, all healthcare disparities should be measured in terms of relative differences in nonreceipt of care. Thus, NCHS would have found that the requirement substantially increased disparities. I will give further attention to the Morita study below in connection with the recent NCHS reversal of position that effectively repudiated a decade of National Healthcare Disparities Reports and other research that relied on the NCHS earlier recommendation.

Referencing a NQF potential interest in the effects of pay-for-performance programs on health and healthcare disparities, the letter also directed President Corrigan's attention to a [webpage](#)²³

²¹ Informal communications to NQF personnel or the Disparities Standing Committee include recent emails in connection with my comments on the July 21, 2017 Draft Report and an email inviting NQF personnel involved with the *Cultural Competency Technical Report* to attend my September 25, 2012 American University Department of Mathematics and Statistics Colloquium titled "[The Mismeasure of Group Differences in the Law and the Social and Medical Sciences.](#)"

²² 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-e552.

²³ This is the Pay for Performance subpage of Measuring Health Disparities page of [jpscanlan.com](#).

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 15

discussing implications of the failure to understand the ways measures tend to be affected by the prevalence of an outcome with regard to appraising effects of general reductions in mortality on relative differences in mortality rates and the effects of incentive programs on absolute differences between rates of receiving appropriate care.

With regard to the former matter, the webpage contained my [comment](#) on a study by Pickett *et al.*²⁴ that found that the “back to sleep” program, which generally reduced SIDS deaths, led to increased relative racial and SES difference in SIDS deaths. The comment explained that such a pattern is to be expected with any program that generally reduces an adverse outcome.²⁵ The Pickett study was discussed in the NQF January 15, 2017 Final Report (at 24-25), but without recognition that any program that generally reduces an adverse outcome will tend to increase relative differences in rates of experiencing it. In consequence of that failure of understanding, the report speculated as to reasons why racial differences in SIDS rates increased after the program was implemented. But such speculations can rarely be of value, and can be extremely misleading, when not undertaken with recognition of the way measures employed tend to be affected by the prevalence of an outcome. See the section of "Race and Mortality Revisited" (at 339-341) titled “Illogical Expectations and Unfounded Inferences.”

With regard to pay-for-performance and healthcare disparities, the page contained several comments pertaining to the finding in Werner *et al.*²⁶ that an incentive program that generally increased coronary artery bypass grafting (CABG) rates was associated with an increase in the absolute differences between black and white CABG rates. The white rate had increased from 3.6% to 8.0% while the black rate increased from 0.9% to 3.0%. These figures are set out in Table 3 with measures of difference.

Table 3. White and Black CABG Rates Before and After Implementation of a CABG Report Card, with Measures of Differences (from Werner *et al.* *Circulation* 2005)

Period	White Rate	Black Rate	W/B Receipt Ratio	B/W non-Receipt Ratio	Abs Difference (perc. pnts)	W/B Receipt Odd Ratio
1	3.6%	0.9%	4.00	1.03	2.7	4.11
2	8.0%	3.0%	2.67	1.05	5.0	2.81

²⁴ Pickett KE, Luo Y, Lauderdale DB. Widening social inequalities in risk for sudden infant death syndrome. *Am J Public Health* 2005;95:97-81.

²⁵ I also discuss the Pickett study at page 10 of my [“The Misinterpretation of Health Inequalities in the United Kingdom.”](#) British Society for Population Studies 2006 Conference, Southampton, UK (Sept. 19-20, 2006).

²⁶ Werner, RM, Asch DA, Polsky D. Racial profiling: The unintended consequences of coronary artery bypass graft report cards. *Circulation* 2005;111:1257–63.

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 16

As reflected in several of the references in the Pay for Performance page, NQF Disparities Standing Committee Chair Marshall Chin and colleagues regarded the increase in absolute differences to indicate that incentive programs tended to increase healthcare disparities and therefore argued for such programs to consider effects on healthcare disparities.²⁷ Disparities Standing Committee member Jose Escarce, who would presumably have relied on relative differences in receipt of the procedure to measure disparities, which is the approach he employed in the 2004 article mentioned above and which was probably the most common approach at the time, would have found that the program decreased disparities.

The NCHS was at the time in the process of deciding to measure healthcare disparities in terms relative difference in non-receipt and thus would, under that approach, have found an increase in racial disparities. Persons who examined the matter in terms of odds ratio, would have found a decrease in disparities.

President Corrigan responded by letter of October 29, 2009, advising that NQF took the issues I raised seriously and that she would be working with the NQF Performance Measurement team to address them. President Corrigan also advised that she would provide a complete response to my letter as soon as possible.

I sent a similar [letter](#) to the President and CEO of the Robert Wood Johnson Foundation (RWJF) Risa Lavizzo-Mourey on April 9, 2010, also emphasizing the implications of the measurement issues I raised with regard to pay-for-performance issues, while discussing the recent New England Journal of Medicine commentary co-authored by (current) NQF Board Chair Bruce Siegel.²⁸ Referencing my [comment](#) on the article, I suggested that it would be a serious mistake to implement programs providing monetary incentives for addressing health or healthcare disparities until there exist more satisfactory measures of healthcare disparities than are currently being employed. I received a response from RWJF President Lavizzo-Mourey similar to the one I had received from President Corrigan, indicating interest in the subject.

²⁷ The current page is slightly different from the one that existed at the time of the letter to President Corrigan. The page at the time had 14 references including two (items 5 and 14) that were comments on articles co-authored by Professor Chin. The current version includes materials subsequently created, including one comment on an article by Professor Chin (item 15) and one comment on an article by (current) NQF Board Chair Bruce Siegel (item 16). The comments all criticize the discussion of effects of incentive programs on disparities and the promotion of incentive programs to address disparities without recognition of the ways measures tend to be affected by the prevalence of an outcome. In contrast to Professor Chin who had read an increase in the absolute difference during times of increases in an uncommon outcome as evidence the improvements in healthcare tend to increase disparities, Dr. Siegel had read a decrease in the absolute difference during times of increase in a common outcome as evidence that improvements in healthcare tend to decrease disparities.

²⁸ Siegel B, Nolan L. Leveling the field – ensuring equity through National Health Care Reform. *N Engl J Med* 2009;361:2401-2403.

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 17

I did not hear further from NQF President Corrigan or RWJF President Lavizzo-Mourey. But, apparently, at some point in the months following the exchange with President Corrigan, NQF, using funds from a RWJF grant, contracted with the Disparities Solutions Center of Massachusetts General Hospital to produce the *Commissioned Paper*, a draft of which was made available for comment in July 2011.

As discussed, the document (including in its draft form) specifically recognized that a relative difference and the absolute difference could yield opposite conclusions about the directions of changes in disparities. The report specifically discussed the Werner study mentioned above, noting that while the absolute difference between rates increased, the relative difference in the receipt of the procedure decreased. And, though it did so somewhat obscurely, the document recognized the possibility that the relative difference in a favorable outcome and the relative difference in the corresponding adverse outcome could change in opposite directions.

But even though the lead author of the *Commissioned Paper* had co-authored the principal NCHS document recognizing that relative differences in favorable outcomes and relative differences in the corresponding adverse outcomes tend to change in opposite directions as the prevalence of an outcome changes,²⁹ the *Commissioned Paper* contained no such recognition. Nor did it contain any recognition of that measures tend to change solely because the prevalence of an outcome changes and that appraisals of the effects of policies on disparities must attempt to determine the extent to which an observed change in a measure is simply a function of a change in the prevalence of an outcome.

My comments on the draft *Commissioned Paper*, the response to those comments, and subsequent actions of NQF reflected in the *Consensus Standards* have already been adequately discussed.

My first effort to have the *Commissioned Paper* withdrawn after it was finalized may be found in an October 9, 2012 [letter](#) to Harvard University (at 43-44). The letter, written in conjunction with a methods works at the University's Center for Quantitative Social Science,³⁰ was principally a criticism of health and healthcare disparities research at Harvard Medical School and Harvard School of Public Health for failure to recognize the ways the measures employed in such research tend to be affected by the prevalence of an outcome and, save for the *Commissioned Paper*, a failure to recognize even that different measures in fact often yield opposite conclusions about the directions of changes in disparities.

²⁹ See Keppel Kenneth G., Pamuk Elsie, Lunch John, *et al.* [Methodological Issues in Measuring Health Disparities](#), *Vital Health Stat* 2005;2 (141).

³⁰ See "[The Mismeasure of Group Differences in the Law and the Social and Medical Sciences](#)," Applied Statistics Workshop at the Institute for Quantitative Social Science at Harvard University (Oct. 17, 2012).

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 18

After then addressing withdrawal of the document with the authors themselves (who declined to do so), I sought withdrawal of the *Commissioned Paper* in the aforementioned letter to NQF President Miller and other officials of organizations involved with the *Commissioned Paper*. That letter can speak for itself and some of its points are addressed above.

I received no response from NQF or RWJF. But the research integrity officers of Harvard Medical School and Massachusetts General Hospital did respond by [letter](#) of December 12, 2012. In the letter, the officers indicated that apart the institutions did not independently assess the merits of work of faculty members apart from ensuring that there is no research misconduct. They also stated that they regarded the issues I raised to involve a difference of scientific opinion not rising to the level of research misconduct, and that therefore the institutions would take no further action on the matter beyond informing the authors of the issues I raised. The letter did not indicate whether it had been copied to entities funding the *Commissioned Paper*.

As discussed in "Race and Mortality Revisited" (at 344), I do not believe that the failure of a measurement document even to address that measures tend to be affected by the prevalence of an outcome to involve a simple difference of scientific opinion. Having accepted funding from NQF to produce the document the failure, the authors' failure to address issues even after I brought them to authors' attention was inexcusable. That same would hold for the NQF's failure to address the issues having accepted federal funds to provide guidance on the role of measurement in reducing health and healthcare disparities.

In any case, with regard to the recent documents, NQF is in a position to elicit from the Disparities Standing Committee and NQF staff involved in the recent work whether they are aware that measures tend to be affected by the prevalence (or even that in fact different measures commonly yield opposite conclusions as to the directions of changes in disparities over time), as well as whether the guidance in recent document can be of value without consideration of the effects of the prevalence of an outcome on measures of disparities (and without any indication of recognition that different measures in fact commonly yield opposite conclusions about directions of changes in disparities).

NQF is also in a position to address with CMS whether publication (or continued publication) of the recently produced NQF document will further the agency's interest in promoting sound science regarding the effects of policies on health and healthcare disparities.

Illustrations of some of the confusion in analyses health and healthcare issues

Apart from a seeming unawareness of the possibility that different measures can yield opposite conclusion as to direction of changes in disparities, the recent NQF documents on health and healthcare disparities would appear be premised on the belief that persons quantifying health and healthcare disparities must be doing so in a sound manner and that there is some consistency in the approaches employed by such persons. Any such belief is manifestly incorrect.

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 19

Appreciation of the confusion in this area requires an understanding of the actions of the NCHS since 2004. Beginning in 2004 NCHS recognized that health and healthcare improved relative differences in (increasing) favorable outcomes tended to decrease while relative differences in the corresponding (decreasing) adverse health and healthcare outcomes tended to increase. At the time, health disparities usually were measured in terms of relative differences in adverse outcomes, which caused improvements in health to usually be associated with increasing disparities. On the other hand, healthcare disparities usually were measured in terms of relative differences in favorable outcomes (as in the case of the study co-authored by Professor Escarce), which caused improvements in health care usually to be associated with decreasing disparities. See my "[Race and Mortality](#)," *Society* (Jan./Feb. 2000), which was the principal basis for NCHS statisticians' recognition of the pattern by which the two relative differences tend to change in opposite directions as the prevalence of an outcome changes, and which discusses a prior recognition by the NCHS director.

Given that the forces causing favorable adverse outcome rates of advantaged and disadvantaged groups to differ are the same forces that cause the corresponding adverse outcomes to differ, NCHS should have recognized that the pattern called into question the utility of either relative difference for quantifying the strength of those forces or serving as a guide to evaluating factors that affect those forces. Rather, the agency simply determined that, while continuing to measure health disparities in terms of relative differences in adverse outcomes, it would now also measure healthcare disparities in terms of relative differences in adverse outcomes (nonreceipt of appropriate care).

As a result of that decision, improvements in health continued to be associated with increasing health disparities. But now improvements in healthcare would also tend to be associated with increasing healthcare disparities.

The belief of NCHS that it could arbitrarily choose a measure that would tend to say that disparities were increasing over one that would tend to say that disparities were decreasing reflected a fundamental misunderstanding of why society devotes resources to the study health and healthcare disparities – that is, to understand underlying processes and to inform policies that address the forces causing the outcome rates of advantaged and disadvantaged groups to differ. That should be especially evident when, as with the recent NCHS work, the focus is on incentivizing policies that will tend to reduce disparities.

In 2015-16, NCHS reversed its policy with regard to healthcare disparities and such disparities would now again be measured in terms of relative differences in favorable healthcare outcomes. So once again improvements in care would tend to be associated with decreased disparities. But the agency continued to fail to show recognition of the problematic nature of either relative difference for quantifying either health or healthcare disparities. Throughout this process, the agency has continued to fail to recognize the ways absolute differences between rates tend to be affected by the prevalence of an outcome or that it is impossible to understand the effects of policies on the forces causing outcome rates of advantaged and disadvantaged groups to differ

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 20

without consideration of the effects of the prevalence of an outcome on the measures employed in analyses of disparities issues.

The history of NCHS actions on this subject are summarized in the FCSM Paper and "Race and Mortality Revisited" (with regard the initial actions of NCHS) and in "The Mismeasure of Health Disparities" (with regard to the recent NCHS reversal of position as to healthcare disparities). Readers should bear in mind that history – as well as the fact that few people analyzing health and healthcare disparities are aware of it and that the great majority of health and healthcare disparities research shows no awareness whatever that the choice of measures can affect determinations of directions of changes in disparities – as they consider the discussion that follows.

Tables 4 through 6 appear in one or more of the three items just mentioned and I present them with the columns used in the FCSM paper even though not all of the columns important to the discussion. For simplicity the rate ratio columns refer to favorable and adverse outcomes for receipt and nonreceipt of the type of care at issue. The final column, EES (for estimated effects size) shows a measure that is theoretically unaffected by the prevalence of an outcome. It involves deriving from a pair of outcome rates the difference between the means, in terms of percentage of a standard deviation, of the hypothesized normal distributions of each group's risk of experiencing an outcome and its opposite. Further discussion of the measure, including its strengths and weaknesses, may be found in the three items just mentioned.

Table 4 is based on the data on which the NCHS relied in explaining its recognition that determination of whether health and healthcare disparities were increasing or decreasing would commonly turn on whether one relied on the relative difference in the favorable outcome or the relative difference in the corresponding adverse outcome.³¹ The table shows that during a period of general increases in mammography, the relative difference between white and Hispanic rates of experiencing the increasing outcomes (receipt of mammography) decreased, while the relative difference between white and Hispanic rates of experiencing the decreasing outcome (nonreceipt of mammography) increased.

Table4. Changes in Mammography Rates of Whites and Hispanics between 1990 and 2002, from Keppel et al. 2005, with Disparity Measures

Year	Wh Mam Rt	Hi Mam Rt	Wh/Hisp Fav Ratio	Hisp/Wh Adv Ratio	Abs Diff (perc pts)	Wh/Hisp Fav Odd Ratio	EES
1990	52.70%	45.20%	1.17	1.16	7.5	1.35	0.19
1998	68.00%	60.20%	1.13	1.24	7.8	1.40	0.21

According to the approach the NCHS adopted at the time, NCHS would have determined that the disparity increased. Presumably, if the efficacy of a program aimed at addressing the white-

³¹ See Keppel K., Pamuk E., Lynch J., *et al.* 2005. [Methodological issues in measuring health disparities](#). *Vital Health Stat* 2005;2 (141).

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 21

Hispanic mammography disparity were at issue, in this situation the programs would have been deemed to increase the disparity. According to the position the NCHS later adopted in 2015-16, the program would have been deemed to reduce the disparity.

Table 5 is based on a study by Harper et al.³² that reported in its abstract a very large increase in the relative difference between mammography rates of the highest and lowest income groups during a period of substantial increase in mammography rates.

Table 5. Changes in Mammography Rates of Highest and Lowest Socioeconomic Groups between 1987 and 2004, from Harper et al. 2009, with Disparity Measures

Year	Highest SES Mam Rt	Lowest SES Mam Rt	Hi/Low Fav Ratio	Low/High Adv Ratio	Abs Diff (perc pts)	Hi/Low Fav Odds RaOio	EES
1987	36.30%	17.20%	2.11	1.30	19	2.74	0.60
2004	77.40%	55.20%	1.40	1.98	22	2.78	0.62

As discussed in the FCSM Paper (at 16-17), the text of the study went on to clarify that, in reliance on NCHS guidance, the study was relying on the relative difference in adverse outcomes to measure healthcare disparities, and that the increase in relative difference in mammography rates reported in the abstract was actually an increase in the relative difference in nonreceipt of mammography. Few readers of the Harper study, however, would understand that the reason NCHS adopted that approach was recognition that relative differences in receipt and nonreceipt of appropriate healthcare commonly change in opposite directions or grasp that the relative difference in receipt of mammography had actually decreased substantially.

As discussed, NCHS would now regard the disparity to have decreased substantially rather than increased substantially. The EES indicates that, to the extent we can effectively measure the disparity, it increased slightly.

Table 6 is based on the Morita study that I had mentioned in the 2009 letter to NQF President Corrigan. As discussed, the implementation of a school-entry vaccination requirement that dramatically increased vaccination rates resulted in a substantial decrease in the relative racial difference in receipt of vaccination but a substantial increase in the relative racial difference in the failure to receive vaccination. It also shows the common pattern where the increase in vaccination rates led to an increase in the absolute difference between black and white rates where rates were quite low (grade 5) and a decrease in the absolute difference where rates were fairly high (grade 9). The EES suggests that the disparity declined substantially.

³² Harper S, Lynch J, Meersman SC, et al. Trends in area-socioeconomic disparities in breast cancer screening, mortality, and survival among women ages 50 years and over (1987-2005). *Cancer Epidemiol Biomarkers Prev* 2009;18(1):121-131.

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 22

Table 6. Hepatitis B Vaccination Rates for Whites and Blacks In Grades 5 and 9 Before and After Implementation of School-Entry Vaccination Requirement, from Morita et al, 2008, with Disparity Measures

Grd	Year	Program	White VacRt	Black VacRt	Wh/Bl Fav Ratio	Bl/Wh Adv Ratio	Abs Diff (perc pts)	Wh/Bl Fav Odds Ratio	EES
5	1996	Pre	8%	3%	2.67	1.05	5	2.81	0.47
5	1997	Post	46%	33%	1.39	1.24	13	1.73	0.34
9	1996	Pre	46%	32%	1.44	1.26	14	1.81	0.37
9	1997	Post	89%	84%	1.06	1.45	5	1.54	0.24

The authors gave no indication of an awareness of NCHS guidance to measure healthcare disparities in terms of relative differences in nonreceipt of care (or awareness of the possibility that the relative difference in nonreceipt of vaccination could or in fact did change in the opposite direction of the relative difference in receipt of vaccination). They simply relied on the relative difference in receipt of vaccination to measure racial disparities and concluded that disparities had declined substantially. NCHS would instead have found a substantial increase in disparities. Now, however, the NCHS would agree with the findings of the Morita authors who ignored NCHS's earlier guidance, just as it would disagree with the findings of the Harper authors who followed that guidance.

Another illustration of the anomalies arising from following and not following NCHS guidance is discussed in my March 8, 2016 [letter](#) to the Stanford Center on Poverty and Inequality. The letter (at 4-6) discusses a situation in a recent Center on Poverty and Inequality report where authors analyzed the size of insurance disparities across states, while, in reliance on NCHS guidance, measuring disparities in terms of relative difference in uninsurance rates. They then drew inferences about processes on the basis of the comparative size of the disparities so measured. As should be implicit in the discussion above, and as discussed in "Race and Mortality Revisited" (at 339-341), one commonly draws opposite, or at least very different, inferences, about processes based on the comparative size of disparities depending on whether one examines relative differences in favorable outcomes or relative differences in the corresponding adverse outcome. In this instance, the authors relied on the NCHS guidance to examine insurance disparities in terms of relative differences in the adverse outcome while apparently unaware that NCHS had already reversed the guidance.

I mentioned above that "Race and Mortality Revisited" (at 333) had discussed as a particularly egregious waste of resources a study by the Institute for Medicine and Public Health of the Vanderbilt University Medical Center that sought to evaluate the effectiveness of quality improvement in reducing healthcare disparities, while showing no recognitions that different measures could (or systematically tend to) yield opposite conclusions about whether quality improvements increase or decrease disparities. The study was funded by the Agency for

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 23

Healthcare Research and Quality (AHRQ) and the agency bears at least as much responsibility for the waste of resources as the researchers conducting the study.

AHRQ has several times indicated that, in accordance with approach to Healthy People 2010 adopted by NCHS in 2005, it measures healthcare disparities in terms of relative differences in adverse outcomes. But AHRQ has never indicated an awareness that NCHS recognized that the relative differences in receipt of appropriate care and nonreceipt of appropriate care tend to change in opposite directions as appropriate care rates generally increase. And AHRQ has shown little awareness that determinations of directions of changes in disparities over time often turn on the measure chosen. See my letter to July 1, 2015 [letter](#) discussing the National Healthcare Disparities Reports generally and explaining that AHRQ confusion over how it was intending to measure disparities led to the situation where the 2010 report highlighted as some of the largest reductions in disparities between two points in time situations where the agency would also regard the disparities be much larger at the end of the period than at the beginning of the period.

The Center for Medicare and Medicaid Services (CMS) (the agency funding the NQF project that is the subject of this letter) has lately given substantial attention to the measurement of health and healthcare disparities, though not to my knowledge showing an awareness that different measure could, or would tend to, yield opposite conclusions about the directions of changes in disparities. The agency's [2015 National Impact Assessment of the Centers for Medicare and Medicaid Services \(CMS\)](#) discusses the approach of AHRQ in the National Healthcare Disparities Reports and notes (at 168) that AHRQ used a 10 percentage point difference between the reference group and the study group to identify a disparity, while CMS has chosen to use a 5 percentage point difference. In other words, the CMS intended to adopt an approach that was more likely to indentify a disparity than AHRQ. In fact, however, AHRQ used a 10 percent difference (in either the favorable outcome or the adverse outcome if either would yield such a difference), not a 10 percentage point difference. And many of the disparities that the National Healthcare Disparities Reports have identified as among the largest, which often involve differences of several hundred percent, would not reach a 5 percentage point threshold.

The confusion about disparities measurement is nicely illustrated in the three articles and a commentary in an August 18, 2005 issue of the *New England Journal of Medicine*. A study by Vaccarino *et al.*³³ relied on relative differences in favorable healthcare outcomes (though relative differences in adverse outcomes for health status issues) with regard to outcome rates that were not changing much in overall prevalence during the period examined; and, as commonly happens when overall prevalence does not change much, the study found little to remark on with respect to changes in disparities over time.

³³ Vaccarino V, Rathore SS, Wenger NK, et al. Sex and racial differences in the management of acute myocardial infarction, 1994 through 2002. *N Engl J Med* 2005;353:671-682.

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 24

A study by Jha *et al.*³⁴ relied on absolute differences between rate in examining racial disparities in rates of receiving certain uncommon procedures among Medicare patients that were generally increasing in overall prevalence; and, as commonly happens when outcome rates in the ranges at issue are generally increasing, the authors usually found increasing disparities. Had the authors employed the approach Professor Escarce and colleague had employed a year earlier for outcomes of similar prevalence among a like population for a somewhat overlapping time frame, the authors would have generally found the disparities to be decreasing.

A study by Trivedi *et al.* relied on absolute differences between rates in examining adequacy of care (which included both treatment and control of conditions) where adequacy of care rates (especially as to treatment) were at generally high levels and increasing; and, as commonly happens in such circumstances, the authors found absolute difference between rates usually to be decreasing (especially as to treatment).³⁵ A commentary³⁶ discussed the various findings and their perceived implications and stressed the need for more health disparities research and action to reduce such disparities. As was common in 2005, as it is now, neither the commentary nor any of the articles mentioned anything about ways different measures might yield different conclusions as to directions of changes in the disparities or the way any measure might be affected by general changes in the outcome being examined.

The Jha and Escarce studies also fit into an extreme illustration of the confusion in this area. Both studies, in the main,³⁷ found what typically occurs in the circumstances of an increase in an uncommon outcome: (a) decrease in the relative differences in rates of receipt of procedure; (b) an increase in the relative difference in nonreceipt of the procedure; and (c) an increase in the absolute difference (with only (c) being contingent on the fact that the outcomes are uncommon). But, because the authors measured disparities differently, they reported opposite conclusions.

More recently, a study by Le Cook *et al.*,³⁸ unaware the Jha and Escarce studies had in fact shown very similar patterns of changes in measures, discussed the contrasting conclusions

³⁴ Jha AK, Fisher ES, Li Z, Orav EJ, Epstein AM. Racial trends in the use of major procedures among the elderly. *N Engl J Med* 2005;353:683-691.

³⁵ Trivedi AN, Zaslavsky AM, Schneider EC, Ayanian JZ. Trends in the quality of care and racial disparities in Medicare managed care. *N Engl J Med* 2005;353:692-700. See my [Comment on Trivedi JAMA 2006](#) regarding the authors' later effort to explain different patterns as to treatment and control, which made very reasonable points, but without consideration of the generally lower rates of control compared with treatment.

³⁶ Lurie N. Health disparities – Less talk. more action. *N Engl J Med* 2005;353:727-729.

³⁷ There were some departures from these patterns. There might be things to be learned from these departures. But learning such things is only possible when one understands the patterns described here.

³⁸ Lê Cook B, McGuire TG, Zuvekas SH. Measuring trends in racial/ ethnic health care disparities. *Med Care Res Rev.* 2009 Feb; 66(1):23-48.

Attachment A to March 15, 2019 letter to National Quality Forum

Shantanu Agrawal, MD, MPhil
President and CEO
National Quality Forum
August 29, 2017
Page 25

without any consideration of the fact that two studies had relied on different measures. It then opined about the possible reasons for what were deemed to be different results, while suggesting that a study over a longer time frame might be revealing. See the [Spurious Contradictions](#) subpage of [Measuring Health Disparities](#) page of [jpscanlan.com](#).

In sum, confusion over how to measure health and healthcare disparities has resulted in the waste of many billions of research dollars as well as the implementation of policies based on conclusions that have no sound statistical basis. NQF's current project could make a substantial contribution of correcting this situation. But what the organization has so far produced as part of this project, by obscuring rather than revealing the serious measurement issues that must be addressed in all health and healthcare disparities research, will only exacerbate the existing situation.

Sincerely,

/s/James P. Scanlan

James P. Scanlan

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

October 22, 2009

Janet M. Corrigan, PhD, MBA
President and Chief Executive Officer
National Quality Forum
601 Thirteenth Street, NW
Suite 500 North
Washington, DC 20005

Re: Measuring Healthcare Disparities

Dear Dr. Corrigan:

I recently reviewed the National Quality Forum's 2008 document styled "National Voluntary Consensus Standards for Ambulatory Care—Measuring Healthcare Disparities." The document crucially fails to appreciate a serious problem in health and healthcare disparities research arising from the fact that researchers rely on various standard measures of differences between outcome rates to appraise the size of disparities without recognizing that each such measure tends to be systematically affected by the overall prevalence of an outcome.

Most notably, the more common an outcome the smaller tends to be the relative difference in experiencing it and the larger tends to be the relative difference in failing to experience it. Thus, as procedures like mammography or immunization increase in overall prevalence, relative differences in mammography and immunization rates tend to decline while relative differences in rates of failing to receive mammography or immunization tend to increase. Absolute differences and odds ratios tend to also to change as the overall prevalence of an outcome changes, though in a more complicated way. Roughly, as uncommon outcomes become more common, absolute differences between rates tend to increase; as common outcomes become even more common, absolute differences between rates tend to decline. Differences measured by odds ratios tend to change in the opposite direction of absolute differences between rates. Thus, none of these measures can alone indicate whether health or healthcare disparities are changing in a meaningful sense.

Janet M. Corrigan, PhD, MBA

October 22, 2009

Page 2

More than a hundred references explaining these patterns may be found on the [Measuring Health Disparities](#)¹ page of [jpscanlan.com](#) (MHD) and the nuances of the patterns may be found on the [Scanlan's Rule](#) page of the same site. Several key references are listed in note 2 below.² The [Solutions](#) sub-page of MHD describes a method for measuring health and healthcare disparities that is not affected by the overall prevalence of an outcome and the [Solutions Database](#) sub-page provides a downloadable database with which to implement the approach. While the approach is imperfect in a number of respects, it remains far superior to the near universal practice of relying on one or another standard measure of differences between outcome rates without regard to the way the measure tends to be affected by the overall prevalence of the outcome.

The item listed as (e) in note 2, which comments on an award-winning 2008 study by Morita *et al.*, is particularly illustrative of the disarray in the area. Morita and colleagues examined the effect of a school-entry Hepatitis B vaccination requirement on racial and ethnic disparities in vaccination rates. The requirement substantially increased overall vaccination rates. Relying on relative differences in vaccination rates, the study found that racial and ethnic disparities in vaccination rates decreased dramatically. The comment explains that the National Center for Health Statistics (NCHS), which insists on relying on relative differences in adverse outcomes (here, the failure to be vaccinated), would have found dramatic increases in disparities. It also explains why researchers or entities that employ other approaches would have reached still different conclusions as to directions of changes in disparities over various time frames from those that Morita and colleagues reached or that NCHS would have reached.

As discussed in references (b) and (e) of note 2, the NCHS position mentioned above is a misguided response to reference (c). But lately, in the United States,³ and more so in

¹ The underlining of various references in the text of this document reflects the fact, in order to facilitate review of those references, links to the references are provided in an electronic copy of this letter posted on the Letters sub-page of the Measuring Health Disparities page of [jpscanlan.com](#).

² (a) Scanlan JP. [Can we actually measure health disparities?](#) *Chance* 2006;19(2):47-51 ; (b) Scanlan JP. [Measuring health disparities.](#) *J Public Health Manag Pract* 2006;12(3):293-296 (responding to Keppel KG, Percy JN. Measuring relative disparities in terms of adverse events. *J Public Health Manag Pract* 2005;11(6):479-483) ; (c) Scanlan JP. [Race and mortality.](#) *Society* 2000;37(2):19-35; (d) Scanlan JP. Measurement Problems in the National Healthcare Disparities Report, presented at American Public Health Association 135th Annual Meeting & Exposition, Washington, DC, Nov. 3-7, 2007 ([PowerPoint Presentation](#); [Oral Presentation](#); [Addendum](#)); (e) Scanlan JP. [Study illustrates ways in which the direction of a change in disparity turns on the measure chosen.](#) *Pediatrics* Mar. 27, 2008 (responding to 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-e552).

³ (a) Mechanic D. Disadvantage, inequality and social policy. *Health Affairs* 2002;21(2):48-59; (b) Mechanic D. Who shall lead: Is there a future for population health? *J Health Politics, Policy and Law* 2003;28(2):421-442; (c) Mechanic D. Population health challenges for science and society. *Milbank Quarterly* 2007;85(3):553-559.

Janet M. Corrigan, PhD, MBA

October 22, 2009

Page 3

Europe,⁴ there has been increasing recognition of the existence of the patterns I describe and of their pertinence to the understanding of health and healthcare disparities.

Finally, I note that pay-for-performance is a subject of considerable interest to the National Quality Forum, though I do not know whether it has yet addressed the subject of the potential implications of pay-for-performance on healthcare disparities. In any event, efforts to tie pay-for-performance to healthcare disparities will be seriously misguided until disparities measurement issues have been satisfactorily resolved. See the [Pay for Performance](#) sub-page of MHD.

I hope your organization will give careful attention to these issues in its further efforts to provide guidance on reviewing or measuring health and healthcare disparities.

Sincerely,

/s/ James P. Scanlan

James P. Scanlan

⁴ (a) Carr-Hill R, Chalmers-Dixon P. [The Public Health Observatory Handbook of Health Inequalities Measurement](#). Oxford: SEPHO; 2005 (171-172); (b) Houweling TAJ, Kunst AE, Huisman M, Mackenbach JP. [Using relative and absolute measures for monitoring health inequalities: experiences from cross-national analyses on maternal and child health](#). *International Journal for Equity in Health* 2007;6:15; (c) Eikemo TA, Skalicka V, Avendano M. [Variations in health inequalities: are they a mathematical artefact?](#) *International Journal for Equity in Health* 2009;8:32; (d) Bauld L, Day P, Judge K. Off target: A critical review of setting goals for reducing health inequalities in the United Kingdom. *Int J Health Serv* 2008;38(3):439-454.

HARVARD MEDICAL SCHOOL



December 12, 2012

James P. Scanlan
Attorney at Law
1529 Wisconsin Avenue, NW
Washington, DC 20007

Re: Letter of October 26, 2012

Dear Mr. Scanlan,

I am writing to let you know the resolution of the concerns you raised to President Faust, Dean Flier, and Mr. Slavin about *Commissioned Paper: Healthcare Disparities Measurement*

A preliminary assessment was conducted by Massachusetts General Hospital and Harvard Medical School, and concluded that the allegations you raised represent a difference of scientific opinion, and do not rise to the level of, or provide sufficient evidence of research misconduct to warrant proceeding to an inquiry of the matter.

We have brought this matter to the attention of the authors, and, as with all differences of opinion, it is the responsibility of faculty members to evaluate whether your observations warrant revisions to the paper. Absent concerns with falsification or fabrication of data, Harvard Medical School and Massachusetts General Hospital do not independently assess the merits of any individual paper, leaving such interchange to the natural discourse in the scientific literature as an inherent and healthy part of the development of the scientific record.

Thank you for bringing this matter to our attention. While this may not be the outcome you had hoped for, I want to assure you that this was a thorough, fair and objective process.

Sincerely,

Gretchen Brodnicki
Dean for Faculty and
Research Integrity
Harvard Medical School

F. Richard Bringhurst, M.D.
Research Integrity Officer
Massachusetts General Hospital