We published data showing that current soil health indicator (SHI) assessments do not consistently detect differences in a range of soil management practices implemented in North Carolina soils. Van Es and Karlen reanalyzed our data and asserted that it validates SHI correlation to crop yields and sensitivity to management as measured by the Comprehensive Assessment of Soil Health (CASH). We respond to van Es and Karlen with a more representative analysis of our data showing that individual SHI measurements are not predictive of crop yield from the 30-yr North Carolina agronomic trial. Regressions for aggregate stability ($r^2 = 0.07$) and P ($r^2 = 0.18$) show that neither SHI sufficiently predicts corn yield for this dataset and show no obvious pattern based on tillage intensity. Relationships between corn (Zea mays L.) yield and most biological SHI had $r^2 \leq 0.18$, with only soil protein being moderately predictive of corn yield ($r^2 = 0.43$). The CASH index to assess overall soil health by integrating physical, chemical, and biological SHI measurements into a single value of soil health is also not predictive of corn yield in the trial ($r^2 = 0.12$). It is possible that current sampling and analytical procedures for assessing soil health do not consistently detect differences in productivity from soils with regional differences in land and ecological resources. We believe that calibrating SHI assessments to quantifiable agroecological outcomes instead of statistical rankings will reduce bias across regions and create a more inclusive framework for quantifying soil health.

Abbreviations: CASH, Comprehensive Assessment of Soil Health; SHI, soil health indicator.

We appreciate interest in our work that inspired van Es and Karlen (2019) to reanalyze our data in effort to gain more information, but their reanalysis misses important aspects of the original analysis in Roper et al. (2017). In our response to van Es and Karlen (2019) we want to provide more information about our original research objectives in Roper et al. (2017) which were to: (i) compare results from different soil tests, (ii) assess the ability for SHI to differentiate soil management effects on soils in North Carolina, and (iii) assess relationships between soil health and crop yields.

The original analysis includes data from three agronomic trials across North Carolina that ranged in duration from 15 to 30 yr and with treatments ranging in management intensity from no-till organic management to annual deep tillage with a moldboard plow. When comparing soil health tests, we learned that the Haney soil health test assessed all soils, except the moldboard plowed treatment, as having adequate soil health, whereas the CASH assessed only the no-till organic management treatment as having adequate soil health. We concluded that interpretations of soil health from these tests were too contradictory to provide a satisfactory answer about how management had affected soil health in the agronomic trials, and that more investigation is needed to develop a suitable soil health assessment protocol.
The van Es and Karlen (2019) reanalysis is specific to the CASH and focuses on the second and third objectives of Roper et al. (2017), which involve assessments of the ability to differentiate soil management practices based on how they affect SHI and for SHI to serve as predictors of crop yields. In their reanalysis, van Es and Karlen conclude that SHI included with the CASH are sufficient for differentiating management while emphasizing certain statistical differences in SHI results among tillage treatments (van Es and Karlen, 2019, Table 7 within). Many statistical differences in van Es and Karlen (2019) are similar to those in Roper et al. (2017, Table 6 within), with the distinction that the van Es and Karlen consolidate treatment groups in a manner that reduces the gradient in management intensity present in the original analysis. Another distinction is that van Es and Karlen have minimized the fact that some differences in SHI response to management are not present for similar tillage treatments at other locations. Because of inconsistencies in SHI responses, we concluded that CASH procedures for assessing soil health were not sufficient for differentiating management of North Carolina soils. Inconsistent sensitivity to management for SHI may possibly occur because of inherent limitations of soil response based on regional factors. Before the publication of Roper et al. (2017), the CASH framework was adjusted to assess soils with greater consideration of regional differences in land and ecological resources. The reanalysis by van Es and Karlen (2019) with raw metrics from individual CASH SHI measurements does not appreciably change the outcome.

Our third objective, relating SHI measurements to crop yield, is a major focus of the van Es and Karlen reanalysis, and we agree that crop yield, as well as other quantifiable management outcomes should have a significant role in assessments. In Roper et al. (2017), we state that "it may be useful to consider potential crop yields in soil health assessments for agronomic management" because long-term crop yields from conservation tillage systems were generally more favorable than yields from tilled soils, but we also caution that CASH interpretations were inconsistent with crop yields. We disagree with the way that van Es and Karlen reanalyzed yield data (van Es and Karlen, 2019, Fig. 1 and 2 within) because it consolidates regression data points into mean values for treatment groups that obscure variability in the relationship between SHI measurements and corn yields. We repeated the analyses of corn yields from the North Carolina piedmont agronomic trial in Reidsville with a more consistent representation of treatment groups (Fig. 1). We present the data labeled with tillage intensity (conservation tillage with no-till or in-row subsoiling, moderate tillage with a single pass of chisel or disk plow, and intense tillage with a pass of chisel or moldboard plow followed by a pass with disk plow) and include 30-yr average yield for each research plot (replicate) rather than grouped averages.

Biological SHI included with the CASH at the time of the experiment were organic matter (OM), CO2 respiration, autoclaved citrate extractable (ACE) protein index, and permanganate oxidizable C (active C), and these are all shown in Fig. 1. Aggregate stability and P are the only physical and chemical SHI included because of their closer association with biological soil processes relative to other SHI in their respective categories. Regressions for aggregate stability ($r^2 = 0.07$) and P ($r^2 = 0.18$) show that neither SHI sufficiently predicts corn yield for this dataset nor is there an obvious pattern based on tillage intensity. For all biological SHI except OM, there is slight separation with most intense tillage soils measuring at lower biological values and most conservation tillage soils measuring at higher biological values. Due to large variability in yield at intermediate

![Fig. 1. Regression of long-term corn yield with select soil health indicator measurements of the Comprehensive Assessment of Soil Health (CASH) for soils from the North Carolina Piedmont agronomic trial in Reidsville. Corn yield is the average of 15 crop years from 1985-2015. Each point represents a plot in the trial and plots are labeled based on tillage intensity. Conservation tillage = no-till or in-row subsoiling, moderate tillage = a single pass of chisel or disk plow, and intense tillage = a pass of chisel or moldboard plow followed by a pass with disk plow.](image-url)
biological values, however, most relationships between biological SHI and corn yield have $r^2 \leq 0.18$, and the only SHI with a moderately predictive relationship is ACE soil protein with $r^2 = 0.45$. A similar trend is observed with soybean [Glycine max (Merr) L.] yields from the same agronomic trial (data not shown). In contrast to the analyses in van Es and Karlen (2019, Fig. 1 within), our results show that these SHI measurements are not predictive of corn yield because we include all replicates in the regression rather than mean values of treatment groups. It is possible that current procedures for sampling and analyzing soils for health assessments may be confounded by factors that introduce variability into soil analyses (sampling depth, seasonal variability, soil morphology, etc.) and therefore do not consistently detect differences in productivity from these soils.

In addition to measurements of individual SHI, the CASH provides an index created from scoring functions that convert raw values from SHI measurements into indexed values between 0 and 100 based on a statistical distribution of measured values for soils in the CASH database. Within the index, a score of 0 represents very low soil health and 100 represents very high soil health. Indices for the 12 SHI included with the CASH are averaged to create an overall soil health value for each soil sample. Part of what makes the CASH framework unique is the combination of physical, chemical, and biological soil property measurements into a single index to assess soil health based on rankings in the CASH database, but the index was minimized in the van Es and Karlen (2019) reanalysis in favor of raw values. The concept of integrated soil health assessment was important in the original experiment (2019) reanalysis in favor of raw values. The concept of integrated soil health assessment was important in the original experiment (2019), which is presented as a contrast to the conclusions of Roper et al. (2012). This seems to be a primary conclusion of van Es and Karlen (2019), which is presented as a contrast to the conclusions of Roper et al. (2017). We do not dispute this. Comments made in Roper et al. (2017) were not made to disparage others from studying SHI, but instead to encourage researchers to continue improving on methods of quantifying relationships between soil management practices and associated outcomes.

We believe that comparisons of soils to create high and low rankings for measured values of SHI have obscured the agroecological context of soil health. If a goal of soil health management is to create sustainable agroecosystems, then soil health metrics should be calibrated to provide favorable agroecological outcomes for a given soil instead of using indices derived from statistical rankings that lack context to the intended goal. Bias in soil health interpretations can likely be reduced if soil health metrics are interpreted using known thresholds for agroecological productivity, such as the amount of CO₂ respiration associated with N mineralization (Picone et al., 2002) or soil penetration resistance that is known to restrict root growth (Vepřaskas and Waggoner, 1989). These types of interpretations would likely be similar to crop nutrient recommendations that are correlated and calibrated to crop yield expectations.

Even if we learn that North Carolina soils generally measure lower on the soil health scale than, for example, Northeastern US soils, there should still be an improvement in our understanding of whether particular management practices applied to North Carolina soils have enhanced agroecological productivity. Unfortunately, this improvement in understanding is not provided by the current structure of the CASH. There has been discussion about the ability of SHI to differentiate the effects of management practices on soil health in Roper et al. (2017), van Es and Karlen (2019), and other research (e.g., Congreves et al., 2015; Morrow et al., 2016; Rinot et al., 2019), but there should be more discussion about quantifying the agroecological relevance of the measured values of SHI considered for assessments.

Researchers have found that measurements of some proposed SHI can dynamically respond to soil management practices (Oades, 1984; Doran, 1987; Idowu et al., 2008; Culman et al., 2012). This seems to be a primary conclusion of van Es and Karlen (2019), which is presented as a contrast to the conclusions of Roper et al. (2017). We do not dispute this. Comments made in Roper et al. (2017) were not made to disparage others from studying SHI, but instead to encourage researchers to continue improving on methods of quantifying relationships between soil management practices and associated outcomes.

We believe that comparisons of soils to create high and low rankings for measured values of SHI have obscured the agroecological context of soil health. If a goal of soil health management is to create sustainable agroecosystems, then soil health metrics should be calibrated to provide favorable agroecological outcomes for a given soil instead of using indices derived from statistical rankings that lack context to the intended goal. Bias in soil health interpretations can likely be reduced if soil health metrics are interpreted using known thresholds for agroecological productivity, such as the amount of CO₂ respiration associated with N mineralization (Picone et al., 2002) or soil penetration resistance that is known to restrict root growth (Vepřaskas and Waggoner, 1989). These types of interpretations would likely be similar to crop nutrient recommendations that are correlated and calibrated to crop yield expectations.

Even if we learn that North Carolina soils generally measure lower on the soil health scale than, for example, Northeastern US soils, there should still be an improvement in our understanding of whether particular management practices applied to North Carolina soils have enhanced agroecological productivity. Unfortunately, this improvement in understanding is not provided by the current structure of the CASH. There has been discussion about the ability of SHI to differentiate the effects of management practices on soil health in Roper et al. (2017), van Es and Karlen (2019), and other research (e.g., Congreves et al., 2015; Morrow et al., 2016; Rinot et al., 2019), but there should be more discussion about quantifying the agroecological relevance of the measured values of SHI considered for assessments.

![Fig. 2. Regression of corn yields and Comprehensive Assessment of Soil Health (CASH) overall index scores for soils in the North Carolina piedmont agronomic trial in Reidsville. Corn yield is the average of 15 crop years from 1985–2015. Each point represents a plot in the trial and plots are grouped according to tillage intensity. Conservation tillage = no-till or in-row subsoiling, moderate tillage = a single pass of chisel or disk plow, and intense tillage = a pass of chisel or moldboard plow followed by a pass with disk plow. The entire scale of the CASH index is shown to emphasize the range of the measured values compared to the entire range of the index. Descriptors for the 2015 CASH overall index are: 0–40 = very low, 40–55 = low, 55–70 = medium, 70–85 = high, 85–100 = very high.](www.soils.org/publications/sssaj)
Soils collected from agronomic systems with 30 yr of no-till were still considered by the CASH to be constrained by physical and biological conditions despite no-till being a fundamental recommendation for improving soil health. If an essential management practice like no-till does not receive a positive soil health assessment after 30 yr, this raises questions about methods used to collect and analyze soils for quantitative health assessments and the effect of soil health practices in different agroecosystems. Continued research on SHI should be guided by objectives of sustainable soil management and agroecological outcomes. In Roper et al. (2017) we found that the soil health assessments we evaluated did not consistently differentiate the effects of a range of soil management practices on North Carolina soils. van Es and Karlen reanalyzed the data to produce an interesting addition to the original dataset, but their reanalysis does not provide a convincing case that the CASH methodology of quantifying soil health is adequate for improving our understanding of the agroecological context of soil health management. We encourage our colleagues to continue improving upon the foundation of soil health research. By sharing data with others, the Cornell Soil Health Lab, who developed the CASH, and other institutions amassing data on various soil properties, can create a more inclusive framework for quantifying soil health. Thank you again for your interest in our research and we hope to see more research from others soon.

REFERENCES

Soil Science Society of America Journal