

RICHARD BURTON BELZER, PHD

PO. Box 319
Mount Vernon, VA 22121
rbbelzer@post.harvard.edu

1 February 2013

Dr. John Faust
Chief, Community Assessment and Research Section
Office of Environmental Health Hazard Assessment
1515 Clay Street, Suite 1600
Oakland, CA 94612

Subject: Comments on January 2013 Draft *California Communities Environmental Health Screening Tool* (CalEnviroScreen)¹

Dear Dr. Faust:

I am writing to provide comments on one of several issues that have only come to light since OEHHA provided access to its summary data, which it did for the first time two weeks ago.² Both the July 2012 and January 2013 drafts are implicated.³

OEHHA's Scores and Rankings Are Arbitrarily Distorted Because Averages Were Obtained from Skewed Indicator Distributions

OEHHA calculates scores and ranks for 1,768 ZIP Codes by averaging percentiles across six "Exposure" indicators, four "Environmental Effects" indicators, and seven "Population Characteristics" indicators. Scores are the weighted product of these averages, each divided by 10. Unfortunately, the group averages are distorted because most of OEHHA's indicators are severely lopsided (i.e., "skewed").

This is shown in Figures 1a through 1c below. Each bar plots the value of the conventional skewness statistic. Five of the six Exposure indicators, all four Environmental Effects indicators, and four of the seven Population Characteristics indicators are highly (and in several cases, extraordinarily) skewed. Orange squares denote the average value of each indicator's distribution, expressed as a percentile. A normally distributed indicator would have a mean equal to the 50th percentile. These means range from the 52nd percentile (PM_{2.5}) to the 88th percentile (TRI). For two PC indicators (asthma and limited English proficiency ["linguistic isolation"]), the average value exceeds the 60th percentile.

¹ The preparation of these comments was supported by the California Chamber of Commerce. The analysis and opinions expressed herein belong solely to the author.

² California Environmental Protection Agency Office of Environmental Health Hazard Assessment (2013a).

³ California Environmental Protection Agency Office of Environmental Health Hazard Assessment (2012a, California Environmental Protection Agency Office of Environmental Health Hazard Assessment (2013b).

Figure 1a: Five of Six 'Exposure' Indicators Are Highly Skewed

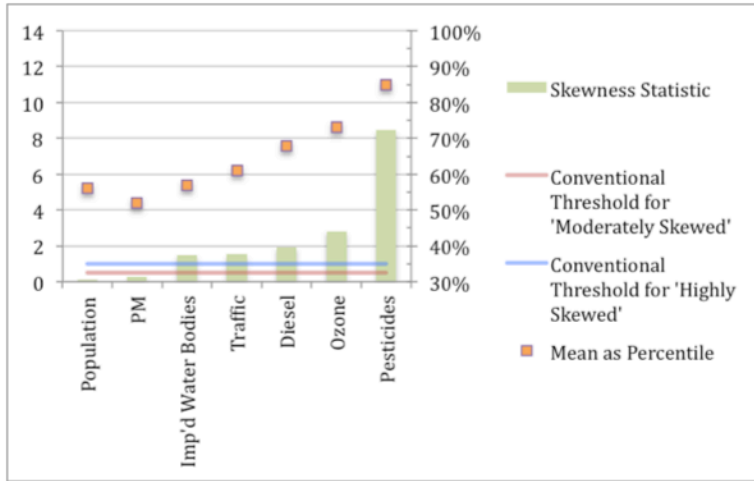


Figure 2b: All Four 'Environmental Effects' Indicators Are Extremely Skewed

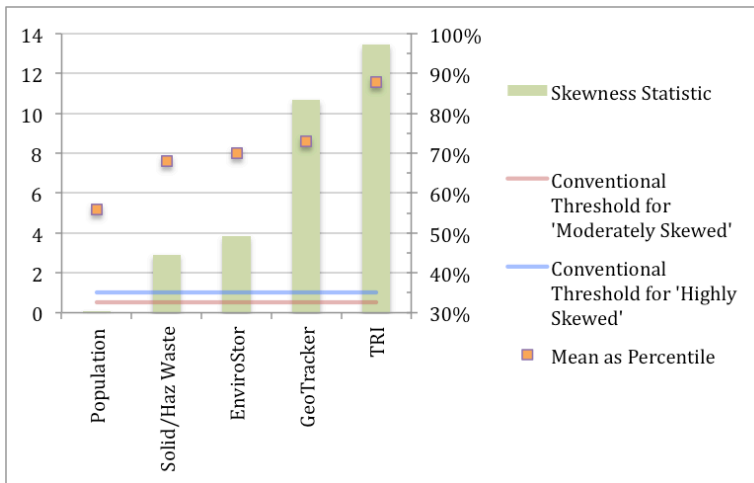


Figure 1c: Four of Seven 'Population Characteristics' Indicators Are Highly Skewed; One is Extremely Skewed

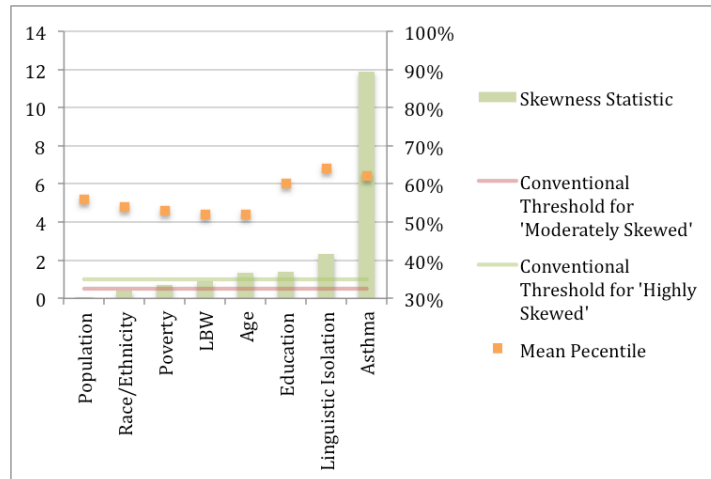


Figure 2: Percentile-based Ranking Is Neither Highly Correlated Nor Proportional to Standard Score-based Ranking

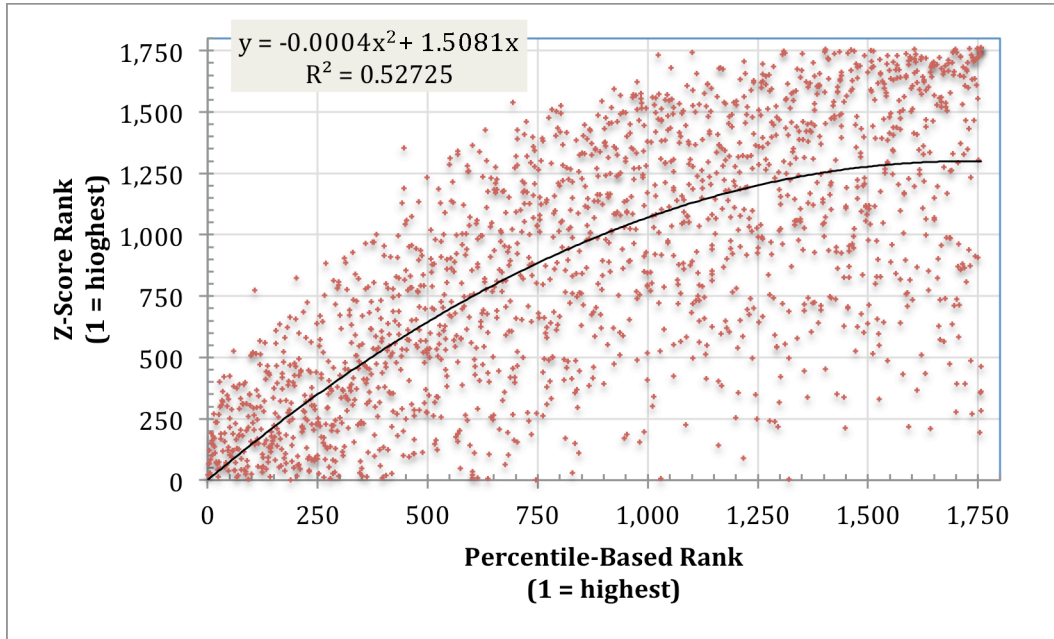


Figure 3: Standard Score-Based Ranks for Top 30 ZIP Codes Ranked by Averaging Percentiles

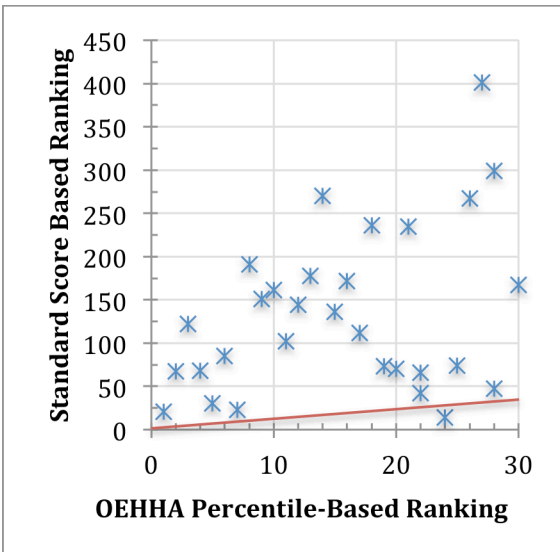
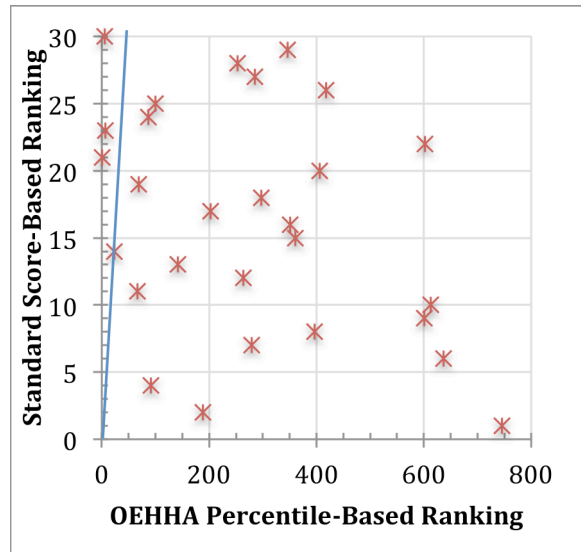


Figure 4: Percentile-Based Ranks for Top 30 ZIP Codes Ranked by Averaging Standard Scores



Belzer Comments on January 2013 Draft California Communities Environmental Health Screening Tool (CalEnviroScreen)

1 February 2013

Page 4 of 7

When a mix of unskewed and skewed indicators is averaged, each indicator no longer gets the same weight. Skewed indicators get weighted more heavily, and extremely skewed indicators get extraordinarily large weights. Meanwhile, the signal transmitted by indicators that are not skewed (such as PM_{2.5}) is overwhelmed.

I. OEHHA's Rankings Are Unreliable and Arbitrary, Especially for ZIP Codes that OEHHA Ranks at the Top

It is useful to compare OEHHA's percentile-based ranks with ranks obtained by using standard scores (i.e., "Z-scores"). Converting raw data into Z-scores maintains their rank-order, but it adjusts for the relative size of their means and the width of their standard deviations. If skewness did not matter, scores would be proportional and ZIP Code rankings under both schemes would be very nearly the same. Knowing a ZIP Code's percentile-based rank would enable one to predict, with a high degree of accuracy, its standard score-based rank.

Figure 2 shows that percentile- and standard score-based ranks are not highly correlated. In addition, the best fitting relationship is not proportional. Ranking errors are especially pronounced for ZIP Codes at the tails because that is where skewness imparts the greatest bias. This poses an obvious problem because OEHHA intends to focus public attention exactly where its ranking scheme is most unreliable.

For the 30 ZIP Codes to which OEHHA assigns the highest ranks, Figure 3 shows that there is hardly any correlation at all between ranks obtained by averaging percentiles and averaging standard scores. If it were not for statistical anomalies such as skewness, both averaging schemes would yield similar results and standard score-based ranks would be clustered along the red line. This is clearly not the case. Only four of OEHHA's top 30 stay in the top 30. Seventeen of OEHHA's top 30 ZIP Codes get standard score-based ranks greater than 100.⁴ The scores and ranks of OEHHA's top 30 Zip Codes are clearly arbitrary.

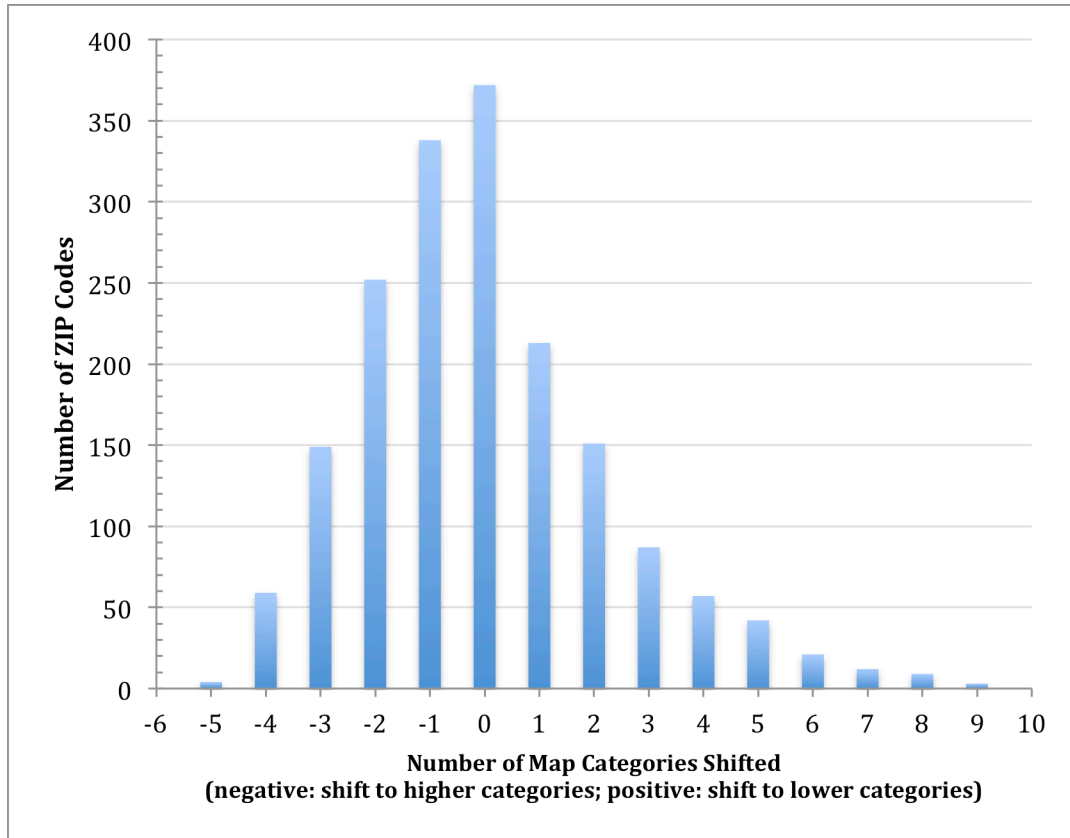
A similar result is obtained when looking at the top 30 ZIP Codes ranked by averaging standard scores, as shown in Figure 4. Were no statistical anomalies such as skewness present, percentile-based ranks for the top 30 would be clustered along the blue line. Obviously, they do not. Only four of the top 30 ZIP Codes ranked by standard scores make OEHHA's top 30, and 23 of them get percentile-based ranks greater than 100.⁵

For individual ZIP Codes, and especially those in OEHHA's top 30, scores and ranks are wholly unpredictable. ZIP Codes that OEHHA ranks 1, 2, and 3, decline in rank to 21, 67, and 122 when standard scores are used instead of percentiles. Similarly, ZIP Codes that get the top three ranks with standard scores are ranked by OEHHA at 746, 189, and 1320. The reason is simple: OEHHA's scores and ranks are distorted by skewed indicators, and this distortion is most severe for ZIP Codes getting the highest scores because this is where skewed indicators impart the greatest bias.

⁴ The R² of a bivariate regression is 0.13.

⁵ The R² of a bivariate regression is 0.22, and the coefficient has the wrong sign.

Figure 5: Minimum Number of ZIP Codes that Would be Assigned to Lower, the Same, or Higher Color-Coded Categories in OEHHA's Maps



II. OEEHA's Maps Are Similarly Unreliable and Arbitrary

In its draft reports, OEHHA presents maps that show visually how scores vary across the State's 1,768 ranked ZIP Codes. For each indicator and relevant composites, maps are color-coded with 10 shades corresponding to decile ranking. Color-coding conveys powerful visual signals.

But how trustworthy are those visual signals if the scores and ranks underlying them are arbitrary? We can obtain additional clues by calculating the minimum number of ZIP Codes that would be color-coded differently if standard scores were averaged instead of percentiles. We also can determine the minimum number of ZIP Codes whose color-coding would change by more than one decile.⁶

Figure 5 shows that categories (and map shades) would be unchanged for at most 372 of the State's 1,768 (21%) ZIP Codes. Color-coding would change for 80% of California ZIP Codes. Of the 1,396 ZIP Codes that would change color, about half would be assigned five or more deciles away.

⁶ For technical reasons, this illustration understates, probably by a lot, the true number of ZIP Codes that would move one or more deciles.

Belzer Comments on January 2013 Draft California Communities Environmental Health Screening Tool (CalEnviroScreen)

1 February 2013

Page 6 of 7

The take-home message is clear: OEHHA's model produces arbitrary scores and ranks just because many of OEHHA's indicators are highly skewed, and some of them are extraordinarily skewed.

III. Indicator Skewness Must Be Removed, but Removing It Won't Be Enough

CalEnviroScreen was purportedly designed to give each indicator in a common category the same weight. The data OEHHA finally disclosed show that this is not the case. By averaging a mix of skewed and unskewed indicators, CalEnviroScreen violates this design principle and transmits multiple biases, all of them hidden from the public. Skewed indicators get disproportionately greater weight, and extraordinarily skewed indicators get exceedingly disproportionate weight.

Skewness and other statistical abnormalities must be removed. Failing to normalize each and every indicator *before* averaging will ensure that CalEnviroScreen produces unreliable and arbitrary scores and ranks. For each ZIP Code, and especially those allegedly at the top, scores and ranks will be artifacts of whichever skewed indicator distributions predominate.

Solving the skewness problem is, of course, just the tip of the iceberg. As you know, I participated in the September 2012 Academic Workshop⁷ and identified numerous scientific, technical, and economic problems with the indicators OEHHA initially proposed, OEHHA's scientifically indefensible decision to use ZIP Codes as the geographic unit of analysis, and perhaps dozens of other fatal flaws. I supported these opinions with extensive public comments.⁸

Based on my review of the January 2013 draft, it appears that OEHHA has not yet made the corrections necessary to overcome the defects I noted in these comments, nor has it explained why they are unimportant. In addition, the latest draft includes a new indicator—Diesel PM—that relies on data predating USEPA's Ultra Low Sulfur Diesel rules, which model year 2007 and later engines are required to use and which USEPA says reduce emissions by more than 90%. Inexplicably, the January 2013 draft does not explain why OEHHA believes it is scientifically appropriate to knowingly include obsolete data in its scoring scheme.

Respectfully submitted,



Richard Burton Belzer, PhD

⁷ California Environmental Protection Agency Office of Environmental Health Hazard Assessment (2012b).

⁸ Belzer (2012); Draft California Communities Environmental Health Screening Tool (CalEnviroScreen): Proposed Method & Indicators; CIPA Work Group & Public Review Draft.

Belzer Comments on January 2013 Draft California Communities Environmental Health Screening Tool (CalEnviroScreen)

1 February 2013

Page 7 of 7

References

Belzer RB. 2012. Comments on *California Communities Environmental Health Screening Tool (CalEnviroScreen): Proposed Method & Indicators* (July 30, 2012 External Review Draft). Available: <http://oehha.ca.gov/ej/pdf/103012/BusinessComment3Belzer.pdf> [accessed January 30, 2013].

California Environmental Protection Agency Office of Environmental Health Hazard Assessment. 2012a. Draft California Communities Environmental Health Screening Tool (CalEnviroScreen): Proposed Method & Indicators; CIPA Work Group & Public Review Draft. Available:

http://www.oehha.ca.gov/ej/pdf/DraftCalEnviroScreen073012.pdf?utm_source=07%2F30%2F12&utm_campaign=CI+notice&utm_medium=email [accessed August 26, 2012].

California Environmental Protection Agency Office of Environmental Health Hazard Assessment. 2012b. *Academic Workshop on the Draft California Communities Environmental Health Screening Tool: Overview of Proposed Methods and Indicators*. Oakland, Calif.: California Environmental Protection Agency Office of Environmental Health Hazard Assessment, 24.

California Environmental Protection Agency Office of Environmental Health Hazard Assessment. 2013a. Availability of Data Related to the Second Public Review Draft of the California Communities Environmental Health Screening Tool. Available: <http://oehha.ca.gov/ej/cipa011613.html> [accessed January 23, 2013].

California Environmental Protection Agency Office of Environmental Health Hazard Assessment. 2013b. 2nd Draft California Communities Environmental Health Screening Tool (CalEnviroScreen): Proposed Method & Indicators; CIPA Work Group & Public Review Draft. Available: <http://www.oehha.ca.gov/ej/pdf/CalEnviroScreen2ndPublicReviewDraft010313.pdf> [accessed January 13, 2013].

cc: Ken Alex, Director, Governor's Office of Planning and Research
Matthew Rodriguez, Secretary for Environmental Protection
Miriam Barcellona Ingenito, Deputy Secretary for Environmental Policy and Community Programs, CalEPA
Arsenio Mataka, Assistant Secretary for Environmental Justice and Tribal Affairs
George Alexeeff, Director, Office of Environmental Health Hazard Assessment