Meaningful, Useful, and Legitimate Information in the Use of Index Numbers in Decision Making

Fred S. Roberts

Helen M. Roberts

Alexis Tsoukias

Index Numbers and Decisions

- Often decisions are driven by information summarized in an index number.
- Under what conditions can we trust and use the information contained in an index number?
- Will explore this with two examples:
  - Body mass indices
  - Air pollution indices

credit: National Park Service, Wikimedia Commons

credit: FatM1ke Wikimedia Commons

Meaningfulness

- Suppes (1959), Suppes & Zinnes(1963): A statement using scales of measurement is ***meaningful*** if the truth or falsity remains unchanged if admissible (allowable) transformations of scale are made.
- Admissible transformations: Kilograms to Pounds, degrees Centigrade to degrees Fahrenheit, meters to feet, …
- Meaningless statements can be misleading: They depend on a somewhat arbitrary choice of parameters such as unit or zero point.
- Example: Today's temperature is twice yesterday's. This may be true in degrees Fahrenheit and false in degrees Centigrade.
- There are situations where an alternative definition is needed, for example when not every version of a scale can be obtained from another by an admissible transformation. But we will disregard those subtleties.

Usefulness

- Information also needs to be ***useful***: It needs to satisfy the demands for which it has been requested and the demands of the person or organization that is expected to use it for some decision purposes.
- ***This introduces a subjective dimension that is not formally captured like meaningfulness.***
- We will explore properties of this concept that will, hopefully, lay the groundwork to make it more precise.
- We will get at this by studying statements using index numbers that are useful and useless.

Legitimacy

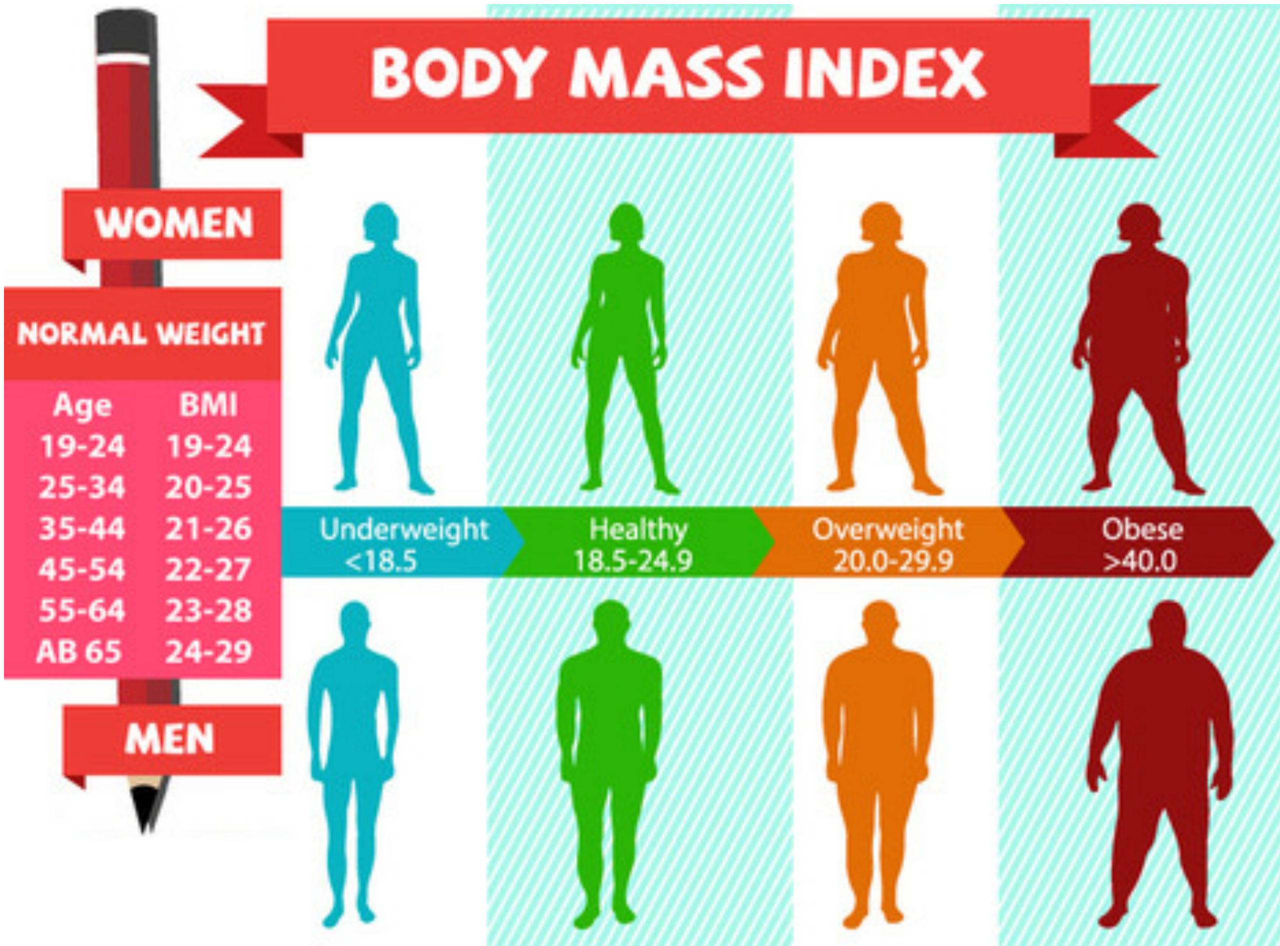
- Information also needs to be ***legitimate***: The way it is collected, elaborated upon, and used needs to satisfy cultural, historical, organizational, and legal constraints underlying the whole decision process.
- ***As with usefulness, this is a subjective dimension that is not formally captured like meaningfulness.***
- We will explore properties of this concept that will, hopefully, lay the groundwork to make it more precise.
- We will illustrate statements using index numbers that are legitimate and illegitimate.
- We will also look for which combinations of meaningful/meaningless, useful/useless, and legitimate/illegitimate might occur.

Scale Type

Credit: Toby Hudson, Wikimedia commons (no changes)

- Following Stevens (1946, 1951, 1959), we will refer to different types of scales of measurement as defined by the class of admissible transformations.
- ***Ratio scale:*** the admissible transformations are multiplication by a positive constant (e.g., mass, length)
- ***Interval scale****:* the admissible transformations are multiplication by a positive constant and change of unit (e.g., temperature)
- ***Ordinal*** scale: the admissible transformations are all monotone increasing functions (e.g., ranking of minerals by hardness)

BODY MASS INDEX



Credit: DrV-Amar, Wikimedia commons (no changes)

Obesity/Body Mass Index

- Body fatness or adiposity is an indicator of potential medical problems
  - High blood pressure, high cholesterol, Type 2 diabetes, coronary heart disease, stroke, and some types of cancers.
- Some of methods of measuring body fat are skinfold thickness measurements, underwater weight measurement, bioelectrical impedance, and dual-energy x-ray absorptiometry.
- ***Existing ways to measure body fatness using such measurement methods can be expensive or require specially trained personnel, and methods are difficult to standardize. That might make them useless.***

Credit: Wikimedia Commons, NIH, public domain

Obesity/Body Mass Index

- ***The difficulty to standardize might also make the metrics illegitimate as well as useless:***
  - One person doing the measurements might come up with a different index than another person doing the measurement.

Skinfold body fat measurement

credit: U.S. Marine Corps, Wikimedia Commons

Obesity/Body Mass Index

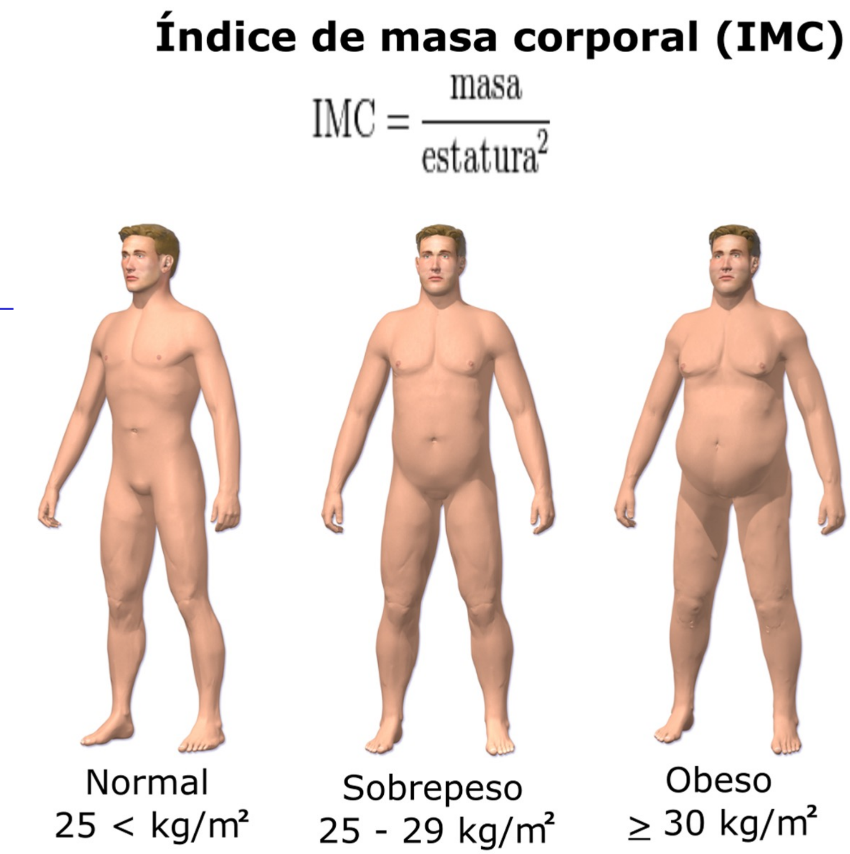
- However, if the ways to measure body fatness lead to a ratio scale, then it is *meaningful* to make comparisons such as body fatness of one person is 10% higher than body fatness of another.
  - Skinfold thickness, underwater weight measurement do seem to define ratio scales
  - This kind of comparison would be meaningless for interval scales, e.g., temperature.
- ***Meaningfulness depends on the scales used to describe the data, not on the procedure used to gather the data or the characteristics of the population the statement using metrics or indices is describing or directed at.***

BMI

$$BMI = \frac{W}{H^2}$$

W is weight in Kg, H = height in meters

- A person with BMI ≥ 30.0 is considered ***obese***
- Someone with 25.0 ≤ BMI ≤ 29.9 is considered ***overweight***



credit: BruceBlaus De la traducción: Ortisa, Wikimedia Commons, no changes

BMI

- In contrast to measures of fatness, BMI is easy to measure and standardize.
- In that sense, it is useful.
- But: does it relate to adiposity?
- Does it do what it was designed to do: predict certain kinds of diseases such as cardiovascular disease?
- So, usefulness must be determined.
- ***Usefulness includes several components: ease of use, but also appropriateness for intended use.***

Comparisons Using BMI

- Are comparisons of BMI meaningful? If *BMI(x)* is the body mass index of individual *x* and *BMI(x,t)* is BMI of *x* at time *t,* it is meaningful to say that

- ***BMI(x) > BMI(y)*** (1)
- ***BMI(x) = 2BMI(y)*** (2)
- ***BMI(x,t) = 1.2*BMI(x,t-1)*** (3)

- Because multiplication of weight *W* by a positive constant and of height *H* by a positive constant doesn't change truth or falsity.
- However, it is not meaningful to say that *x* is obese, i.e., that *BMI(x)* ≥ 30, without specifying units used.
- (Though units are usually understood to be the standard units if they are not mentioned.)

BMI for Different Populations

- If an ***athlete*** has a BMI ≥ 30, he or she would be called obese.
- That statement is meaningful, but it may not be so useful for athletes because the BMI might be higher because of increased muscularity rather than increased body fatness.
- ***Thus, for certain populations, BMI is less useful than for others.***
- In other words, perhaps there is even need for understanding ***degree of usefulness***.

Credit: User:Fitness training for life wikimedia commons (no changes)

BMI for Different Populations

- *BMI is interpreted differently for children than for adults.*
- ***Same formula used, but amount of body fat changes with age, and it differs between boys and girls.***
- So, the guidance as to what defines overweight or obesity changes.
- The value of BMI defining obesity depends on a reference population of children of a given age and sex, with obesity defined as having a BMI at or above the 95^{th} percentile in this population.
- If x is a boy of age 12 and y is a boy of age 13, it is meaningful to say that $BMI(x) > BMI(y)$.
- But this is not a useful comparison since it is comparing apples to oranges.

BMI for Different Populations

- Similarly, if *x* is a boy of age 12 and *y* is a girl of age 12, the statement *BMI(x) > BMI(y)* is both meaningful and not useful.
- ***In short, usefulness can depend on the population a statement refers to, and it is possible that a statement that is similar might be useful for some comparisons and not for others.***

Credit: Tony Alter from Newport News, USA wikimedia commons (no changes)

BMI for Different Populations

- *To say that a boy of age 12 is obese is a meaningful statement.*
- If *BMI(x)* is at least as high as that of 95% of other boys of age 12 in the reference population, this statement remains true or false if weight and height are multiplied by appropriate positive constants.
- The statement is also, presumably, useful: It can suggest medical interventions
  - Diet, exercise, medication

Credit: Wellcome blog post (archive)
Wikimedia Commons, no changes

BMI for Different Populations

- The choice of a 95% threshold is related to usefulness, not meaningfulness.
- A 75% threshold could just as easily be used and the resulting class of obese boys of age 12 would be much larger.
- But now the conclusion of obesity might not be as useful since it might not trigger the need for behavioral or medical intervention.

BMI for Different Populations

- Concluding that an individual is obese can lead to behavioral or medical intervention
- Such medical-based decisions may disregard cultural factors: In some societies, obesity has been valued.
- Historically, some ***Tahitians*** valued obesity and fattened up people to make them more sexually attractive.
- Historically, ***the Nauru*** fattened up young women because fat was associated with fertility and beauty.
- Thus, for some cultures, the conclusion that a person is obese would not lead to medical intervention, but instead might be valued. ***For these cultures, the conclusion from use of the BMI is not legitimate.***

Credit: FatM1ke wikimedia commons (no changes)

BMI for Different Populations

- ***Christian Scientists*** generally view disease and illness as a mental issue, not a physical one, and so resort to prayer rather than medicine to cure disease.
- *Some believe that prayer will help in weight loss.*
- However, this does not make the conclusion that they are obese illegitimate.
- ***But it does make the resulting recommendation to use medicine illegitimate, because it violates religious principles.***

Sensitivity/Specificity

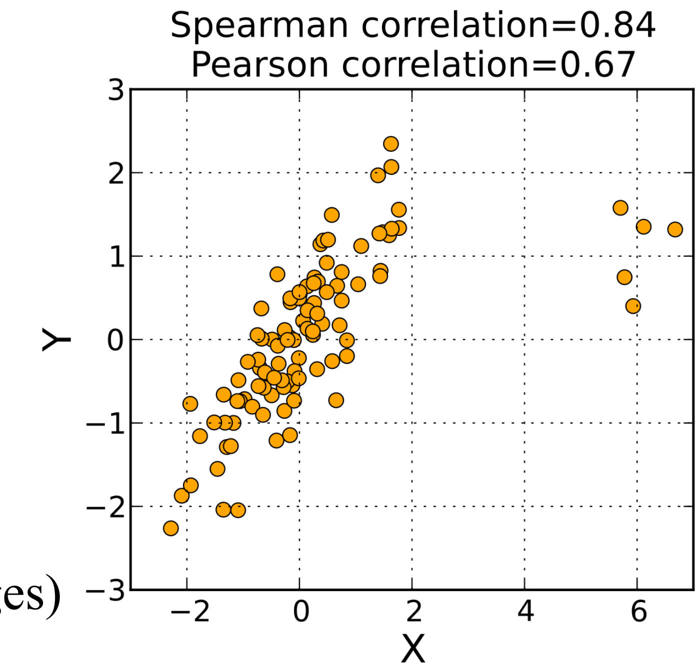
- The ***sensitivity*** of a test = number of true positives (number correctly identified as positive) over the number of positives.
- The ***specificity*** = the number of true negatives (number correctly identified as negative) over the number of negatives.
- Smalley, et al. (1990): For men, if % body fat = %BF is accurate, the sensitivity of the BMI index = 44.3%.
  - ***Less than half of the men identified as obese in terms of %BF were identified as obese by BMI.***
- For men, the specificity of BMI = 90.1%.
  - ***90.1% of men not obese were correctly identified as not obese by BMI.***
- These conclusions are certainly meaningful.

Sensitivity/Specificity

- The suggestion is that BMI is not very useful in identifying men who are actually obese, but quite useful in identifying men who are not obese.
- So, while it is meaningful to say that a man is obese or that a man is not obese, the usefulness of these two conclusions differs.
- ***It is another peculiarity of the concept of usefulness that a conclusion and its negation, while both meaningful, can differ in terms of usefulness.***

Relationship between BMI and Other Metrics

- If BMI is going to be useful as a measure of obesity, it will need to be closely related to %BF.
- One way to test this is to see if they are ***correlated***.
- For the widely-used correlation coefficients, since BMI and %BF are both measured on ratio scales, the correlation between them doesn't change under admissible transformations.
- So, it is meaningful to say that the correlation between BMI and %BF is a given number.
- However, is it useful?



Credit: Skbkekas wikimedia commons (no changes)

Relationship between BMI and Other Metrics

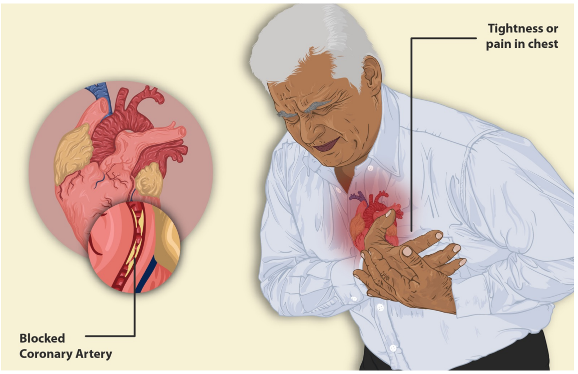
- Wellens, et al. (1996) compared BMI to %BF, using Spearman rank correlation ρ.
- They calculated $\rho = 0.79$ for women, $\rho = 0.65$ for men.
- ***Since correlation is invariant under admissible transformations, it is meaningful to say that the correlation for women is higher than for men.***
- ***But is a correlation of 0.65 high enough to be able to conclude that BMI is a useful proxy for %BF?***
- Might there be different thresholds for “high” for different populations, e.g., adults vs. children?

Relationship between BMI and Other Metrics

- Some feel that in physics, a correlation must be at least 0.95 or at most -0.95 to be useful, in chemistry at least 0.9 or at most -0.9, in biology at least 0.7 or at most -0.7, and the social sciences at least 0.6 or at most -0.6.
- Is medicine more like physics?
- *Deciding on a medical treatment may require a correlation of 0.9, whereas deciding on what telephone to buy may only require a correlation of 0.4.*
- One of the challenges is to understand what correlations are high enough to make a correlation useful in different applications.
- ***Usefulness of a conclusion involving index numbers can depend on the decision the conclusion is used for.***

BMI as a Predictor of Disease

- ***Biomarkers*** are thought to be predictive of disease.
  - *FBG* (fasting blood glucose, which is related to diabetes); *HDL* (high-density lipoprotein cholesterol, which is connected to heart disease); *SBP* (systolic blood pressure); *TG* (triglyceride).
  - Changes in the levels of these biomarkers are four of the most important physiological consequences of obesity.
- Willet, et al. (2006) calculated Pearson correlation *r* between BMI and such biomarkers.
- For men, for HDL vs. BMI, $r = –0.32$, for FBG vs. BMI, $r = 0.19$
  - *An increase in BMI was (weakly) correlated with decrease in HDL (higher HDL is better).*
  - *An increase in BMI was (weakly) correlated with an increase in FBG.*



Credit: https://www.myupchar.com/en wikimedia commons (no changes)

BMI as a Predictor of Disease

- So, would your physician suggest you lower your BMI because of these calculations?
- How useful these conclusions are will depend on assumptions about your behavior, priorities, etc.
- In both cases it could depend on the level of your BMI (and other medical conditions).
- Or on how results will be applied: Change your diet? Start you on medication?
  - Correlation may be high enough to suggest the former but not the latter.
- ***The resulting decision depends not on the meaningfulness of the results, but on some judgments of usefulness of the results and about how the results will be applied.***

BMI as a Predictor of Disease

- Willet, et al. (2006) studied use of ***bioelectrical impedance analysis (BIA)*** as a way to obtain %BF and compared it to BMI as predictor of biomarkers FBG, HDL, SBP, and TG.
- They concluded that BMI is a better, cheaper, and easier measure to use and its correlations with these biomarkers were comparable to those of BIA.
- ***Using BMI would be much more useful because of ease of calculation though not because of usefulness in terms of correlation with a metric – there are a variety of interpretations of usefulness.***
- As observed before: Ways to measure degree of usefulness may be worth developing.

BIA to measure body fat

Credit: Department of Defense wikimedia commons (public domain)

AIR POLLUTION INDICES

Credit: Kentaro IEMOTO, Wikimedia commons no changes

Pollutants

- There are a variety of pollutants such as carbon monoxide (CO), hydrocarbons (HC), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), ozone (O_3), and (various kinds of) particulate matter (PM).
- When air pollution measurement was in its infancy, there was a goal of finding a pollution index based on the levels of emissions of different kinds of pollutants.
- *This could help in comparing strategies for pollution control and take into account tradeoffs.*
- *It might also help in making recommendations for individual behavioral responses to different pollution conditions.*
- The issues were and still are rather complex.

Using Weight of Pollutants

- ***Simple idea: measure total weight of emissions of pollutant i over a fixed period of time in a given volume of air and sum over i to define a pollution index.***
  - The pollutant concentration is measured in units such as milligrams per cubic meter (which is usually used for particulate matter) or parts per million by volume (ppm) (which is usually used for CO or O_3) or parts per billion by volume (ppb) (which is usually used for NO_2 and SO_2). (There are conversions that take the metric in milligrams per cubic meter and change it to ppm by volume, and vice versa.0
- Let ***e(i,t,k)*** = total weight of emissions of pollutant *i* over the *t*th time period and due to the *k*th source or measured in the *k*th location.
- Let ***A(t,k)*** be the sum over all *i* of *e(i,t,k)*.

Using Weight of Pollutants

- Using the *weight-based index A*, Walther (1972) concluded things like:

(1) Transportation is the largest source of air pollution, with stationary fuel combustion (especially by electric power plants) second largest.

(2) Transportation accounts for over 50% of all air pollution.

(3) CO accounts for over half of all emitted air pollution.

- Are conclusions such as these meaningful?

Credit: Salvatore Arnone wikimedia commons (no changes)

Using Weight of Pollutants

- Conclusions such as these are meaningful if we use a standard measure of pollutant concentration such as milligrams per cubic meter or ppm or ppb by volume.
- For they can be written as:

$$A(t,k) > A(t,k')$$

$$A(t,k_r) > \sum_{k \neq k_r} A(t,k)$$

$$\sum_{t,k} e(i,t,k) > \sum_{t,k} \sum_{j \neq i} e(j,t,k).$$

- And all the scales are ratio scales.
- But are they useful? And legitimate?

Using Weight of Pollutants

- ***A unit of mass of CO is far less harmful than a unit of mass of NO_2.***
- This suggests that simply summing as in *A(t,k)* is not a useful measure of pollution.
- At one point, U.S. Environmental Protection Agency standards based on health effects for a 24-hour period allowed 7800 units of CO, 330 units of NO_2, 788 of HC, 266 of SO_2, 150 of PM (Environmental Protection Agency, 1971).
- These are ***Minimum acute toxicity effluent tolerance factors*** (MATE criteria) or ***tolerance factors.***
- The tolerance factor is the level at which adverse effects are known or thought to occur.

Using Weight of Pollutants

- There are other issues.
- Some of these pollutants are more serious in the presence of others, e.g., SO_2 are more harmful in the presence of PM.
- Also, the products of chemical reactions of the different pollutants can be damaging. Oxidents such as ozone are produced by HC and NO_2 reacting in the presence of sunlight.
- These measures disregard both of these complications, which ***suggests that using the pollution index A, for example to set standards for emissions or make other air pollution policy decisions, fails the usefulness criterion.***

Using Weight of Pollutants

- Using a variety of subindices (as here) is common with index numbers.
- ***But conclusions from index numbers, though meaningful, can be useless if they disregard the kinds of interactions/interdependencies among the factors measured by the subindices as the ones here.***

Credit: Welp.sk wikimedia commons (no changes)

Using Weight of Pollutants

- What about legitimacy?
- ***While the conclusions we have discussed are useless, they seem to be legitimate:*** The way they are obtained doesn't seem to violate cultural, historical, organizational, or legal constraints.

credit Wikimedia Commons (public domain)

An Index that Considers Health Effects

- Let $\tau(i)$ be the tolerance factor for the i^{th} pollutant. Let the ***severity factor*** be $1/\tau(i)$.
- One idea is to weight the emission levels (in mass) by the severity factor and get a weighted sum.
- This amounts to using the indices $1/\tau(i)*e(i,t,k)$ and summing these to get $B(t,k)$.
- This called ***Pindex***:

$$B(t,k) = \sum_i \frac{1}{\tau(i)} e(i,t,k)$$

- Pindex was introduced in the San Francisco Bay Area in the 1960s when they first tried to seriously measure pollution

credit Jordan McQueen jordanfmcqueen Wikimedia Commons (public domain)

An Index that Considers Health Effects

- Under Pindex, transportation is still the largest source of pollutants, but now accounting for less than 50%.
- Stationary sources fall to fourth place.
- CO drops to the bottom of the list of pollutants, accounting for just over 2% of the total.
- It is easy to see that these conclusions are again meaningful as long as all emission weights are measured in the same units.
- But, are they useful or legitimate?

Smog in downtown Los Angeles

Credit: en:User:Downtowngal wikimedia commons (no changes)

An Index that Considers Health Effects

- Pindex amounts to the following: For a given pollutant, take the percentage of a given harmful level of emissions that is reached in a given period of time, and add up these percentages over all pollutants.
  - The sum can be greater than 100% as a result.
- *If 100% of the CO tolerance level is reached, this is known to have some damaging effects.*
- *But Pindex implies that the effects are equally severe if levels of five major pollutants are relatively low, say 20% of their known harmful levels.*
- ***It is therefore doubtful that this index of pollution gives useful results.***

An Index that Considers Health Effects

- In the early days of air pollution measurement, reported severity factors differed from study to study.
- One reason was that air quality standards were not all laid out for the same time period; some for one hour, some for eight hours, etc.
- There were differing opinions as to how to extrapolate the standards to the same time period, e.g., 24 hours.
- ***Thus, using Pindex again failed on the usefulness criterion, since the ways it was measured were inconsistent.***

An Index that Considers Health Effects

- What about legitimacy?
- Using Pindex does not necessarily seem illegitimate, just as using the weight-based index *A* was not.

Air Quality Index AQI

- The AQI has been issued by the U.S. Environmental Protection Agency since 1976.
- Variants of the AQI are now used around the world.
- The AQI focuses on health effects that an individual might experience within a few hours or days after breathing polluted air.

Credit: AirNow

Air Quality Index AQI

- ***The AQI assigns a number between 1 and 500 for AQI subindices for each of five pollutants: PM, CO, SO_2, NO_2, and ozone O_3.***
- The subindices are calculated by converting measured pollutant concentrations (e.g., in micrograms per cubic meter or ppm or ppb) to a uniform index ***based on the health effects associated with a pollutant.***
- Those health benchmarks are established by the Environmental Protection Agency and updated regularly.
- ***The overall AQI is reported as the highest of the AQI subindices.***
- One day (or hour) it could be due to ozone, another to CO.

Air Quality Index AQI

The level for each pollutant is reported in one of six categories of increasing seriousness.

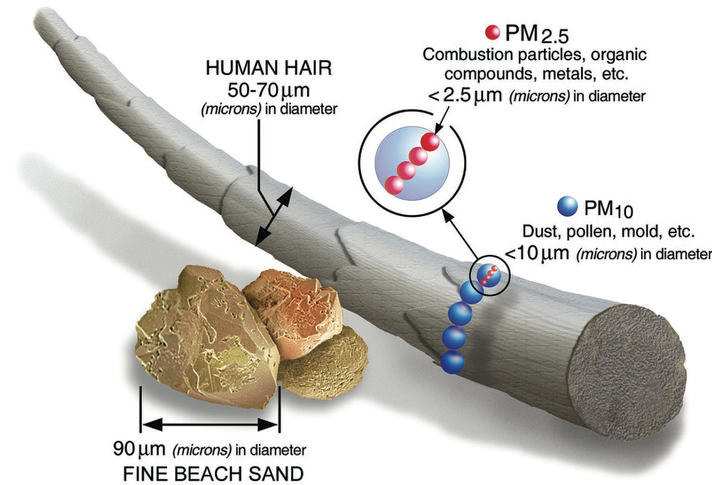
Daily AQI Color	Levels of Concern	Values of Index	Description of Air Quality
Green	Good	0 to 50	Air quality is satisfactory, and air pollution poses little or no risk.
Yellow	Moderate	51 to 100	Air quality is acceptable. However, there may be a risk for some people, particularly those who are unusually sensitive to air pollution.
Orange	Unhealthy for Sensitive Groups	101 to 150	Members of sensitive groups may experience health effects. The general public is less likely to be affected.
Red	Unhealthy	151 to 200	Some members of the general public may experience health effects; members of sensitive groups may experience more serious health effects.
Purple	Very Unhealthy	201 to 300	Health alert: The risk of health effects is increased for everyone.
Maroon	Hazardous	301 and higher	Health warning of emergency conditions: everyone is more likely to be affected.

Air Quality Index AQI

- It is certainly meaningful to say that the AQI for a given pollutant *i* is in a more serious category today than it was yesterday.
- But what if the AQI score for ozone of 209 was highest yesterday and the AQI for CO of 230 was highest today?
- Is it meaningful to say that the overall air quality AQI was higher today than yesterday?
- To answer this questions requires some explanation of how the AQI subindices are calculated.

Air Quality Index AQI

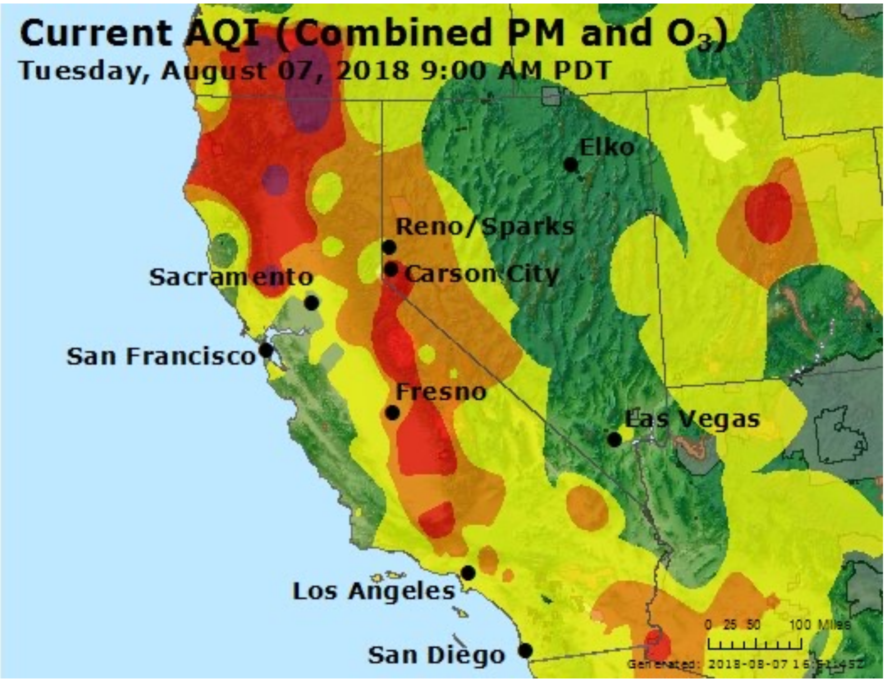
- The Environmental Protection Agency has created AQI values for the different pollutants from known information about health effects.
- For example, we know that the moderate level of health effects for some kinds of PMs ($PM_{2.5}$) ranges between 12.1 mg/m^3 to 35.4 mg/m^3.
- The moderate level of AQI for any pollutant ranges from 51 to 100, so 12.1 mg/m^3 corresponds to an AQI value of 51 for PM, and 35.4 mg/m^3 corresponds to an AQI value of 100.



Credit: EPA wikimedia commons (public domain)

Air Quality Index AQI

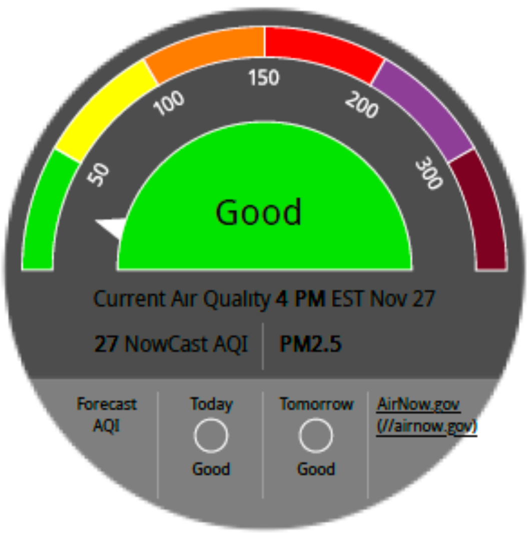
- For mg/m^3 values between 12.1 and 35.4, the AQI values are obtained by interpolation between 51 and 100.
- The same kind of thing works for other pollutants.
- In other words, scales are set up so that a given AQI for ozone is in the same "place" relative to health effects as the same AQI for CO and for ozone.



Credit: EPA wikimedia commons (public domain)

Air Quality Index AQI

- Volume defines a ratio scale.
- Changing volume measurement from cubic meter to cubic feet or cubic meter to cubic Kilometer would not result in changes of the mapping between concentration and AQI values, whether at the boundaries or in between via interpolation.
- ***So, the conclusion that the overall air quality AQI was better today than yesterday seems to be meaningful.***
- However, is AQI useful for decision making?



Credit: AirNow

Air Quality Index AQI

- The Environmental Protection Agency offers guidance.
- For example, one can ask: If the AQI forecast for tomorrow is 120, should I go out to exercise tomorrow?
- For ozone, consider a score of 101-150, which is unhealthy for "Sensitive Groups."
- ***Sensitive Groups*** "include people with lung disease such as asthma, older adults, children and teenagers, and people who are active outdoors."
- **<u>For Sensitive groups</u>**: *"Make outdoor activities shorter and less intense. Take more breaks. Watch for symptoms such as coughing or shortness of breath. Plan outdoor activities in the morning when ozone is lower."*
- **<u>For people with asthma</u>**: *"Follow your asthma action plan and keep quick-relief medicine handy."*
- **<u>For "Everyone else:</u>** *"Consider making outdoor activities shorter and less intense."*

Air Quality Index AQI

- So, simply giving one of six categories for a given pollutant, which are ranked on an ordinal scale, is both meaningful and useful.
- ***Note that there is a difference in how to use the air pollution scores depending on the person using them.***
- For example, a person with asthma as opposed to a healthy, young adult.
- The recommendations seem legitimate for the same reason that recommendations using weight-based index *A* and Pindex were.

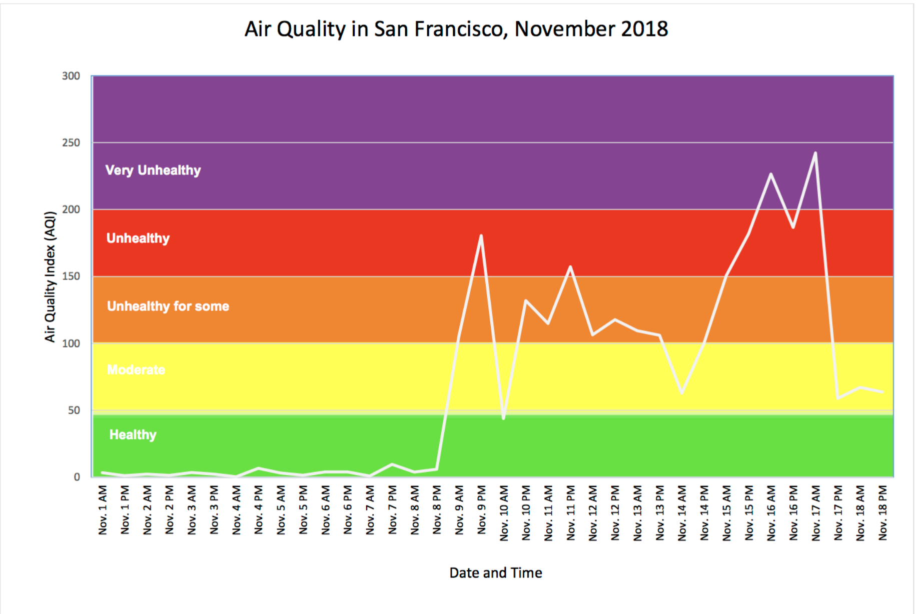


Figure 1. Air Quality in San Francisco, California for the month of November 2018. Based on data from Purple Air sensors. Morrison, L, et al., 2018. *The BMJ*.

credit: Lily Morrison University of California San Francisco, Wikimedia Commons (no changes)

Giving Score of Worst Pollutant

- The AQI reports overall air quality class (from green to maroon) as that of the pollutant that has the highest AQI.
- ***Basing overall air quality only on the pollutant with the highest AQI can lead to problems.***
- Compare two policies:
  - One that is expected to produce AQI scores for the five pollutants of interest PM, CO, SO_2, NO_2, and ozone O_3 of **(25, 25, 301, 25, 25)**
  - One to produce AQI scores of **(300, 300, 300, 300, 300)**
- The worst score of the first puts this in the "Hazardous" (maroon) category, with other pollutants being in the "good" category.
- The worst in the second puts this at the high point of the "Very Unhealthy" (purple) category for all pollutants. Isn't this much worse than the first?
- ***To say that the air is worse in the first case than in the second might be meaningful, but useless (in the sense of being misleading).***
- But it is presumably legitimate.

AQI Comparisons

- *Is it meaningful to say that the overall AQI today is 20% higher than it was yesterday, or twice as high as yesterday?*
- It is since volume is measured on a ratio scale. But is this useful?
- If AQI was 50 yesterday and it doubles to 100 today, the air goes from good to moderate.
- But if AQI was 100 yesterday and it doubles to 200, the air goes from moderate to unhealthy.
- ***So, the doubling conclusion has different interpretations for different levels, and this conclusion seems to be useless.***

AQI Comparisons

Smog in Taiwan

- Taiwan uses the U.S. EPA version of the AQI.
- Consider air pollution in the Kaoping region of Taiwan.
- Cheng, et al. (2004) concluded that between 1997 and 2001, the *average annual AQI declined from 68.5 to 62.0.*
- This is a meaningful conclusion: If we change the volume scale, the AQI is unchanged, and so the average is also unchanged.
- *It is meaningful to say that the average AQI in one year is less than it was in an earlier year.*
- *It is even meaningful to say that it is 20% less.*
- But, just as with the conclusion about doubling of AQI, or decrease of AQI by 20% being useless, so is the conclusion that the average has decreased by 20%.
- ***Great care needs to be taken when using index numbers to justify policy changes.***

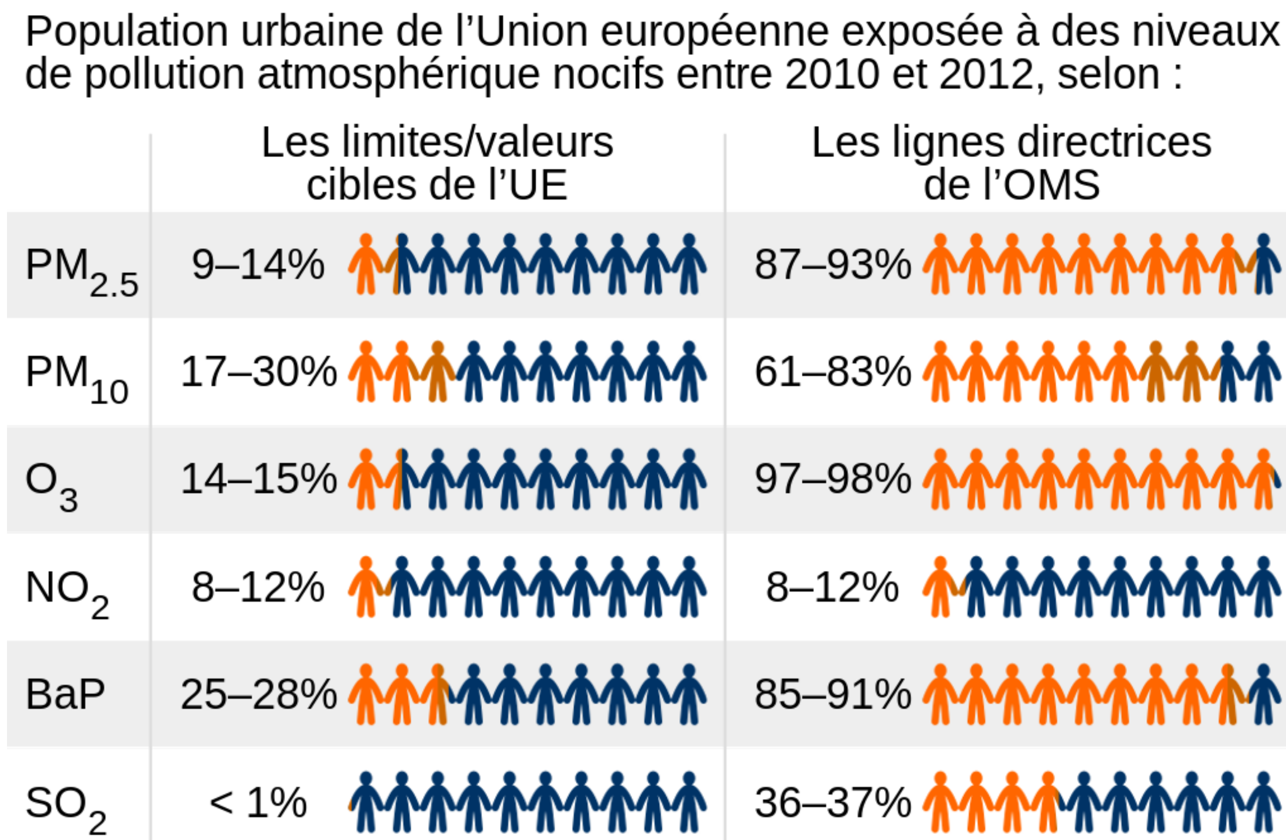
credit: Johntarantino1, Wikimedia Commons (no changes)

AQI Legitimacy

- Variants of the AQI are in use around the world.
- The World Health Organization (WHO, l'OMS) has published air quality guidelines for Europe since 1987, using a variant of AQI.
- Per WHO: AQI is intended to be used worldwide, with the following proviso: "Air quality standards are an important instrument of risk management and environmental policy, and should be set by each country to protect the health of its citizens. ***The standards set in each country will vary according to specific approaches to balancing risks to health, technological feasibility, economic considerations and other political and social factors.*** This variability will depend on the country's level of development, capability in air quality management and other factors. The guidelines recommended by WHO acknowledge this heterogeneity and recognize in particular that, in formulating policy targets, governments should consider their own local circumstances carefully before using the guidelines directly as legal standards."

AQI Legitimacy

- ***This suggests that different conclusions using AQI might be legitimate in one country and not in another.***
- Cheng, et al. (2004) make a similar point.
- They observe that the AQI is used by a variety of countries, but there are "differences in standard concentrations, average times, calculations, and statistical analysis"



credit: European Environmental Agency, The RedBurn Wikimedia Commons (no changes)

AQI: Ambiguity and Eclipsicity

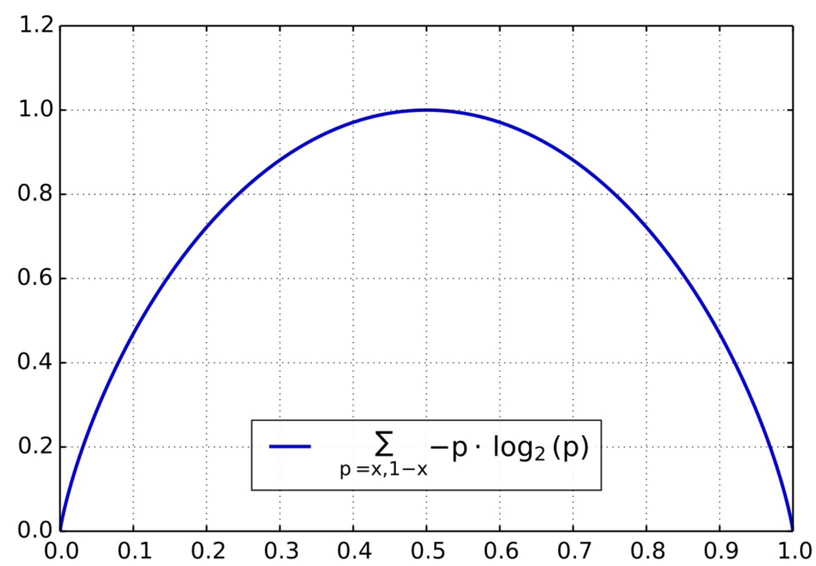
- How can we minimize ***ambiguity*** of conclusions from air pollution indices, *situations when an index reports air to be highly polluted when it is not?*
- How can we minimize ***eclipsicity*** of such conclusions, *situations when highly polluted air is reported as less so?*
- Both potentially render conclusions from air pollution measurement useless.
- ***Developing indices for level of ambiguity and eclipsicity would provide a way to determine the degree to which indices of air pollution are useful, and would also help in determining ways to minimize ambiguity and/or eclipsicity.***

AQI: Ambiguity and Eclipsicity

- Consider again the vector of AQI scores for the five pollutants PM, CO, SO_2, NO_2, and ozone O_3 and compare the two cases **(100, 100, 100, 100, 100)** and **(10, 10, 10, 10, 100)**
- ***Both cases would give an overall AQI of 100, but the air in the latter case is surely much better since it is only ozone that is at that level and all the other pollutants have very low levels.***
- In one approach to reduce ambiguity and eclipsicity, Cheng, et al. (2004) propose a correction to AQI producing a revised index RAQI.
- Among other things, the correction multiplies the AQI as measured by the maximum of the AQIs of the five pollutants by a factor involving the average value of the AQI of the five pollutants and by a second factor involving the Shannon entropy.

AQI: Ambiguity and Eclipsicity

- Averaging the five AQI values would definitely distinguish between the two cases of (100, 100, 100, 100, 100) and (10, 10, 10, 10, 100).
- ***Using the Shannon entropy is intended to reduce the overall AQI score when there is a varying distribution of AQI values over the different pollutants.***
- The Shannon entropy, invented for information theoretic applications, quantifies (in expected value) the information contained in a message (in this case a message about air pollution levels) in units such as bits.



credit: Geek3 Wikimedia Commons (no changes)

Air Stress Index

- In contrast to AQI, an ***air stress index*** takes an annual perspective.
- One air stress index considers the number of times C_i in a year that the concentration of a given pollutant i in the air exceeds some standard for that pollutant.
- It compares that to a reference value R_i giving the number of times a year that is permitted in some directive or guideline.
- This air stress index is the average of the ratio of C_i to R_i:

$$ASI = \sum_{i=1}^{p} \frac{C_i}{R_i}$$

Air Stress Index

- Since both C_i and R_i are just counts, we have what is called an ***absolute scale*** (the only admissible transformation is the identity).
- All kinds of statements involving *ASI* are meaningful.
- If *ASI(t)* is *ASI* for year *t*, then for example it is meaningful to say that *ASI* increased by 20% year over year:

$$ASI(t) = 1.2*ASI(t-1)$$

- *This would be a wakeup call about air pollution and would suggest that some mitigations be put into effect, so it is useful in that sense.*

Air Stress Index

- To determine what mitigations might be needed would require drilling down and finding out which ratio C_i/R_i went up significantly year over year.
- ***Note that ASI is useful for policy. It is not intended to be used to provide real-time advice on short-term health effects as AQI is.***

credit: National Institute of Health, Wikimedia Commons (public domain)

Air Stress Index

- We might want to set the goal of reducing C_i, for example the number of days for which levels of a given pollutant, e.g., ozone, are in bad categories, e.g., Orange or worse.
- How can we tell if a given policy change has achieved a given reduction?
- *Consider the date by which the number of such days exceeds 100 for the first time.*
- We would like this day to be later in the year.
- Suppose in Year *t-1* it is June 30, and in Year *t* it is July 19.
- *This is a 10% improvement from 200 days to 180 days.* But, is this meaningful?

Air Stress Index

- In the U.S., the federal "fiscal year" begins October 1, not January 1. If we use the fiscal year, the improvement is from 292 days to 272 days, about 7%.
- So, the 10% improvement conclusion is meaningless – unless we specify the beginning of the year.
- ***This shows that we need to be careful to specify additional information before drawing conclusions that we can use to make or check policy.***
- If the beginning of the year (the zero point) is specified, then the 10% improvement conclusion is probably not only meaningful but useful as well, at least for some purposes.
- It shows how much progress we are making.

Closing Comments

- *A variety of combinations of meaningful/meaningless, useful/useless, and legitimate/illegitimate can occur:*
  - There are conclusions that are **meaningful, useful and legitimate**. Many of the examples given fall in this category.
    - E.g., use of BMI to determine obesity, at least for some cultures.
    - E.g., use of AQI to make individual decisions about reducing activity under certain air pollution conditions.
  - There are conclusions that are **meaningful, useful, and illegitimate.**
    - E.g., use of BMI to determine obesity that is illegitimate for some cultures.
  - There are conclusions that are **meaningful, useless, and legitimate.**
    - E.g., use of the weight-based air pollution index *A* or the variant Pindex or AQI to make policy decisions in certain cases.

Closing Comments

- There are conclusions that are **meaningful, useless, and illegitimate.**
  - E.g., comparisons of %BF when the latter is measured using skinfold thickness measurements, underwater weight measurement, bioelectrical impedance, or other methods that are expensive to obtain, difficult to measure, and difficult to standardize.
- **We doubt that meaningless conclusions can be useful. However, they can be legitimate** if the use of an index doesn’t violate social, cultural, or regulatory norms

Closing Comments

- It may be difficult to formalize usefulness or legitimacy but it may be useful to develop indices of degree of usefulness or even of legitimacy.
- Since there are different components of usefulness or legitimacy, the indices may need to be multi-dimensional.
  –

Questions?

Fred Roberts
froberts@dimacs.rutgers.edu

Command, Control, and Interoperability
Center for Advanced Data Analysis