University of Waterloo

W. H. Cherry

MEASURING: *Shrinking kilogram sows mass confusion in labs*

The article EM0303 reprinted below is a reminder of the fundamental importance of *standards* to the concept of *measuring.*

- ∗ The article EM9810 about *self-esteem*, reprinted on pages HL42.2 and HL42.3, provides a contrast to EM0303 by illustrating measurement issues that can arise in the *absence* of standards; such issues are often of concern with measuring in the human sciences.
- ∗ The article EM0210 *New scoring system would have iced the gold for the Canadians* about judging figure skating, reprinted on pages HL42.4 and HL42.5, is another instance of measurement issues arising in the absence of standards.
- ∗ The article EM0208 *Better off red* about the touted health benefits of drinking red wine, reprinted on pages HL42.5 and HL42.6, involves overlapping (statistical) issues of causation and measuring – see also the article EM9411 *The attack on red wine's hearty reputation* on the overleaf side (page HL13.2) of Statistical Highlight #13.

EM0303: The Globe and Mail, May 27, 2003, pages A1, A14

Shrinking kilogram sows mass confusion in labs

Quest for exact kilogram involves counting atoms

BY OTTO POHL
BRAUNSCHWEIG, GERMANY

The kilogram is getting lighter, scientists say, sowing potential confusion over a range of scientific endeavours.

The kilogram is defined by a platinum-iridium cylinder, cast in England in 1889. No one knows why it is shedding weight, at least in comparison with other reference weights, but the change has spurred an international search for a more stable definition.

"It's certainly not helpful to have a standard that keeps changing," said Peter Becker, a scientist at the Federal Standards Laboratory in Braunschweig, an institution of 1,500 scientists dedicated entirely to improving the ability to measure things precisely.

The kilogram's apparent loss of 50 micrograms (less than the weight of a grain of salt) is enough to distort scientific calculations.

Mr. Becker is leading a team of international researchers seeking to redefine the kilogram as a number of atoms of a selected element. Other scientists are developing a competing technique to define the weight using a complex mechanism known as the watt balance.

The final recommendation will be made by the International Committee on Weights and Measures, created by international treaty in 1875. The agency keeps the international reference kilogram in a heavily guarded safe in a château outside Paris.

It is visited once a year, under strict security, by the only three people to have keys to the safe. The weight change has been noted on the occasions it has been removed for measurement.

The race is well under way to determine a new standard, although at a measured pace, since creating reliable measurements is such painstaking work.

The kilogram is the only one of the seven base units of measurement that still retains its 19th-century definition. Over the years, scientists have redefined units such as the metre (first based on the earth's circumference) and the second (conceived as a fraction of a day). The metre is now the distance light travels in 1/299,792,458th of a second, and a second is the time it takes for a cesium atom to vibrate 9,192,631,770 times. Each can be measured with remarkable precision, and, equally important, can be reproduced anywhere. The kilogram was conceived as the mass of a litre of water, but measuring that accurately proved very difficult. Instead, an English goldsmith was hired to make a platinum-iridium cylinder that would be used to define the kilogram.

It is visited once a year, under strict security, by the only three people to have keys to the safe.

..... creating reliable measurements is such painstaking work.

A total of 80 copies of the reference kilogram have been created and distributed to signatories of the metric treaty. The sometime colourful history of these small metal cylinders underscores how long the world has used the same definition of the kilogram.

Some of the metal plugs were issued to countries that later vanished, such as Serbia and the Dutch East Indies. The Japanese had to surrender theirs after the Second World War. Germany has acquired several, including the one issued to Bavaria in 1889 and the one that belonged to East Germany.

To update the kilogram, Germany is working with scientists from countries including Australia, Italy and Japan to produce a perfectly round one-kilogram silicon crystal. The idea is that by knowing exactly what atoms are in the crystal, how far apart they are and the size of the ball, the number of atoms in the ball can be calculated. That number then becomes the definition of a kilogram.

To separate the three isotopes of silicon, Mr. Becker and his team are turning to old nuclear-weapons factories from the former Soviet Union, where centrifuges once used to make highly enriched uranium are able to produce the required purity of silicon.

"We need so many nines," Mr. Becker said, and the Soviet uranium processors are one of the only places to get them. "With the Russians, we're getting about four of them" – 99.99 per cent pure silicon 28.

A test crystal has been produced, and Dr. Arnold Nicolaus, another scientist at the German standards laboratory, is responsible for measuring whether it is perfectly round. He has measured the crystal in a half-million places to determine its shape.

It's probably the roundest item ever made by hand. "If the earth were this round, Mount Everest would be four metres tall," Dr. Nicolaus said.

An intriguing characteristic of this smooth ball is that there is no way to tell whether it is spinning or at rest. Only if a grain of dust lands on the surface is there something for the eye to track.

But scientists from the United States, England, France and Switzerland say the challenge of calculating the exact number of atoms in a silicon crystal is too imprecise with today's technology, so they are refining a technique to calculate the kilogram using voltage.

"Measuring energy is easier than counting atoms," said Dr. Richard Steiner, a scientist at the National Institute of Standards and Technology in Gaithersburg, Md.

New York Times Service

(*continued overleaf*)

EM9810: The Globe and Mail, May 16, 1998, page D9

PSYCHOLOGY Self-esteem loses its lustre

BY KIRK JOHNSON
The New York Times

SELF-ESTEEM had it all, once. Common sense and research showed that people who did best in life felt good about themselves, and it seemed a short leap to conclude that the reverse must also be true: If successful people enjoyed high self-esteem, then high self-esteem would foster success.

Students would do better if they had more self-esteem. Employees would work harder. Girls who felt inferior because of poor body image or math phobia would gain from self-esteem training.

By 1986, when California created a commission to bolster self-image as a statewide goal, the concept had become a popculture phenomenon. Celebrated in the media, in politics and in schools, self-esteem had become an end in itself – a commodity, like expanded memory for a personal computer, that could be installed in a do-it-yourself upgrade.

But self-esteem is having image problems these days.

Research is indicating that it is not in and of itself a strong predictor of success. Criminals and juvenile delinquents, it turns out, often have high self-esteem. New movements in education have blamed the emphasis on self-esteem for students' failures in learning.

"Self-esteem became mixed up in a whole series of issues, and people wanted it over," says Arthur Levine, president of Teachers College at New York's Columbia University.

But self-esteem is by no means dead, researchers say. The deep tradition of individualism guarantees that esteem or one of its many variants – from positive thinking to self-efficacy – will persist. What has changed is that self-esteem as an idea and a societal force has lost its unified champions and to a great degree its ability to be succinctly defined.

Like IQ tests and SAT scores, self-esteem has become but one gauge to indicate success in college or life.

"A dozen years ago, research was showing heavily positive things about high self-esteem," says Roy Baumeister, a psychology professor at Cleveland's Case Western Reserve University. "Since then, questions have been raised about the size of effects, the direction of effects and whether in fact it's a mixed blessing to even have high self-esteem."

Even before William James, the Harvard professor, philosopher and psychologist, invented the term self-esteem in the late 1800s, people have worshipped the concept of self. The will to do, to achieve, to improve, was stained into the culture along with its icons – Horatio Alger, Dale Carnegie, rags to riches.

There were other ideas of psychology, such as alienation, with its darker, more European nuances. But it was bright-eyed, optimistic self-esteem that caught on.

It was not until the 1960s, when new analytic tools were developed to measure self-esteem, particularly the Rosenberg scale, developed by Morris Rosenberg at the University of Maryland, that the idea really took root. The scale, based on 10 questions, created what every academic craves – numerical measurability.

At the same time, some segments in society, especially the feminist movement, seized on the idea that low self-esteem among many girls and women could and should be raised through training.

But the zenith for self-esteem, often identified as the California project, was probably also the beginning of its undoing.

Confronted with the fact that self-esteem had become a goal for public schools and society, rather than a result of achievement, researchers realized they had no proof that the path of logic really worked that way. If you improved a person's self-image, did that translate into better behaviour? By contrast, did life's losers really have poor self-esteem?

"The idea of self-esteem is so ingrained in our culture, it's presumed to be a real thing inside the human condition," says Timothy Owens, an associate professor of sociology at Indiana University. "But it got overblown on both ends – society's and the academy's."

In psychology, the idea has gained ground that there is no coherent self at all as people generally think of it, but rather a series of selves, like mirrors that reflect different aspects of an individual's connection to the world.

"The critical notion of the unified self whose levels can be fixed – that idea has ended," says Kenneth Gergen, a psychology professor at Swarthmore College in Pennsylvania.

At the same time, the idea that high self-esteem is the exclusive province of those with admirable achievements has been rejected as simply wrong. Studies of gang members and criminals found their self-esteem – reinforced by peers or lawlessness – to be as high as that of any overachiever.

Another study crushes the idea that welfare women become pregnant to boost their self-esteem. Other studies found distortions in how self-esteem statistics have been gathered.

People, researchers conclude, are not the simple calculators of worth

Gauging Self-Esteem

The Rosenberg Self-Esteem Scale is based on 10 questions. Respondents are asked to **strongly agree, agree, disagree** or **strongly disagree** with these items.

1. On the whole, I am satisfied with myself.
2. At times I think I am no good at all.
3. I feel that I have a number of good qualities.
4. I am able to do things as well as most other people.
5. I feel I do not have much to be proud of.
6. I certainly feel useless at times.
7. I feel that I am a person of worth, at least the equal of others.
8. I wish I could have more respect for myself.
9. All in all, I am inclined to feel that I am a failure.
10. I take a positive attitude toward myself.

Half the questions are phrased positively and half negatively. For the **positively** phrased questions – numbers **1, 3, 4, 7 and 10** – score as follows: Strongly agree, 4 points; agree, 3; disagree, 2; strongly disagree, 1. For **negative** questions – numbers **2, 5, 6, 8 and 9** – reverse the scoring so that strongly agree is worth 1 point, and so on. The maximum is 40 points, the minimum 10.

Most people in the general population score in the 30-to-40 range. A much smaller number are in the 20s. A score of 10 to 20 is often associated with clinical depression, according to Timothy Owens of Indiana University.

Sources: Conceiving the Self *by Morris Rosenberg (Basic Books, 1979); Timothy Owens.*

University of Waterloo W. H. Cherry

MEASURING: *Self-esteem loses its lustre* (continued 1)

that the Rosenberg scale might indicate. Some people achieve great things in life while gripped by feelings of worthlessness. Others set low standards and feel great.

In the end, says Albert Bandura, a psychology professor at Stanford University, in his 1997 book *Self Efficacy*: *The Exercise of Control*, "self-esteem affects neither personal goals nor performance."

Self-esteem became a cornerstone of the feminist critique of society beginning in the 1960s. But many feminist thinkers now say that trying to raise the self-esteem of girls and women may be self-defeating.

"Early feminism went in the direction of saying, 'Give every girl assertiveness training so she can get in there and compete', and that made some sense at the time", says Judith Jordan, a psychologist at Harvard Medical School and director of training at the Stone Center for women's studies at Wellesley College. "Now we say, 'Why are we accepting that as the norm, why not change the norm?'"

Judith Butler, a professor of rhetoric and comparative literature at the University of California at Berkeley, adds: "If you take an 11-year-old girl who's got failing self-esteem because her friends have name clothes or fancy haircuts, the real issue is to get that 11-year-old girl to think critically about how femininity is defined in the culture."

One of the great paradoxes of self-esteem is that, in politics at least, conservatives are the people who came to oppose it. An idea that had, in earlier incarnations, been draped with all the trappings of bootstrapping individual self-improvement became, instead, associated with liberal values and political correctness, perhaps because of its surge to prominence in the unruly 1960s.

When education came under attack in the 1990s because of declining achievement, says Columbia's Prof. Levine, conservatives led the charge against "all the things that looked different, whether those things were elective curriculum or whole language, or self-esteem or multiculturalism – all became items to be attacked because they represented departures from the tried and true of years past."

The movement in education toward higher standards – visible everywhere from New York City public schools to efforts in Washington to create U.S. education standards – has shifted the emphasis toward achievement. The new view looks toward the world, rather than the self, and toward results rather than motivation.

Richard Elmore, a professor of education at Harvard who has been working with the New York City public school system, says self-esteem – originally posited as a way to help underprivileged students – mutated into a kind of crutch that explained and even reinforced low achievement and low performance.

"For most teachers, self-esteem is a theory they invent to cover the fact that they have low expectations for kids", he says. By contrast, the premise of an experimental program in a district that includes much of Manhattan's poor Lower East Side, is to train teachers to "accept no excuses" and to expect the same level of achievement required of any middle-class student.

Deanna Burney, a former principal who helped put the Lower East Side program together, says: "We've led a lot people astray about self-esteem – children feel good about themselves when they can read and write."

1 A major statistical theme of the article EM9810 reprinted on the facing page HL42.2 and above is *measuring* – in this instance, a human characeristic (*i.e.*, a *response variate*) described as *self-esteem*. A measuring *instrument* is the *questionnaire* given at the lower right of the facing page HL42.2; what (at least formally) are the *other* three components of the measuring process under discussion?

- Which two components of this measuring process are likely to be the *same* entity in practice? Discuss briefly how this identity might affect the performance of the measuring process.

2 Discuss critically the questionnaire, given at the lower right on the facing page HL42.2, from the perspective of obtaining *truthful* responses.

- To which category of error would *un*truthful responses give rise?
  - Outline whether, under conceptual repetition, untruthful responses would be likely to lead to measuring *inaccuracy* or measuring *imprecision* or both.

3 Compare and contrast the statistical issues raised by measuring processes based on questioning humans to quantify level of:

- a psychological characteristic (such as self-esteem);
- knowledge of a subject area (such as science or geography) – see also Statistical Highlights #93 and #36;
- a particular class of activities (such as leisure pursuits or sexual behaviour);
- consumer spending on:
  - durable household items;
  - alcohol and tobacco;
  - food;
- an unacceptable behaviour (such as illegal drug use, theft or cheating on tests).

4 *Causation* is another statistical issue in the article EM9810 reprinted on the facing page HL42.2 and above (as well as in the article EM0208 reprinted on pages HL42.5 and HL42.6); outline how this issue arises in the article EM9810.

- Which (two) paragraphs of the article EM9810 best capture the causation issue? Explain briefly.

The article EM9810 reprinted on the facing page HL42.2 and above is also used in Figure 8.8e of the STAT 220 Course Materials and Figure 3.5e of the STAT 332 Course Materials; the other three articles EM0303, EM0210 and EM0208 reprinted in this Highlight #42 are used in the STAT 231 Course Materials in, respectively, Chapters 6, 6 and 11.

For Question 5 on page HL42.5: ∗ **Estimating:** a process which uses statistical theory to derive the distribution of an *estimator* and data to calculate an (interval) *estimate*.

(*continued overleaf*)

EM0210: **The Globe and Mail, February 16, 2002, page F6**

New scoring system would have iced the gold for the Canadians

It's the eternal question when it comes to skating competitions: Can the bias of judges be detected and eliminated? **STEPHEN STRAUSS** finds that something called the Rasch measurement may do the trick

While much of the television viewing world gagged when Jamie Salé and David Pelletier were judged Olympic silver but not gold medal skaters this week, a small subset of statisticians was probably giving a silent yippee.

Another scoring controversy means another batch of numerical papers attempting yet again to resolve figure skating's eternal question: How do we deal with that rogue Russian judge?

This is not exaggeration: In the early 1990s, an entire session of an American Statistical Association's annual meeting was devoted to ice-skating judging.

The "Russian judge" in this construct is not literally a Russian, but simply any judge whose impartiality is overwhelmed by one or another non-skating factors. In the era of the Iron Curtain, it was widely assumed by Western skaters that Eastern European judges had colluded to fix the events for Russian or other Communist country skaters.

So the question becomes: Can one make a scoring system so refined that it catches vote-rigging and compensates for it? The answer appears to be both an unqualified "yes" and a qualified "maybe not."

The unqualified "yes" is as follows. Computerized statistical techniques that are specifically designed to uncover and compensate for judging prejudice already exist. And some have already been applied in a *post facto* sense to figure skating.

For instance, in 1997, Marilyn Looney, who is in the department of physical education at Northern Illinois University, took the highly disputed results of the 1994 women's Olympic skating final and subjected them to what is known as the Rasch measurement.

This statistical weighting formula, first proposed by Danish statistician Georg Rasch in the 1960s, has achieved great currency in educational testing, in which it is used to smooth out the effects of hard and easy graders on medical or dental certification exams.

What the program tries to do is break down the judging process into discrete elements that take into consideration the judges' idiosyncrasies and the program's technical difficulty. Maverick ratings can be flagged and their influence on the final scores calculated.

Looney reported that in a Rasch-adjusted universe second-place winner American Nancy Kerrigan should have beaten out Ukrainian Oksana Baiul for the gold medal.

What the program tries to do is break down the judging process into discrete elements that take into consideration the judges' idiosyncrasies and the program's technical difficulty. Maverick ratings can be flagged and their influence on the final scores calculated.

... the confidence you have in any adjustment is a function of the confidence you have that you have truly measured judge prejudice in the first place. Differences in educational judging severity might make use of hundreds of thousands of scoring patterns.

Four judgements – two each for the short and long programs – on 27 skaters may simply not be a large enough sample size for a computer to make readjustments free from its own statistical biases.

In answer to the question of what a Rasch-type analysis would have meant for the Salt Lake City Olympics, Mike Linacre, a U.S. statistician who was Looney's consultant for the earlier work, has done a simple computation of the "prejudice pattern" at work in the judging.

Taking into consideration the average score for each program, each task, each judge and each competitor, Linacre produced for *The Globe and Mail* a statistical picture of bias. Judges from four countries – Russia, France, Ukraine and Poland – were overwhelmingly lenient on Russian skaters and just as overwhelmingly tough on Canadian skaters.

"You can see there is a block vote by four judges and it is also surprising to see that the Polish and the Ukrainian judges were more lenient on the Russians than on their own skaters," Linacre says.

What the initial pass at the Rasch program does not tell you is why this pattern occurs.

In 1996, Montreal statisticians Bryan Campbell of Concordia University and John Galbraith of McGill reported that their investigation of Olympic figure skating stretching from 1976 to 1994 indicated that all judges tended to favour skaters from their own country, but that nationalism was inversely related to the ability of the skaters being judged.

If a skater was one of the top six in his or her class in the world, that excellence twigged national prejudice. However, if "you didn't have a shot at any of the medals, the judges appeared to be unbiased," Galbraith said.

In spite of all these caveats, the question remains: Could better number crunching have caught and corrected the prejudice pattern that Linacre turned up? According to at least some statisticians, if the Salt Lake City scores had been sent through a computerized Rasch prejudice filter, the newest Olympic figure-skating fiasco might never have occurred.

"The outcome could have been adjusted for judges' severity in real time after all the scores were in. We do it all the time in educational testing," says Mary Lunz of Measurement Research Associates, a Chicago company that uses the Rasch approach in medical certifications.

The result would have been a process in

(*continued*)

MEASURING: *Better off red?* (continued 2)

which even the judges who judged could not have known for sure who had won until after their bias was corrected for.

And what would a program that adjusted for human error – either intended or accidental – have produced in medal rankings? "Thought you would ask this. Yes, a Canadian victory," says Linacre, who is working for a company that is developing Rasch software.

But maybe not, other statisticians say.

McGill's Galbraith points out that the confidence you have in any adjustment is a function of the confidence you have that you have truly measured judge prejudice in the first place. Differences in educational judging severity might make use of hundreds or thousands of scoring patterns.

Four judgments – two each for the short and long programs – on 27 skaters may simply not be a large enough sample size for a computer to make readjustments free from its own statistical biases.

All of this ignores the fact that massive cheating simply skews everything. "If all the countries put in the fix, I don't know of a model that can pick that up", Looney says.

Finally, while a more sophisticated statistical approach might make it harder to cheat, it may not resolve what has been put forward as the fundamental judging problem in skating. Scoring figure skating, the old saying goes, is like asking someone after seeing a ballet: Who won?

This is something that University of Chicago economist Gilbert Bassett has thought hard about. In 1994, in conjunction with colleague Joseph Persky, he did an analysis of the present International Skating Union scoring system. Their examination was inspired by what Persky thought was nonsensical statistical flaws in an ice-skating competition in which his daughter had participated.

"We initially thought, 'What a crazy system they have in place'. But after looking hard at it, we arrived at a completely different take on the matter," Bassett says.

They concluded the present system, in which, no matter how many low rankings he received, the skater who gets the majority of the first-place votes wins, is less prone to bias manipulation than scoring systems like the one used in gymnastics that knock out the top and bottom scores, or other ones that take into consideration total marks.

"Although we find no historical evidence that skating officials ever had this in mind, they have picked a system particularly well-suited to ..... serve as a method of statistical estimation", the U.S. pair wrote at the time.

As a fallout of their initial work, Bassett subsequently was part of a group of statisticians who had figure-skating judges explain to them what the judging judged. What emerged was a sort of culture of incomprehension. The statisticians easily understood how to interpret the numbers, but they didn't at all understand how the numbers were arrived at in the first place.

"Figure skating seems to me to be unjudgable, but these guys came in and talked as if they were Inuit and they had five different words for snow. As if there was objective material that anyone with reasonable abilities could say that was the guy who won, he says with amazement attaching to every word.

"To me, beauty is in the eye of the beholder."

5 Discuss critically the statement in the second paragraph of the right-hand column above: *... they have picked a system particularly well-suited to ..... serve as a method of statistical estimation,"* bearing in mind *our* two definitions (*e.g.*, from page HL6.2 of Statistical Highlight #6 or pages HL38.1 and HL38.2 of Statistical Highlight #38):

∗ **Measuring:** the process used to determine the value of a variate. (See the bottom of page HL42.3 for 'Estimating'.)

EM0208: The Globe and Mail, February 12, 2002, page R5

Better off red?

In moderation, alcohol can be good for your heart, but what about red wine's anticancer claims? **ANDRÉ PICARD** uncorks the truth about the wine's alleged full-bodied benefits

Louise Larouche loves a fine wine. She can take up to an hour breathing in the ever-changing bouquet before taking a sip, and spends countless hours studying vintages.

"For me, every glass of wine is a journey, a discovery. It's pure pleasure", she said.

As president of the Disciples of Bacchus, an exclusive Quebec City wine-lovers' club, she can indulge herself, and her hobby, in rarefied company.

But Ms. Larouche says the intellectual stimulation she gets from wine drinking is matched by the physical benefits.

Years ago, she had dangerously high levels of cholesterol in her blood. From her reading, she knew that drinking wine was good for the heart.

Ms. Larouche started drinking two glasses of wine daily. Her low-density lipoproteins (LDL), the bad cholesterol, fell. Her high-density lipoproteins (HDL), the good cholesterol, rose.

"I'm the living proof that wine is good for you", Ms. Larouche said.

That alcohol, consumed in moderation, is good for the heart is no longer in dispute.

Whether it comes from beer, wine or spirits, alcohol raises HDL. It is also an anticoagulant. That means that a drinker's arteries are less likely to clog, and a drinker's blood less likely to create clots, the principal causes of heart attacks and strokes.

But wine, and red wine in particular, is accorded all sorts of other attributes, many of which are in dispute.

Wine is said to delay the onset of Alzheimer's, prevent ulcers, provide some protection from cancer and even extend lifespan.

But all these touted benefits come from research that demonstrates associations, not cause and effect. As a result, the experts are cautious about their recommendations.

"We're careful to not advocate drinking, because the evidence isn't there to do so", said Andreas Wielgosz, a spokesman for the Heart and Stroke Foundation of Canada.

"But, if you do drink, do so in moderation and it certainly won't do any harm".

Not exactly a ringing endorsement. But, in recent years, wine consumption has increased dramatically, at least in part due to health claims.

Dr. Wielgosz, who is also a professor of medicine and community health at the University of Ottawa, said that, personally, he will drink the occasional glass of wine, but certainly not on a daily basis.

What he does do religiously is eat five to 10 servings of fruits and vegetables a day, and exercise regularly.

(*continued overleaf*)

"Eating a diet rich in fresh fruits and vegetables is the most universal message we have. The evidence is much stronger than for wine", Dr. Wielgosz said.

But there is no doubt that, given a choice between a lovely cabernet sauvignon and an extra helping of broccoli, many people will opt for the former.

The enthusiasm for wine as a heart-healthy drink has its roots in the so-called French paradox. Years ago, Dr. Serge Renaud of the Paris-based research institute INSERM, published eye-popping research showing that, despite similar fat intake, the rate of heart attacks in France was about one-third of what it was in the U.S. To explain the difference, he identified, as a key factor, the French custom of drinking wine with meals.

Subsequently, however, other researchers noted that a more significant distinction may be the fact that the so-called Mediterranean diet favoured by the French features many more fresh fruits and vegetables, more dark fish (rich in omega-3), and more olive oil than the typical North American diet.

The leisurely pace at which they dine is also indicative of a far less stressful lifestyle, a major contributor to heart disease.

Still, much of the debate about the health benefits of wine have been at the biochemical level. There are a number of elements in wine that have garnered scientific attention.

One is polyphenols, which are found largely in the skin of red grapes. Researchers have found that the more polyphenols in a drink – and red wine has the most – the lower the level of endothelin-1 in the blood. Endothelin-1 is a substance that causes arteries to constrict and, when too much is produced by the body, it can cause arteries to clog up and harden.

Polyphenols also have an anti-oxidant effect. (There are, in the body, free radicals, oxygen atoms that damage cells; anti-oxidants, like vitamin E, prevent that cell damage.)

This has led some to claim that wine has an anti-aging quality. There is even a popular line of French beauty products that are made with wine.

The evidence that wine extends lifespan, however, comes from a Danish study. Researchers there followed 13,000 people for more than 12 years, and found that moderate drinkers were half as likely to die as non-drinkers. That research, however, looked at alcohol consumption, and did not single out wine as particularly beneficial.

Another important element in wine is bioflavonoids, which also have anti-oxidant qualities. Bioflavonoids are found in grapes, but in even more significant quantities in berries, particularly strawberries.

"There's much less risk in over-consuming strawberries than wine", Dr. Wielgosz noted wryly.

Quercitin is yet another compound in red wine that has been singled out. It helps dilate blood vessels, and prevent blood clots.

But the element that has garnered the most attention is probably resveratrol.

Red grapes produce resveratrol to protect themselves from fungus, but there is evidence that it can provide some protection against cancer. In laboratory experiments, it has inhibited the development of cancer in animals.

But all these touted benefits come from research that demonstrates associations, not cause and effect. As a result, the experts are cautious about their recommendations.

In people, the most compelling evidence that wine protects against cancer comes from the Harvard Nurses Study. In that long-term research, women who drank moderately (1½ drinks a day) had a lower incidence of cancer overall, and of breast cancer in particular.

Resveratrol is also an anti-oxidant, and helps decrease the stickiness of blood that can cause clotting. In animal studies, it has been shown to inhibit inflammation, which also contributes to heart disease.

Dr. Haydn Pritchard, a biochemist and professor of pathology at the University of British Columbia, said, despite the vast number of claims about the benefits of wine, a "cautious approach is necessary because the evidence really isn't that strong".

Another huge point of contention is whether the benefits attributed to wine are due to the drink itself, or to the lifestyle of wine drinkers.

An expert in lipids, he said the evidence that moderate alcohol consumption improves HDL is good. Still, he noted that moderate exercise also boosts HDL as much, if not more, than a stiff drink.

But Dr. Pritchard, who likes the occasional beer, said "there is no substantive evidence that there is something special about red wine".

"This is not, as some people would have us believe, Drano for the veins".

He said that, in recent years, knowledge about anti-oxidants has improved greatly, and the "anti-oxidant story is falling apart". For example, it is now becoming clear that consuming vitamin E in tablets confers virtually no anti-oxidant benefit, that those effects come from consuming fruits and vegetables, not just by deriving some compounds.

Dr. Pritchard said that if all the benefits attributed to wine were true, then grape juice and dealcoholized wines would be equally effective. But numerous studies have demonstrated that the improvements in heart health come only to drinkers of real wine, showing it is the alcohol that is the key ingredient.

This can also be demonstrated with an extreme example. Severe alcoholics rarely get heart disease. They also tend to have fabulous cholesterol readings: Their HDL are very high, and LDL very low. And because they often suffer from malnutrition, they have very little clotting, meaning their risk of stroke is almost non-existent.

"But alcoholism is a dangerous place to go", Dr. Pritchard notes, because alcohol, consumed in excess, destroys the liver, killing more efficiently than heart disease. Excess alcohol can also harm the heart's pumping function.

Alcohol is also laden with calories. A glass of wine, a beer or a shot of liquor all contain between 100-150 calories. When a person consumes alcohol frequently, they tend to build up abdominal fat (a so-called beer belly) that is quite dangerous for the heart. Alcohol can also block the body's ability to burn fat.

Alcohol can also be dangerous to people with pre-existing heart conditions. Those who suffer from hypertension (high blood pressure) should be very careful, as should patients who have suffered heart failure. Alcohol also interacts with medication, and should not be taken with most drugs.

Another huge point of contention is whether the benefits attributed to wine are due to the drink itself, or to the lifestyle of wine drinkers.

It is well-established that, for a variety of reasons, people with higher incomes have better health outcomes, including less heart disease. It is also clear that there is a preponderance of wine drinkers in these higher income groups.

Even if the more-well-to-do prefer wine, the evidence is that a $10 bottle of wine has as much health benefit as the pricey bottle of Lalande de Pommerol that Ms. Larouche prefers.

But even she, a wine aficionado, believes that the real benefit of wine may be esoteric, that it may come from the relaxed, stress-free atmosphere that a good meal with wine can create.

"There's nothing worse than sharing a $350 bottle of wine in unpleasant company," Ms. Larouche said. "On the other hand, sharing a $10 bottle of wine with good friends is delightful, and wonderful for the heart".