Mr. Harrington, Self-Rated Health and the Canadian Chicken

M. Harrington, l'état de santé auto-déclaré et le poulet canadien

by ROBERT G. EVANS
Professor of Economics
University of British Columbia
Vancouver, BC

Abstract

The simplest way to find out how healthy people are is to ask them. The question: "Would you rate your health as excellent, very good, good, fair or poor?" is quick and easy (and cheap) to administer, and correlates well with more objective indicators of health (such as subsequent death). But there are significant cross-national differences in response patterns. Canadians are much less likely than Americans to provide extreme responses – excellent or poor. International comparisons yield more striking examples. Americans rate their health among the highest in the OECD, despite mortality measures that are among the worst. The Japanese, with the world's best mortality measures, rate their health near the bottom. Can self-reports be standardized for these cultural effects?

Résumé

La façon la plus simple de s'enquérir de l'état de santé de quelqu'un est de le lui demander. La question : « Comment évalueriez-vous votre santé : excellente, très bonne, bonne, passable ou médiocre? » est rapide et facile (et peu dispendieuse) à administrer et corrèle bien avec des indicateurs de santé plus objectifs (comme le décès ultérieur). Mais on observe des différences internationales importantes dans les réponses fournies. Les Canadiens sont beaucoup moins susceptibles que les Américains de fournir des réponses extrêmes – excellente ou médiocre. Des comparaisons internationales livrent des exemples encore plus frappants. Les Américains se considèrent comme étant parmi les plus en santé des pays de l'OCDE, et ce, malgré des mesures de mortalité figurant parmi les pires. Les Japonais, qui ont les meilleurs taux de mortalité au monde, auto-évaluent leur santé comme étant parmi les pires. Les auto-déclarations peuvent-elles être normalisées de manière à tenir compte de ces effets culturels?

———

"If you can't measure something, you can't understand it. If you can't understand it, you can't control it. If you can't control it, you can't improve it." (Harrington and McNellis 2006)

"Why did the Canadian chicken cross the road? To get to the middle."

A BIT OVER 10 YEARS AGO, JACK WILLIAMS MADE A REMARK THAT STAYED with me, and came back when I was looking at OECD data on aggregate measures of country health status. It was at a CHERA (Canadian Health Economics Research Association) meeting, and concerned the North American GUSTO trial (Global Utilization of Streptokinase and Tissue Plasminogen Activator [tPA] for Occluded Coronary Arteries). That trial compared these two alternative "clot-busting" agents and two modes of administration for patients with acute myocardial infarct. tPA was at that time much more expensive, and was accordingly more commonly used in the United States than in Canada. Did it yield better outcomes?

The GUSTO trial indicated that it might. In particular, a substudy of the trial focusing on quality of life post-AMI "… suggested, but did not prove, that the more aggressive pattern of health care in the United States may have been responsible for better quality of life for American patients" (Williams et al. 1995).

But a question remained: "[A]re there societal or cultural differences in self-ratings of health that may account for the differences in quality of life of Canadians and Americans?" (Williams et al. 1995).

The answer seems to be "yes," and that was Jack's point at the CHERA meeting.

Williams and his colleagues have studied the results from all the population-based surveys of self-rated health (SRH), on both sides of the border, since the late 1970s. A clear pattern has emerged. When faced with a five-part scale – excellent, very good, good, fair or poor – Canadians are significantly less likely than Americans to rate their health as either "excellent" or "poor." Just as our national self-image would suggest, Americans are more likely than Canadians to go to extremes.

There is much more to their analysis – examination of subgroups in both populations, standardization for individual characteristics where data are available – but the dominant message remains. As Williams et al. noted, however, the SRH questions in these surveys were not identical. Canadians were asked: "When comparing your health to others' … "; in the United States, that qualifier was absent. The authors' conclusions "would be strengthened if they were confirmed in a study specifically designed to compare responses" (Williams et al. 1995).

They have been. Williams and Agha (2004) subsequently analyzed data from two surveys that did ask identical questions on the two sides of the border, the United States National Health Interview Survey for 1994 and the National Population Health Survey for Canada in 1994/95. The findings, unfortunately never written up for publication, confirmed that Canadians tend to avoid the extreme ends of the health range. (The findings showed a number of other things as well, such as the correlation of SRH status with race/ethnicity in the United States but not in Canada, and the weaker relationship to income in Canada, but our interest here is in cultural influences on SRH.)

> **… Canadians are significantly less likely than Americans to rate their health as either "excellent" or "poor." Just as our national self-image would suggest, Americans are more likely than Canadians to go to extremes.**

In that same year, Sanmartin et al. (2004) published results from the JCUSH (Joint Canada/United States Survey of Health), a survey specifically designed for cross-border comparability. They too found that overall, and in each of three age categories, higher proportions of Americans than of Canadians reported their health to be either excellent or poor (Sanmartin et al.: Table A-1).

One might expect that similar differences, or others, would show up within Canada as well. After all, Canada is a very large country with a diverse population. Might there not also be similar cultural differences in response patterns at provincial boundaries? Williams et al. (1995) did find differences in reporting patterns across provinces, but "the differences between the Canadian and American surveys are more

marked than the variations within Canada. The primary difference between American and Canadian ratings still holds."

It is a bit of a stretch, but one might interpret these findings as evidence for the existence of a country called Canada as something more than a geographical expression or a commercial inconvenience. They may, perhaps, be placed alongside other findings, from very different sources, indicating that the border really does matter.

Michael Adams (1997), for example, compiled the results from a number of population surveys to identify clusters of Canadian attitudes and social values – "tribes." He concluded that these clusters showed a distinct pattern of separation between Americans and Canadians. Despite the undeniable fact that our two populations are probably more similar than those of any other pair of countries one could pick, Canadians and Americans do report, on average, quite significant differences in their values and in the way they view the world.

On a more mundane level, Helliwell (2002) found that despite the common impression that the main trade flows in North America run north–south, the extent of interprovincial trade dwarfs the (nonetheless very large) cross-border flows. From the Montreal Annexation Manifesto (and riots) of 1849 to the vigorous advocacy of "deeper economic integration" by Michael Hart (2007) at the Institute for Research on Public Policy (IRPP) conference on the Canadian Priorities Agenda, spokesmen for commercial interests have called for closer economic union – or just plain union – with the United States.[1] If the border is effectively erased in the interests of commerce, Helliwell's – and Adams's – findings may fade away. But that hasn't happened yet.

These are issues much bigger than differences in SRH. And in fact, a cultural difference in willingness to use the ends of a five-part scale matters very little, if at all, for most purposes. The Canadians who prefer not to claim "excellent" or "poor" health status turn up again as "very good" or "fair." (Unfortunately, few surveys have a category for "not too bad, eh?") So if one aggregates the five-level scale into three, or even two, as many analysts do, the cross-border difference disappears. Chalk it up to American exuberance, or Canadian reticence, and carry on.

But as in everything else, differences that are subtle across our southern border become much larger when one looks farther across the world. Table 1 shows SRH from the countries of the OECD (2006), represented by the proportion of the population surveyed that report their health as "excellent," "very good" or "good." This is matched with two measures of mortality, (crude) life expectancy at birth (both male and female) and potential years of life lost (PYLL) due to premature death.[2] Figures 1

---

1. The rhetoric of regulatory coordination and harmonization can be expressed more succinctly as "Whatever you say, boss."

2. Details of the calculation are provided in OECD (2006).

TABLE 1. Correlation of health status measures

	LE	SRH	PYLL
Australia	80.6	81.9	2,443
Austria	79.3	73.5	2,374
Belgium	78.8	76.9	3,053
Canada	79.9	88.2	2,669
Czech Republic	75.8	62.2	2,911
Denmark	77.6	77.9	3,081
Finland	78.8	67.1	2,385
France	80.3	79.5	2,542
Germany	78.6	66.1	2,513
Greece	79.0		2,200
Hungary	72.8	43.2	4,399
Iceland	81.0	78.7	2,395
Ireland	78.3	86.5	2,788
Italy	79.7	59.8	2,179
Japan	82.1	38.7	1,969
Korea	77.4	45.6	2,716
Luxembourg	78.0	91.0	2,495
Mexico	75.2	67.8	
Netherlands	79.2	77.1	2,772
New Zealand	79.2	89.6	3,108
Norway	79.9	79.9	2,492
Poland	75.0	46.8	3,477
Portugal	77.4	31.3	2,985
Slovak Republic	74.1	34.4	2,628
Spain	80.5	68.3	2,187
Sweden	80.6	72.4	2,141
Switzerland	81.2	85.8	2,301
Turkey	71.2	55.0	
United Kingdom	78.5	74.5	2,762
United States	77.5	88.6	3,719
AVG.	78.3	68.6	2,703
MED.	78.8	74.5	2,542

and 2 plot each country's SRH against the actual mortality experience, and show mean and median values.

Trend lines plotted on each figure show that, in aggregate, the SRH data are correlated with the different measures of mortality. But the relationship is not strong, and there are some remarkable outliers.

Most prominently, the United States and Japan seem to be mirror images of each other. The famously long-lived Japanese have the lowest level of life years lost to premature death; the United States is close to the top. Its PYLL score places it among the former Soviet societies of Eastern Europe, far worse off than Western Europe, Canada or the South Pacific.

Yet when asked to rate their own health, the Japanese take a very dim view. Over half of those surveyed considered their health to be only "fair" or "poor." A half-dozen countries show these dismal ratings, but the others all have below-average mortality statistics as well, and half of them are in Eastern Europe. The Japanese just seem depressed.

The ebullient Americans, on the other hand, cheerfully place themselves right at the

top of the OECD rankings, with almost 90% rating their health as "good," "very good" or "excellent." The United States is among a half-dozen "optimistic" countries – including Canada, Ireland, Luxembourg, New Zealand and Switzerland – in which over 85% of those surveyed report their health as "good" or higher. But in the others, these ratings are consistent with the PYLL scores. In the United States, they are not.

FIGURE 1. Self-rated health (percentage reporting "good" or higher) plotted against potential years of life lost (PYLL), OECD countries
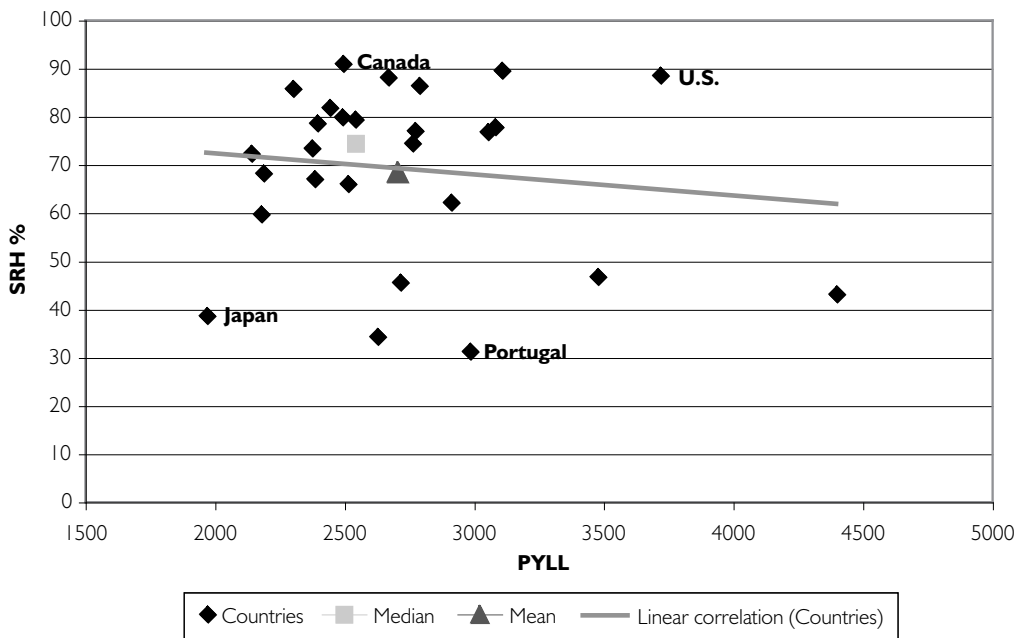

Life expectancies show a similar pattern. There does appear to be a better fit between national life expectancies and SRH status, but the Japanese and American anomalies remain. And if one were to divide the scatter-plots into four quadrants – above- or below-average life expectancy or PYLL, above- or below-average SRH – one would not find the country observations grouped in the logical pattern. If anything, the contrary.

How to interpret this phenomenon? Perhaps the high PYLL in the United States is a result of the high mortality from firearms or other forms of violence, and from HIV/AIDS. People who die early account for more PYLL but are not available later to answer questionnaires about their health status. They may be genuinely in good health until their sudden removal from the potential survey population.

And since the excess mortality pressure in the United States is correlated (nega-

tively) with socio-economic status, it is preferentially weeding out some of those who, if they had survived, would have been in below-average health. Brutally put, kill off your poor early, and they cannot complain. Whether this process can account for any significant share of the American anomaly is, however, an empirical question.

FIGURE 2. Self-rated health (percentage reporting "good" or higher) plotted against average life expectancy (ALE), OECD countries
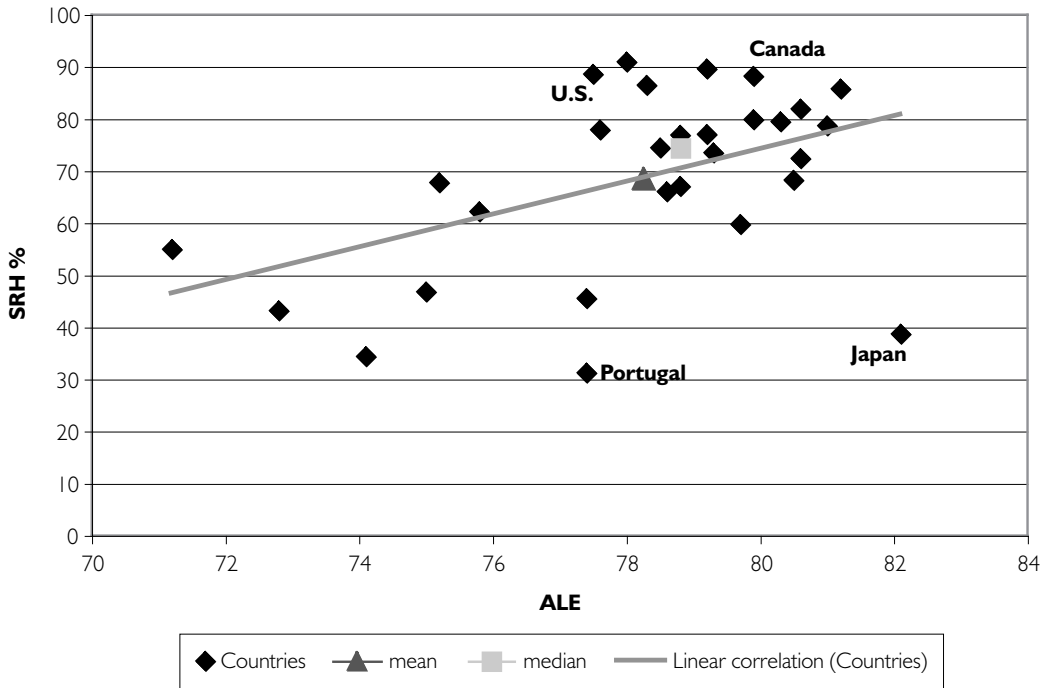
These SRH data are not adjusted for differences in population age structure. European populations are older than those of the United States and Canada, and this fact may partially explain their willingness to report poorer health. Canada and the United States, with younger populations, are both at the top of the SRH list. But Canada's average life expectancy is also close to the top, while the Americans are well below average and behind us by a whopping 2.4 years. They are also 40% higher on PYLL – 1,050 extra potential life years lost per 100,000 population.

Other anomalies: Life expectancies are pretty much equal in Portugal, Korea, Denmark and the United States. Yet their SRH runs from 31.3% "good" or higher, through 45.6% and 77.9% to 88.6%. Their PYLL scores are all above average, but quite similar, apart from the United States.

The comparison between the United States and the United Kingdom is particularly interesting in light of the findings of Banks et al. (2006). These authors found self-report of the presence of particular health conditions to be quite accurate, or at least consistent with biological data. Those surveyed knew what their medical problems were. But the translation of this knowledge into self-ratings of health is obviously very different on the two sides of the Atlantic. The British come out well below the Americans – 74.5% reporting "good" or better compared with the 88.6% of Americans. Yet their life expectancy is a year longer and their PYLL score is lower by 957 years per 100,000, consistent with the finding by Banks et al. that the slice of the British population they studied was in fact significantly healthier than the corresponding Americans.

These discrepancies, which have not gone unnoticed by students of health status measurement, illustrate a more general point emphasized by Corin (1994) with specific reference to the diagnosis of schizophrenia. Measurement instruments – including those much more sophisticated and expensive than good old E, VG, G, F, P – are designed in a particular cultural context (e.g., North America). They may be measuring something quite different, or perhaps nothing meaningful at all, when used somewhere else (e.g., central Africa). Not only the language but also the underlying concepts of health and disease may be radically different.

But if differing responses among cultural groups reflect different attitudes towards extreme statements rather than actual differences in quality of life, then they do not tell us anything about differences in the effects of intervention.

The observation that Japan, and to a lesser extent Korea, are extreme outliers in the OECD data may reflect more than simple difficulties in finding appropriate words in Asian languages to translate questions. The SF-36, the most widely used single measure of health status, has very rigorous translation protocols to ensure question comparability. But in Asian languages these failed to capture deep cross-cultural differences in the ways that health is conceptualized or described (Michael Wolfson, personal communication).

There are both good and less good reasons for continuing the search for the Holy Grail of a common metric for SRH. One good reason was illustrated above by the GUSTO trial. Whether tPA rather than streptokinase treatment resulted in better quality of life for patients, as perceived by themselves, is not a trivial question. But if differing responses among cultural groups reflect different attitudes towards extreme statements rather than actual differences in quality of life, then they do not tell us

anything about differences in the effects of intervention. Insofar as interventions are increasingly offered as improving quality of life rather than saving lives or effecting "cures," these distinctions become critical to trial design and evaluation.

Second, the growing interest in the determinants of population health, social as well as medical, becomes operational only insofar as we can actually measure population health. And patients' perceptions of their own health are or should be an important dimension of that measure. The indications from research on gradients in population health are that particular diseases may be "epiphenomena" arising from an underlying state of stress or distress; how that underlying state is expressed in disease varies across time and space. Cross-sectional comparisons of SRH, in addition to biomedical measures of particular diseases, may be very powerful provided that they can be meaningfully separated from cultural variations in response patterns.

The problem of cultural diversity of SRH is particularly important for Canada, given the rapid increase in the proportion of foreign-born and the diversity of sources of immigration.

The less good reasons are the ideology of "scientism" or physics-envy, and the commercial possibilities. If "science is measurement," then to be scientific (with the attendant prestige and funding support) one must have some universal concept(s) to measure. The rest is storytelling. And if one has some proprietary instrument for measuring that universal concept, one can sell it. The SF-36, HUI and EQ5d are currently the most widely used and most translated short-form "generic health status" questionnaires. Each has limitations, and no international consensus has emerged on which should be used. Choice is often the result of competitive marketing by originators.

The search for consensus on a cheaper and better short-form health status questionnaire, which will, inter alia, yield comparable results across cultures, is currently being carried forward by the Budapest Initiative (BI), a collaboration of the World Health Organization, UN Economic Commission for Europe, EuroStat and a self-selected group of enthusiastic countries, including the United States, the United Kingdom, Australia, Finland, Italy and Canada.

The problem of cultural diversity of SRH is particularly important for Canada, given the rapid increase in the proportion of foreign-born and the diversity of sources of immigration. Attempts to standardize reporting for the cultural peculiarities of particular countries – Japan, Portugal, the United States – through, for example, estimating "country dummies" to adjust SRH scores – will not be immediately applicable

within countries, though they might be combined with data on national origins to adjust SRH data across time and provinces or regions.

A different initiative, PROMIS (Patient Reported Outcomes Measurement Information System), originated with the US National Institutes of Health and seeks to develop a standard, single-dimension measure of SRH to be used in clinical trial outcomes. The focus is accordingly on consistency of measures across clinical trials with different interventions and outcomes, rather than on consistency across countries or cultural groups.

The two strands will have to come together if SRH data from trials are not to be confounded by cultural differences – this is where we came in. Some collaboration is emerging between BI and PROMIS. But the Holy Grail of a common metric of health status, both for population-based surveys and for the whole clinical trial industry, is still a long way off.

REFERENCES

Adams, M. 1997. *Sex in the Snow: Canadian Social Values at the End of the Millennium.* Toronto: Viking.

Banks, J., M. Marmot, Z. Oldfield and J.P. Smith. 2006. "Disease and Disadvantage in the United States and in England." *Journal of the American Medical Association* 295(11): 2037–45.

Corin, E. 1994. "The Social and Cultural Matrix of Health and Disease." In R.G. Evans, M.L. Barer and T.R. Marmor, eds., *Why Are Some People Healthy and Others Not? The Determinants of Health of Populations* (pp. 93–134). New York: Aldine–de Gruyter.

Harrington, H.J. and T. McNellis. 2006 (May). "Mobilizing the Right Lean Metrics for Success." *Quality Digest* 26(11). Retrieved April 14, 2007. <http://www.qualitydigest.com/may06/articles/02_article.shtml>.

Hart, M. 2007 (March 8). Presentation to the IRPP Canadian Priorities Agenda Conference, Toronto.

Helliwell, J.F. 2002. *Globalization and Well-Being.* Vancouver: University of British Columbia Press.

OECD Health Data. 2006 (June). Paris: Organization for Economic Cooperation and Development.

Sanmartin, C., E. Ng, D. Blackwell, J. Gentleman, M. Martinez and C. Simile. 2004 (June). *Joint Canada/United States Survey of Health, 2002–2003.* Ottawa: Statistics Canada, and Washington, DC: National Center for Health Statistics, Centers for Disease Control and Prevention.

Williams, J.I. and M. Agha. 2004 (April 14). "Self-Rating of Health: National, Racial and Ethnic Differences." Rounds presentation. Toronto: Institute for Clinical Evaluative Sciences.

Williams, J.I., J. Kelly, M. Agha and T. To. 1995 (November 16). *Are There Societal Differences in Rating of Health?* Toronto: Institute for Clinical Evaluative Sciences.