



# Open Research Online

---

The Open University's repository of research publications and other research outputs

## What's the evidence for Lean?

### Journal Item

How to cite:

Dyba, Tore and Sharp, Helen (2012). What's the evidence for Lean? IEEE Software, 29(5) pp. 19–21.

For guidance on citations see [FAQs](#).

© 2012 IEEE

Version: Accepted Manuscript

Link(s) to article on publisher's website:  
<http://dx.doi.org/doi:10.1109/MS.2012.126>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's [data policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

Voice of Evidence

## What's the Evidence for Lean?

Tore Dybå, Helen Sharp

The question of what empirical evidence is, how it is interpreted, and how useful it is for different contexts, are questions of constant importance for software managers and practitioners who are faced with technology adoption decisions (Dybå et al., 2005). Independent of such evidence, however, we cannot expect a technology to be universally good or universally bad, only more (or less) appropriate in some circumstances and for some organizations.

Yet, in *The Machine that Changed the World*, Womack et al. (1990) purported that lean production is the one best way to manufacture cars and everything else: “lean production is a superior way for humans to make things,” so “it is in everyone’s interest to introduce lean production everywhere as soon as possible.”

Given what we just said about the appropriateness of new technologies, these are quite bold claims, and the question is: what’s the evidence to back them up? Also, because lean is one of the software development trends that has attracted a lot of attention lately, it is interesting to take a closer look at the original evidence behind lean.

We will not go into, nor make any claims directly related to lean software development itself, neither will we go into the literature that criticizes the appropriateness or social consequences of lean. We will only take a glimpse behind the curtains at the evidence behind the original concept of lean production, and its popular interpretation, as an example of the inherent challenges of measuring and interpreting evidence for performance differences.

Basically, it all narrows down to the interpretation of the results from the largest international productivity survey in the history of the car industry: the MIT’s International Motor Vehicle Program (IMVP) survey carried out in the late 1980s (Krafcik, 1988; Womack et al., 1990) to compare productivity across car assembly plants. The IMVP survey was a study of car assembler productivity based on the general ratio between inputs and outputs. In this case, between labor hours and number of cars built.

Due, in a large part, to the *The Machine that Changed the World*, the bottom line measures of the IMVP survey have received a lot of attention; like the claimed 2:1 ratio between the best Japanese productivity and worst American performers, and the claim that it had been conclusively demonstrated that lean production had lowered the labor hours to build cars at any level of factory automation. However, the technicalities of calculations and interpretations leading up to these measures have not received the same attention.

In the following, we will examine the empirical evidence and its interpretation from the IMVP survey, its relation to lean production, and what we can learn from it. This discussion is based on Coffey’s (2006) excellent analysis in *The Myth of Japanese Efficiency*. While doing so, we want to emphasize that we differentiate between the outstanding survey itself (Krafcik, 1988) and the popular account and interpretation of it given by Womack et al. in *The Machine that Changed the World*. Like Coffey, our main points are that i) the survey is worth taking seriously, but ii) that it is open to quite different interpretations from those which are generally accepted.

## General issues of the IMVP survey design

The IMVP survey only deals with measurements taken in the later stages of car production in the assembly plant, which typically involve welding, painting, engine assembly, trim, and final assembly. Womack et al. justify this selection highlighting that the goal of the IMVP survey was to compare assembly plants, and that assembly plants are highly comparable. As Coffey points out, given the claims of a 'lean' revolution, this last aspect is rather astonishing since any radical variances in plant layout or organization are discounted from the outset of the survey when accounting for differences in the resources used to build cars:

[A]ssembly plants all over the world do almost exactly the same things, because practically all of today's cars and light trucks are built with very similar fabrication techniques. (Womack et al., p. 76).

Several commentators have criticized the IMVP survey's productivity measurement based on the way it constructs a corrected bottom line through cumulative adjustments of, e.g., direct and indirect workers, standard activities, standard working times, and average cars, using a number of complex indices (e.g. Williams et al., 1994). However, it's easy to get lost in the complexities of both these criticisms and the IMVP methodology itself, and lose sight of the larger picture. We will therefore only look at two major issues highlighted by Coffey: the way in which labor hours are measured, and the estimated levels of assembly plant automation.

## Labor hours and automation

The bottom line productivity measure of the IMVP survey is the number of labor hours used to build a car. The first major issue that Coffey highlights, is an important overlooked point of the IMVP survey - that no allowance was made for work carried out in excess of a *single* standard shift, i.e., all overtime was ignored and only a simple headcount of workers was included.

Nevertheless, the estimated hours of labor input based on the single shift was divided by a daily *output* of cars. To see why this is a major issue when comparing input-output relationships across manufacturing sites in the automobile industry, let's take a closer look at Coffey's example. Let's assume a plant in which there are 1,500 available workers that perform a single shift of 8 hours with no overtime. Suppose that the daily output is 400 cars assembled. The basic productivity measure in this example would then be  $(8 \times 1,500)/400 = 30$  hours per car.

Let's change the assumptions and look at a similar plant with a different configuration of shift work and overtime. Now, let's assume that there is overtime and that all workers currently perform a double shift. Suppose for simplicity that the headcount of available workers at this plant is 750 with the same daily output of 400 cars assembled. An accurate assessment of plant productivity in this example would then be the same as in the previous example:  $(16 \times 750)/400 = 30$  hours per car. But, if no allowance is made for double-shift working, and we assume that all workers in the headcount only work a single shift, this halves to give  $(8 \times 750)/400 = 15$  hours per car.

Any survey that multiplies an employment total by the hours worked in a single non-overtime shift, to get an estimate of labor input to production, is in great danger of introducing systematic bias. Because the car industry is an industry characterized by both complex shift patterns and the use of overtime, this casts serious doubts on the interpretations of the IMVP survey findings presented in *The Machine that Changed the World*. And, especially so since available data shows, e.g., a more than ten hours' difference in the average working week in Japan versus Europe, and since Japanese manufacturers at the time of the IMVP survey were known for 'massive amounts of overtime'.

The second major issue that Coffey points out is the interpretation of the relationship between assembly plant hours used to build a car, and differences in the estimated levels of assembly plant automation. The data obtained by the IMVP survey indicated that assembly plants in Japan were highly automated compared to plants in other regions covered by the survey. Even the least automated of the Japanese assembly plants surveyed still achieved a very high score on automation compared to most other car assembly plants in the world.

All else being equal, this would clearly suggest that automation played a very important role in accounting for Japan's typically low score on the labor hours used per car in the IMVP survey, *vis-à-vis* assembly plants elsewhere in the world. Indeed, in their analyses, Womack et al. found that variation in the levels of plant automation seemed to account for about one-third of the total variation in the hours used to build cars, as measured by linear regression. However, automation was downplayed as a major determining factor, and the IMVP survey findings were accordingly interpreted in terms of a simple comparison between average hours used to assemble a car

According to the IMVP survey, car assembly plants in Europe seemed to perform differently from similarly automated plants in the rest of the world. Due to this fact, Coffey shows how it would have been possible to draw quite different inferences from the same set of data: "Indeed, if the data for Europe had been taken separately, 'evidence' of a lean revolution would have been much harder to find" because plants in Japan would have been distinguished from plants in other regions of the world first and foremost on the basis of automation.

After making allowance for the possible distorting effects of Europe, Coffey's reanalysis shows that automation would now have accounted for around three-quarters of the sampled variation. By most standards this would be characterized as a very large effect size, which is why Coffey concludes that, for plants outside Europe, evidence of organizational superiority (or lean production), as opposed to high automation, would have been very hard to come by.

### **Assessment and interpretation**

From the viewpoint of reinterpreting the IMVP survey findings, therefore, there are two different issues. Coffey showed us that a reasonable case can be made that Japanese car assembly plants looked globally exceptional at the level of site automation, but less so at the level of differences in labor input not ascribable to automation. But more importantly, the detail provided by Krafcik on the design of the survey points to the possibility of a systemic bias in the labor input measure, one, which Coffey points out, might just as reasonably have been expected to influence results both by plants and by regions.

Taking the collated evidence seriously, and giving due credit to the survey architects, Coffey reaches a quite contrary set of conclusions, stating that:

First we dispute that evidence was ever presented to indicate that Japanese car assemblers possessed a discernible organizational advantage over Western competitors once due allowance was made for differences in measured levels of assembly plant automation. Second, we show that the data in question was biased ... on this basis, we suggest a *prima facie* case for considering whether correcting for this bias would in fact suggest that Japanese assemblers were at this point in time struggling to convert worker effort into finished cars, *vis-à-vis* Western competitors. ...

On the basis of the findings from the IMVP survey Coffey thus concludes that Womack et al. may have given an account of the ability of an assembler to convert hours of effort into finished goods that is flawed and a more plausible interpretation is that bias in the measurement of productivity, and level of site automation better explains the results.

**So**, what can we learn from this? Hopefully, we have shown that evidence isn't always what it seems to be and that popular interpretations aren't necessarily the only ones (or even the most accurate ones). It is necessary, therefore, to go behind the bottom line measures to challenge their assumptions, their standardized models, and their use and adjustments of inputs and outputs. Only then can we make sense of the evidence behind a claim and its wider applicability.

## References

D. Coffey, *The Myth of Japanese Efficiency: The World Car Industry in a Globalizing Age*, Edward Elgar Publishing, 2006.

T. Dybå, B.A. Kitchenham, and M. Jørgensen, "Evidence-based Software Engineering for Practitioners," *IEEE Software*, Jan./Feb. 2005, pp. 58-65.

J.F. Krafcik, *Comparative Analysis of Performance Indicators at World Auto Assembly Plants*, M.Sc. Thesis, The Sloan School of Management, Massachusetts Institute of Technology (MIT), 1988.

K. Williams, C. Haslam, J. Williams, and S. Johal, "Deconstructing Car Assembler Productivity," *Int. Journal of Production Economics*, Vol. 34, 1994, pp. 253-265.

J.P. Womack, D.T. Jones, and D. Roos, *The Machine that Changed the World*, New York: Free Press, 1990.