



Thomas, P. J., & Vaughan, G. J. (2015). 'Testing the validity of the "value of a prevented fatality" (VPF) used to assess UK safety measures': Reply to the comments of Chilton, Covey, Jones-Lee, Loomes, Pidgeon and Spencer. *Process Safety and Environmental Protection*, 93, 299–306. <https://doi.org/10.1016/j.psep.2014.11.003>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY-NC-ND

Link to published version (if available):  
[10.1016/j.psep.2014.11.003](https://doi.org/10.1016/j.psep.2014.11.003)

[Link to publication record on the Bristol Research Portal](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via Elsevier at <https://www.sciencedirect.com/science/article/pii/S0957582014001797> . Please refer to any applicable terms of use of the publisher.

## University of Bristol – Bristol Research Portal

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/brp-terms/>

Contents lists available at [ScienceDirect](#)

# Process Safety and Environmental Protection

journal homepage: [www.elsevier.com/locate/psep](http://www.elsevier.com/locate/psep)

IChemE

## Reply to the Letter to the Editor



**‘Testing the validity of the “value of a prevented fatality” (VPF) used to assess UK safety measures’: Reply to the comments of Chilton, Covey, Jones-Lee, Loomes, Pidgeon and Spencer**

Thomas and Vaughan (2014) (hereafter “the Article”) drew attention to the weaknesses in the approach underpinning the “value of a prevented fatality” (VPF) used as a reference for health and safety decisions across the UK’s process, nuclear and other industries, as well as the NHS (Glover and Henderson, 2010). It is important for the UK’s future health and safety strategy that the lack of evidence for the VPF should be debated openly in the pages of a learned journal. Hence we are pleased that Chilton, Covey, Jones-Lee, Loomes, Pidgeon and Spencer, authors from the Carthy study (Carthy et al., 1999), have attempted to answer our criticisms and that we have been given the opportunity to respond.

Having studied the comments provided by Chilton et al., we believe that our significant reservations concerning the Carthy study and its methods remain entirely valid. We find encouragement in the several admissions by Chilton et al. that our criticisms are justified.

A major problem for the Carthy chained approach is the significant deviation between direct and indirect measurements of the same quantity, which common sense suggests should not occur. Two “values of a prevented injury” (VPI) per individual can be calculated for one specified serious but non-fatal injury cited in the Carthy study, with one personal VPI,  $m_{xi}^{(1)}$ , calculated directly and the other,  $m_{xi}^{(2)}$  through two-injury chaining. The two figures should be the same or similar, but they turn out to be completely different. Moreover, there is barely any correlation between the two in the 167-strong sample.

We are pleased that Chilton et al. acknowledge this in their “Concluding comments”:

“there is a definite and seemingly systematic divergence between direct and indirect estimates which is illustrated by the comparison between  $m_{xi}^{(2)}$  and  $m_{xi}^{(1)}$ .”

Chilton et al. highlight the useful contributions of the Carthy study as:

- (i) “providing evidence which, when blended with judgement, helped consolidate the VPF”, and
- (ii) “demonstrating the need and potential for further work to improve the methodology of stated preference elicitation in respects where limitations still undoubtedly exist.”

and request that their study should not be regarded as “worthless”.

Against that must be set the fact that the two-injury chained method has failed its validation test. Moreover, Chilton et al. admit applying their own judgement in deriving the final VPF value, with the decision space over which their judgement ranged being very large – between £0.5 M and £1.6 M even near the end of the study, for example. No route can be traced from the opinions of the respondents to the recommended VPF figure without it passing first through the filter of the Carthy authors’ judgement. The VPF figure they advance is thus not an objective consolidation of the views of the respondents and must be regarded as compromised and unreliable as a result.

Moreover neither the Carthy study nor Chilton et al. have put forward evidence to show that the wealth of the 167 respondents reflected the range of personal wealths in the nation. Furthermore it will be shown in the reply to the comments on Section 4 of the Article that Chilton et al. have been unable to sustain a valid objection to the Article’s findings on the apparently very low levels of wealth in the sample. Were the methods of the Carthy study correct and consistent, the average wealth of the respondents would have been less than a tenth of the average net wealth of UK adults in 1997. Hence no trust could have been put in the VPF emerging from the Carthy study, even if its underlying method had not failed its validation test.

There is no evidential base for the VPF recommended by the Carthy study grounded in the opinions of a representative set of respondents and hence no evidential basis for the UK’s VPF.

Many of the comments of Chilton et al. on Sections 1, 2, 4 of the Article have been found to be flawed or mistaken. We have pointed out the apparent misunderstandings in our detailed replies, taken Section by Section in the order used by

Chilton et al. Their comments on Sections 3 and 7 consist largely of attempts to salvage credibility for the two-injury chained method in the face of the very significant and admitted disparity between its two measurements of the same quantity. In our opinion, credibility has not been achieved.

### Reply to comments headed “TV’s Section 1: Introduction”

Chilton et al. (2014) say that the 1997 opinion survey was “not the sole basis for the VPF” used in the UK when, by contrast, Chilton et al. (2002) seemed to be claiming credit for the Carthy study:

“In the light of the findings in [Beattie et al. (1998) and Carthy et al. (1999)] the ... DTER [Department for Transport, Environment and the Regions] elected to increase its WTP-based roads VPF to some £1.05 million in 1998 prices.”

Moreover, in their Final Report to the Government’s Inter-Departmental Group for the Valuation of Life and Health, Wolff and Orr (2009) conclude:

“it appears that the Carthy study is now the primary source of VPF figures, adjusted for inflation and changes in GDP.”

The Spackman report (Spackman et al., 2011) also confirms that it is

“the 1997 WTP study on which the current WTP values are based”

making it clear that the Carthy study and the “1997 WTP study” are synonymous. The Spackman report also explains that the “WTP value” (=the Carthy VPF) makes up about 93% of the DfT’s VPF, based on the most recently available figures, with the remaining 7% covering medical and ambulance costs and the net lost output of the victim.

Chilton et al. (2014) come closer to the mark when they claim a “key role” for the Carthy VPF, making it vital that those responsible for the UK’s safety policy should know that the chained method on which the UK’s VPF rests is invalid.

Chilton et al. object to the Article’s quotation from the Spackman report, indicating its recommendation “against any early new full scale WTP study” and quote both the complete sentence and the two preceding sentences. It is difficult to see the words as other than a recommendation, but if there was ever any doubt, these words are backed up in the report’s “Section 8 Conclusions and Recommendations”, subsection “8.5.1. UK transport WTP valuations”:

#### “Recommendations

There should be no near term full revaluation of the WTP element of the VPF or of the VPIs.”

As noted by Chilton et al., the Spackman report also contains the statement:

“We conclude that the chained approach is in principle superior to other stated-preference procedures that have so far been used to estimate WTP-based values of fatality risks.”

If this statement were to be believed, then given the analysis in the Article and the acceptance by Chilton et al. of the limitations of the chained approach (“limitations still undoubtedly exist”), it would seem to raise significant issues about the viability of stated preference methods of any sort.

Chilton et al. dispute the following sentence from the Article:

“They introduced the two-injury chained model, where the response to a serious injury of the second type, e.g. a fatal injury, is deduced after first eliciting from the person a statement of how much he would spend to reduce the probability of a lesser injury of type 1.”

It is a fundamental to the two-injury chained model that, for every individual  $i$ , the marginal rate of substitution,  $m_{1i}$ , of non-injury probability for wealth for injury 1 needs to be found first. Then further questioning is used to establish the ratio of the individual’s desired spending to reduce the chances of the more serious, type 2 injury to his spending to reduce the probability of the type 1 injury. In symbols,  $m_{1i}$  is found first, the ratio,  $m_{2i}/m_{1i}$ , is found next, allowing  $m_{1i} \times m_{2i}/m_{1i}$  to produce, notionally at least,  $m_{2i}$ , the marginal rate of substitution of non-injury probability for wealth,  $w_i$ , for injury 2 for individual  $i$ .

The complaint of Chilton et al. is that two statements need to be elicited from individual  $i$  before  $m_{1i}$  can be calculated, rather than just one: he needs to state the maximum acceptable price (MAP) he is prepared to pay to avert injury 1 and the minimum acceptable compensation (MAC) he would require to make up for his receiving injury 1. This point is explained fully by Thomas and Vaughan in the more extensive description of the two-injury chained method given in the next Section, “Data”, on the next page. Nevertheless it is accepted that omitting the words, “a statement of” in the disputed sentence would render it more accurate, although we feel that this is a rather small point of little importance to the main criticisms we have made. And, of course, the two-injury chained approach is summarised in mathematically precise terms in Appendix A of the Article.

### Reply to comments headed “TV’s Section 2: Data”

Chilton et al. are mistaken in saying

“34 cases which TV [Thomas and Vaughan (2014)] describe as ‘inconsistencies in the results’ are those in which the respondent’s stated willingness to pay for a complete cure for a particular injury (MAP) was greater than or equal to the sum that he/she would be willing to accept in compensation for suffering the injury (MAC).”

since 29 are cases where  $MAC \leq MAP$ , and 5 have  $MAC > MAP$ . Moreover, the 34 do not include a further 7 respondents with  $MAC \leq MAP$ .

Chilton et al. are also in error when they assert that:

“the Negative Exponential utility function produces a completely different result from the one that we derived and reported in the data that we sent to them.”

Table 1 compares the values calculated using the method described in Appendix B.4 of the Article for the 34 cases where the Carthy authors over-write a common value for personal VPF,  $m_{xi}^{(2)}$ , across all utility functions. It is clear that, with the exception of respondent 123, the results are essentially the same. They would be exactly the same if the Carthy authors had chosen to solve Eqs. (ix) and (x) of their Appendix B using a high-precision numerical iterative technique rather than using their derived approximations (accurate to about a half a percent).

**Table 1 – The common value from the Carthy study and the corresponding value from the Negative Exponential utility function.**

Respondent i	$\phi_{Xi}$	Personal VPF from 2-part chaining from injury X (£)	
		Negative Exponential (Thomas and Vaughan, 2014, Appendix B.4)	Common value Carthy study
6	0.3820	712,802	713,715
7	0.0724	78,445	78,669
12	0.7796	1,105,436	1,108,890
13	1.0000	19,980,000	19,980,000
24	0.1607	1,375,907	1,390,571
27	0.0724	44,125	44,251
39	0.6544	904,648	912,975
58	0.6106	241,403	242,614
59	0.0164	37,754	37,762
70	0.5547	1,509,739	1,498,500
77	0.8192	2,717,044	2,747,250
85	0.3051	53,704	53,824
86	0.6106	262,858	264,180
88	0.0724	313,778	314,674
91	0.3820	962,283	963,036
97	0.3820	17,164	17,173
99	0.7796	9,024	9,052
100	0.6106	262,858	264,180
107	0.8468	1,342,156	1,348,650
108	2.0000	1,441,252	1,441,357
109	1.6180	160,381	160,589
113	0.1607	280,801	283,790
114	1.0000	5,046	5,045
122	0.2451	297,978	297,480
123	1.7549	446,968	792,540
128	1.0000	499,500	499,500
130	1.0000	50,200	50,201
145	0.7182	34,713	34,557
148	0.0164	19,387	19,391
152	0.3820	12,830,440	12,837,151
161	0.7796	2,210,872	2,217,780
163	0.2451	35,999	35,937
164	1.7549	1,191,914	1,189,920
165	1.7549	595,957	594,960

Respondent 123 has  $MAC > MAP$ , contradicting the contrary suggestion by Chilton et al., so his personal VPF should be as listed by Thomas and Vaughan in Table 1.

Table 1 shows also the value of the parameter,  $\phi_{Xi}$ , needed to solve Eq. (6) of the Article. It may be noted that when  $MAC = MAP$ ,  $\phi_{Xi} = 1$ . From Appendix B.4 of the Article,  $\phi_{Xi} = e^{\beta_i x_{Xi}}$  so that  $\beta_i = (\ln \phi_{Xi}) / x_{Xi}$ . Since  $x_{Xi} > 0$ , it follows that  $\beta_i = 0$  when  $MAC = MAP$  and  $\phi_{Xi} = 1$ . Such an eventuality is, in fact, prohibited by Carthy et al. in Eq. (vii) of their Appendix B, but apparently reinstated in the first paragraph of their Section 3.3. Chilton et al. endorse the Carthy reinstatement of  $\beta_i = 0$ , saying

“we applied a linear utility of wealth function whenever  $MAP = MAC$ ”

But noting the definition of the Negative Exponential utility function:

$$U_i(w_i) = -e^{-\beta_i w_i} \tag{1}$$

and substituting  $\beta_i = 0$  gives:

$$U_i(w_i) = -1 \text{ for all wealths, } w_i \tag{2}$$

This does, indeed, give a straight-line graph, but the line is horizontal, implying a utility function for wealth that is independent of wealth, which is a contradiction in terms. Thus

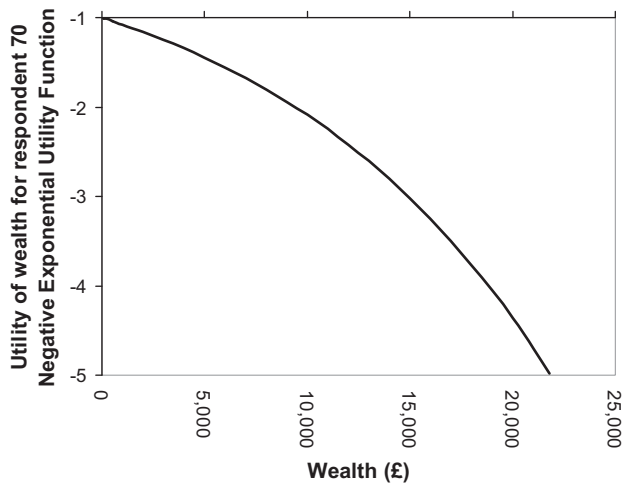
$\beta_i = 0$  needs to be excluded, although the Carthy authors did not seem to be fully aware of the fact, nor do Chilton et al. appear to have realised it.

The situation is, if anything, worse for cases where  $MAC < MAP$ , since this entails  $0 < \phi_{Xi} < 1$ . Applying  $\beta_i = (\ln \phi_{Xi}) / x_{Xi}$  with  $x_{Xi} > 0$  implies  $\beta_i < 0$ . Chilton et al. seem to be under the erroneous impression that this condition converts the Negative Exponential utility function, concave when  $\beta_i > 0$ , into a convex utility function:

” a strictly convex, positive exponential utility of wealth function whenever  $MAP > MAC$ .”

Not so. The exponent,  $\lambda_i w_i$ , where  $\lambda_i = -\beta_i > 0$ , will be positive, but the negative sign preceding the exponential will remain. In so far as the resulting function can be described as a utility function at all, it will remain a Negative Exponential utility function. For what it is worth, the new function will remain concave (not convex as suggested by Chilton et al.), but it is inadmissible as a utility function, since it predicts that utility will decrease monotonically with wealth. See Fig. 1, which shows the (improper) Negative Exponential utility function for respondent 70 from Table 1, who has  $MAC < MAP$ .

To summarise, the Carthy study substituted a common value for personal VPF for 34 respondents across the other three utility functions. That common value was calculated from the Negative Exponential utility function using



**Fig. 1 – Negative Exponential utility function with  $\beta_i < 0$  (respondent 70).**

an approximate method. The Negative Exponential utility function was, however, being operated beyond its range of legitimacy whenever  $MAC \leq MAP$ , which was the case for 29 out of the 34 cases under discussion, even though numerical results could be generated. The attempted justification offered by Chilton et al. for applying a common personal VPF derived from the Negative Exponential utility function across all utility functions disappears for the other 5 cases, where their assumption that  $MAC \leq MAP$  does not hold.

Despite the fact that average VPFs apparently coming from the Constrained Power, the Logarithmic and the Negative Inverse utility functions have a large number of their constituent, personal VPFs supplied from the Negative Exponential utility function, no indication of this was carried in their labels. See Tables 2, 3, 6 and 7 of the Carthy study. While Chilton et al. have now pointed to the first paragraph of Section 3.3 of the Carthy study as a statement that the Carthy authors chose to operate in this way, and this is accepted, nevertheless it is inherently misleading to file results under the wrong label: deliberate misattribution is poor practice. Nor did the cited paragraph indicate the large scale of the practice, with the average VPFs for the other utility functions carrying a 22% contribution from the Negative Exponential utility function. Such a large common contribution ought, as a minimum, to be reflected in the labelling of the VPFs, clarifying that the four utility function streams are not independent. Thus, for example, the results listed as “Logarithmic” ought to carry the label, “Logarithmic/Negative Exponential”.

It is not clear why the Carthy authors resorted to using results from the Negative Exponential utility function when the other utility functions can generate similar results to those of the *ultra vires* Negative Exponential utility function when they, too, are pushed beyond their legitimate limits.

But most of the 34 cases under discussion should have been excluded from consideration because of the inability of the Carthy method to cope when  $MAC \leq MAP$ , as should the additional 7 cases with  $MAC \leq MAP$ . The number of respondents from the sample with which the Carthy method can cope reduces to either 120 or 121, depending on the utility function used, down by about a quarter on the already low starting figure of 167. Neither Carthy et al. (1999) nor Chilton et al. comment on the small sample size, to be considered at the end of the next section.

## Reply to comments headed “TV’s Section 4: The wealth of respondents”

Chilton et al. are in error when they suggest that the wealth offset,  $\beta$ , may take some arbitrary value. If the four utility functions used in the Carthy study are to measure the same wealth characteristic, essential for consistency and comparability, then the wealth offset,  $\beta$ , must be zero.

The Article gives the Negative Exponential utility function as:

$$U_i(w_i) = -e^{\beta_i w_i}, \beta_i > 0 \quad (\text{B.28})$$

with the subscript,  $i$ , added to identify it with the  $i$ th individual. Including a wealth offset,  $\beta^{(i)}$ , for individual  $i$  modifies the form of Eq. (B.28) to:

$$U_i(w_i) = -e^{\beta_i(w_i - \beta^{(i)})} \quad \begin{array}{l} \beta_i > 0 \\ 0 \leq \beta^{(i)} < w_i \end{array} \quad (3)$$

where the condition,  $0 \leq \beta^{(i)} < w_i$ , is stipulated in the Carthy study. Comparing Eqs. (B.28) and (3) shows that

$$\beta^{(i)} = 0 \quad \text{for all individuals, } i \quad (4)$$

for the Negative Exponential utility function.

Characterising the four utility functions by

$$U_i(w_i) = F_i(w_i - \beta^{(i)}) \quad (5)$$

where  $F_i(\cdot)$  is an increasing function of its argument. The second condition accompanying Eq. (3) implies

$$w_i - \beta^{(i)} \geq 0 \quad (6)$$

so that the lowest value of utility will occur as  $w_i \rightarrow \beta^{(i)}$ . This implies that the wealth offset,  $\beta^{(i)}$ , is the limiting value of wealth that gives the individual his lowest utility, a result independent of utility function.

While a different numerical value for utility will be obtained for each of the four utility functions, the wealth the individual feels he needs before it brings him any utility, his wealth offset,  $\beta^{(i)}$ , will be a characteristic of him not of the function used to model his utility. Hence the assertion by Chilton et al. that

“there is absolutely no reason why the shift parameter,  $\beta$ , should be the same in the two cases [Negative Inverse and Logarithmic]”

is incorrect. The wealth offset,  $\beta^{(i)}$ , being a characteristic of the individual, will hold across all utility functions, and since it is specified by the authors of the Carthy study as  $\beta^{(i)} = 0$  for all  $i$  for the Negative Exponential utility function, that value,  $\beta^{(i)} = 0$ , must apply to the other three utility functions.

While the above has proved that  $\beta^{(i)} = 0$  for all  $i$ , a further comment is appropriate on the lack of realism of a non-zero figure for wealth offset. Suppose that a person possesses a wealth,  $w_i = \beta^{(i)}/2$ . Condition (6) is now violated, and no utility can be found for the person, unless Eq. (5) is replaced with

$$U_i(w_i) = \begin{cases} F_i(0^+) & \text{for } w_i \leq \beta^{(i)} \\ F_i(w_i - \beta^{(i)}) & \text{for } w_i > \beta^{(i)} \end{cases} \quad (7)$$



where the argument,  $0^+$ , allows the Lognormal and Negative Inverse utility functions to return a value.

Assuming Eq. (7) holds, a wealth,  $w_i = \beta^{(i)}/2$ , will cause the individual to experience his lowest possible utility. Now suppose his wealth increases by 50% to  $w_i = 0.75\beta^{(i)}$ . Will he experience any benefit? Apparently not. But such a result goes against normal economic wisdom that, *ceteris paribus*, more is preferred to less. It also goes against a basic assumption of economic utility theory, namely diminishing marginal utility (e.g. [Lipsey and Chrystal, 1995](#)). For a person who is risk averse but who has a positive wealth offset will have a marginal utility that starts and continues at zero as his wealth increases towards his wealth offset. His marginal utility will then rise discontinuously to its highest value just above the wealth offset and only then undergo the steady decline with wealth that is characteristic of diminishing marginal utility.

Considerable doubt is thus cast on the realism of expecting the wealth offset,  $\beta^{(i)}$ , to take any positive value. Certainly Daniel Bernoulli included no offset when he introduced the Logarithmic utility function ([Bernoulli, 1738, 1954](#)). Moreover the UK Treasury has adopted a Logarithmic utility function without offset ([Treasury, 2011](#)). A positive wealth offset also violates a condition laid down in the Carthy study:

“Now suppose that the individual prefers more wealth to less and is financially risk-averse so that  $(\forall w)U'(w) > 0$ ,  $U''(w) < 0$  where  $U''(w)$ , denotes the second derivative of  $U(\cdot)$ ”

To summarise, Chilton et al. have sustained no valid objection to the findings of Section 4 of the Article concerning the very low values of wealth apparently revealed in the Carthy study.

Moreover, the validity of averaging the personal VPFs to give a nationwide VPF (Eq. (A.36) of the Article) depends critically on the sample reflecting very closely the probability density for wealth in the nation as a whole, including people with significant wealth as well as those with middling and low wealth. But no evidence has been advanced in the Carthy study nor in the response by Chilton et al. to show that the wealth of the 1997 respondents reflected the full range of wealths in the UK. It is not even clear that the authors of the Carthy study collected the wealth data needed to justify a survey-based VPF.

It is doubtful that 167 people, still less the 120 for whom personal VPFs could reasonably be computed, would be enough to represent adequately the full spread of wealths in the UK.

### Reply to comments headed “TV’s Section 3: Testing the validity of the two-injury chained method”

Chilton et al. object to the clause, “the degree of linear correlation is almost non-existent”, although the rest of the sentence defines what is meant in numerical terms:

“Moreover, the degree of linear correlation is almost non-existent: while the square of the correlation coefficient,  $R^2$  should obey  $R^2 = 1.0$  or be close, the actual value is  $R^2 = 0.0719$ .” (The Article)

Even though they themselves describe it as “quite weak” in their next paragraph, the contention from Chilton et al. seems to be that the degree of linear correlation cannot be “almost non-existent” when some degree of positive correlation is likely. It is important to address this statistical misunderstanding.

What may be described as the “scatter plot” of Fig. 2 of the Article is based on the Constrained Power utility function, where the value of  $R^2$  is  $0.0719 \Rightarrow R = 0.268$ . Given the high number of data points on the graph, such a value of  $R$  is indeed unlikely to be the chance outcome of an uncorrelated process. However, the degree of correlation will be very low indeed, in the sense that only 7.19% of the variation in the  $m_{Xi}^{(2)}$  values can be attributed to the best linear relationship with the  $m_{Xi}^{(1)}$  values, while 92.81% will be associated with other, unidentified factors. Similar scatter plots arise from the Logarithmic and Negative Inverse utility functions. In both these cases the correlation coefficient is high enough to indicate some degree of correlation, but the amount of variation in the  $m_{Xi}^{(2)}$  values that can be attributed to the linear relationship with the  $m_{Xi}^{(1)}$  values is 4.43% under the Logarithmic utility function and just 1.95% for the Negative Inverse utility function. A scatter plot is also possible based on the Negative Exponential utility function, but now there is a good chance (>5%) that the observed correlation,  $R = 0.089$ , is consistent with a real value  $R = 0$ . Even if the correlation did exist, it would account for less than 1% of the observed variation in  $m_{Xi}^{(2)}$ . Thus the degree of linear correlation between  $m_{Xi}^{(2)}$  and  $m_{Xi}^{(1)}$  is extremely low, irrespective of which utility function is used, as stated in the Article.

Chilton et al. do not dispute that there is a significant disparity between the two VPI values for each individual. But in their desire to restate the discrepancy so that it looks better, they resort to the data censoring that was characteristic of the Carthy study. This time respondent 43 is to have his views disregarded in addition to respondent 51, whose views were censored in the Carthy study. Then, if “just these two observations are set to one side”, as Chilton et al. put it, the regression slope falls from 8.24 to 3.27. No justification is given for disregarding the views of respondents 43 and 51 beyond the convenience that dropping respondents associated with high  $m_{Xi}^{(2)}$  values reduces the apparent disparity.

Of course a slope of 3.27 when it should be 1.0 is still a major discrepancy, and the next argument of Chilton et al. is that if the VPI for the same injury is found for the same person in two different ways, the two VPI values should no longer be required to be the same. Hence rather than expressing the requirement mathematically as:

$$m_{Xi}^{(2)} = m_{Xi}^{(1)} \quad (8)$$

as suggested the Article, Chilton et al. suggest instead

$$m_{Xi}^{(2)} = f(m_{Xi}^{(1)}) \quad (9)$$

where  $f(\cdot)$  is a nonlinear function such that

$$f(m_{Xi}^{(1)}) \neq m_{Xi}^{(1)} \quad (10)$$

According to Chilton et al.:

“The truth is that the relationship is much better described as nonlinear”

despite this view seeming to go against what they wrote just three paragraphs before:

“In theory, for a respondent with error-free deterministic preferences, these two estimates would be identical, or at least very similar.”

This sentence is noteworthy in its own right for its attempt to pass onto the respondent the responsibility for any discrepancies between his two very different VPIs for the same injury. According to Chilton et al., it is not the measurement processes that are getting it wrong, rather these are accurate, but the respondent keeps changing his mind and making mistakes. This idea seems to be another unjustified judgement on the part of the researchers, and lies behind their frequent use of the word “deterministic” as a disclaimer. It would be interesting to know what evidence there is for a person’s deep feelings on serious injury changing greatly in the 30–60 min of his or her interview.

It is difficult to dismiss the possibility that a new measurement and filtering method based on complicated theory might fare badly when put into practice and, surprisingly, the breakdown of the chained method in practice appears to be the next argument advanced by Chilton et al. They devote paragraphs 6–15 of their Section headed “TV’s Section 3: Testing the validity of the two-injury chained method” to an attempt to demonstrate that the more direct,  $m_{Xi}^{(1)}$ , estimate of the VPI is likely to be more reliable than the version,  $m_{Xi}^{(2)}$ , found through two-part chaining. Intriguingly, they base part of their argument on the answers provided by respondent 51, whose opinion they suggested dismissing two paragraphs earlier, and whose views were indeed censored in the Carthy study, but whose opinion is now sufficiently typical to be used as an exemplar.

Broadly, Chilton et al. claim that the hospitalisation and the recovery period for injury X are each 4–5 times longer than for injury W (actual ratios: 4.67–7.0 hospitalisation; 4.5–6.0 recovery) and that they would expect individual VPI ratios for the two injuries to conform to these ratios. Although not quoted by Chilton et al., the large sample standard deviation, 2.52, gives some appreciation of the high degree of scatter on the sample mean, 3.8, of the VPI ratio under the Constrained Power utility function. On the basis that 3.8 should be regarded as close to the ranges (although actually lying outside them), Chilton et al. suggest that  $m_{Xi}^{(1)}$  is likely to be a fair estimate of the personal VPI for injury X.

Examining this claim further, by definition

$$m_{Xi}^{(1)} = \frac{m_{Xi}^{(1)}}{m_{Wi}} m_{Wi} = \psi_{Wi} m_{Wi} \tag{11}$$

Now approximate  $\psi_{Wi}$  by the mean VPI ratio:

$$\psi_{Wi} = \frac{m_{Xi}^{(1)}}{m_{Wi}} \approx \overline{\left( \frac{m_{Xi}^{(1)}}{m_{Wi}} \right)} \tag{12}$$

Since the average ratio,  $\overline{m_{Xi}^{(1)}/m_{Wi}}$ , seems to conform very roughly to a linear model of the periods of hospitalisation and recovery, and  $m_{Wi}$  is also assumed to be about right, then, goes the apparent argument,  $m_{Xi}^{(1)}$  will be roughly correct. (The very large scatter on the individual values of  $m_{Xi}^{(1)}/m_{Wi}$  suggests significant problems with this reasoning.)

Having attempted to show that  $m_{Xi}^{(1)}$  is relatively dependable, Chilton et al. seek now to establish that  $m_{Xi}^{(2)}$  is a poor estimate of the personal VPI for injury X. They say that  $m_{Xi}^{(2)}$  values are problematic because:

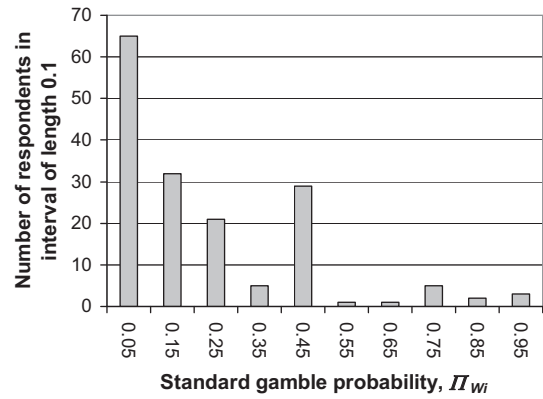


Fig. 2 – Histogram of the standard-gamble probabilities.

“SG [standard-gamble] responses are liable to produce much higher ‘multiples of badness’ factors than MAC and MAP responses.”

The calculational route to obtaining  $m_{Xi}^{(2)}$  via “chaining” may be expressed mathematically as:

$$m_{Xi}^{(2)} = \rho_{Wi} m_{Wi} \tag{13}$$

with the standard-gamble multiplier,  $\rho_{Wi}$ , for the individual given by:

$$\rho_{Wi} = \frac{1 - \theta_W}{\Pi_{Wi} - \theta_W} \tag{14}$$

where the respondent chooses the standard-gamble probability,  $\Pi_{Wi}$ , for the failure probability of an operation with a potentially better outcome than the reference operation which has a failure probability,  $\theta_W = 0.01$ . Fig. 2 gives a histogram of probabilities chosen. Chilton et al. note that

“small perturbations among low SG responses can result in very substantial effects on the multipliers.”

which may be verified easily by differentiation. This feature is regarded as producing “strong asymmetries”, in the sense that equal differences,  $\Delta \Pi_{Wi}$ , in standard-gamble probability will lead to unequal differences,  $\Delta \rho_{Wi}$ , in the standard-gamble multiplier. They inveigh against the high number of large “badness multiples”,  $\rho_{Wi}$ , resulting from the standard-gamble calculation as compared with the much lower number of large ratios,  $\psi_{Wi} = m_{Xi}^{(1)}/m_{Wi}$  calculated from the “direct” VPIs:

“whereas there were only five individuals exhibiting a  $m_{Xi}^{(1)}/m_{Wi}$  ratio greater than 10, there are sixty-five whose SG responses entail badness multiples greater than 10”

(These are essentially the respondents making up the mode in Fig. 2) A little later they conclude that

“We suggest that it is the multiplication of the  $m_{Wi}$  figures by an inverse function of small SG probabilities that is mainly responsible for the nonlinear departure of  $m_{Xi}^{(2)}$  from  $m_{Xi}^{(1)}$ .”

In fact, replacing the standard-gamble multiplier with a term linearly decreasing in standard-gamble probability,  $\Pi_{Wi}$ , would lead to equal differences in standard-gamble probability producing equal differences in standard-gamble multiplier, as desired Chilton et al., but what would be the justification?

Given that it is integral to the chained method, it is surprising that Chilton et al. have argued that the standard-gamble

multiplier becomes unworkable when faced with the opinions of real people (“in the real world of stated preference surveys”). They suggest that some questions, such as choosing the standard-gamble probability,  $\Pi_{W_i}$ , are “vulnerable to noise, error or bias”.

Chilton et al. concede finally:

“This is undoubtedly a problem for the use of the chained approach”

They follow this up with a request that their 1997 study should not be regarded as worthless and that “some understanding for the nature of the problem” should be shown.

But for the two-injury chaining method to provide a credible figure for the VPF, two techniques need to work correctly:

- (i) that for finding a person’s VPI based on his MAC and MAP (the “direct” method)
- (ii) that of “chaining” from the VPI to the VPF using the standard-gamble multiplier.

Since the directly derived VPI figures for injuries W and X are regarded by Chilton et al. as fair, if the VPI for X found by chaining from W is very different from the direct version, which it is, the chained VPI must be wrong, with the fault lying in the chaining process and the standard gamble multiplier, according to the logic of Chilton et al.

If all that was at stake were the VPI for injury X, then a satisfactory substantiation of  $m_{X_i}^{(1)}$  would be sufficient. But the Carthy study is based on the premise that the VPF figure,  $m_{D_i}^{(1)}$ , cannot be derived directly. Reliance must therefore be placed on the chained VPF,  $m_{X_i}^{(2)}$ , found by multiplying  $m_{X_i}^{(1)}$  by the standard-gamble multiplier. But if the latter is unreliable or wrong, no reliance can be placed on the individuals’ chained VPFs,  $m_{X_i}^{(2)}$ , nor on the consolidated VPF taken as an average of the individual figures.

In summary, the Carthy VPF can be regarded as valid only by ignoring the failure of the chaining method in its validation test, which runs counter to the philosophy and practice of science (Popper, 1934).

### Reply to comments headed “TV’s Section 7: Censoring the data”

Chilton et al. propose a new argument. Since the average ratio of the chained VPI to the direct version,  $\frac{m_{X_i}^{(2)}}{m_{X_i}^{(1)}}$ , is very high (a factor of about 6, although note the huge standard deviation of 15, Table 3 of the Article) and since  $m_{X_i}^{(1)}$  is judged by Chilton et al. to be a fair figure, then the average chained VPI,  $m_{X_i}^{(2)}$ , will be too high, and, by analogy, the average, chained VPF,  $m_{D_i}^{(2)}$ , will be too big also. Therefore, apparently, the analyst should be given the freedom to reduce  $m_{D_i}^{(1)}$  through censoring and the application of his judgement. One way is to “trim out” some of the highest personal, chained VPFs. Another is to adopt the median in preference to the mean.

According to Chilton et al., the analyst can compensate for the poor performance of the two-injury chained method as long as he can make the right “judgement calls”. They concede that

“It is true that the details of some of the judgment calls in Carthy et al. (1999) are not spelled out as clearly as they

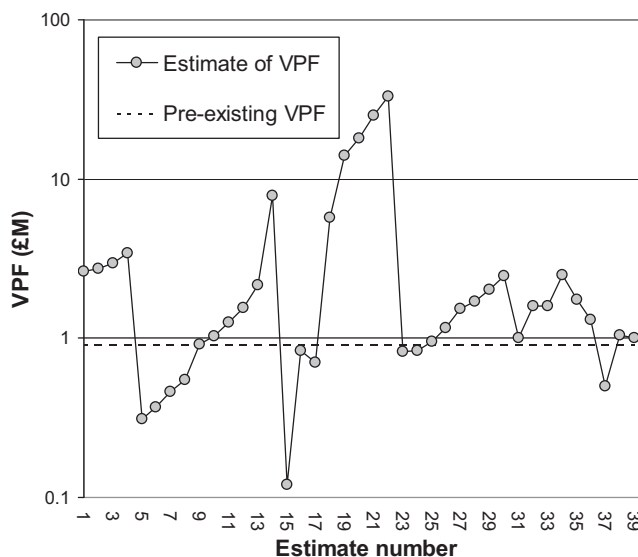


Fig. 3 – The 39 steps to the VPF.

might have been and it is fair for TV to draw attention to this.”

The Article explains that when the distribution of personal VPFs is approximately lognormal with a large spread, an analyst who feels free to choose either the mean or the median or some combination will have huge leeway for applying judgement to interpret his results. The range of VPF figures over which the Carthy authors felt free to judge was enormous, from £120,000 to £33,000,000, as shown in Fig. 3. VPFs of £0.5 M and £1.6 M were still under discussion in the 3rd last paragraph of the “Concluding comments” of the Carthy study. 39 steps were needed before the Carthy authors closed on their final judgement of £1,000,000, a VPF described by Wolff and Orr (2009) as “very close to the existing one”, then £902,500.

While censoring Respondent 132’s view was shown in the Article to be untenable because it was not an outlier, Chilton et al. now suggest that the high personal VPF of Respondent 132 might have arisen from a transcription error. The value of his standard-gamble probability,  $\Pi_{X132}$ , might not have been 0.0011, as recorded, but 0.0015 instead:

“There is one case for which the multiplier is shown on our spreadsheet as 9990, which might have been the result of an error in entering data: we assign that case a multiplier of 1998 instead”

No justification seems to accompany this new suggestion, surfacing 17 years after the Carthy study was carried out, but its effect is to reduce the personal VPF of Respondent 132 by a factor of five.

Chilton et al. say “TV may be reluctant to engage in trimming” as if censoring data or not was a matter of personal preference. This is perhaps the strangest aspect of their “defence”. In effect, they are saying that their judgement is better than the data. But as a general point, once procedural errors such as mistakes in transcription have been corrected, there is no justification for rejecting a view from an opinion survey or according it a reduced weighting simply because it differs significantly from the rest. Such a procedure would be defensible only if it could be proved that the person concerned was unqualified to offer an opinion.

The issue is analysed in Thomas (2014), which shows that, when seeking to consolidate people’s views on the size of a



numerical quantity, the analyst cannot defend himself against a charge of lack of objectivity if he uses a consolidation method that violates the criterion of structural view independence. The sample mean satisfies the criterion but “trimmed means” do not.

### Reply to comments headed “Our concluding comments”

Chilton et al. concede that

“there is a definite and seemingly systematic divergence between direct and indirect estimates which is illustrated by the comparison between  $m_{xi}^{(2)}$  and  $m_{xi}^{(1)}$ .”

Requesting again that the Carthy study should not be regarded as “worthless”, they suggest that the useful contribution of the Carthy study lies predominantly in:

- (i) “providing evidence which, when blended with judgement, helped consolidate the VPF”, and
- (ii) “demonstrating the need and potential for further work to improve the methodology of stated preference elicitation in respects where limitations still undoubtedly exist.”

Claim (i) contains an admission by Chilton et al. that their recommended VPF was reliant on the exercise of their own judgement.

Claim (ii) contains a welcome acknowledgement by Chilton et al. of their method’s limitations.

The views of Chilton et al. on the limitations of stated preference techniques in general are also of interest.

### References

- Beattie, J., Covey, J., Dolan, P., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N., Robinson, A., Spencer, A., 1998. *On the contingent valuation of safety and the safety of contingent valuation: Part 1 – caveat investigator*. *J. Risk Uncertain.* 17, 5–25.
- Bernoulli, D., 1738, 1954. *Exposition of a New Theory on the Measurement of Risk*, *Econometrica*, 22, 1, 23–36. Translated by Dr. L. Sommer from “Specimen Theoriae Novae de Mensura Sortis”, *Commentarii Academiae Scientiarum Imperialis Petropolitanae, Tomus V* (Papers of the Imperial Academy of Sciences in Petersburg, vol. V), pp. 175–192, Available at <http://www.jstor.org/stable/1909829> (accessed February 2014).
- Carthy, T., Chilton, S., Covey, J., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N., Spencer, A., 1999. *On the contingent valuation of safety and the safety of contingent valuation: Part 2 – the CV/SG chained approach*. *J. Risk Uncertain.* 17, 187–213.
- Chilton, S., Covey, J., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N., Spencer, A., 2002. *Public perceptions of risk and preference-based values of safety*. *J. Risk Uncertain.* 25 (3), 211–232.
- Chilton, S., Covey, J., Jones-Lee, M., Loomes, G., Pidgeon, N., Spencer, A., 2015. Response to Thomas and Vaughan. *Process Saf. Environ. Prot.* 93, 293–298.
- Glover, D., Henderson, J., 2010. *Quantifying health impacts of government policies: a how-to guide to quantifying the health impacts of government policies*, Department of Health, available at: [www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/216003/dh\\_120108.pdf](http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216003/dh_120108.pdf) (accessed October 2014).
- Lipsey, R.G., Chrystal, K.A., 1995. *An Introduction to Positive Economics*. Oxford University Press, Oxford, UK.
- Popper Sir, K.R., 1934. *Logik der Forschung*, translated as *The Logic of Scientific Discovery*, 1959, Revised edition, 1980. Hutchinson, London.
- Spackman, M., Evans, A., Jones-Lee, M., Loomes, G., Holder, S., Webb, H., 2011. *Updating the VPF and VPIs: Phase 1: Final Report [to] Department for Transport*, NERA Economic Consulting, Available at: <http://assets.dft.gov.uk/publications/pgr-economics-rdg-updatingvpfvpi-pdf/vpivpfreport.pdf> (accessed May 2014).
- Thomas, P.J., 2014, January. *Structural view independence: a criterion for judging the objectivity of economic parameters measured by opinion survey*. *Measurement* 47, 161–177.
- Thomas, P.J., Vaughan, G.J., 2014. Testing the validity of the “value of a prevented fatality” (VPF) used to assess UK safety measures. *Process Saf. Environ. Prot.* <http://dx.doi.org/10.1016/j.psep.2014.07.001>
- Treasury, 2011, 2011. *The Green Book, Appraisal and Evaluation in Central Government*. The Stationery Office, London, Available at: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/220541/green\\_book\\_complete.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220541/green_book_complete.pdf)
- Wolff, J., Orr, S., 2009. *Cross-Sector Weighting and Valuing of QALYs and VPFs. A Report for the Inter-Departmental Group for the Valuation of Life and Health*, Final Report, 8 July 2009, <http://www.ucl.ac.uk/cpjh/docs/IGVLH.pdf>

P.J. Thomas \*

G.J. Vaughan

School of Mathematics, Computer Science and Engineering, City University London, Northampton Square, London EC1V 0HB, United Kingdom

\* Corresponding author.

E-mail address: [pjt3.michaelmas@gmail.com](mailto:pjt3.michaelmas@gmail.com) (P.J. Thomas)

Available online 27 November 2014

<http://dx.doi.org/10.1016/j.psep.2014.11.003>

0957-5820/© 2014 The Institution of Chemical Engineers.

Published by Elsevier B.V. All rights reserved.