

## **The New Worklife Expectancy Tables' Critique: A Rejoinder**

Gary R. Skoog and David C. Toppino\*

### **I. Introduction**

The New Worklife Expectancy Tables (Vocational Econometrics Inc. (VEI), 1987, 1991, 1995, 1998, 2002) (henceforth "the Tables") and the VALE program which embeds them continue to be the subject of controversy among forensic economists (FE) and litigators. A critique, written by these authors (Skoog and Toppino, 1999, henceforth "S&T", below), provoked a response, by Gibson and Tierney (2001) (henceforth "G&T", below), and subsequently has led to a public debate at the American Rehabilitation Economics Association meeting in Reno, ongoing comment and debate on the NAFE listserv and, perhaps most importantly, has resulted in additional research. James Rodgers (2001) has examined the Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP) to determine whether it is possible to produce meaningful worklife tables using either data source and concluded in the negative. Thomas Hale (2001), writing for the Bureau of Labor Statistics (BLS) in its own publication, the *Monthly Labor Review* (MLR), echoed many of the points made in the S&T article. Ciecka, Rodgers and Skoog (2003, forthcoming; "CRS", below) will expand upon several of the original S&T points. Ciecka and Skoog (2001) have shown the lack of scientific discipline in "molding," an essential element in the use of the Tables.<sup>1</sup> The present authors monitor and contribute to the National Association of Forensic Economics listserv (below, the "NAFE-L"), an active internet community of forensic economists, and have seen no defense of these Tables. To the contrary, we have observed that the NAFE-L community generally repudiates the Tables and their use in litigation.

This rejoinder is deemed necessary, not because there is doubt about matters of substance in the community of professional economists, but because of assertions appearing in plaintiffs' lawyers' briefs in Daubert and state court challenges to the Tables. These briefs erroneously contend that our original critique was "effectively and comprehensively rebutted."<sup>2</sup> While to most forensic economist readers of the G&T response this is a grossly inaccurate appraisal, we were not afforded the opportunity to provide a rejoinder in the

---

\*Respectively, De Paul University, Department of Economics and <http://legaleconometrics.com>; and the principal of Vocational Economic Analysts, Sherman Oaks, CA. The authors acknowledge without implication for any remaining errors, the comments of James Ciecka, James Rodgers and Tom Ireland. We are especially indebted to Robert Male and an anonymous reviewer for many contributions in both content and expositional clarity.

<sup>1</sup>G&T, p. 309 "Use of the Tables" offers a basis from which the expert can "mold" an opinion.

<sup>2</sup>Plaintiff's Motion Opposing BNSF's Motion to Exclude Dennis J. O'Donnell, PhD., in *Nielson v. BNSF*, Cause No. CV-00-109-BLG-RFC, United States District Court For the District of Montana, Billings Division, p. 12.

*Journal of Forensic Economics (JFE)* when the G&T response was published. Some therefore seem to confer more credibility upon the response than it deserves. This rejoinder is also apparently needed since many of the arguments in the original paper were either not understood by Gibson and Tierney, or were ignored. Therefore, in the spirit of helping G&T (and others) fully understand all of the serious flaws in the Tables, we here offer additional clarifications and elaborations.

In the 10 sections of this rejoinder we examine the major points brought up in G&T. We will show that their positions are unsupported, as though grounded upon quicksand. We also comment on recent developments and add several empirical references in support of our original thesis. Hopefully, this rejoinder will sufficiently focus the debate so that readers will no longer be diverted from the main issues by the selective choice of topics and their treatment in the G&T response.

## **II. Intent: The CPS Was Not Designed to Measure the Effects of Any Kind of Disability**

G&T spend two of their eight pages on our statement that “the CPS was never intended as a tool to measure the existence or impact of disability.” They claim that our “contention that the CPS was not intended to identify work disability is clearly wrong.” They cite, as do we, two 1980’s publications of the U.S. Census (1983, 1989) on “work disability.” G&T appear to be asserting:

1. if a government publication reports a statistical cross-tabulation, that it must reflect “intent” in design of the survey instrument; and,
2. that the published statistics associating lower employment with the “work disabled” must have the causal interpretation they wish to give them.

We thought it was obvious that neither presumption is correct, and ask the reader’s forgiveness at having to elaborate upon what to G&T was not self-evident.

The “intent” of any survey constitutes indirect evidence regarding its suitability for a specific purpose. It is very difficult to design a survey that will truly measure what is intended. It is even more difficult to be able to validly use survey responses for purposes that were not contemplated in its original design. When scientific (read valid and reliable) surveys are constructed and refined, the questions and methods should undergo cognitive and psychometric analyses to determine if the participants are actually responding to what the questioners think they are asking. Field workers are also trained and tested for inter-rater reliability. Designing valid and reliable questions and methodologies for a concept as fraught with ambiguity as impairment or disability is exceedingly difficult. It is most unlikely that the CPS survey accomplished what G&T have attempted to bestow upon it, lacking as it does the necessary specific intent and design elements. If the CPS has in fact accomplished the tasks that G&T have asserted, and if this can be demonstrated through appropriate follow-up testing, then we would agree that data from it could be used as asserted by G&T, regardless of the government’s “original intent.” However, this is not the case and the “wishful” assertions of G&T do not make it so.

For example, as long as one has an appropriate (probability or stratified) sample, one may compute statistics and apply standard statistical theory without each statistic having been contemplated in advance by the sample design. Transition probabilities for the increment-decrement model fall into this category. G&T find this necessary characteristic merely “interesting” (p. 310) and then avoid any serious attempt to deal with it. One of our main points (see the Hamel quotation, (p. 247) and our paraphrase and emphasis, (p. 248) included introduction of the concepts of “bias” and “measurement error.” These are elaborated upon below, under the term “sample selection bias.” While the entire sample of the CPS data is not biased, various *sub-samples* may be. For the purpose of calculating employment probabilities of groups of “work-disabled,” the VEI sub-samples are non-random and suffer from sample selection bias.

The S&T article and the G&T response have offered two very different theories about governmental intent, with implications for the usability of the CPS data. Who was right? Hale’s (2001) article “*The Lack of a Disability Measure in Today’s Current Population Survey*” implies that, if the government meant to measure disability through the CPS, it failed miserably. Hale states: “The current CPS questions on work limitation and disability income did not begin with *any* definition of disability” (p. 39, italics in original). This is a clear statement regarding government intent-and evidence in favor of the S&T theory.

G&T quote the Census P23-127, page 1-3 as follows: “...The purpose of adding questions about disability status was to identify those persons who should be asked about their receipt of disability income” (p. 311). This is precisely our point, and we are grateful to G&T for helping us make it. It also reflects G&T’s lack of knowledge concerning scientific survey research. Survey questions may be dichotomized into “characteristics questions”—those used to produce publishable statistics resulting in various counts of attributes—and “screeener questions,” those used to cast a wide net from which additional questions will separate out individuals for particular additional questions.<sup>3</sup> Here, the CPS interest or intent was not whether one is suffering a lawsuit related condition,<sup>4</sup> or is even impaired, much less “work disabled.” The CPS interest *was* in whether the respondent was collecting disability related income. Thus, the questions were not designed to be able to study the prevalence of *work disability* in the entire population or even in meaningful subpopulations. This is precisely what VEI purports and needs to measure. Researchers should only cross-tabulate with “characteristics questions,” *not* “screeener questions.” The failure to understand this point and the need to seriously address the relevant concept of “prevalence” is absolutely fundamental to the valid use of any data set. We provide additional elaboration in the next section.

Regarding the second major point several paragraphs earlier about causal interpretation: this issue is the lack of exogeneity of a self-declared disability, or more precisely the endogeneity of this variable. Bound (1995) finds that

---

<sup>3</sup>Tom Hale has been helpful in emphasizing this distinction.

<sup>4</sup>In a lawsuit, we observe *the condition or impairment*—not “work disability”; the latter is already conclusory. We should more accurately be speaking about impaired worklife tables rather than disabled worklife tables so as not to prejudice opinions from the start.

“when self-reported measures are used, health seems to play a larger role and economic factors a smaller one than when more objective measures are used” (p. 106). Stern (1989) finds evidence of endogeneity in the CPS (p. 361), but also finds other problems with the questions: “evidence that some respondents do not understand the questions, possibly because the questions are unclear” (p. 362). Citing Sickles and Taubman (1986), Stern also points to another potential source of endogeneity—“boredom or general lack of activity causes health deterioration when out of the labor force” (p. 362). We touched on such considerations in our point 1 where we said that the “health problem” answer is “not validated with other objective measurements of disability” (p. 245). Indeed, Hale in personal correspondence has indicated that a CPS household head’s 15-year-old child might have made this declaration rather than the allegedly disabled person, suggesting the extent of noise in these data.

### **III. More G&T Misunderstandings: Monthly Survey/ March Supplement, Prevalence, and Sample Selection Bias of the VEI Estimate of Means—“Descriptive Statistics”**

As we stated,

Neither the Current Population Survey nor its more extensive March demographic supplement (the basis of the Tables’ worklife probability calculations) are meant for, or collected for, the task of measuring the employment status of persons with disabilities. (p. 245)

We were called “clearly wrong.” (G&T, p. 311) However, in light of the distinction between “screener” and “characteristics” questions above, we were clearly correct. G&T’s lengthy appeal to the “venerable” 1983 Census P23-127 document offers no contradiction as they intend; it is just irrelevant, addressing as it does the desirability of household surveys in disability research.

The S&T article quoted Hamel (p. 247) about concerns he had regarding the lack of specific disability questions in the monthly survey and the March supplement. G&T attempt to paint Hamel’s criticisms as being directed only to the “*monthly* survey, without recognition that The Tables use the March supplement” (p. 312). We were aware of this, and that is why we quoted Hamel as stating *of the supplement*:

Even those data, however, would not provide overall estimates of the disabled population or work force. *One problem is that there is no information on persons with disabilities that do not limit the kind or amount of work they can do.*

G&T asserted that “S&T confuse the issues” but it is they who are confused. First of all, they obfuscate by focusing only on the first sentence, while it is the second italicized sentence that destroys the statistical legitimacy of the simple averages found in their Appendix C, the calculations at the heart of the VEI Tables and VALE’s Worklife Probability calculations.

G&T seem to associate the phrase “overall estimates of the disabled population or work force” with “prevalence.” They then dismiss “prevalence” using a

quotation from the 1989 Census P23-160 publication: “CPS data are not the best source for prevalence estimates.” If by “prevalence” they mean a measure of what percentage of the U.S. population has ADA-like conditions, then of course the CPS cannot answer this. As noted above, surveys rarely (if ever) do a valid job of answering questions for which they do not collect appropriate data. G&T say that their Tables instead use “descriptive statistics of the work-disabled population.” This is a confusing and seriously misguided statement. Once again it shows G&T’s fundamental lack of knowledge concerning scientific research. First, “descriptive” is a dismissive word for a statistic in scientific and academic circles, as in *here are my structural equations, estimates, standard errors, and hypothesis tests, and, by the way, if you are interested, here are some descriptive statistics—the means of the dependent and independent variables, to give you a feel for the sample*. Science is interested in structural estimates with causal interpretations. For their Tables to have any validity, their statistics must be based on sensible (read statistically *consistent* at a minimum, if not statistically *unbiased* or *efficient*) structural parameter estimates. They must answer the question *if one has this education, sex, age, and impairment/condition, what is the conditional probability such a person will be employed?*

To estimate the latter probability, VEI and G&T cannot escape the concept of prevalence—although not, to be sure, ADA prevalence. They need to find a representative (random or stratified) sample from each such sub-population of interest (those with the same age, sex, education group, and with a same or very similar condition or impairment) and they must be able to validly estimate the subpopulation mean.<sup>5</sup> This is indeed a prevalence ratio—the percentage of the sub-population that is employed. Roncs (1981), Hamel (1994), S&T (1999) and Hale (2001), all make, and spell out, this point in varying degrees of detail. The complete failure of VEI and G&T to understand these fundamental issues, let alone address them scientifically, renders the Tables useless.

#### IV. Two Wrongs Do Not Make a Right and Other Odds and Ends

In citing the U.S. Census (1983) publication which predated the original 1987 Tables by four years, we (hopefully) could not have left readers with the impression that VEI invented these disability definitions, as indicated in G&T at page 313 under “source of criteria.” When “veteran’s disability” became a seventh prong in the disability definition, as opposed to being implicitly included among the six mentioned in the P23 publications, is entirely unimportant. What is interesting about the veteran’s disability category, but is not present in the G&T response, is that it serves as a poster child for heterogeneity. As Gibson noted in his 2001 paper, the probability of employment of “disabled” Veterans at 91% mimics the percentage of the overall population. Its inclusion biases the non-severely disabled tables upwards.

---

<sup>5</sup>Perhaps it is G&T’s view that means or averages are lowly or mere “descriptive” statistics, and that no thought need be given to their estimation. This status may be elevated by noting that they are a special case of regressions. In any event, in the presence of a non-random sample, as emerges here with sample selection bias, the consistent estimation of the mean is anything but pedestrian, since the garden variety sample average VEI report is biased and inconsistent.

We never claimed that VEI was the *only* group to misuse the CPS data. The upshot of the G&T footnote 6 (p. 313) is that we should add Burkhauser et al. (2001) to the list. Indeed, two wrongs do not make a right and these academics have come under criticism from Hale for this very reason. However a careful reading of Burkhauser, et al., (2001) shows a different and more cautious approach than that used by VEI. They do not use the levels of the data, but their first difference or the trend. They do this in the hope, but not the certainty, that whatever may cause biases in the levels will be stable over time so that inferences about trends will be more valid. This signals an awareness of the problems we have been discussing. They also perform a preliminary cross validation analysis with National Health Institute Survey (NHIS) data. The net result is that VEI and G&T claim much more than Burkhauser (since they estimate levels), and they deliver much less (providing no cross validation).

To briefly respond to some other issues raised by G&T, we offer the following comments. Use of the LPE model (G&T, p. 311) by VEI should not be accepted as a source of pride. The LPE model's failure to use all of the state (status) information, as well as its falsifiable and falsified predictions about equalities across transition probabilities will be treated elsewhere. The hypothesis tests of the six-year averages mentioned by G&T (p. 315) were not meant to discern trends from cycles, but rather to illustrate the gross errors that result from ignoring trends in forecasting—and to demonstrate that such trends were consistent with their data. Many scientific researchers would say that 6 or 10 data points, with 2 of the degrees of freedom used up in estimation will not result in enough power to reject any null hypothesis. Moreover when so small a data set finds a trend—or any departure from an independent and identically distributed (i.i.d.) specification—to be statistically significant, one had better pay attention. G&T misread our focus on trend as opposed to cycle or some other departure; the point is that *any* departure from their i.i.d. specification causes them problems. The VEI practice of “averaging everything in sight” can be ruinous if no attention is paid to the underlying statistical model. The lack of lifetime employer-employee relationships cited by G&T (p. 315) is not accompanied by evidence, let alone evidence pertaining to people with disabilities. Since most forensic economists treat unemployment in the base level of income, G&T's point is irrelevant, even if true. The chronic and severe disability issue mentioned by G&T (p. 314) would most likely appear in conjunction with the severely disabled. However, we are correspondingly told that those tables should be abandoned in such a case—overridden with the value of 0, causing new sets of problems. A thorough discussion of this will be taken up in CRS (2003).

#### **V. The Vague Health Problem and General Lack of Validity and Reliability**

We said the health problem question was vague (p. 245, point 3), based on a common sense reading of the questions. Even if this question were understood perfectly by field worker and responder, the other problems we have highlighted would not permit use of these Tables. The G&T reply is that “81%

of those responding positively to this question also responded positively to one of the other six questions,” (p. 314.) We have six responses:

1. Hale (2001, p. 39) lists, in addition to the lack of validity, “uncertainties” about the question, including the temporal occurrence, the temporary or permanence of the condition, failure to report, and cross-cultural variation.
2. Hale has asked for, but never been given access to, the construct validity and psychometric testing of the reliability of these questions.
3. Table 3-4 in Mathiowetz (2000) quotes Haber (1990) in giving the percentage of the U.S. population classified as having a work disability between the years of 1966 and 1988. The percentage varies from 8.6% in the March, 1984 CPS to 17.2% from the SSA studies of 1966 and 1978. Evidently there are real problems regarding a definition of this concept—and this was pre-ADA.
4. Table 3-5, quoted from McNeil (1993) linked the U.S. Census 1990 long form and the Content Re-interview Survey, a cross validation follow-up, which asked about the “limited in kind or amount of work or prevented from working” questions. Only 68% of those answering that they had the limitation on the long form indicated the limitation in the re-interview, demonstrating a lack of reliability with the same kind of question used in the CPS.
5. Sections VI-VIII below are devoted to problems in the SIPP survey, thought previously to be the “gold standard,” since its disability questions are “characteristics” questions. It is highly unlikely that surveys where the disability questions are mere “screeners” would be any more reliable or valid.
6. The Workshop on Survey Measurement of Work Disability: Challenges for Survey Design and Method, which met in Washington, DC, May 27-28, 1999 would not have been necessary but for these severe reliability and validity problems, nor would the corresponding Executive Order discussed below have been issued.

Hale (2001), further points out

The Bureau of Labor Statistics recognizes that an adequate measure of the employment status of persons with disabilities may not exist, either from the CPS in its current form, or from other surveys....The Executive Order (13078) requires that meaningful measures of individuals with disabilities be accurate and reliable. The first steps necessary to produce meaningful statistics are being undertaken. (p. 40)

Regarding the March supplement questions relied upon by VEI, Hale writes:

At first glance, these questions seem to provide a reasonable means of identifying the population with disabilities. However, a closer examination leads to the conclusion that the questions lack validity as identifiers of persons with disabilities. In the case of disability, a valid

question would be written in a way such that the people who are envisioned to have a disability would identify themselves when asked the question. Understanding the validity characteristics of questions is key to understanding what the data mean. (p. 39)

Hale appears to agree with S&T in concluding that the key CPS health reason and job limitations screener questions are vague, lack reliability and validity, and are in need of revision if they are to be used to identify a work or any other disabled population.

## VI. John McNeil Closely Read

G&T (p. 312) paraphrase a passage from an affidavit signed by John McNeil (2001) as a form of validation for their use of CPS data in the Tables. Unfortunately, they are again guilty of misinterpreting the comments of government employees (in this instance a retired one), and the absence of validation research as if it helped their position. In the very paper mentioned in McNeil's affidavit (2000) he wrote:

The CPS is designed to measure changes in employment status, but there are no questions in that survey that are intended to identify individuals with disabilities. It is true that there are questions each month that identify individuals who are reported to be not in the labor force because they are disabled, and there are questions each March that ask if anyone in the household has a condition that prevents them from working or which limits the kind or amount of work they can do, but these items were not intended to, nor do they, identify the population that would be considered to have a disability *under the ADA definition*. When SIPP is viewed as an instrument for measuring changes in the employment rate of individuals with disabilities, certain weaknesses are apparent. First, as described above, there are questions about the reliability of the disability data obtained in SIPP. *It is most unlikely that this problem is unique to SIPP*, but it is a problem that affects SIPP. (p. 6)

Regarding the first words we have italicized above, the qualification, while true, is not necessary, as Hale and others have pointed out—*no disability population, "work", or ADA, or other, is identified in the CPS*. Thus, statements in the first paragraph of the McNeil affidavit do nothing to bolster G&T's assertions because they do not speak to the real issue, articulated in the previous sentence in italics. The second italicized words actually appear to *impeach the implications drawn from the McNeil's affidavit*—since they suggest that the CPS has escaped being determined to lack reliability and validity only because no one has looked! Since no one has looked, McNeil can safely say he cannot *know* of a reason the CPS data could not be used—but his wording suggests that he suspects it. Such parsing of words to produce misleading inferences is a low point in this discussion and sad commentary on the process engaged in here. Of course, it is not surprising that no validity and reliability testing has been

attempted for the screener questions of the CPS, since researchers are not supposed to produce counts and ratios of counts from “screener” questions.

In a 1993 paper McNeil, quoted by Mathiowetz (2000), questioned the validity of the SIPP data and the “health problem” question generally, by pointing out that in failing to take into account environmental factors, people who could work may say they could not because of lack of transportation. Also, people who did not work might say they could if the environment were different. Question ambiguity begets invalidity. In 1993, McNeil argued for the lack of validity, and his 2000 Western Economics Association paper demonstrated unreliability. Additional studies attributable to McNeil, documenting the lack of reliability, will be cited in Section VII. In light of these studies, we wonder what McNeil was thinking about in constructing the second paragraph of his affidavit.

McNeil’s second paragraph is slippery for another reason. The last U.S. government worklife table, produced in a different government agency (the BLS), was published in 1986. Its author, Shirley Smith (1986), has long since left, and her position has remained unfilled. Mr. McNeil worked and retired from the Census Bureau, and he has claimed no professional involvement in mathematical demography or worklife table construction. He may as well have said he knew of no reason why the CPS statistics *could* be used for this purpose. More telling, his affidavit is silent about the point we raised (S&T, p. 245) which has not been answered: the printed disability publications cited in the G&T response are old (1983 and 1989) contemporaries of Bulletin 2254 and the original Tables. In his three published disability studies in the last decade, for 1991-2 (P70-33), for 1994-5 (P70-61) and for 1997 (P70-73), McNeil abandoned the CPS data and switched to the SIPP data. Why? As we have pointed out previously, there are numerous statistical reasons that would mandate this change. This abandonment of the CPS for work disability was not a “McNeil only” decision—a Hale, Hayghe and McNeil (1998) paper used the SIPP data, presumably because it asked its disability questions of everyone in the sample, avoiding the sample selection bias inherent in estimating “work disability” from the combination of the monthly and March supplement CPS data. For this statistical reason, and its breadth of coverage, we recommended the SIPP as a possible basis for useful disability research. However, since our paper, we have learned of severe reliability issues in the disability module of the SIPP mentioned above, and must now caution against its use. Thus, even though it is not our burden, we have cited here additional evidence establishing once again the lack of reliability or validity for these CPS questions.

## VII. Testing for Validity, Reliability, and Sources and Rates of Error

Unlike other sections of this response, the comments offered here are of a general nature, and meant to introduce forensic economists (“FE’s”) to concepts which, for many, may be new terrain. The words “reliability” and “validity” were two terms of art that did not have much meaning for FE’s before Daubert,

Joiner and Kumho. Now, FE's may be asked in depositions about the reliability and validity of their methods.

For two reasons U.S. government statistics have been under the radar screen of testifying experts. First, the Courts' doctrine of "judicial notice" has meant that government statistics are presumed valid and reliable. Secondly, under Rule 703, if experts routinely in the course of their profession rely upon a source, it was deemed acceptable. While these remain arguments in favor of government statistics, in a post-Daubert world where everything is subject to challenge, these arguments may no longer carry the day.

There are tensions and contradictions aplenty. It is quite possible to lose sight of the comment attributed to statistician John Tukey that "the best is the enemy of the good," and to permit overaggressive advocacy to set the bar so high that no statistic can ever be admitted. An antidote may be the "delusive exactness" doctrine in Pfeifer (1983, p. 552), and the common sense position that, like prediction, opinion testimony requires by its nature some use of "rough and ready" (Pfeifer, p. 546) efforts or approximations. A related tension involves expert testimony itself. One can define the problem so that no experts exist. For this discussion, by combining requirements in survey research, disability subject matter considerations, psychology, psychometrics and general statistical theory, one might claim a need for 5 PhD's before one gets to the application at hand. Worklife expectancy construction is itself a cross disciplinary subject contributed to by econometricians, actuaries and mathematical demographers.

Borrowing heavily from Mathiowetz (2000), within the field of survey methodology, errors may be classified as follows:

#### THE SURVEY RESEARCH PERSPECTIVE

##### Errors of Non-observation

1. Sampling Error
2. Coverage Error
3. Non-response Error

##### Errors of Observation

1. The Questionnaire
2. The Respondent
3. Interviewers

#### THE PSYCHOMETRIC PERSPECTIVE

Validity is defined as "the correlation between the true score and the respondent's answer over several trials" (Groves, 1991, p. 8)

Different validity measures are:

1. content (face validity)—comprehensiveness and relevance of questions asked;

2. construct—extent to which the questions get at the correct underlying concept;
3. concurrent—a “criterion related” measure of the correlation of the item measured to another item, event, or behavior at the same point in time;
4. predictive—same as 3 but the second event is at time t+1;
5. criterion—connection with the “gold standard,” if one exists.

Reliability is defined as “the ratio of the true score variance to the observed variance, where variance refers to variability over persons in the population and over trials within a person,” (Bohrnstedt, 1983).

Different reliability measures include:

1. internal consistency (coefficient alpha, or Cronbach’s Alpha)—the extent to which all items in a scale measure the same underlying concept; only applicable to multi-Likert scales
2. test-retest -same person tested under same measurement conditions
3. inter-rater—the consistency of outcomes with respect to different raters

We must ask whether the CPS disability questions have undergone “appropriate” testing for validity and reliability. Indeed, in light of the previous outline, what testing is even appropriate? Hale (2001, p. 39) writes:

The work limitation and income questions in the March supplement might identify a subset of the disability population (an untested empirical question), but they are not likely to capture the larger population with disabilities.

The word “subset” implies that sample selection bias remains, and the “untested” adjective implies that reliability and validity testing have not been done. In offering new and allegedly “scientific” results, it is the burden of G&T and VEI to produce this testing, rather than to simply assert that it must have been done because “Questions in the Current Population Survey were developed at great expense and considerable scrutiny...”(p. 316). Returning momentarily to the McNeil 2000 paper (p. 17-18), we find:

. . . the major conclusion of this paper regarding measurement over time of the employment status of individuals with disabilities is that there is currently no satisfactory vehicle for producing such a measure. In fact, the Presidential Order dated March 13, 1998 established a National Task Force on Employment of Adults with Disabilities and instructed the Bureau of Labor Statistics and the Census Bureau, in cooperation with other selected Federal agencies, to design and implement a statistically reliable and accurate method to measure the employment rate of adults with disabilities.

Concerning the Task Force, in the publication *National Disability Policy: A Progress Report, November 1999–November 2000* by The National Council on Disability, we read:

On the other hand, some current and major methods used to track employment changes over time (e.g., the Current Population Survey or CPS), according to key government researchers, are not appropriate tools to use to measure employment of people with disabilities over time. For example, there are no questions in the CPS that identify the group that is commonly thought of as the disability population...therefore, because current CPS questions do not provide a valid measure of the disability population in the first place, researchers' conclusions about the employment rate trend for the disabled population and their underlying trends are not empirically supported. (p. 16)

Thus, it seems very apparent to knowledgeable researchers that valid and reliable statistics for people with disabilities do not exist at present.

### VIII. Is the Work Limitation Temporary or Permanent?

S&T pointed out (p. 243), that the issue of permanency of work limitation proves fatal for use of the Tables when used to compute the difference of work-life expectancy between the pre-accident and post-accident states. The VEI methodology assumes a "worklife expectancy" constructed on the assumption that, but for the accident, plaintiff was not, and never would have become disabled. Further, because of the incident, plaintiff is, and will remain, disabled forever. The former assumption is always inappropriate—but for the accident, there was the potential for other disabling (and non-compensable) injury. Neither is the latter assumption a valid generalization, and it cannot be shown to be valid using CPS data.

G&T's section "temporary in nature" (p. 314) serves, in their view, to allegedly "debunk this speculation" (S&T's assertion that people move into and out of disability states). G&T's arguments on this issue are not the least bit persuasive. First, their definition of "non-severe disability" is ill-suited to *ever* moving from the "not severely disabled" state. If the respondent has *ever* left a job for health reasons, he/she is *forever* in this disability category, by definition. (One might say that the definition lacks even *face validity*.) G&T write that:

... persons with a work disability are more likely to move from employed to not employed and are less likely to move from not employed to employed showing deteriorating, rather than improving, employment. (p. 314)

These may be correct assertions. If valid and reliable data of this nature were available it could be useful in a model that expands the state space to include activity/inactivity status interacted with, say, three disability states, within the framework of an increment-decrement model. Such a suggestion raises some interesting logical and modeling questions that the first author has taken up elsewhere. But this is *not* their model. The VEI model does not logically allow any such transitions.

Rather, their model makes the assumption of no disabled/non-disabled transitions except once, because of the event giving rise to the lawsuit—a counterfactual. Their second point "debunking this speculation" is an admission

that their model is false, and that such transitions do occur. In response they simply claim that the effect is small while making no attempt to describe or quantify its magnitude. Furthermore, their first “fact” about transition into higher unemployment for the “worklife disabled” has two problems:

1. it unfortunately requires abandonment of generally accepted worklife definitions, which do not factor in unemployment; and
2. for sample selection reasons G&T are not entitled to make this assertion for the relevant sub-population until they correct for this problem; this is unfortunately impossible from the CPS as designed.

John McNeil (1998) has computed several tables comparing SIPP measures at times 1 and 2 (“T1” and “T2”, below) to check *reliability* of the responses. Table 3 looks specifically at work disability status, and finds that from the 92/93 to the 94/95 “waves” in the SIPP, of those with a non-severe limitation at time T1, 14.3%, or 1/7, report no limitation at T2 one year later; 72.5% stay the same, and 13.2% move to a severe limitation.<sup>6</sup> Of those with a severe limitation at T1, 6.1% move to no limitation while 15.3% progress to a non-severe limitation by T2. This represents considerable state movement and is a stinging refutation of the VEI/G&T model assumption of a static disabled state. Since this is a representative national sample with among the best disability questions, one would expect to find similar or greater lack of “state” reliability in the CPS data, if the latter’s self selection bias could be overcome. Thus significant and relevant evidence has been found against the assumption of no disability state transitions that are built into the Tables. These facts also refute G&T’s charges that the S&T paper was “only a polemic. The arguments presented are mere red herrings, offering no meaningful statistical support or analyses” (p. 316). Such a charge is more correctly leveled against the VEI Tables and their response to our critique.

### IX. Three Econometric Problems That Will Not Go Away

Borrowing from the econometric terminology in CRS (2003), regarding the “work disability” definition of VEI (and the Census before 1990), the estimation of probabilities of employment from the Current Population Survey is riddled with sample selection bias, heterogeneity and endogeneity. These are serious statistical problems. Probable solutions in other contexts have been cited by Nobel Prize winners on at least three occasions.<sup>7</sup>

In his 2000 Nobel lecture, Professor James J. Heckman cited all three of these concepts:

---

<sup>6</sup>The chance that one remains in the non-severely disabled for  $n$  periods is  $(6/7)^n$ ; thus, for someone 20 years away from retirement, the chance of The Tables correctly classifying this individual is *de minimus*.

<sup>7</sup>Sample selection bias within the context of limited dependent variables was recognized and corrected for in estimation in Heckman (1979) and cited in his Nobel announcement in 2000. Endogeneity, or the lack of exogeneity of regressors, was addressed in the simultaneous equations literature developing out of the Cowles Commission, prominent members of which were Tjalling Koopmans and Tyrgeve Haavelmo, Nobel winners in 1975 and 1989, respectively.

1. "Accounting for *heterogeneity* and diversity and its implications for economics and econometrics is a central message of this lecture and a main theme of my life's work." (p. 675);
2. "Assumption 1 ( $X, W$ ) independent of  $U$ ," (p. 682), is a definition of *exogeneity*, where  $X$  are assets, demographic structural variables, etc. and  $W$  is the wage, and  $U$  is an unobservable, a statistical disturbance or "error term" in a structural equation, from the point of view of the economist (p. 679); "As we shall see, unobservables play a big role in micro-econometrics. There is much evidence that unobservables are empirically important. Modern micro-econometrics is devoted to accounting for them" (p. 679). Here,  $U$  would reflect the particular kind of disability—more information than can be affordably collected in a survey, as well as other unobserved person specific effects, perhaps including accommodation, healing and environmental factors;
3. "*Selection bias* arises in estimating structural models with partially observed outcomes. But the problem of selection bias is more general and can arise when a rule other than simple random sampling is used to sample the underlying population that is the object of interest. The distorted representation of a true population in a sample as a consequence of a sampling rule is the essence of the selection problem" (p. 690).

Reasons for the departure from random sampling have been discussed repeatedly.

The New Worklife Expectancy Tables and VALE are intended to be used to determine the worklife of an individual who claims an injury, disorder of a medical nature or a diagnosis—in short a *condition*, pursuant to a lawsuit, where the condition may or may not separate him or her from the overall population. S&T used the word "impairment" to indicate this kind of specific physical or mental condition. More accurately we could have used the word "condition." Our usage, however, was more in keeping with modern conventions than VEI, who refer to this situation as a "disability." One common professional application (following Nagi, 1965) uses "disability" as the limitation in the ability to carry out regular activities, and is a function of the underlying condition and environmental variables. Our point has always been that to be able to separate an individual with a condition from the overall population, it is necessary to estimate the desired characteristic (participation or employment) from a random sample of individuals *with this condition*. Estimation of participation or employment probabilities from a population of people with the condition *and* other conditions will not pass statistical muster, since the estimator will be an average over a heterogeneous population different from the subject individual. This is the heterogeneity problem. Estimation of participation or employment probabilities from a population with the condition who *also* report difficulty working will obviously not produce an unbiased or consistent estimator, a conclusion which follows because of the under-sampling of those with the condition who have no difficulty working. This is a textbook illustration of the sample selection problem, and is intrinsic to the VEI definition of "work disability."

In other words, if we compute an average employment probability of those with the condition and a corresponding average for those without the condition, and interpret the difference in employments as being attributable to the condition, we are implicitly assuming that the only causation possible runs from the condition to the employment probability. On theoretical grounds, we would expect the presence of a condition to potentially effect employment (what is referred to in economics as a structural equation). But we must understand that in the CPS the condition is not determined by an objective observer, such as an independent medical examiner. Rather it is self-declared, by the person (or household member). Consequently, for at least part of the population, and so for the whole sample, there would be the (rebuttable) presumption of another structural relation running “backwards” from employment to self-declared “work disability” since there are economic incentives for those not wishing to work to declare a disability. There are also social incentives—the “justification hypothesis.”

Another mechanism resulting in feedback or reverse causation would obtain if better health emerged *from* labor force participation. In fact, this is a valid assumption given the proper re-employment fit post-injury. With a second structural relation, having the condition and participation become jointly simultaneously determined. In different terms, “having the condition” is no longer an exogenous variable in the first equation, the one VEI estimates by averaging, which is a special case of ordinary least squares. For the same reason that regressions of quantity on price do not generally estimate either supply or demand curves, the differences in measured employment rates will not represent the structural parameters VEI asserts. Instead, they represent a mixture of the responses of employment to the condition *and* the declaration of the condition to employment. Both endogeneity and grossly unreliable data are consistent with the findings reported in Baker, et al., (2001), that “more than 50 per cent of individuals who report having a disease in the NHPS (National Population Health Study) have no corresponding doctor’s diagnosis in the OHIP (Ontario Health Insurance Plan) records” and “more than 50 percent of the individuals who have a doctor’s diagnosis in the OHIP data fail to report having the disease in the NPHS” (p. 3).

It is important to recognize that these are three different problems, and that, if several studies were to rule out one of the three objections, the other two would remain. The burden is on the moving party, VEI, to test their sample and estimators in light of these objections. Instead, they have attempted to shift their burden onto these researchers. This is evidenced in their introduction (p. 310) and in their conclusion (p. 316). Asserting something, even repeatedly, does not make it so. Indeed, their conclusion terms this as “unfortunate and ironic, for The Tables provide the only published scientific (sic) basis from which to derive worklife expectancies for persons with permanent partial work disabilities” (p. 316). Our analysis suggests that their “scientific basis” requires wishful leaps of faith and the abandonment of scientific, economic and statistical theory. That only one organization would produce and promote Tables that are constructed with such invalid and unreliable methods speaks well of the rest of the profession and does *not* provide justification for their use.

### X. Conclusion

We have considered the G&T response by elaborating upon our original criticisms. We cite additional authority and evidence, and refer to other academic papers where similar conclusions have been reached. Whether one accepts appeal to authority, the logic of econometric and statistical theory, or empirical results establishing the lack of validity and reliability, the statistical methodology employed to construct the Tables does not meet any semblance of professional standards for economics, econometrics, or statistical research, let alone those for survey or psychometric research. In the end, the Tables and VALE are demonstrably invalid and unreliable uses of the underlying CPS database.

### References

- Baker, Michael, Mark Stabile, and Catherine Deri, "What Do Self-Reported, Objective, Measures of Health Measure?," National Bureau of Economic Research, Working Paper 8419, August, 2001.
- Bohrstedt, G. "Measurement," in Rossi, P. Wright, K. and Anderson, A., eds. *Handbook of Survey Research*. New York: Academic Press, 1983.
- Bound, John, "Self-Reported versus Objective Measures of Health and Retirement," *Journal of Human Resources*, 1991, Vol. 36(1) p. 106-138.
- Burkhauser, Richard V., Mary C. Daly, Andrew J. Houtenville, and Nigar Nargis, "Economic Outcomes of Working-Age People with Disabilities over the Business Cycle: An Examination of the 1980's and 1990's," March, 2001 working paper.
- Ciecka, James, James Rodgers and Gary Skoog, "The New Gamboa Tables: A Critique," paper presented at the WEA meetings, Seattle, WA, 2002 and submitted for publication.
- \_\_\_\_\_, Thomas Donley, and Jerry Goldman, "A Markov Process Model of Work-Life Expectancies Based on Labor Market Activity in 1997-98," *Journal of Legal Economics*, 2000, 9(3), 33-68.
- \_\_\_\_\_, and Gary Skoog, "An Essay on the New Worklife Expectancy Tables and the Continuum of Disability," *Journal of Forensic Economics*, 2001, 14(2), 135-140.
- Gamboa, A.N., "The New Worklife Expectancy Tables: Revised 1998 By Gender, Level of Educational Attainment, and Level of Disability," Louisville, KY: Vocational Econometrics, Inc., 1998.
- \_\_\_\_\_, "The New Worklife Expectancy Tables: Revised 2002 By Gender, Level of Educational Attainment, and Level of Work Disability," Louisville, KY: Vocational Econometrics, Inc., 2002.
- Gibson, David S., and John P. Tierney, "Disability and Worklife Expectancy Tables: A Response," *Journal of Forensic Economics*, Fall 2000, 13(3), 309-318.
- \_\_\_\_\_, "Daubert, Disability, and Worklife Expectancies," circulated to VEI subscribers, 2001. Available for download at: <http://www.vocecon.com/daubertwle.pdf>.
- Groves, R., *Survey Errors and Survey Costs*, New York: John Wiley and Sons, 1959.
- Haber, L., "Issues in the Definition of Disability and the Use of Disability Survey Data," in Levin, D., Zitter, M., and Ingram, L., eds, *Disability Statistics: An Assessment*, Washington, DC: National Academy Press, 1990.
- Hale, Thomas W., Howard V. Hayghe, and John M. McNeil, "Persons with Disabilities: Labor Market Activity, 1994," *Monthly Labor Review*, September 1998, 3-12.
- \_\_\_\_\_, "The Lack of a Disability Measure in Today's Current Population Survey," *Monthly Labor Review*, June 2001, 38-40.

- Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, 47(1), 153-61.
- \_\_\_\_\_, "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture," *Journal of Political Economy*, 2001, 1094, 673-748.
- Mathiowetz, Nancy, "Methodological Issues in the Measurement of Work Disability," Chapter 3 in N. Mathiowetz, and G. Wunderlich (eds.), *Survey Measurement of Work Disability*, Washington, DC: Institute of Medicine, 2000.
- McNeil, J., "Census Bureau Data on Persons with Disabilities: New Results and Old Questions about Validity and Reliability," Paper presented at the 1993 annual meeting of the Society for Disability Studies, Seattle, Washington. 1993.
- \_\_\_\_\_, "Selected 92/93 Panel SIPP Data, Time1=October 94-January 94, Time 2 =October 94-January 1995, unpublished tables," Washington, DC: Bureau of the Census, 1998.
- \_\_\_\_\_, "Employment, Earnings and Disability," paper presented at the 75th Annual Conference of the Western Economic Association, July 3, 2000, Vancouver, Canada.
- Nagi, S., "Some Conceptual Issues in Disability and Rehabilitation," in Sussman, M.B., ed., *Sociology and Rehabilitation*. Washington, DC: American Sociological Association, 1965.
- National Council on Disability, *National Disability Policy: A Progress Report, November 1999–November 2000*, Washington, DC: June 14, 2001.
- Rodgers, James R., "Exploring the Possibility of Worklife Expectancies for Specific Disabilities," *The Earnings Analyst*, 2001, 4(1), 1-36.
- Rones, Philip, "Can the Current Population Survey be used to Identify the Disabled?," *Monthly Labor Review*, June 1981, 37-38.
- Sickles, Robin and Taubman, Paul, "An Analysis of the Health and Retirement Status of the Elderly," 1986, *Econometrica* 54(6), pp.1339-56.
- Skoog, Gary R., "A Worklife Expectancy: Aggregation Theory, With Special Reference to Disability," presented at the WEA meeting, Seattle, WA, 2002.
- \_\_\_\_\_, and David Toppino, "Disability and the New Worklife Expectancy Tables from Vocational Econometrics, 1998: A Critical Analysis," *Journal of Forensic Economics*, 1999, 12(3), 239-254.
- Smith, Shirley J., "Worklife Estimates: The Effects of Race and Education," *Bulletin 2254*, Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics, February, 1986.
- Stern, S., "Measuring the Effect of Disability on Labor Force Participation," *Journal of Human Resources*, 1989, 24, 361-395.
- Toppino, David C., and William Reed, "A Proposal for Estimating Work Life Disability in the Disabled State," presented at the AREA symposium, Reno, NV, 2002.
- U.S. Bureau of the Census, Current Population Reports, P23-127, *Labor Force Status and Other Characteristics of Persons with a Disability, 1982*, Washington, DC: U.S. Government Printing Office, 1983.
- U.S. Bureau of the Census, Current Population Reports, P23-160, *Labor force Status and Other Characteristics of Persons with a Disability 1981 to 1988*, Washington, DC: U.S. Government Printing Office, 1989.
- Daubert v Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).
- General Electric Co. v. Joiner, 522 U.S.136, 1997.
- Jones & Laughlin Steel Corp. v. Pfeifer, 103 S. Ct . 2541 (1983).
- Kumho Tire Company v. Carmichael, 526 U.S. 137 (1999).