HARVARD MEDICAL SCHOOL
DEPARTMENT OF HEALTH CARE POLICY

RONALD C. KESSLER, PH.D.
PROFESSOR
180 LONGWOOD AVENUE
BOSTON, MA 02115-5899

TEL: (617) 432-3587
FAX: (617) 432-3588
E-MAIL: kessler@hcp.med.harvard.edu

March 8, 2003

To: Persons interested in HPQ scoring
Fr: Ron Kessler
Re: Scoring the HPQ

The HPQ contains questions about absenteeism, quality and quantity of work while on the job (often referred to as "presenteeism"), and critical incidents on the job. The latter includes work-related successes, failures, and accidents.

Based on calibration of the self-reported HPQ questions about absenteeism and presenteeism in a series of empirical studies described by Kessler et al. (2003), computerized imputation rules have been developed for adjusting self-reported measures of these outcomes to be more accurate reflections of true absenteeism and presenteeism. These adjustments are implemented in the HPQ master data file software.

It is also possible to calculate unadjusted HPQ scores using the methods described in this memo. Unadjusted scores are good approximations of adjusted scores.

Absenteeism

The HPQ asks a number of questions about hours and days missed from work. Hours are the main focus, as a "day" of work means very different things for people who are salaried versus non-salaried, full-time versus part-time, and those who work regular hours versus those who work split shifts and different numbers of hours on different days.

The first questions ask about typical hours expected to work each week and hours actually worked in the past 7 days. Subtracting expected hours from actual hours might be considered the most accurate way to calculate absolute absenteeism in the HPQ due to the short recall period. Similarly, the ratio of hours missed to hours expected might be considered the most accurate way to calculate relative absenteeism in the HPQ. However, these measures have two problems. First, the shorter recall period reduces statistical power compared to a measure that asks about a longer recall period. Second, the shorter recall period introduces some amount of bias into estimates due to the fact that people who are ill on the day of first receiving the HPQ survey are more likely than other respondents to postpone completion until they feel better. We can see this in the fact that payroll record measures of sickness absence in our calibration study samples consistently

show slightly lower rates of sickness absence on the week of completing the HPQ survey than on other weeks of the month.

As a result of these problems, the HPQ also asks about absenteeism over the past 28 days (four weeks). Despite the comments two paragraphs earlier about the problem with asking about days rather than hours of work, we begin these questions by asking respondents to add up all the days in the past 28 when they were at work, on vacation, on sick leave, etc. The task is to allocate days to each category and to have the total add to 28. It takes a bit of work to get this right. We created this task on purpose to force respondents into active memory search. We then ask a question about hours worked in the past 28 days as the very next question capitalizes on the active memory search that was required to answer the "days" question.

Researchers might want to use information in the "days" question to calculate either absolute or relative measures of absenteeism days. An absolute measure, as described two paragraphs above, calculates number of missed workdays, while a relative measure calculates the ratio of days missed over total days either at work or missed from work. It should be noted that a decision is needed as to whether only days defined as sickness-absence days should be counted as being "missed" or if all days missed from work should be counted. Our preference is to count all days. This is true for two reasons. First, more and more companies are offering consolidated benefits to employees that bundle sickness-absence days and vacation days. In situations of this sort, there is no real distinction between sickness absence days and any other days of missed work. Second, even when benefits are not consolidated, sick people often use vacation days as substitutes for sickness days when they have used up all their sickness days.

For reasons described above, we are more interested in hours worked over the past 28 days than in days worked. As with the hours and days measures described above, either absolute or relative absence measures can be calculated using the information obtained earlier in the questionnaire about hours expected to work.

Two problem areas involving missing data should be noted here. First, some respondents report that they have no expected number of work hours. They have a job to do and are rewarded for performance no matter how many hours it takes. External sales workers are a prototypical example. In a situation of this sort, we cannot calculate sickness absence based on hours. While we could make such a calculation based on days (as such respondents typically tell us the number of days they were sick over the past month), it is unclear whether this is legitimate in light of the fact that these workers have already told us that time spent working is not a central defining characteristic of productivity on their jobs. Researchers working with the HPQ consequently need to decide whether to delete such workers from all their analyses of absenteeism or to use them in analyses in which days off work are the outcomes. Note that it is also possible to carry out analyses of days missed twice – once with these workers excluded and the second time with them included – and to compare results in order to evaluate the implications of excluding this set of workers.

The second problem area involves the measure of hours worked in the past 28 days. Accurate responding to this question is intellectually challenging. The respondent has to remember days worked over a four-week recall period, calculate number of hours worked on those days, and then do some mental arithmetic to arrive at a calculation of total hours worked over the four weeks. Because of the complexity of this question, we find that a substantial proportion of respondents omit responding to this question, while others appear to be estimating rather than calculating. The latter conclusion is based on the fact that answers typically come in round numbers, a pattern that is usually indicative of estimating.

We have attempted to simplify the task of giving an accurate answer to the "hours in the past 28 days" question by providing a few examples of exemplar calculations. The percent of respondents who answer the question has increased in surveys that added these examples. In telephone surveys, furthermore, we use standard strategies of question decomposition to help respondents answer this question.

The ideal situation is for researchers to use a full range of absenteeism outcome measures in their empirical studies: both absolute and relative measures of hours in the past 7 days, hours worked in the past 28 days, and days missed in the past 28 days. The documentation of consistency of results across this range of measures can be used to assess insensitivity of substantive results to the unique limitations of the various measures.

Presenteeism

In the same way that we use the question about days missed in the past 28 days to encourage active memory search before asking about hours worked in the past 28 days, we use a two-part approach to measure presenteeism in the HPQ. The first part begins with a series of Likert scale questions that ask respondents how often, during their working hours, they had decrements in quantity and quality of work. These questions were purposefully phrased in rather general terms in an effort to make the questions relevant to all respondents. More concrete questions (e.g., problems lifting or falling behind in answering phone calls) run the risk of being much more relevant to some types of jobs than others, leading to differential precision as a function of job type. Presenteeism scales composed of concrete questions are to be preferred when the research focuses on a single occupation. When the goal is to make broader statements or to compare across occupations, though, more abstract questions of the sort asked in the HPQ are to be preferred.

These general questions are then followed with a series of yes-no questions about critical incidents that are described in the next section of this memo. The critical incident questions, like the Likert scale questions, are designed to encourage active memory search about good and bad performance over the past 28 days.

These memory-priming questions are then followed by a series of self-anchoring scale questions in which the respondent is asked to rate the average person working in their job

on a 0-10 scale of work performance (worst to best), rank themselves in terms of their usual performance, and, finally, to rank themselves over the past 28 days during the time they were at work.

As described below, responses to these questions can be used to calculate self versus other scores. However, ipsative data of this sort can also be gathered directly. This is done in the HPQ by asking two-part unfolding question about whether the respondent considers his/her work over the past 28 days better, worse, or about the same as the typical worker on the same job. If either better or worse, the respondent is then asked to rate whether that is a lot, some, or only a little better or worse.

As with absenteeism, presenteeism scores can be constructed in several different ways using responses to the above questions. First, responses to the Likert scale questions for a strong first factor in factor analysis, making it possible to create a presenteeism scale from responses to these variables based either on factor-based scaling, factor-weighted scaling, or Item Response Theory scaling.

Next, responses to the 0-10 self versus other questions can be used to calculate absolute and relative measures of presenteeism based on logic very similar to that used in the calculation of absenteeism scores. Absolute presenteeism can be calculated as the difference between the score for self over the past 28 days and the score for the average worker in the same job. A relative presenteeism score can be computed as the ratio of self versus other scores.

Responses to the two-part unfolding question, finally, can be used as an alternative coarsely coded measure of presenteeism.

Critical incidents

The HPQ includes a question about whether the respondent had a workplace accident at any time in the past 12 months. This comparatively long recall period is used because industrial accidents are rare and a long recall period is needed to learn about enough instances for meaningful analysis unless the sample is extremely large. Fortunately, due to the fact that workplace accidents are rare, they are also recalled with good accuracy, minimizing the problem of recall bias that is often associated with questions that have a long recall period.

As briefly mentioned in the last section, we also ask about critical incidents in the past 28 days. These include big successes, big failures, and mistakes that could have caused workplace accidents or injuries. These questions were asked because the costs of low performance at work cannot be captured entirely by questions about duration in relation to employee salary. When the Exxon Valdez ran ashore on the coast of Alaska, the pilot might have been having a day that would be considered 0 on a 0-10 scale of work performance, but the salary-equivalent monetized value of that performance decrement doesn’t begin to capture the enormous financial cost of that accident to the company. The

big success and big failure questions are designed to allow these costs to be captured in the HPQ.

The enormous variety of big successes and big failures is so great that we have included opportunities for open-ended responses in the HPQ to describe the specifics of these situations. We find that reports of these critical incidents are sufficiently rare, that there is no great burden in reading these responses and collaborating with employers in making evaluations of the financial implications of these incidents for the company. (This is always done using hypothetical scenarios in order to guard the confidentiality of respondents.)

Monetizing work performance reports

It is often of interest to monetize work performance reports in order to generate estimates of the financial impact of illnesses on the workplace. As noted recently by Murray et al. (in press), a serious effort along these lines requires deep institutional knowledge that is seldom available to researchers unless they have close collaboration with employers. Rough heuristic estimates can be made, though, by using information about employee salaries, which is collected in the HPQ, to estimate the salary-equivalent costs of lost time and reduced performance. As such measures can never be more than rough approximations in the absence of institutional knowledge, no best method of monetizing can be proposed. The HPQ reports use one set of approaches, but others are equally reasonable in light of our currently primitive level of development in this area.

References

Kessler, R.C., Barber, C., Beck, A., Berglund, P., Cleary, P.D., McKenas, D., Pronk, N., Simon, G., Stang, P., Ustun, T.U., Wang, P. (2003). The World Health Organization Health and Work Performance Questionnaire (HPQ). Journal of Occupational and Environmental Medicine, 45, 156-174.

Murray, J. F., Nicholson, S., Pauly, M., and Berger, M. L. (in press). Investing in Health to Boost Employee Productivity: The Employer's Perspective In Health and Work Productivity: Emerging Issues in Research & Policy. Eds. Kessler, R. C. and Stang, P. E. University of Chicago Press.