



Relations among questionnaire and laboratory measures of rhinovirus infection

B. Barrett*, R. Brown*, R. Volland*, R. Maberry* and R. Turner#

ABSTRACT: Due to high incidence and quality-of-life impact, upper respiratory infection substantially impacts on population health. To test or compare treatment effectiveness, a well-designed and validated illness-specific quality-of-life instrument is needed.

Data reported in the current study were obtained from a trial testing echinacea for induced rhinovirus infection. Laboratory-assessed biomarkers included interleukin (IL)-8, nasal neutrophil count (polymorphonuclear neutrophils (PMN)), mucus weight, viral titre and seroconversion. The questionnaires used included the general health short form (SF)-8 (24-h recall version), the eight-item Jackson cold scale, and the 44-item Wisconsin Upper Respiratory Symptom Survey (WURSS).

In total, 399 participants were inoculated with rhinovirus and monitored over 2,088 person-days. Statistically significant associations were found among nearly all variables. Between-questionnaire correlations were: WURSS–Jackson=0.81; WURSS–SF-8=0.62; and Jackson–SF-8=0.60. Correlations with laboratory values were as follows: WURSS–mucus weight=0.53; Jackson–mucus weight=0.55; WURSS–viral titre=0.37; Jackson–viral titre=0.46; WURSS–IL-8=0.31; Jackson–IL-8=0.36; WURSS–PMN=0.31; and Jackson–PMN=0.28. Neither WURSS nor Jackson yielded satisfactory cut-off scores for diagnosis of infection.

Symptomatic and biological outcomes of upper respiratory infection are highly variable, with only modest associations. While Wisconsin Upper Respiratory Symptom Survey and Jackson questionnaires both correlate with biomarkers, neither is a good predictor of induced infection. The inclusion of functional and quality-of-life items in the Wisconsin Upper Respiratory Symptom Survey does not significantly decrease the strength of association with laboratory-assessed biomarkers.

KEYWORDS: Common cold, quality of life, questionnaires, rhinovirus, upper respiratory infection, validation

The common cold is caused by viral infection of the upper respiratory tract. Rhinoviruses cause between 25 and 60% of cold episodes [1–4]. A weakness in cold research is due to the lack of well-developed and validated outcome measures. While many laboratory-measured biomarkers are available, very little work has focused on patient-oriented quality-of-life measures.

The Jackson scale is widely used. This index assesses eight symptoms (sneezing, nasal obstruction, nasal discharge, sore throat, cough, headache, chilliness and malaise), using a three- or four-point response range [5–7]. No items assess functional or quality-of-life domains. Validity, reliability and responsiveness have not been thoroughly assessed.

The Wisconsin Upper Respiratory Symptom Survey (WURSS) [8] was developed as an

evaluative illness-specific quality-of-life outcomes instrument, to measure change over time in domains most important to cold sufferers [9–13]. The 44-item WURSS (WURSS-44) comprises 32 items assessing symptoms, 10 functional items and two global assessment items, using 7-point Likert-type response ranges. Formal validity testing of WURSS supports reliability, responsiveness and external validity [14]. A short form (WURSS-21) is currently undergoing validation. WURSS is free of charge for nonprofit and educational use, but must be licensed by for-profit users (www.fammed.wisc.edu/wurss/).

WURSS scores associate more strongly with general health-related quality-of-life (short form (SF)-8) [15] and with the Jackson scale than either of these measures do with each other [14]. Using the minimal important difference (MID) framework [16–18], the current authors estimated that a

AFFILIATIONS

*University of Wisconsin, Madison, Madison, WI, and

#University of Virginia, Charlottesville, VA, USA.

CORRESPONDENCE

B. Barrett

Dept of Family Medicine
University of Wisconsin Medical School

777 S. Mills
Madison WI 53715
USA

Fax: 1 6082635813

E-mail: bruce.barrett@fammed.wisc.edu

Received:

January 07 2006

Accepted after revision:

April 08 2006

SUPPORT STATEMENT

The trial from which current data were obtained was supported by a USA National Institutes of Health (NIH) grant R01 AT001146 (National Center for Complementary and Alternative Medicine (NCCAM), Gaithersburg, MD, USA). R. Turner is the principle investigator. B. Barrett is the principle investigator on another trial supported by NIH NCCAM (R01-AT-1428-01), and is supported by the Robert Wood Johnson Foundation Generalist Physician Faculty Scholars program.

B. Barrett, R. Maberry and R. Brown hold partial copyrights on the Wisconsin Upper Respiratory System Survey instrument (which is free for educational and nonprofit use, but can be licensed through the Wisconsin Alumni Research Foundation by for-profit entities www.fammed.wisc.edu/wurss/). R. Turner has a number of grants and consultancies, but none are deemed a conflict of interest.

European Respiratory Journal

Print ISSN 0903-1936

Online ISSN 1399-3003

two-armed clinical trial using the WURSS-21 would require 74 participants to detect MID, compared with 92 for the WURSS-44, 124 for Jackson and 156 for SF-8 [14]. While encouraging, these findings should be verified in other settings, and are limited by the lack of assessment of associations with biomarkers.

MATERIALS AND METHODS

The data reported in the present study came from an induced rhinovirus (RV)-cold echinacea trial, which was reported as negative [19]. College-age participants susceptible to RV-39 were randomised to one of three types of echinacea, either as prevention or treatment, and were housed in hotel rooms from inoculation (day 0) until discharge on day 5. Protocols were approved by Virginia and Wisconsin ethics committees.

Interleukin (IL)-8, neutrophil count and viral culture were obtained from daily nasal wash, and were analysed using previously reported methods [20]. Mucus weights came from nasal tissue that was pre-weighed, distributed, then collected and re-weighed. Serum was collected on day 0 and ~3 weeks later to assess serological response. Jackson symptoms were elicited by study nurses twice a day using a four-point response range. Daily scores were defined as the higher of the two nurse-assisted ratings. The WURSS-44 was scored once daily by all participants. The first two batches of participants ($n=150$) scored the SF-36 [21–23] (4-week recall). When the 24-h recall version of the SF-8 [15] became available, the current authors substituted this more appropriate instrument, which was used in the last four batches ($n=249$).

The present study was guided by a conceptual framework in which biomarkers and self-reports are imperfect measures of underlying illness domains. For example, IL-8, neutrophil count and mucus weight reflect nasal inflammation, while viral culture and serology indicate infection. Self-reports reflect various illness domains, such as nasal congestion, sore throat or cough, or difficulties with thinking, breathing or carrying out daily activities.

One aim of the current study was to determine whether WURSS would predict biomarkers as efficiently as Jackson. As WURSS includes domains not specific to colds, it was possible that WURSS would correlate less strongly. Conversely, the expanded severity response range of WURSS might better measure underlying (continuous) domains, and hence yield tighter correlations. The authors were also interested in the abilities of WURSS and Jackson to discriminate between infected and noninfected participants.

Statistical analysis began with tabular and graphical portrayal of all variables for each study day. Outlying and missing data were assessed. Bivariate analyses were conducted using scatterplots, Pearson correlations and linear regressions.

To assess the relationships between Jackson, WURSS-44 and other measures, a bioequivalence approach was selected using the two one-sided test method described by SCHUIRMANN [24] and PHILLIPS [25]. Let ρ_1 be the correlation of Jackson with biomarkers, ρ_2 be the correlation of WURSS with biomarkers, and let δ_L and δ_U be the lower and upper bounds of

bioequivalence. The null hypothesis of nonequivalence is:

$$H_0: \rho_1 - \rho_2 < \delta_L \text{ or } \rho_1 - \rho_2 > \delta_U \quad (1)$$

and the alternative hypothesis of bioequivalence is:

$$H_a: \delta_L < \rho_1 - \rho_2 < \delta_U \quad (2)$$

A conservative range of acceptable difference of (δ) 5–15% was chosen [24, 25]. Due to numerous correlational contrasts, error rates were adjusted using the sequential false discovery rate approach for multiple hypothesis testing by BENJAMINI and HOCHBERG [26]. This yields a quantity representing the expected proportion of false-positive findings among all rejected hypotheses.

To assess diagnostic accuracy of Jackson and WURSS in discriminating between those with and without infection, several approaches were tried, beginning with simple logistic regression. Due to imbalance (only 12% were not infected), the current authors proceeded to an exact logistic regression approach, using PROC LogXact [27]. Finally, a learning linear discriminant function (LLDF) modelling strategy was used [28].

RESULTS

Participants were enrolled in six batches starting in May 2002 and ending in March 2004. A total of 419 participants were challenged with RV-39. Of these, two withdrew and 18 were excluded from the analysis because either: 1) nasal lavage culture demonstrated other pathogens, or 2) serum antibodies at entry suggested recent exposure to RV-39. Therefore, the data set included 399 people followed over 2,088 person-days. Of these, 350 (88%) demonstrated evidence of RV-39 infection (positive culture or seroconversion).

In general, data were consistent with the current authors' conceptual model and previous reports. Nasal and throat symptoms were more prevalent than cough, headache or fever. A majority of participants rated symptoms as absent, very mild or mild on most days of the trial, yielding a skewed response distribution. Overall, there were very little missing data; hence, the current authors chose not to impute for the analyses portrayed here. However, there were significant outlying data, especially among biomarkers.

Figure 1 portrays central tendency and variability over time. Both Jackson and WURSS show gradual increases from day 0 to day 2, with maximum scores on day 3. While mucus weights follow a similar pattern, viral titre, neutrophil (polymorphonuclear neutrophils (PMN)) count and IL-8 are not as predictable. A logarithmic y-axis was chosen because of variability and skewing.

Figure 2 portrays bivariate relationships on day 3. Day 3 was chosen because overall severity is greatest, providing best estimates of associations. While all Pearson correlations are significant at $p < 0.01$, strength of association varies (table 1). Days 0 and 1 were excluded because too few people had developed infections. Day 5 was excluded because the WURSS and SF-8 data were not collected and SF-36 was excluded because very little association was seen. This was not surprising as the SF-36 refers to health over the past 4 weeks,

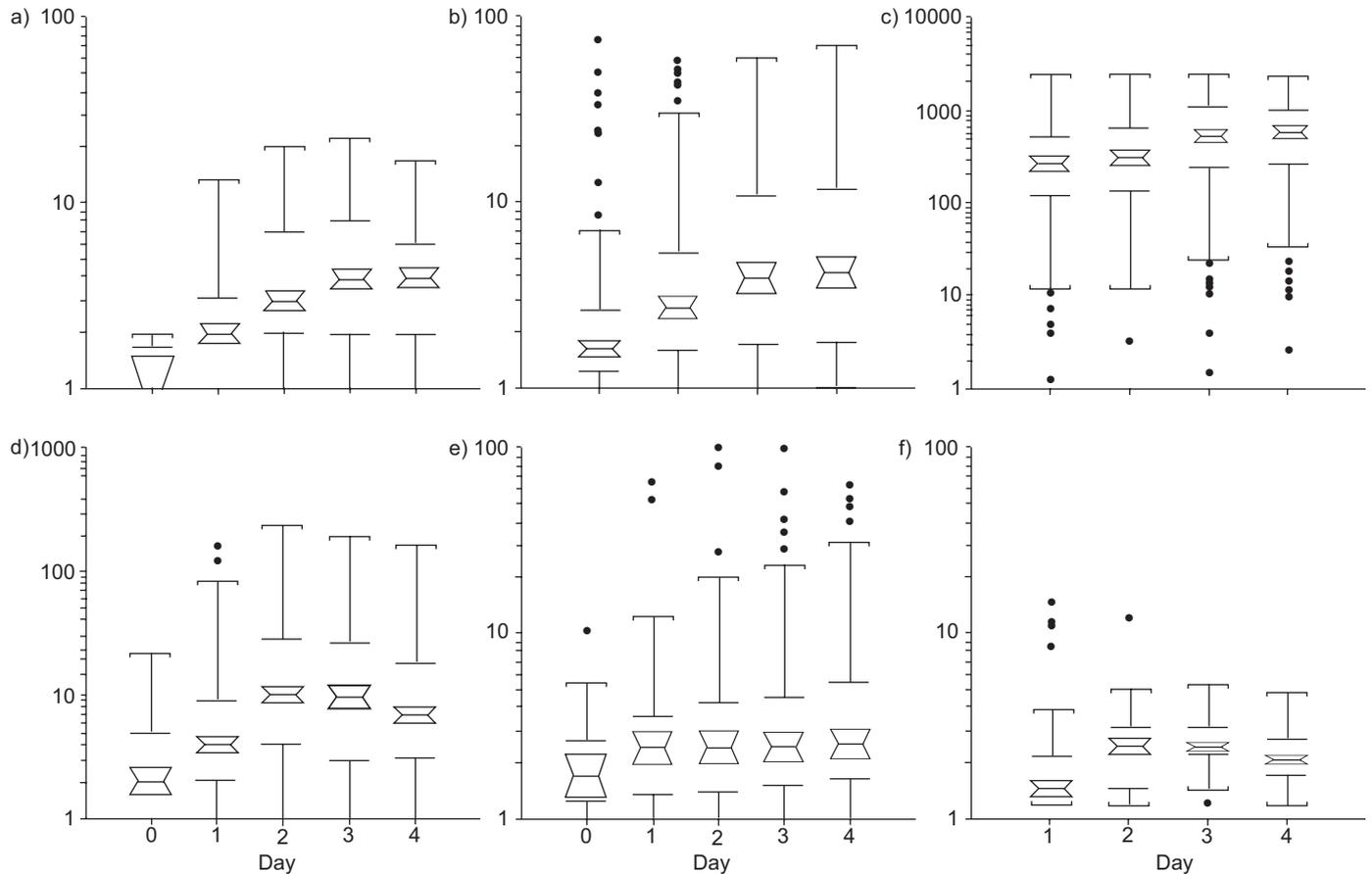


FIGURE 1. Central tendency and variability of primary measures over time of a) Jackson cold scale, b) mucus weight, c) interleukin (IL-8), d) Wisconsin Upper Respiratory Symptom Survey, e) nasal neutrophil count, and f) virus titre. IL-8 is a measure of change from baseline (day X IL-8 minus baseline IL-8). Boxes portray the median ± 1.57 (interquartile range (IQR))/ n^{-2} and thus can be compared with assessed difference at the $p=0.05$ level of significance (not accounting for multiple comparisons). Horizontal lines immediately above and below the boxes indicate the 25 and 75 percentiles, respectively. The other horizontal lines indicate the last actual data point within 1.5 (IQR) from the 25 and 75 percentiles. ●: outlying data points. $n=350$.

and is unlikely to be affected by a few days of mild cold symptoms.

Not unexpectedly, the strongest associations were between WURSS and Jackson, with coefficients ranging 0.76–0.84 (average=0.81). Also, not surprisingly, the physical domain of the SF-8 correlated more strongly with WURSS and Jackson than the mental domain. Associations between questionnaires and biomarkers yielded coefficients ranging 0.46–0.64 for mucus weight, 0.30–0.50 for viral titre, 0.26–0.40 for IL-8, and 0.22–0.39 for nasal neutrophils (PMN). There were no indications that associations among measures varied systematically over time.

Bioequivalence assays for assessing whether Jackson and WURSS were equally good at predicting laboratory measures suggested equivalence across a 5–15% range of acceptable difference. While minor trends suggested that Jackson might better predict viral titre and IL-8, and WURSS might better predict SF-8, these were not statistically significant. For mucus weight and PMN, some days one instrument was favoured, while other days the other was favoured. Collectively, analyses suggested that WURSS and Jackson were equally good (or equally bad) at predicting biomarkers.

While WURSS and Jackson are most commonly used to evaluate illness severity over time, it is conceivable that they could be used to diagnose infection. Using conventional binary statistical theory, the current authors sought WURSS and Jackson cut-offs that would maximise sensitivity and specificity, using seroconversion and/or positive culture as reference. No adequate cut-offs could be found. The present authors then progressed to LLDF models, limiting possibilities to first-order equations. For WURSS, the best prediction rule yielded a sensitivity of 85% and a specificity of 44%. For Jackson, the best rule yielded a sensitivity of 81% and a specificity of 66%. Neither equation was judged useful enough to portray here, or to prospectively test in future studies.

DISCUSSION

The common cold syndrome is characterised by variability rather than central tendency. While there are indisputable links between infection, inflammation, symptoms and quality-of-life impact, the degree of association among these domains is limited. Even in this tightly controlled, induced rhinovirus-infection model, variability greatly outweighs central tendency.

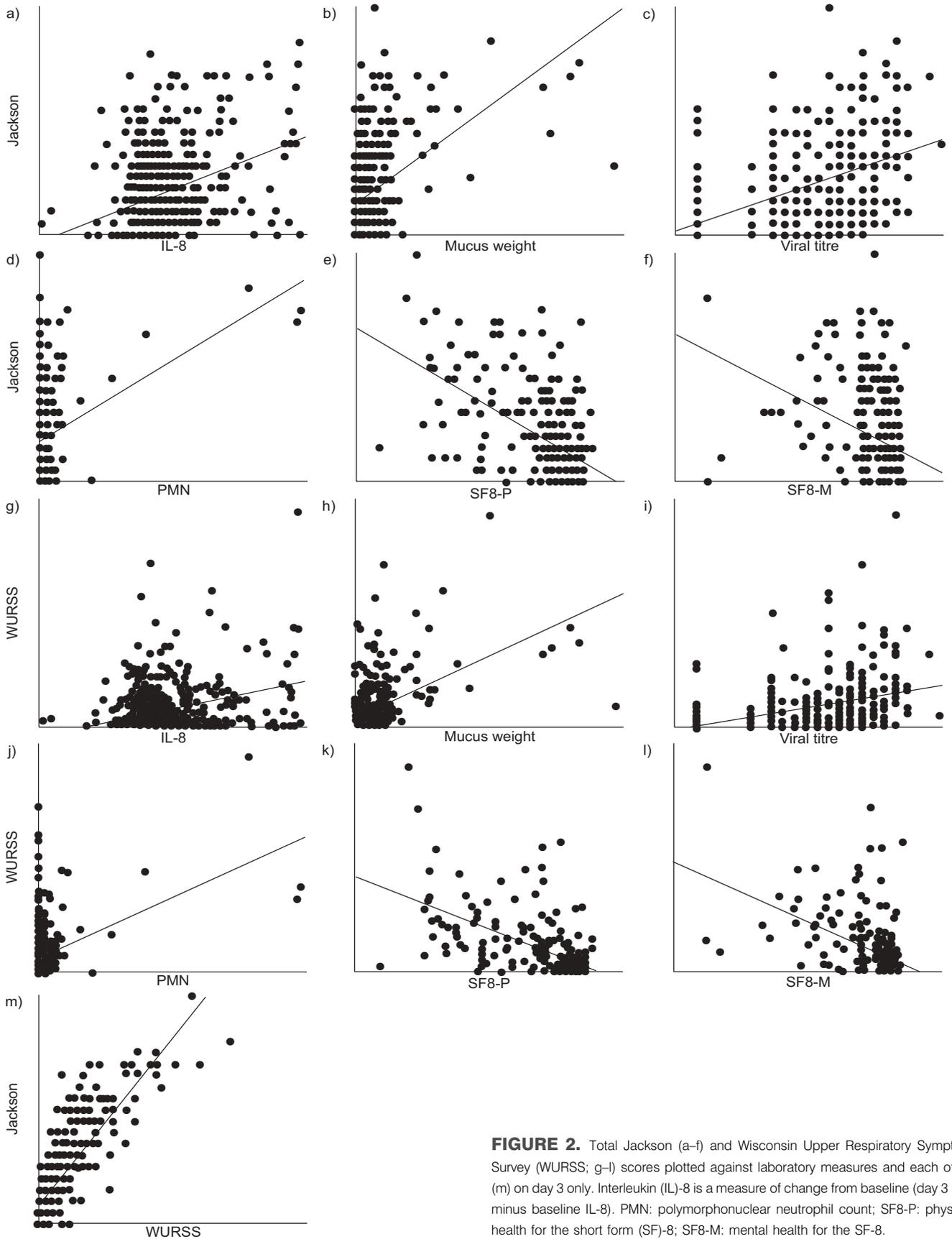


FIGURE 2. Total Jackson (a–f) and Wisconsin Upper Respiratory Symptom Survey (WURSS; g–l) scores plotted against laboratory measures and each other (m) on day 3 only. Interleukin (IL)-8 is a measure of change from baseline (day 3 IL-8 minus baseline IL-8). PMN: polymorphonuclear neutrophil count; SF8-P: physical health for the short form (SF)-8; SF8-M: mental health for the SF-8.

TABLE 1 Correlations among variables

	Day 2		Day 3		Day 4	
	WURSS-44	Jackson	WURSS-44	Jackson	WURSS-44	Jackson
SF8-P	-0.580 (-0.67– -0.50)	-0.577 (-0.66– -0.49)	-0.649 (-0.72– -0.58)	-0.621 (-0.70– -0.54)	-0.638 (-0.71– -0.56)	-0.605 (-0.68– -0.53)
SF8-M	-0.446 (-0.54– -0.34)	-0.351 (-0.46– -0.24)	-0.444 (-0.54– -0.34)	-0.440 (-0.54– -0.34)	-0.514 (-0.60– -0.42)	-0.459 (-0.56– -0.36)
Mucus weight	0.512 (0.44–0.58)	0.465 (0.39–0.54)	0.611 (0.55–0.67)	0.637 (0.58–0.70)	0.465 (0.39–0.54)	0.564 (0.50–0.64)
Virus titre	0.429 (0.35–0.51)	0.500 (0.43–0.58)	0.389 (0.31–0.47)	0.491 (0.42–0.57)	0.303 (0.21–0.39)	0.401 (0.32–0.48)
IL-8[#]	0.316 (0.23–0.41)	0.339 (0.25–0.43)	0.352 (0.27–0.44)	0.401 (0.32–0.48)	0.257 (0.16–0.35)	0.333 (0.25–0.42)
PMN	0.394 (0.31–0.48)	0.308 (0.22–0.40)	0.315 (0.23–0.40)	0.234 (0.14–0.33)	0.224 (0.13–0.32)	0.288 (0.20–0.38)
Jackson	0.841 (0.82–0.87)	NA	0.833 (0.80–0.86)	NA	0.764 (0.73–0.80)	NA

Data are presented as correlation coefficient (95% confidence interval). WURSS-44: Wisconsin Upper Respiratory Symptom Survey 44-item; SF8-P: physical health for the short form (SF)-8; SF8-M: mental health subscales for the SF-8; IL: interleukin; PMN: polymorphonuclear neutrophil count; NA: not applicable. [#]: scores used in this analysis represent changes from baseline. All Pearson correlation coefficients were individually significant at $p < 0.01$. All day correlation contrasts (WURSS versus Jackson) were nonsignificant at $p > 0.05$. $n = 399$, except for variables of SF8-P and SF8-M where $n = 248$.

While observed correlations among questionnaire and laboratory measures were not due to chance, very little of the biomarker variability could be explained by questionnaire scores.

Associations among the questionnaire instruments were stronger. Correlation coefficients of 0.76–0.84 between WURSS and Jackson were remarkably similar to corresponding coefficients of 0.73–0.93 found in the WURSS primary validation study [14] based on 1,681 person-days of community-acquired colds. Correlations with the SF-8 in that study (coefficients from -0.60– -0.84 for WURSS and -0.55– -0.78 for Jackson) were slightly stronger than those seen here, perhaps partially due to the inherent difficulties in rating health-related quality of life when confined to a hotel room for 5 days. Overall, the current authors interpret the similar degrees of association between the two studies to support external validity of all three instruments [29].

WURSS and Jackson were indistinguishable in their ability to predict biomarkers. Bioequivalence methods were reasonably powered to detect a difference if one indeed existed. The current authors were reassured by these results, as there was concern that the expanded range of WURSS might reduce associations with biological domains.

It was not possible to use WURSS or Jackson to derive useful rules to predict infection. Perhaps this should not be surprising. Previous studies have reported that 25–35% of demonstrable infections occur in people who deny symptoms [30, 31]. Conversely, 20–40% of people with classic upper respiratory infection symptoms fail to yield an aetiological agent when subjected to the most up-to-date viral culture and PCR testing methods [1, 32, 33]. Reporting error and placebo effects may also be involved. In one study, 22% of sham-inoculated participants reported cold symptoms [34]. Perhaps a larger sample with higher symptom scores (and more people without infection) would yield better prediction rules.

This leads to the other limitations. Sample size limitations, random error and perhaps systematic biases may be present. Data from college-aged volunteers with rhinovirus-induced

colds should not be generalised to community-acquired colds in the general population. It is possible that administration of echinacea (or placebo) may have influenced the data, even though no treatment effects were demonstrable.

Notwithstanding these limitations, the present authors feel the presented results are a significant addition to the existing knowledge base. It is now known that the Wisconsin Upper Respiratory System Survey and the Jackson scale perform in similar manners in induced rhinovirus infection. While neither is superior in terms of predicting biomarkers, both do correlate significantly with laboratory-assessed measures. Given the fact that the Wisconsin Upper Respiratory System Survey includes quality-of-life domains important to cold sufferers, and appears to measure change over time better than Jackson (reported elsewhere [14]), the current authors recommend it as the best currently available illness-specific quality-of-life outcomes instrument for the common cold.

ACKNOWLEDGEMENTS

The authors would like to thank P. Beasley who coordinated the human volunteer portions of the study with assistance from M. Potter. They would also like to thank M. Mundt who assisted with database management and A. Nies who entered the WURSS and general health SF-8 data.

REFERENCES

- 1 Arruda E, Pitkäranta A, Witek TJ, Doyle CA, Hayden FG. Frequency and history of rhinovirus infections in adults during autumn. *J Clin Microbiol* 1997; 35: 2864–2868.
- 2 Fox JP, Cooney MK, Hall CE. The Seattle virus watch. V. Epidemiologic observations of rhinovirus infections, 1965–1969, in families with young children. *Am J Epidemiol* 1975; 101: 122–143.
- 3 Gwaltney JM. Rhinoviruses. In: Evans AS, Kaslow RA, eds. *Viral Infections of Humans: Epidemiology and Control*. New York, Plenum Medical Book Company, 1997; pp. 815–838.
- 4 Turner RB. The treatment of rhinovirus infections: progress and potential. *Antiviral Res* 2001; 49: 1–14.

- 5 Jackson GG, Dowling HF, Spiesman IG, Boand AV. Transmission of the common cold to volunteers under controlled conditions. *Arch Intern Med* 1958; 101: 267–278.
- 6 Jackson GG, Dowling HF, Anderson TO, *et al.* Susceptibility and immunity to common upper respiratory viral infections—the common cold. *Ann Intern Med* 1960; 55: 719–738.
- 7 Jackson GG, Dowling HF, Muldoon RL. Present concepts of the common cold. *Am J Public Health* 1962; 52: 940–945.
- 8 Barrett B, Locken K, Maberry R, *et al.* The Wisconsin Upper Respiratory Symptom Survey (WURSS): new research instrument for assessing the common cold. *J Fam Pract* 2002; 51: 265–273.
- 9 Guyatt GH, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chron Dis* 1987; 40: 171–178.
- 10 Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *J Clin Epidemiol* 2001; 54: 1204–1217.
- 11 Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. *Controlled Clin Trials* 1991; 12: Suppl. 4, 142S–158S.
- 12 Kaplan RM, Feeny D, Revicki DA. Methods for assessing relative importance in preference based outcome measures. *Qual Life Res* 1993; 2: 467–475.
- 13 Kirshner B, Guyatt GH. A methodological framework for assessing health indices. *J Chron Dis* 1985; 38: 27–36.
- 14 Barrett B, Brown R, Mundt M, *et al.* The Wisconsin Upper Respiratory Symptom Survey is responsive, reliable, and valid. *J Clin Epidemiol* 2005; 58: 609–617.
- 15 Ware JE, Kosinski M, Dewey JE, Gandek B. How to score and interpret single-item health status measures: a manual for users of the SF-8 health survey. Lincoln, RI, QualityMetric, 2001.
- 16 Guyatt GH, Bombardier C, Tugwell P. Measuring disease-specific quality of life in clinical trials. *CMAJ* 1986; 134: 889–894.
- 17 Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. *J Clin Epidemiol* 1989; 42: 403–408.
- 18 Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; 10: 407–415.
- 19 Turner RB, Bauer R, Woelkart K, Hulsey TC, Gangemi JD. An evaluation of *Echinacea angustifolia* in experimental rhinovirus infections. *N Engl J Med* 2005; 353: 341–348.
- 20 Turner RB, Weingand KW, Yeh CH, Leedy DW. Association between interleukin-8 concentration in nasal secretions and severity of symptoms of experimental rhinovirus colds. *Clin Infect Dis* 1998; 26: 840–846.
- 21 Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992; 30: 473–483.
- 22 McHorney CA, Ware JE, Raczek AE. The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1998; 31: 247–263.
- 23 McHorney CA, Ware JE, Lu JFR, Sherbourne CD. The MOS 36-Item short-form health survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994; 32: 40–66.
- 24 Schuirmann D. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokinetic Biopharm* 1987; 15: 657–680.
- 25 Phillips KF. Power of the two one-sided tests procedure in bioequivalence. *J Pharmacokinetic Biopharm* 1990; 18: 137–144.
- 26 Benjamini Y, Hochberg Y. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J Educ Behav Stat* 2000; 25: 60–83.
- 27 Cytel Software Corporation. Oric0LogXact 5 For SAS Users: Logistic Regression Software Featuring Exact Methods. Cambridge, MA, Cytel Software Corporation, 2003.
- 28 Goldbaum MH, Sample PA, Chan K, *et al.* Comparing machine learning classifiers for diagnosing glaucoma from standard automated perimetry. *Invest Ophthalmol Visual Science* 2002; 43: 162–169.
- 29 McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires. Oxford & New York, Oxford University Press, 1996.
- 30 Gwaltney JM. The use of experimentally infected volunteers in research on the common cold. In: Skoner DP, ed. *Asthma and Respiratory Infections*. New York, Marcel Dekker, Inc., 2001; pp. 103–127.
- 31 Turner RB, Wecker MT, Pohl G, *et al.* Efficacy of tremacamra, a soluble intercellular adhesion molecule 1, for experimental rhinovirus infection. *JAMA* 1999; 281: 1797–1804.
- 32 Gwaltney JM. Clinical significance and pathogenesis of viral respiratory infections. *Am J Med* 2002; 112: Suppl. 6A, 13S–18S.
- 33 Monto AS. Epidemiology of viral respiratory infections. *Am J Med* 2002; 112: Suppl. 6A, 4S–12S.
- 34 Turner RB, Witek TJ, Riker DK. Comparison of symptom severity in natural and experimentally induced cold. *Am J Rhinology* 1996; 10: 167–172.