ERJ Express. Published on April 26, 2006 as doi: 10.1183/09031936.06.00002606

Relations among questionnaire and laboratory measures of rhinovirus infection

Bruce Barrett MD PhD, Roger Brown PhD, Rick Voland PhD, Rob Maberry, Ronald Turner MD

Re-submitted to: *European Respiratory Journal*

March 31, 2006

Dr Turner is at the University of Virginia, Charlottesville, Virginia, USA

All other authors are at the University of Wisconsin, Madison, Madison, Wisconsin, USA

Comments to:

Bruce Barrett MD PhD
Department of Family Medicine
University of Wisconsin Medical School
777 S. Mills
Madison WI 53715
USA

608-263-2220 office
263-5813 fax
257-1646 home
bruce.barrett@fammed.wisc.edu

Running title **Measures of rhinovirus infection**

Word count = 1,820
Abstract = 185
Tables = 1
Figures = 2

Relations among questionnaire and laboratory measures of rhinovirus infection

Abstract

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

Methods Data reported here come from a trial testing echinacea for induced rhinovirus infection. Laboratory-assessed biomarkers include interleukin-8 (IL-8), nasal neutrophil count (PMN), mucus weight, viral titer and seroconversion. Questionnaires include the general health SF-8 (24 hour recall version), the 8-item Jackson cold scale, and the 44-item Wisconsin Upper Respiratory Symptom Survey (WURSS-44).

Results 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-SF8=0.62; Jackson-SF8=0.60. Correlations with laboratory values: WURSS-mucus-weight=0.53; Jackson-mucus-weight=0.55; WURSS-viral titer=0.37; Jackson-viral titer=0.46; WURSS-IL-8=0.31; Jackson-IL-8=0.36; WURSS-PMN=0.31; Jackson-PMN=0.28. Neither WURSS nor Jackson yielded satisfactory cutoff scores for diagnosis of infection.

Conclusions Symptomatic and biological outcomes of URI are highly variable, with only modest associations. While WURSS and Jackson both correlate with biomarkers, neither is a good predictor of induced infection. Inclusion of functional and quality-of-life items in WURSS does not significantly decrease the strength of association with laboratory-assessed biomarkers.

Keywords: common cold, psychometrics, quality-of-life, questionnaires, rhinovirus, symptom measurement, upper respiratory infection, validation

Introduction

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 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 8 symptoms (sneezing, nasal obstruction, nasal discharge, sore throat, cough, headache, chilliness, and malaise), using a 3 or 4 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 WURSS-44 includes 32 items assessing symptoms, 10 functional items, and 2 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 undergoing validation. WURSS is free-of-charge for non-profit and educational use, but must be licensed by for-profit users http://www.fammed.wisc.edu/wurss/ .

WURSS scores associate more strongly with general health-related quality-of-life (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] we estimated that a two-armed clinical trial using the WURSS-21 would need 74 participants to detect MID, compared to 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

Data reported here come from an induced-rhinovirus-cold echinacea trial, reported as negative.[19] College-age participants susceptible to rhinovirus-39 (RV-39) were randomized to one of three types of echinacea, either as prevention or treatment, and were housed in hotel rooms from inoculation (Day 0) until discharged on Day 5. Protocols were approved by Virginia and Wisconsin ethics committees.

Interleukin-8 (IL-8), neutrophil count, and viral culture came from daily nasal wash, and were analyzed using previously reported methods.[20] Mucus weights came from pre-weighed, distributed, then collected and re-weighed nasal tissues. Serum was collected Day 0 and approximately 3 weeks later to assess serological response. Jackson symptoms were elicited by study nurses twice daily using a 4-point response range. Daily scores were defined as the highest 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-hour recall version of the SF-8[15] became available, we substituted this more appropriate instrument, which was used in the last four batches (N=249).

This 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 study aim was to determine whether WURSS would predict biomarkers as well as Jackson. Because WURSS includes domains not specific to colds, it was possible that WURSS would correlate less strongly. On the other hand, the expanded severity response range of WURSS might better measure underlying (continuous) domains, hence yield tighter correlations. We were also interested in the abilities of WURSS and Jackson to discriminate between infected and non-infected 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 relationships between Jackson, WURSS-44, and other measures, we selected a bioequivalence approach, using the TOST method described by Schuirmann[24] and Phillips.[25] Let ρ_1 be the correlation of Jackson with biomarkers, and ρ_2 the correlation of WURSS with biomarkers, and δ_L and δ_U the lower and upper bounds of bioequivalence. The null hypothesis of nonequivalence is:

$$H_0 : \rho_1 - \rho_2 < \delta_L \text{ or } \rho_1 - \rho_2 > \delta_U$$

and the alternative hypothesis of bioequivalence is:

$$H_a : \delta_L < \rho_1 - \rho_2 < \delta_U$$

We chose a conservative range of acceptable difference of (δ) from 5%-15%.[24;25] Due to numerous correlational contrasts, we adjusted error rates using the sequential Benjamini and Hochberg[26] false discovery rate approach for multiple hypothesis-testing. 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), we proceeded to an exact logistic regression approach, using PROC LogXact.[27] Finally, we used a learning linear discriminant function (LLDF) modeling strategy.[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. Two withdrew, and 18 were excluded from analysis because either: 1) nasal lavage culture demonstrated other pathogens, or 2) serum antibodies at entry suggested recent exposure to RV-39. Therefore, the dataset 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 our 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 we 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 titer, neutrophil (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. See Table. Days 0 and 1 were excluded because too few people had developed infections. Day 5 is excluded because the WURSS and SF-8 data weren't collected. SF-36 is excluded because very little association was seen, not surprising, as the SF-36 asks about health over the past four weeks, 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 from 0.76 to 0.84 (average = 0.81). Also not surprisingly, the physical domain of the SF-8 correlated more strongly with WURSS and Jackson than did the mental domain. Associations between questionnaires and biomarkers yielded coefficients from 0.46 to 0.64 for mucus weight, 0.30 to 0.50 for viral titer, 0.26 to 0.40 for IL-8, and 0.22 to 0.39 for nasal neutrophils (PMN). There were no indications that associations among measures varied systematically over time.

Bio-equivalence assays for assesing whether Jackson and WURSS were equally good at predicting lab measures suggested equivalence across a 5% to 15% range of acceptable difference. While minor trends suggested that Jackson might better predict viral titer and IL-8, and WURSS might better predict SF-8, these were not statistically significant. For mucus weight and PMN, some days favored one instrument while other days favored the other. 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, we sought WURSS and Jackson cutoffs that would maximize sensitivity and specificity, using seroconversion and/or positive culture as reference. No adequate cutoffs could be found. We 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%. We judge neither equation useful enough to portray here, or to prospectively test in future studies.

Discussion

The common cold syndrome is characterized 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. 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 to 0.84 between WURSS and Jackson were remarkably similar to corresponding coefficients of 0.73 to 0.93 found in WURSS's 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 to -0.84 for WURSS and -0.55 to -0.78 for Jackson) were slightly stronger than those seen here, perhaps partially due to inherent difficulties in rating health-related quality-of-life when confined to a hotel room for five days. Overall, we 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 ability to predict biomarkers. Bioequivalence methods were reasonably powered to detect a difference if one indeed existed. We were reassured by these results, as we had been concerned 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 report that 25% to 35% of demonstrable infections occur in people who deny symptoms.[30;31] Conversely, 20% to 40% of people with classic URI symptoms fail to yield an etiologic 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 brings us to 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 generalized 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, we feel the results presented above are a significant addition to the existing knowledge base. We now know that WURSS and Jackson 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 WURSS includes quality-of-life domains important to cold-sufferers, and appears to measure change-over-time better than Jackson (reported elsewhere[14]), we recommend it as the best currently available illness-specific quality-of-life outcomes instrument for common cold.

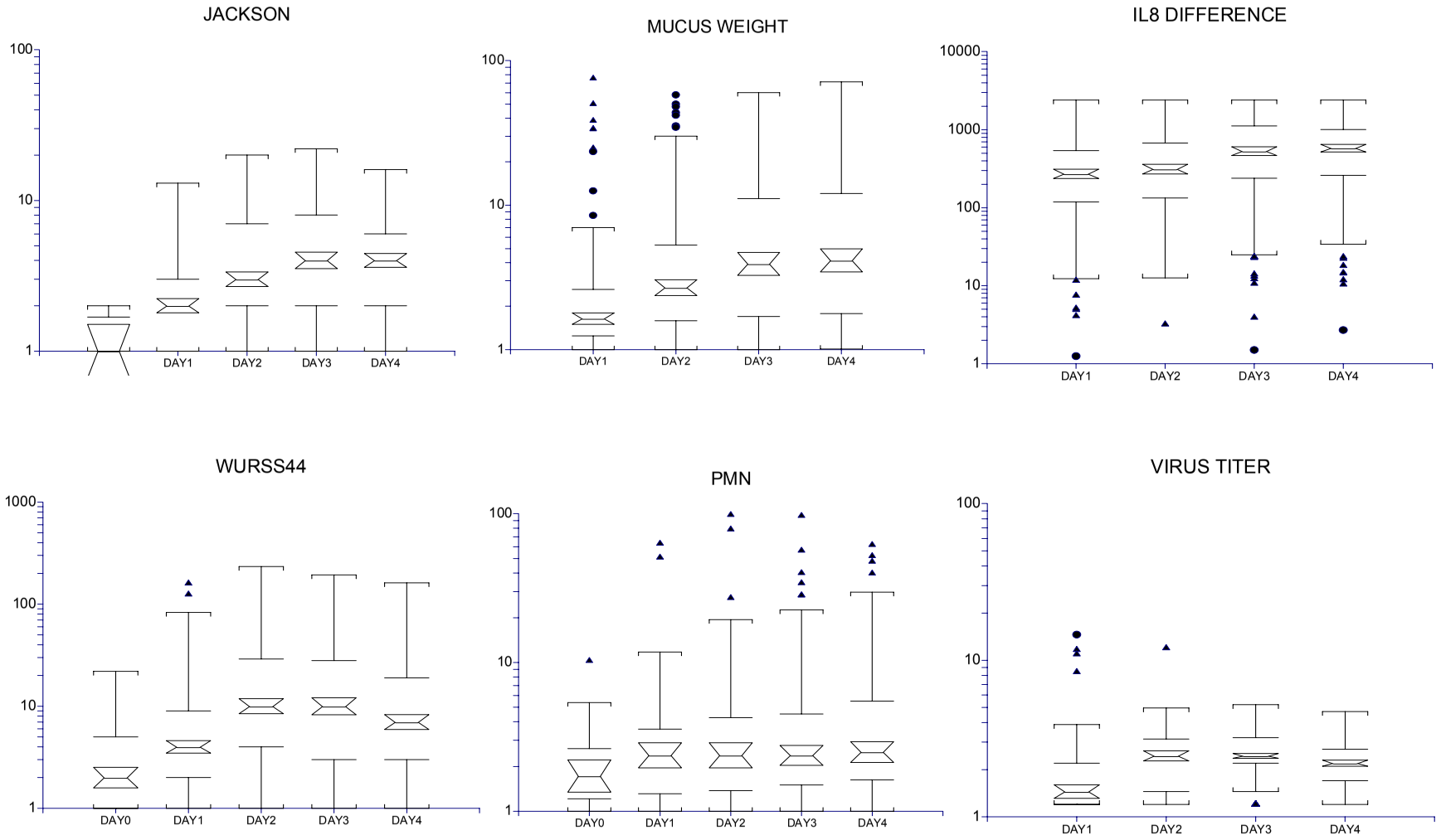
Acknowledgements

The trial from which these data came was supported by U.S. National Institutes of Health grant R01 AT001146 (National Center for Complementary and Alternative Medicine.) Dr. Turner is Principle Investigator. Dr. Barrett is PI 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. Patsy Beasley RN, coordinated the human volunteer portions of the study with assistance from Marilyn Potter RN. Marlon Mundt MS assisted with database management and Ann Nies entered WURSS and SF-8 data. Nevertheless, the article is solely the responsibility of the authors.

Conflicts of interest

There are no substantive conflicts of interest. B. Barrett, R. Maberry and R. Brown hold partial copyrites on the WURSS instrument (which is free for educational and nonprofit use, but can be licensed through the Wisconsin Alumni Research Foundation by for-profit entities http://www.fammed.wisc.edu/wurss/) . Dr. Turner has a number of grants and consultancies, but none are deemed a conflict of interest.

Figure 1. Central tendency and variability of primary measures over time



Notes for **Figure 1**

Data are shown only for those who became infected (N=350)

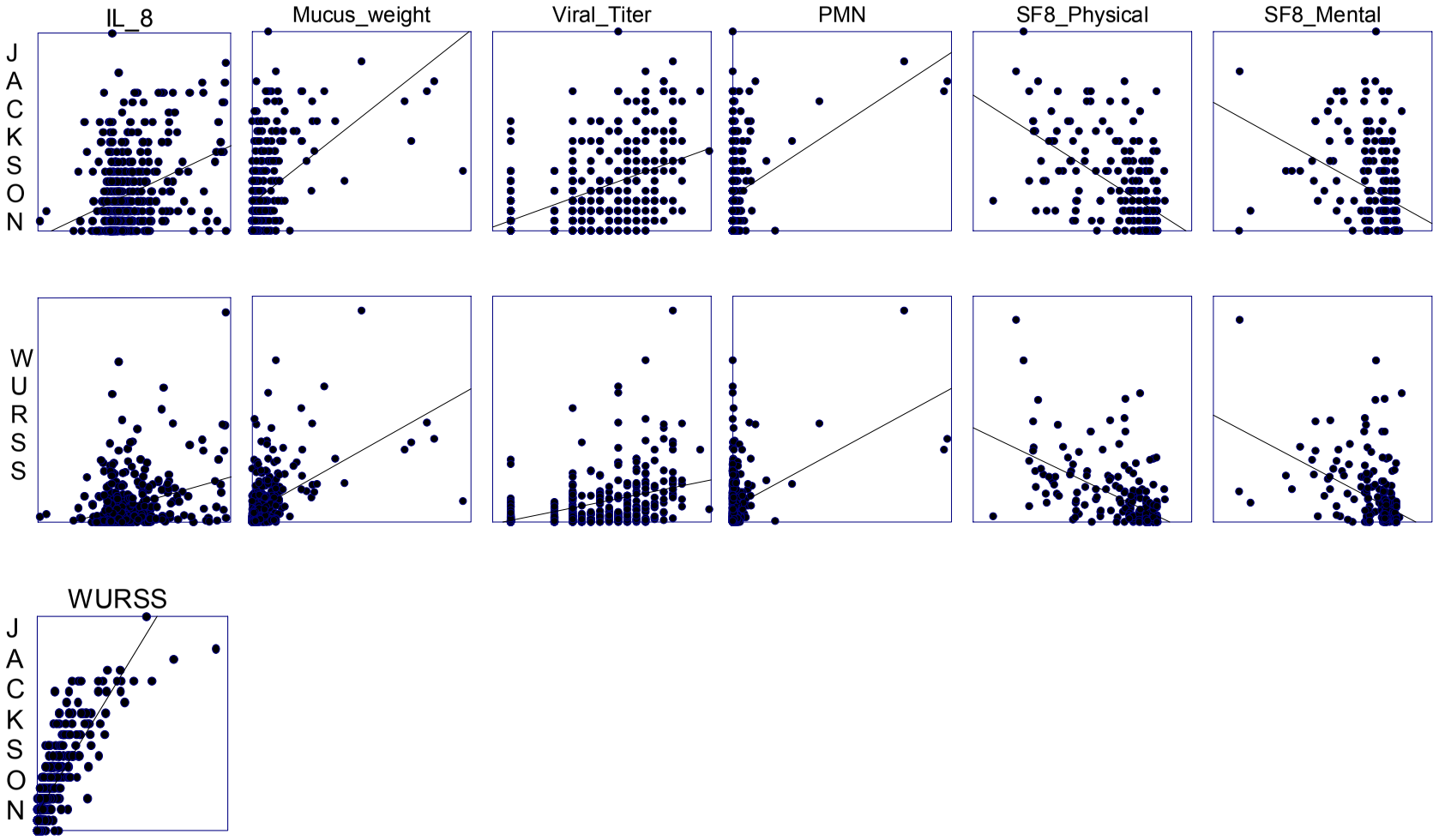
Vertical axes are shown in logarithmic scale, easier to portray and interpret, due to skewing and outlying data

IL-8 is a measure of change from baseline (Day X IL-8 minus baseline IL-8)

Notched boxes portray the median ± 1.57 (interquartile range=IQR) / N^{-2} and thus can be compared to assess difference at the P = 0.05 level of significance (not accounting for multiple comparisons).

The first horizontal lines below and above the notched boxes indicate the 25% and 75% percentiles, respectively. The horizontal lines marking the ends of the vertical lines indicate the last actual data point within 1.5 (IQR) from the 25%ile and 75%ile. The symbols above and below these lines are actual outlying data points.

Figure 2. Total WURSS and Jackson scores plotted against laboratory measures and each other (Day 3)



Notes for **Figure 2**

These scatterplots display data for Day 3 only

IL-8 is a measure of change from baseline (Day 3 IL-8 minus baseline IL-8)

Table Correlations among Variables (with 95% confidence intervals)

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

Notes for **Table**

SF8-P and SF8-M refer to physical health and mental health subscales for the SF-8

IL-8 scores used in this analysis represent changes from baseline

All Pearson correlation coefficients individually significant at $p < 0.01$

All day correlational contrasts (WURSS vs. Jackson) were non-significant at $p > 0.05$ level

N=399, except for variables of SF8-P and SF8-M where N=248

Reference List

1. Arruda E, Pitkäranta A, Witek TJ, Doyle CA, Hayden FG. Frequency and history of rhinovirus infections in adults during autumn. *Journal of Clinical Microbiology* 1997;**35**:2864-8.

2. Fox JP, Cooney MK, Hall CE. The Seattle virus watch. V. Epidemiologic observations of rhinovirus infections, 1965-1969, in families with young children. *American Journal of Epidemiology*. 1975;**101**:122-43.

3. Gwaltney JM. Rhinoviruses. In Evans AS, Kaslow RA, eds. *Viral Infections of Humans: Epidemiology and Control*, pp 815-38. New York: Plenum Medical Book Company, 1997.

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**:267-78.

6. Jackson GG, Dowling HF, Anderson TO, Riff L, Saporta J, Turck M. Susceptibility and immunity to common upper respiratory viral infections-the common cold. *Annals of Internal Medicine* 1960;**55**:719-38.

7. Jackson GG, Dowling HF, Muldoon RL. Present concepts of the common cold. *Am J Public Health* 1962;**52**:940-5.

8. Barrett B, Locken K, Maberry R, Schwamman J, Bobula J, Brown R *et al*. The Wisconsin Upper Respiratory Symptom Survey: Development of an instrument to measure the common cold. *Journal of Family Practice* 2002;**51**:265-73.

9. Guyatt GH, Walter S, Norman G. Measuring change over time: Assessing the usefulness of evaluative instruments. *J Chron Dis* 1987;**40**:171-8.

10. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *Journal of Clinical Epidemiology*. 2001;**54**:1204-17.

11. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. *Controlled Clinical Trials* 1991;**12**:142S-58S.

12. Kaplan RM, Feeny D, Revicki DA. Methods for assessing relative importance in preference based outcome measures. *Quality of Life Research* 1993;**2**:467-75.

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, Safdar N, Dye L, Maberry R *et al*. The Wisconsin Upper Respiratory Symptom Survey is responsive, reliable, and valid. *Journal of Clinical Epidemiology* 2005;**58**:609-17.

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. *Canadian Medical Association Journal* 1986;**134**:889-94.

17. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. *Journal of Clinical Epidemiology.* 1989;**42**:403-8.

18. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: Ascertaining the minimal clinically important difference. *Controlled Clinical Trials* 1989;**10**:407-15.

19. Turner RB, Bauer R, Woelkart K, Hulsey TC, Gangemi JD. An evaluation of Echinacea angustifolia in experimental rhinovirus infections. *New England Journal of Medicine* 2005;**353**:341-8.

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.[see comment]. *Clinical Infectious Diseases* 1998;**26**:840-6.

21. Ware JE,.Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care* 1992;**30**:473-83.

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. *Medical Care* 1998;**31**:247-63.

23. McHorney CA, Ware JE, Lu JFR, Sherbourne CD. The MOS 36-Item short-form health survery (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical 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. *Journal of Pharmacokinetics and Biopharmaceutics* 1987;**15**:657-80.

25. Phillips KF. Power of the two one-sided tests procedure in bioequivalence. *Journal of Pharmacokinetics and Biopharmaceutics* 1990;**18**:137-44.

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. 2003. Cambridge, MA, Cytel Software Corporation.
Ref Type: Computer Program

28. Goldbaum MH, Sample PA, Chan K, Lee T, Blumenthal E, Girkin CA *et al*. Comparing machine learning classifiers for diagnosing glaucoma from standard automated perimetry. *Investigative Ophthalmology & Visual Science* 2002;**43**:162-9.

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*, pp 103-27. New York: Marcel Dekker, Inc., 2001.

31. Turner RB, Wecker MT, Pohl G, Witek TJ, McNally E, George RSt *et al*. Efficacy of tremacamra, a soluble intercellular adhesion molecule 1, for experimental rhinovirus infection. *JAMA* 1999;**281**:1797-804.

32. Gwaltney JM. Clinical significance and pathogenesis of viral respiratory infections. *American Journal of Medicine.* 2002;**112**:Suppl-18S.

33. Monto AS. Epidemiology of viral respiratory infections. *American Journal of Medicine.* 2002;**112**:Suppl-12S.

34. Turner RB, Witek TJ, Riker DK. Comparison of symptom severity in natural and experimentally induced cold. *American Journal of Rhinology* 1996;**10**:167-72.