Wisconsin Primary Care Research and Quality Improvement Forum

Hilton Garden Inn, Middleton, Friday, September 17th

Randomized trials, clinical significance & shared decision-making: Respiratory infection and beyond

Bruce Barrett MD PhD
Associate Professor
Department of Family Medicine
University of Wisconsin - Madison

Acknowledgements

- **Roger Brown PhD, Marlon Mundt PhD, Zhengjun Zhang Phd**
- **Shari Barlow, Michelle Gassman, Tola Ewers, Gay Thomas MS, etc**
- **Mary Beth Plane PhD et al. Department of Family Medicine at the University of Wisconsin School of Medicine and Public Health**
- **National Center for Complementary and Alternative Medicine at the National Institutes of Health (NIH NCCAM)**

  K-23 Patient-Oriented Career Grant

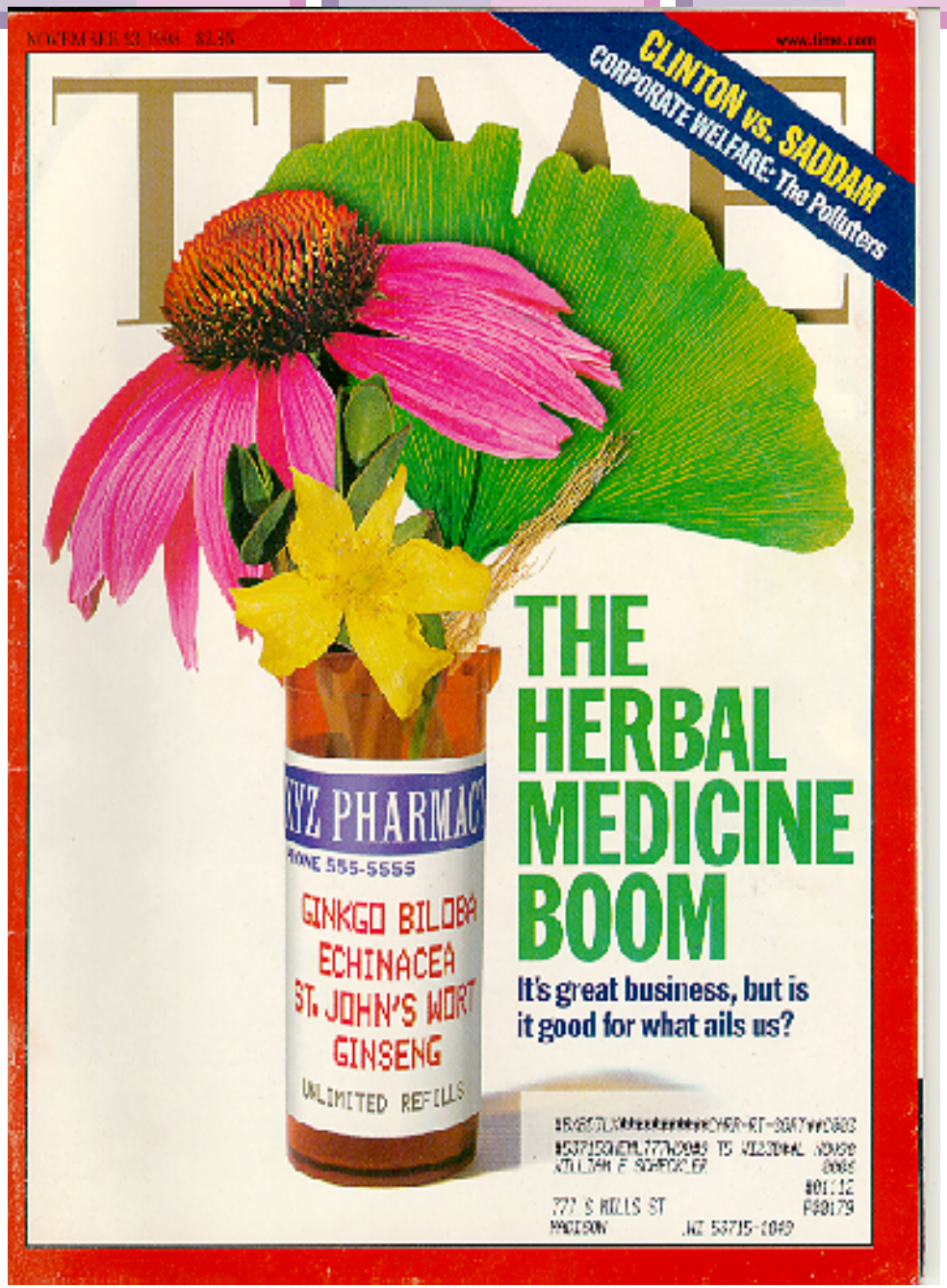
  R01 Physician Echinacea Placebo RCT

  R01 Meditation & Exercise for Prevention of Respiratory Infection

Robert Wood Johnson Foundation Generalist Physician Faculty Scholars Program

- **Gordon Guyatt MD – EBM guru, MID originator & RWJF mentor**

Where it all began (for me)

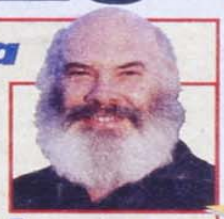

CLINTON vs. SADDAM
CORPORATE WELFARE: The Polluters
THE HERBAL MEDICINE BOOM
It's great business, but is it good for what ails us?
XYZ PHARMACY
PHONE 555-5555
GINKGO BILOBA
ECHINACEA
ST. JOHN'S WORT
GINSENG
UNLIMITED REFILLS

Sun
$1.29/$1.39 Canada February 10, 1998

HERO CAT SAVES DROWNING GIRL

Medical profession applauds safe herb miracle!

ECHINACEA CURES COLDS & FLU



'Try it as a front-line treatment for colds, flu & sore throats'

- TOP NATURAL HEALTH AUTHOR Dr. ANDREW WEIL

THE SECRET DOCTORS ARE ONLY NOW FINDING OUT ABOUT

'Echinacea helps you get over colds & flu much faster than any other treatment'

- PURDUE UNIVERSITY RESEARCH GROUP

'Protects against colds, flu & boosts immune system'

- PRESIDENT CLINTON'S COMMISSION ON DIETARY SUPPLEMENT LABELS

'Supercharge Echinacea's amazing healing power with Goldenseal'

- Dr. MICHAEL MURRAY, BASTYR COLLEGE, SEATTLE

$1.29 / $1.39 Canada


The
Common
Cold

Patient-oriented outcomes

- Health-related quality-of-life (HRQoL) measures are increasingly used as primary outcomes in RCTs
- Items assessing symptomatic and functional domains are self-reported using *general* or *illness-specific HRQoL instruments*
- Patient-Reported Outcome (PRO) questionnaires

McDowell I, Newell C. Measuring health: A guide to rating scales and questionnaires. Oxford & New York: Oxford University Press, 1996

R. Fitzpatrick et al. Quality of life measures in health care. I: Applications and issues in assessment. *BMJ*. 305 (6861):1074-1077, 1992.

A. Fletcher et al. Quality of life measures in health care. II: Design, analysis, and interpretation. *BMJ*. 305 (6862):1145-1148, 1992.

General Health Measures

- EuroQual (EQ-5D)
- Feeling thermometer (100mm VAS)
- Sickness Impact Profile
- Medical Outcomes Survey
- SF-36 (4 week recall)
- SF-12
- SF-8 (24 hour recall)
- Subscales:
- Physical health
- Mental health

Illness specific patient-reported outcomes (PROs)

- Hamilton, Beck (depression, anxiety)
- PHQ-9 (depression)
- Connor's (ADHD)
- Vanderbilt (ADHD)
- Mini-mental status exam (dementia; cognitive function)
- Glascow coma scale
- WOMAC - Western Ontario McMaster Universities Osteoarthritis Index
- AIMS - Arthritis Impact Measurement Scale
- RSOM-31 - Rhinosinusitis outcome measure
- SNOT - SinoNasal Outcomes Test

Wisconsin Upper Respiratory Symptom Survey (WURSS)

- In 1998 there was no HRQoL instrument for common cold
  - **Jackson's 8-item scale assessed symptoms as mild/mod/severe**
- Using local expert opinion and some field testing, we created a questionnaire instrument for use in a 1999 RCT testing echinacea as a cold treatment

B. P. Barrett, R. L. Brown, K. Locken, R. Maberry, J. A. Bobula, and D. D'Alessio. Treatment of the common cold with unrefined echinacea: A randomized, double-blind, placebo-controlled trial. *Annals of Internal Medicine* 137 (12):939-946, 2002.

Wisconsin Upper Respiratory Symptom Survey (WURSS)

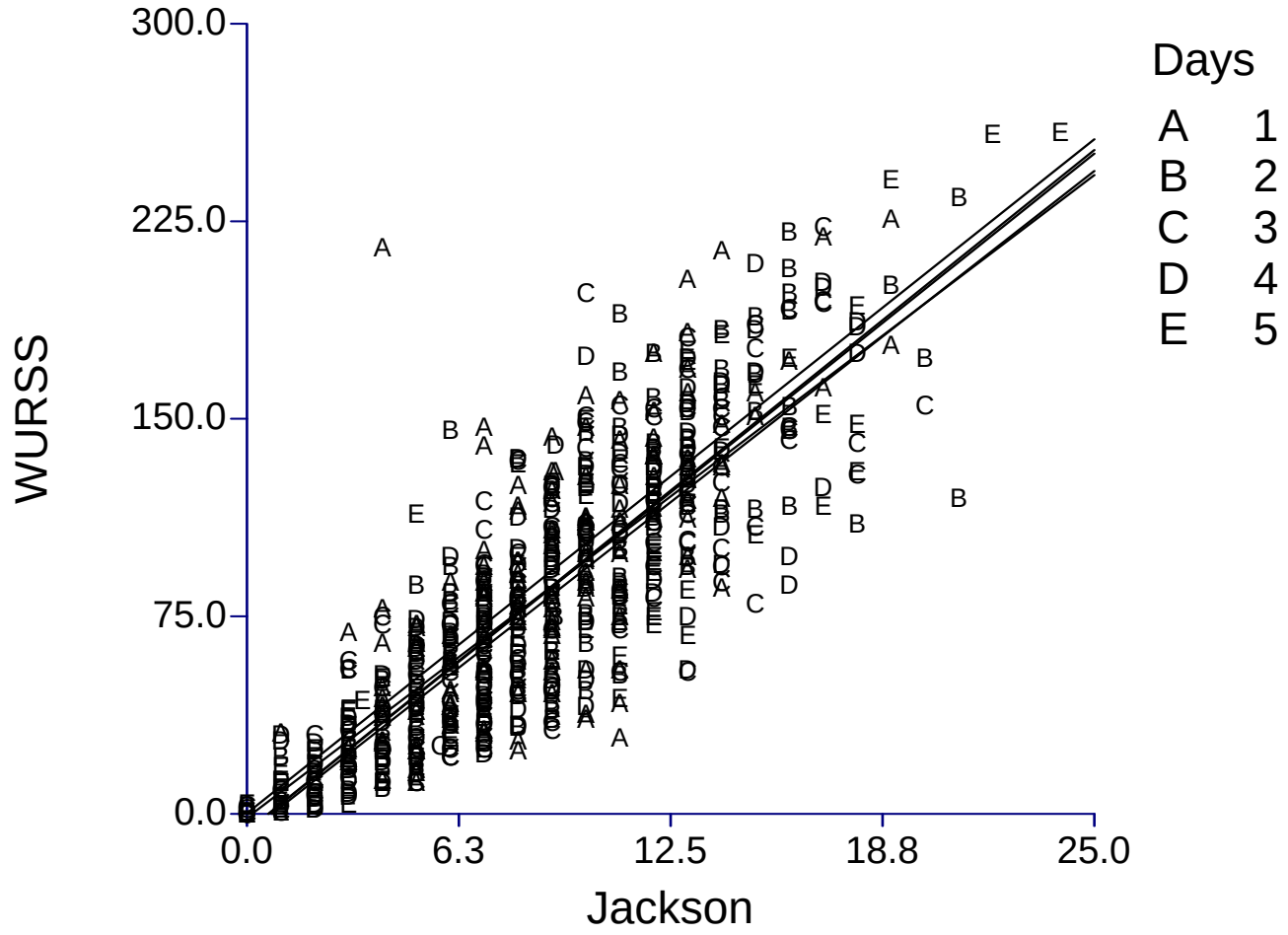
- After the initial echinacea trial, we combined qualitative and quantitative methods to create/develop/validate WURSS
- WURSS is an illness-specific quality-of-life outcomes instrument designed to evaluate cold symptoms, functional impact, and quality of life during cold/flu illness
- Open-ended questions with cold-sufferers gave us >150 terms
- Some 44 were mentioned by multiple partictpants (WURSS-44)

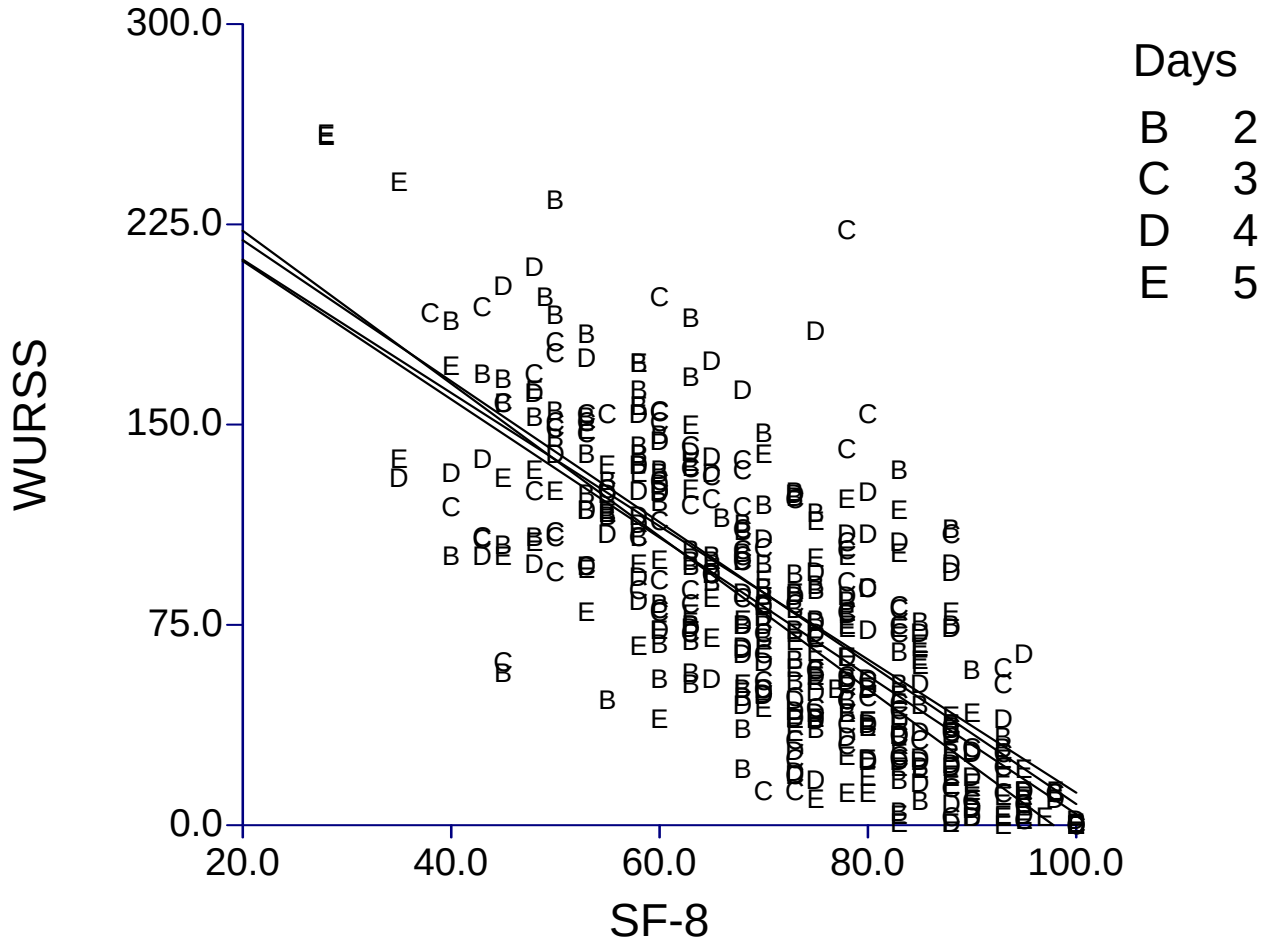
B. Barrett, K. Locken, R. Maberry, J. Schwamman, J. Bobula, R. Brown, and E. Stauffacher. The Wisconsin Upper Respiratory Symptom Survey: Development of an instrument to measure the common cold. *Journal of Family Practice* 51 (3):265-www.jfponline.com, 2002.

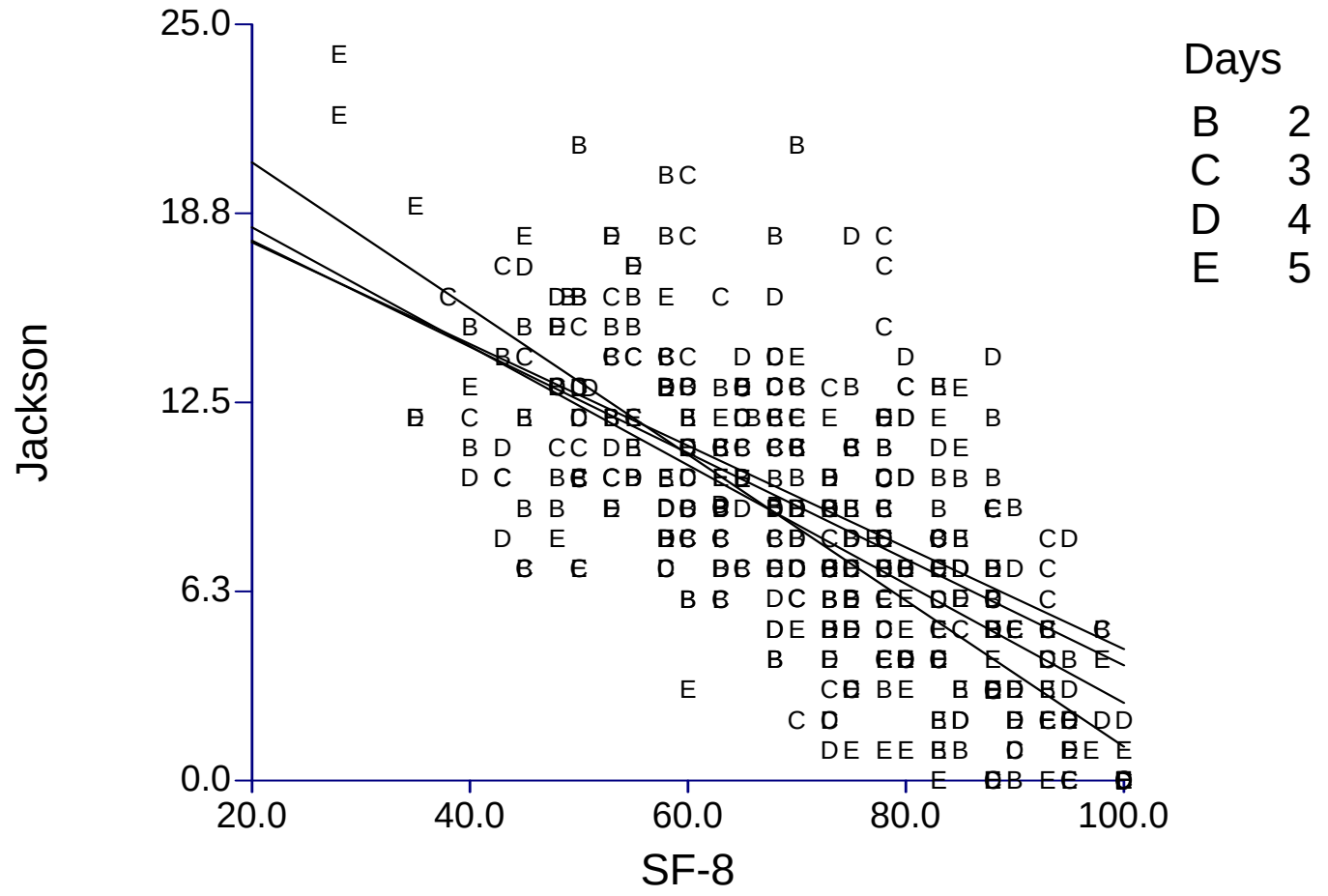
Validity, reliability & responsiveness

- Initial validity/reliability testing
- 151 participants, 1,681 days of cold symptoms
- Frequency, severity, importance-to-patients, responsiveness, and dimensional cohesion
- External (convergent) validity was assessed with comparisons to the Jackson scale and the SF-8 (24 hour recall version)

B. Barrett, R. L. Brown, M. P. Mundt, N. Safdar, L. Dye, R. Maberry, and J. Alt. The Wisconsin Upper Respiratory Symptom Survey is responsive, reliable, and valid. *Journal of Clinical Epidemiology* 58(6): 609-617.


300.0
225.0
150.0
75.0
0.0
WURSS
0.0
6.3
12.5
18.8
25.0
Jackson
Days
A 1
B 2
C 3
D 4
E 5


300.0
225.0
150.0
75.0
0.0
WURSS
20.0
40.0
60.0
80.0
100.0
SF-8
Days
B 2
C 3
D 4
E 5


Days
B 2
C 3
D 4
E 5
Jackson
SF-8
25.0
18.8
12.5
6.3
0.0
20.0
40.0
60.0
80.0
100.0

WURSS-21

- Using importance-to-patients and responsiveness* as guides, we selected items for a short version, the WURSS-21

*responsiveness = sensitivity to change over time, important property for clinical trial outcome measurements

B. Barrett, R. L. Brown, M. P. Mundt, N. Safdar, L. Dye, R. Maberry, and J. Alt. The Wisconsin Upper Respiratory Symptom Survey is responsive, reliable, and valid. *Journal of Clinical Epidemiology* 58(6):609-617, 2005.

Item	Domain	Frequency	Severity*	Importance*	MID	S.E. stable	Responsiveness
1	**Global**	**100**	**3.92 ± 1.17**	**3.79 ± 0.90**	***0.721***	***0.516***	***0.709***
2	**Cough**	**94**	**3.04 ± 1.33**	**3.21 ± 1.10**	***0.389***	***0.839***	***0.300***
3	Cough	80	2.77 ± 1.22	3.13 ± 1.06	0.308	1.283	0.193
5	**Throat**	**90**	**3.19 ± 1.46**	**3.39 ± 0.95**	***0.407***	***1.072***	***0.278***
6	**Throat**	**91**	**3.10 ± 1.37**	**3.16 ± 0.95**	***0.404***	***0.863***	***0.307***
9	**Nasal**	**96**	**3.55 ± 1.47**	**3.60 ± 0.94**	***0.509***	***1.140***	***0.337***
10	**Nasal**	**95**	**2.78 ± 1.27**	**2.72 ± 1.13**	***0.423***	***1.103***	***0.285***
18	**Tired**	**99**	**3.98 ± 1.50**	**3.90 ± 0.91**	***0.627***	***1.276***	***0.392***
26	None	72	2.38 ± 1.23	2.69 ± 1.10	0.216	0.641	0.191
29	**Chest**	**73**	**2.52 ± 1.26**	**3.12 ± 1.00**	***0.279***	***0.767***	***0.226***
32	Tired	97	3.94 ± 1.51	3.98 ± 0.87	0.614	1.808	0.323
34	**Activity**	**85**	**3.09 + 1.41**	**4.04 + 0.97**	***0.431***	***1.327***	***0.265***
39	**Activity**	**85**	**3.02 ± 1.42**	**3.86 ± 0.99**	***0.506***	***0.813***	***0.397***
40	**Activity**	**76**	**3.04 ± 1.46**	**3.37 ± 0.98**	***0.398***	***1.196***	***0.257***
41	**Activity**	**76**	**3.03 ± 1.38**	**3.56 ± 1.00**	***0.437***	***0.724***	***0.363***
42	**Activity**	**87**	**2.78 ± 1.34**	**3.50 ± 1.07**	***0.485***	***1.015***	***0.340***
43	**Activity**	**86**	**3.09 ± 1.42**	**3.91 ± 0.99**	***0.509***	***1.004***	***0.359***

Symptoms	*Symptoms*	*Symptoms*	*Functional impairments*
1. How sick do you feel today?	12. Body aches	23. Swollen glands	**34. Think clearly**
2. Cough	13. Feeling "run down"	24. Plugged ears	35. Speak clearly
3. Coughing stuff up	14. Sweats	25. Ear discomfort	**36. Sleep well**
4. Cough interfering with sleep	15. Chills	26. Watery eyes	**37. Breathe easily**
5. Sore throat	16. Feeling feverish	27. Eye discomfort	**38. Walk, climb stairs, exercise**
6. Scratchy throat	17. Feeling dizzy	**28. Head congestion**	**39. Accomplish daily activities**
7. Hoarseness	**18. Feeling tired**	**29. Chest congestion**	**40. Work outside the home**
8. Runny nose	19. Irritability	30. Chest tightness	**41. Work inside the home**
9. Plugged nose	20. Sinus pain	31. Heaviness in chest	**42. Interact with others**
10. Sneezing	21. Sinus pressure	32. Lack of energy	**43. Live your personal life**
11. Headache	22. Sinus drainage	33. Loss of appetite	**44. Compared to yesterday, I feel...**

The WURSS-21 has now been independently validated in a study of N=230 participants monitored for 2,457 person-days

"Validation of a Short Form Wisconsin Upper Respiratory Symptom Survey (WURSS-21)"

Bruce Barrett MD PhD, Roger Brown PhD, Marlon Mundt PhD,
Gay Thomas MS, Shari Barlow, Alex Highstrom, Mozhdeh Bahrainian MD

Health and Quality of Life Outcomes 7 (76), 2009

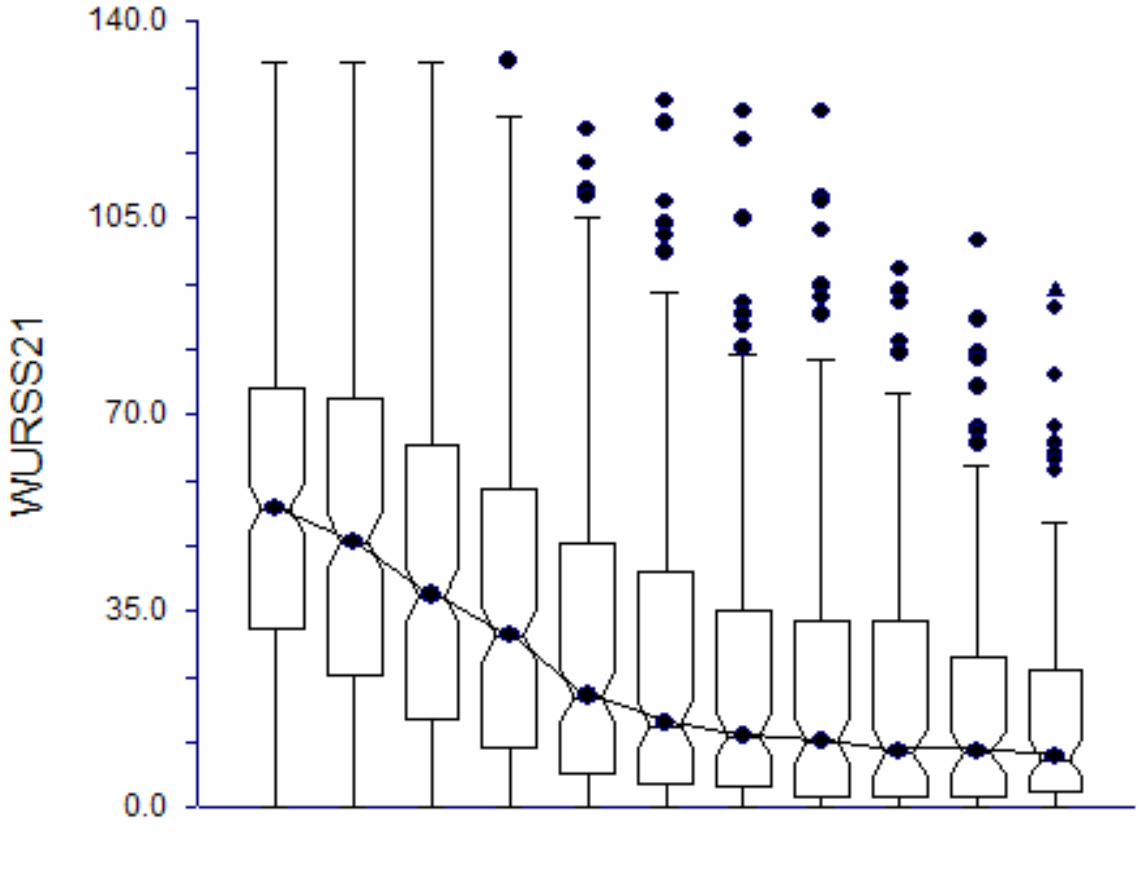
The top and bottom of the notched boxes indicate the 25% and 75% percentiles, respectively.

The center of the notched boxes is the median summed score for that day.

The notches can be compared to assess difference at the P = 0.05 level of significance

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



Health and Quality of Life Outcomes 7 (76), 2009

Frequency, severity, MID, MSE and responsiveness of

WURSS-21 Items

Frequency = present at least once during first seven days

Severity = Mean on 7-point scale averaged over first three days n ± SD

MID = minimal important difference

Responsiveness = MID / $(2*MSE)^{1/2}$

MSE = mean square error

Symptom	Frequency	Severity	MID	MSE	Responsiveness
How sick	100.0	4.13 ± 1.46	0.77	0.78	0.62
Runny nose	98.3	3.70 ± 1.77	0.56	1.48	0.33
Plugged nose	96.5	4.00 ± 1.79	0.57	1.54	0.32
Sneezing	95.7	3.34 ± 1.76	0.50	1.20	0.32
Sore throat	92.6	3.76 ± 1.85	0.49	1.00	0.35
Scratchy throat	96.1	3.82 ± 1.81	0.50	1.25	0.32
Cough	92.2	3.80 ±1.84	0.46	1.76	0.25
Hoarseness	86.1	3.38 ± 2.01	0.41	1.29	0.26
Head congestion	93.0	4.03 ± 1.70	0.64	1.54	0.37
Chest congestion	75.7	3.76 ± 1.88	0.38	0.97	0.27
Feeling tired	99.6	4.33 ± 1.80	0.82	1.41	0.49
Think clearly	91.3	3.53 ± 1.68	0.54	1.02	0.38
Sleep well	93.9	4.17 ± 1.82	0.66	1.69	0.36
Breathe easily	96.5	3.84 ± 1.86	0.60	1.08	0.41
Walk/Climb stairs	89.6	3.75 ± 1.81	0.50	0.88	0.38
Accomplish daily activities	90.4	3.57 ± 1.74	0.57	1.08	0.39
Work outside the home	82.2	3.80 ± 1.84	0.48	1.16	0.32
Work inside the home	87.0	3.52 ± 1.81	0.51	0.80	0.40
Interact with others	86.5	3.50 ± 1.73	0.53	0.93	0.39
Live your personal life	88.3	3.58 ± 1.74	0.58	0.92	0.43

Exploratory factor analysis for WURSS-21 using 2 to 7 dimensions

Dimensions	Chi-square	df	$\chi 2/df$	CFI	TLI	RMSEA	SRMR
2	1547.9	134	11.5	0.986	0.982	0.215	0.064
3	866.5	117	7.4	0.993	0.989	0.167	0.043
4	580.3	101	5.7	0.995	0.992	0.144	0.032
5	381.8	86	4.4	0.997	0.994	0.123	0.023
6	254.8	72	3.5	0.998	0.996	0.105	0.017
7	136.1	59	2.3	0.999	0.998	0.076	0.012

CFI = Comparative Fit Index

TLI = Tuker-Lewis Index

RMSEA = Root Mean Square Error of Approximation

SRMR = Standardized Root Mean Square Residual

Hu and Bentler (1999)[58] suggest the following cut off values for good fit, CFI > .95, TLI > .95, RMSEA < .06, and SRMR < .08

Best fit factorial model for WURSS-21

CFI = Comparative Fit Index

TLI = Tuker-Lewis Index

RMSEA = Root Mean Square Error of Approximation

WRMR = Weighted Root Mean Square Residual

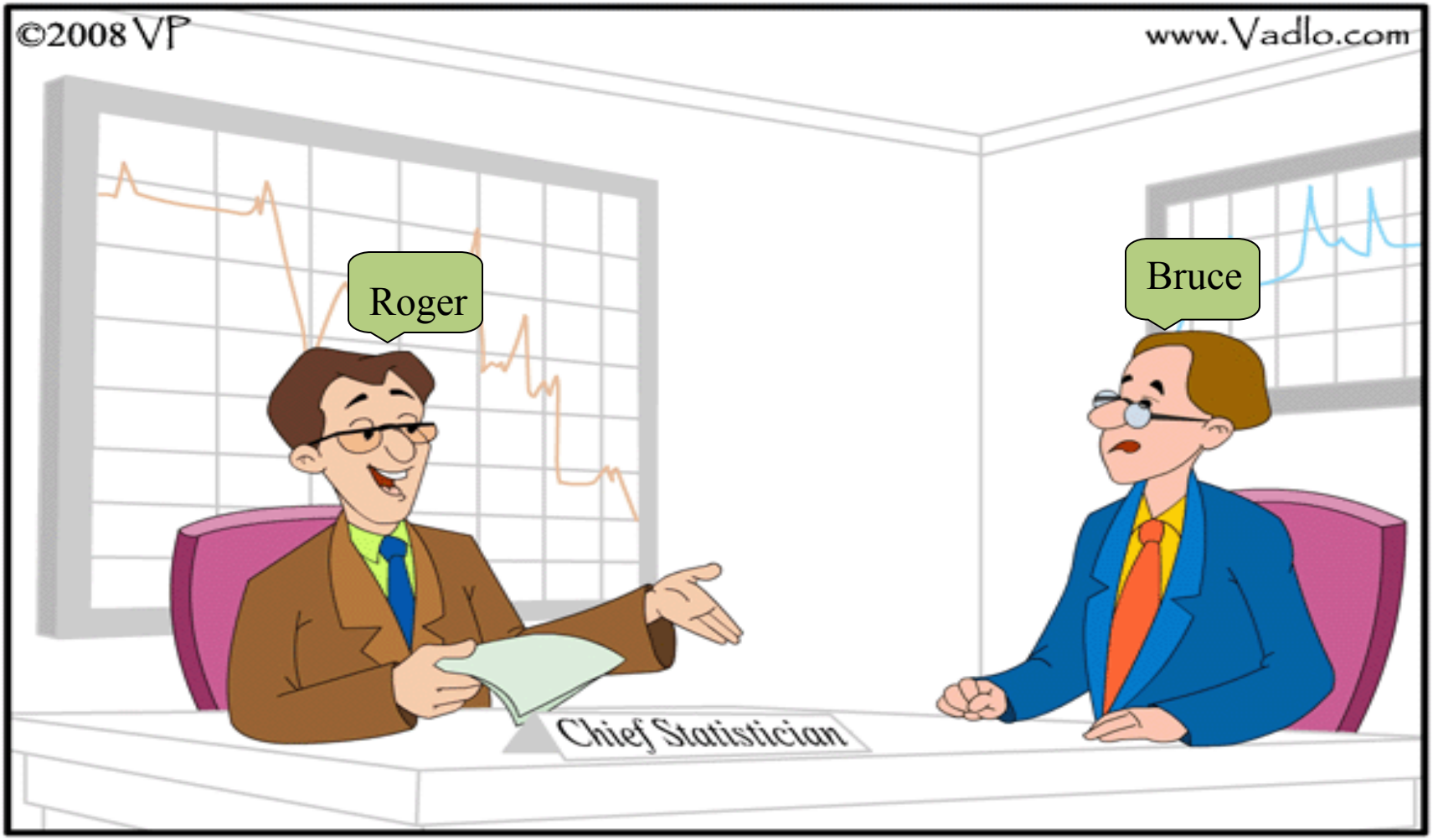
Hu and Bentler suggest the following cut off values for good fit, CFI > .95, TLI > .95, RMSEA < .06, and WRMR < .90

CFA Final model structure of the WURSS-21 at day 3				
Dimensions 3 restricted	Chi-square	df	$\chi 2$/df	CFI
	245.7	37	6.6	0.949
Number of items used = 20		TLI	RMSEA	WRMR
		0.990	0.157	1.074

Time invariance (configural invariance) Days 2 to 7				
Day	CFI	TLI	RMSEA	WRMR
2	0.903	0.978	0.170	1.234
3	0.949	0.990	0.157	1.074
4	0.962	0.993	0.157	1.047
5	0.970	0.995	0.145	0.973
6	0.983	0.995	0.147	1.030
7	0.980	0.995	0.132	0.909

L. T. Hu and P. M. Bentler. Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 6:1-55, 1999.

In tribute to Roger Brown PhD, statistician/psychometrician/methodologist extraordinaire



I can prove it or disprove it, what do you want me to do?

University of California San Francisco (2002-2008); University of Queensland, Australia (2002); University of Wolverhampton, England (2002-04); Case Western University (2003); Harvard University (2003); University of Southern Indiana (2003); Novartis Consumer Health, Hungary (2003); Baylor University in Texas (2003-04); University of Brisbane, Australia (2004-05); Bayer-sponsored URI trial in Italy (2004-06); University of Iowa (2004); Altamira Family Medicine, San Juan, Puerto Rica (2005); UCLA - Harbor UCLA (2005-06); Laboratoires Goemar, Saint Malo, France (2005-06); University of Leuven, Belgium, with field work in Peru (2005-2007); Northwestern University (2005-06); Unilever, the Netherlands (2006); Hindustan Lever Research Centre, Bangalore, India (2006); Baskent University, Ankara, Turkey (2006); California School of Professional Psychology, San Diego (2006); McMaster, Hamilton, Ontario, Canada (2006); University of North Carolina, Chapell Hill (2006); Navy Medical Center Portsmouth (2006); Appalachian State University (2006); Bastyr University, Seattle (2006); Philadelphia Center for Health Equity Research & Promotion (2006); University of Rochester (2007); Duke University Clinical Research Institute (2007); California School of Professional Psychology, Sacramento (2007); Washington University School of Medicine (2007); Georgetown University (2007); East Carolina University (2007); Research Institute for Sport and Exercise Sciences, Liverpool, U.K. (2007); Universiti Teknologi MARA Shal Alam, Malaysia (2007); Daejeon University, South Korea (2007); University of Wales at Bangor, Wales, U.K. (2007); Washington University School of Medicine, St. Louis MO (2007); Georgetown University, Washington DC (2007); SouthHamptom University Hospital, England (2007); Fundacion Santafe, Bogota, Columbia (2007); University of Western Australia (2007); McNeil Consumer HealthCare, Fort Washington, Pennsylvania (2007), University of Iowa, Coralville, Iowa (2007); Research Institute, Lancaster General Hospital, PA (2007); Davao Medical School, Davao City, Philipines (2007); Northwestern University, Chicago (2007); Georgetown University, Washington DC (2007); Pharma Projects, Czech Republic (2007); Castle Hill, Cottingham, U.K. (2007); Northcentral University, Pembroke Pines, FL (2007); Northcentral University, Pembroke Pines, Florida (2007); Queen Mary University, London, England (2007); University Hospital Heidelberg, General Practice and Health Services Research, Heidelberg, Germany (2007); University of California San Francisco (2007); Department of Physiology, Medical University of Vienna, Austria (2007); Université de Montreal (2008); St John Fisher College, Rochester, N.Y. (2008); Massey University / Zespri, Mt Maunganui,New Zealand (2008); Burgmann College, Canberra, Australia (2008); Coventry University, Coventry, U.K.(2008); University of South Carolina (2008); University of Texas, San Antonio (2008); University of Georgia (2008); University of Western Australia (2008); Washington University, St. Louis (2008); James Cook University, Australia (2008). Research Institute, Lancaster General Hospital (2008); University of Jyväskylä, Finland (2008); University of Exeter, U.K. (2008); Griffith University, Gold Coast Campus, Australia (2008); University of Massachusetts, Amherst (2008); Shinobu University, Shinobu, Japan (2008); Georgetown University (2009); Murrayville Family Practice, Langley, Canada (2009); University of Montana, Missoula (2009); Georgetown University, Smithtown, New York (2009); University of Torino, Italy (2009); Sports Medicine, Technical University of Munich, Germany (2009); Faculdade de Educacao Fisica de Limeira, Sao Paulo, Brasil (2009); University of KwaZulu-Nata, Durban, South Africa (2009); Allergy & Immunology, Vzsoké Tatry, Slovakia (2009); University College, London, England (2009); Texas Women's University, Houston, Texas (2009); Clinical Consulting, Quebec, Canada (2009); University of Sumatera Utara, Medan, Indonesia (2009); Family Practice, Trinidad and Tobago (2009); Georgetown University (2009); Biopharmacopae Design International, Quebec, Canada (2009), Seattle Children's Hospital, Seattle, Washington (2009), University of Alberta, Alberta, Canada (2009), Reckitt Benckiser Inc., Long Valley, New Jersey (2009), University of Torino, Italy (2009), Pediatric allergy and immunology; University of Wisconsin Hospitals (2009), Asthma Bhawan, Rajasthan, India (2009), Chinese Clinical Trials Coordination Center, Beijing, China (2009), University Technology MARA, Selangor, Malaysia (2009), Preventive and Rehab. SportMedicine, Technical University of Munich, Germany (2009), Medicus Research, LLC, Midlothian, VA (2009), Northcentral University, Mesa, AZ (2009), GAMD Inc, Los Angeles, CA (2009)

Evidence Based Medicine

- Clinical medicine is increasingly guided by evidence, mostly from randomized controlled trials (RCTs)
- RCTs provide the least biased estimates of the effects of medical interventions (drugs, surgery, counseling, lab tests, screenings, etc)

Guyatt GH, Rennie D. Users' Guides to the Literature: A Manual for Evidence-Based Clinical Practice. Chicago: AMA Press, 2002.

Sackett DL et al. Evidence-based medicine: How to practice and teach EBM. f Edinburgh, London, New York: Churchill Livingstone, 2000.

L. M. Friedman, C. D. Furberg, and D. L. DeMets. *Fundamentals of Clinical Trials*, New York:Springer-Verlag, 1998.

RCTs powered to detect specific effect sizes

- Small (underpowered) RCTs can fail to detect benefits large enough to be considered clinically significant
- Very large (overpowered) trials can detect benefits too small to be considered clinically significant

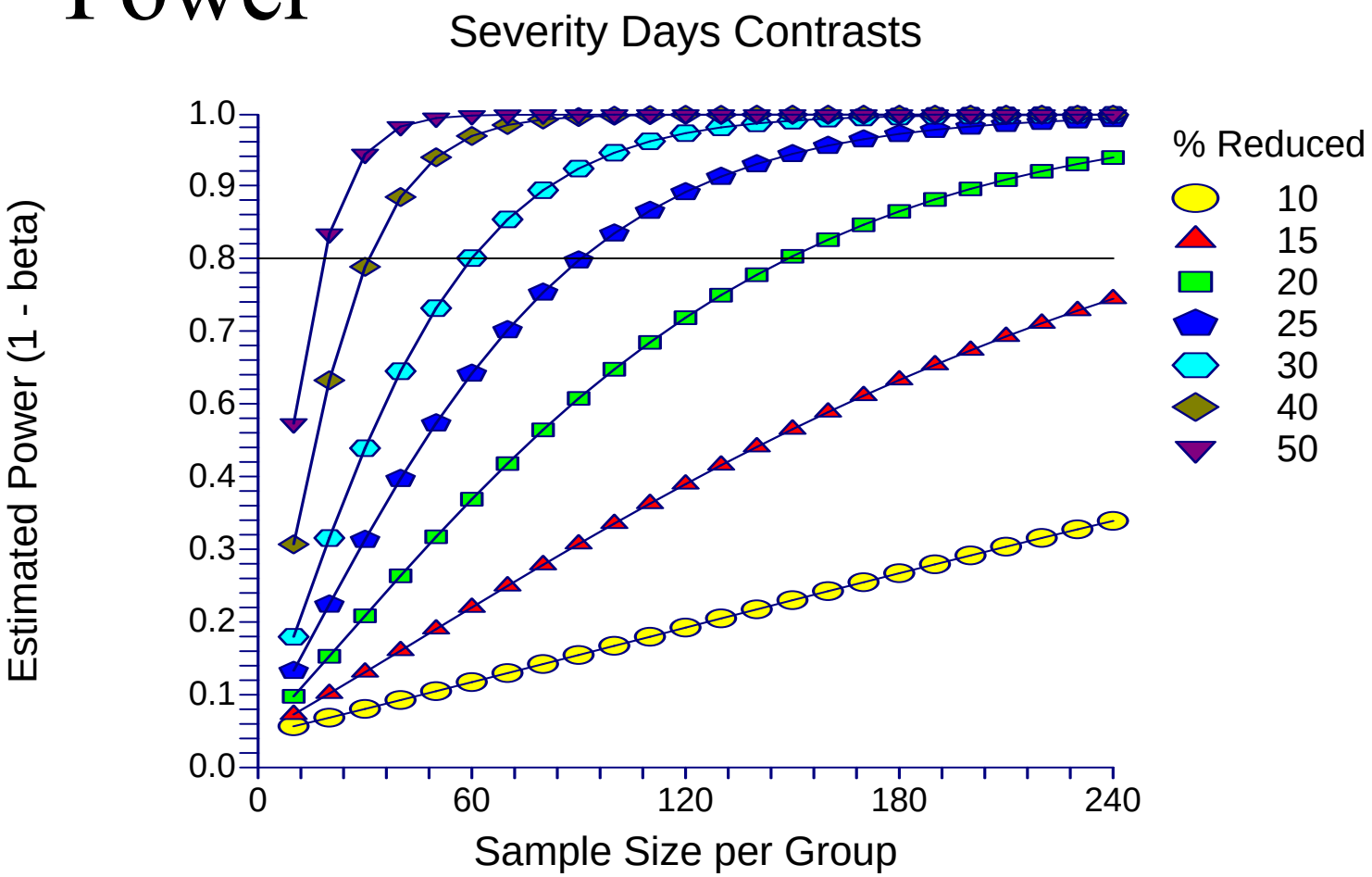
L. M. Friedman, C. D. Furberg, and D. L. DeMets.
Fundamentals of Clinical Trials, New York:Springer-Verlag, 1998.

Correctly powered trials

- RCTs should be powered to detect the smallest effect size that is considered to be clinically significant
- This way, we minimize chances of both false-positive and false-negative trials

M. Egger, G. D. Smith, and D. Altman. *Systematic Reviews in Health Care: Meta-analysis in Context*, London:BMJ Publishing Group, 2001.

Power



240 subjects per group provides more than 80% power ($1 - \beta = .80$) to detect a 20% difference in area-under-the-curve severity, with $\alpha = 0.05$ and assuming one-sided testing and proportionally stable standard deviation

Clinical significance

- "How much of an effect is important?"
- Conventionally, any statistically significant effect was considered important
- Hence, important was equated with detectable
- Clinical & statistical significance were confused

Guyatt GH et al. Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clinic Proceedings.* 2002;**77**:371-83

Mitchell PH. The significance of treatment effects: Significance to whom? *Medical Care* 1995;**33**:AS280-AS285.

Naylor CD,.Llewellyn-Thomas HA. Can there be a more patient-centerd approach to determining clinically important effect sizes for randomized treatment trials? *ournal of Clinical Epidemiology.* 1994;**47**:787-95.

Clinical significance

- Cohen's ***d*** standardized effect size (ES) = absolute effect size / standard deviation (SD)
- SD is a measure of background variability
- ES's up to 0.2 were classified as "small," 0.5 as "medium," and 0.8 as "large"
- Unsatisfactory. Degree of benefit considered important (clinically significant) should not depend on background variability

Cohen J. Statistical Power Analysis for the Behavioural Sciences. London: Academic Press, 1969

Minimal important difference

- MID originally defined as:

  "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troubling side effects and excessive cost, a change in the patient's management"

G. H. Guyatt, C. Bombardier, and P. Tugwell. Measuring disease-specific quality of life in clinical trials. *Canadian Medical Association Journal* 134:889-894, 1986

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

Minimal important difference

- Can be measured by comparing self-reported global improvement to interval changes on HRQoL outcome instruments
- "Compared to yesterday, I feel that my [cold] is..."

  Better The Same Worse

.

G. H. Guyatt, E. F. Juniper, S. D. Walter, L. E. Griffith, and R. S. Goldstein. Interpreting treatment effects in randomised trial. *BMJ*. 316 (7132):690-693, 1998

G. Samsa. How should the minimum important difference for a health-related quality-of-life instrument be estimated? *Medical Care* 39 (10):1037- 1038, 2001.

Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Current Opinion in Rheumatology*. 2002;**14**:109-14

If better	If worse
1. Almost the same, hardly any better at all	1. Almost the same, hardly any worse at all
2. A little better	2. A little worse
3. Somewhat better	3. Somewhat worse
4. Moderately better	4. Moderately worse
5. A good deal better	5. A good deal worse
6. A great deal better	6. A great deal worse
7. A very great deal better	7. A very great deal worse

Minimal important difference

- Minimal important difference (MID) is usually operationalized as the average amount of change between two consecutive HRQoL assessments for everyone who responds “a little better” or “somewhat better”

R. Jaeschke, J. Singer, and G. H. Guyatt. Measurement of health status: Ascertaining the minimal clinically important difference. *Controlled Clinical Trials* 10:407-415, 1989.

Responsiveness

- Also known as sensitivity to change, responsiveness is defined as the ability to detect change over time
- Guyatt's responsiveness index =

$$\frac{\text{Minimal Important Difference}}{(2*\text{MSE})^{1/2}}$$

G. H. Guyatt, S. Walter, and G. Norman. Measuring change over time: Assessing the usefulness of evaluative instruments. *J Chron Dis* 40:171-178, 1987.

G. H. Guyatt, R. A. Deyo, M. Charlson, M. N. Levine, and A. Mitchell. Responsiveness and validity in health status measurement: a clarification. *Journal of Clinical Epidemiology*. 42 (5):403-408, 1989.

MID, Responsiveness, Power

	WURSS-21	WURSS-44	Jackson
MID	9.48	16.7	1.56
Responsiveness	0.80	0.71	0.54
Trial size (# subjects)	74	92	124

(Assuming 2 arms, alpha = 0.05, beta 0.90, 2-tailed testing)

	one-tailed α = 0.005 (two-tailed α =0.01)			one-tailed α = 0.025 (2-tailed α =0.05)			one-tailed α = 0.05 (two-tailed α =0.10)		
ß=	0.05	0.10	0.20	0.05	0.10	0.20	0.05	0.10	0.20
Power	95%	90%	80%	95%	90%	80%	95%	90%	80%
Sample size per group needed to detect day-to-day MID (using Guyatt's responsiveness coefficient)									
WURSS-21	72	60	47	52	43	32	44	35	25
WURSS-44	64	53	42	47	38	28	39	31	22
WURSS-21 - Sample size per group needed to detect between group AUC differences of:									
10%	2348	1961	1540	1712	1385	1035	1426	1129	815
20%	578	483	379	421	341	255	351	278	201
30%	259	217	171	189	153	115	157	124	90
40%	147	123	97	107	87	65	89	71	51
50%	95	80	63	69	56	42	57	46	33
WURSS-44 - Sample size per group needed to detect between group AUC differences of:									
10%	2787	2328	1828	2033	1644	1228	1693	1340	967
20%	697	583	458	508	411	307	423	335	242
30%	312	261	205	227	184	138	189	150	108
40%	176	147	116	128	104	78	107	85	61
50%	113	95	75	82	67	50	69	55	40

Every medical intervention has:

- Potential *benefits* (improved symptoms and functions over time) AND
- Potential *harms* (side effects, adverse events, monetary costs, opportunity costs)

H. C. Sox, M. A. Blatt, M. C. Higgins, and K. I. Marton. *Medical Decision Making*, Boston, London, Oxford:Butterworth-Heinemann, 1988.

M. R. Gold, J. E. Siegel, L. B. Russell, and M. C. Weinstein. *Cost-Effectivenss in Health and Medicine*, Oxford, New York:Oxford University Press, 1996.

What about costs, side effects, and other potential harms?

- MID methods only assess benefit
- Shouldn't RCTs be designed to detect, (and interpreted in the light of) net benefit, where likely positive effects are sufficient to outweigh potential negative effects?

Sufficiently Important Difference

- SID = "the smallest amount of patient-valued benefit that an intervention would require in order to justify associated costs, risks, and other harms"

B. Barrett, R. Brown, M. Mundt, L. Dye, J. Alt, N. Safdar, and R. Maberry. Using benefit harm tradeoffs to estimate sufficiently important difference: the case of the common cold. *Medical Decision Making* 25 (1):47-55, 2005.

B. Barrett, D. Brown, M. Mundt, and R. Brown. Sufficiently important difference: expanding the framework of clinical significance. *Medical Decision Making* 25 (3):250-261, 2005.

SID Methodology

- To estimate the magnitude of benefit that patients consider sufficient to justify costs and risks of treatments
- Common cold as model
- **Benefit harm trade-off** interviews were developed, tested & utilized

Benefit Harm Trade-Offs

- Interviews of people with illness
- Portray evidence-based treatment scenarios
  including benefits, costs, and risks
- Systematically vary one dimension until respondent
  changes orientation from “would take treatment”
  to “would not take treatment” (or vice-versa)

Sampling Framework

- People with self-identified colds in Madison, Wisconsin, responding to community advertisement
- At least 1 of 4 symptoms: sneezing, runny nose, nasal obstruction, or sore throat; Jackson score ≥ 3
- A) Enrolled within 48 hours of first symptom, monitored until cold is resolved. Interviewed at ***intake***, when sick, and again at ***exit,*** after illness resolved N=151
- B) Additionally, 162 people with self-identified colds were interviewed by telephone

Benefit Harm Trade Offs

- First, we describe costs, potential side effects, and possible symptomatic benefits. Then we ask:
- "Would you take this treatment?"

  [Along with qualitative "Why or why not?"]
- Next, we say "On average, colds last about 6 days"

  Then ask, "If this treatment could reduce the length of your cold by 1 day, would you take it?"

Benefit Harm Trade Offs

- If the participant responds “Yes,” we decrease hypothetical duration benefit to “12 hours,” then if still “Yes” to “6 hours,” then if still “Yes” to “Any duration at all.”
- If the participant first responds “No,” we increase hypothetical duration benefit to “2 days,” then “3 days,” then “4 days,” then finally “Any duration at all.”

Tx scenario based on vitamin C evidence

- "A 10-cent vitamin pill must be taken 3 times daily for the first 3 days of your cold. There are no significant risks or side effects to this treatment. Severity of symptoms might be reduced by as much as 20%."

Douglas RM, Chalker EB, Treacy B. Vitamin C for preventing and treating the common cold. *Cochrane Database of Systematic Reviews* 2003.

Tx scenario based on echinacea evidence

- "A 50-cent dropperful of an herbal extract must be taken by mouth 3 times each day for the first 3 days of your cold. Side effects are limited to bad taste. Severity of symptoms might be reduced by as much as 30%."

D. Melchart, K. Linde, P. Fischer, and J. Kaesmayr. Echinacea for preventing and treating the common cold. *Cochrane Database of Systematic Reviews* (1), 2003.

Tx scenario based on zinc evidence

- "A 20-cent lozenge must be dissolved in the mouth every 2 to 3 hours while awake for the first 3 days of your cold. Side effects are limited to bad taste, and occasionally nausea. Severity of symptoms might be reduced by as much as 20%."

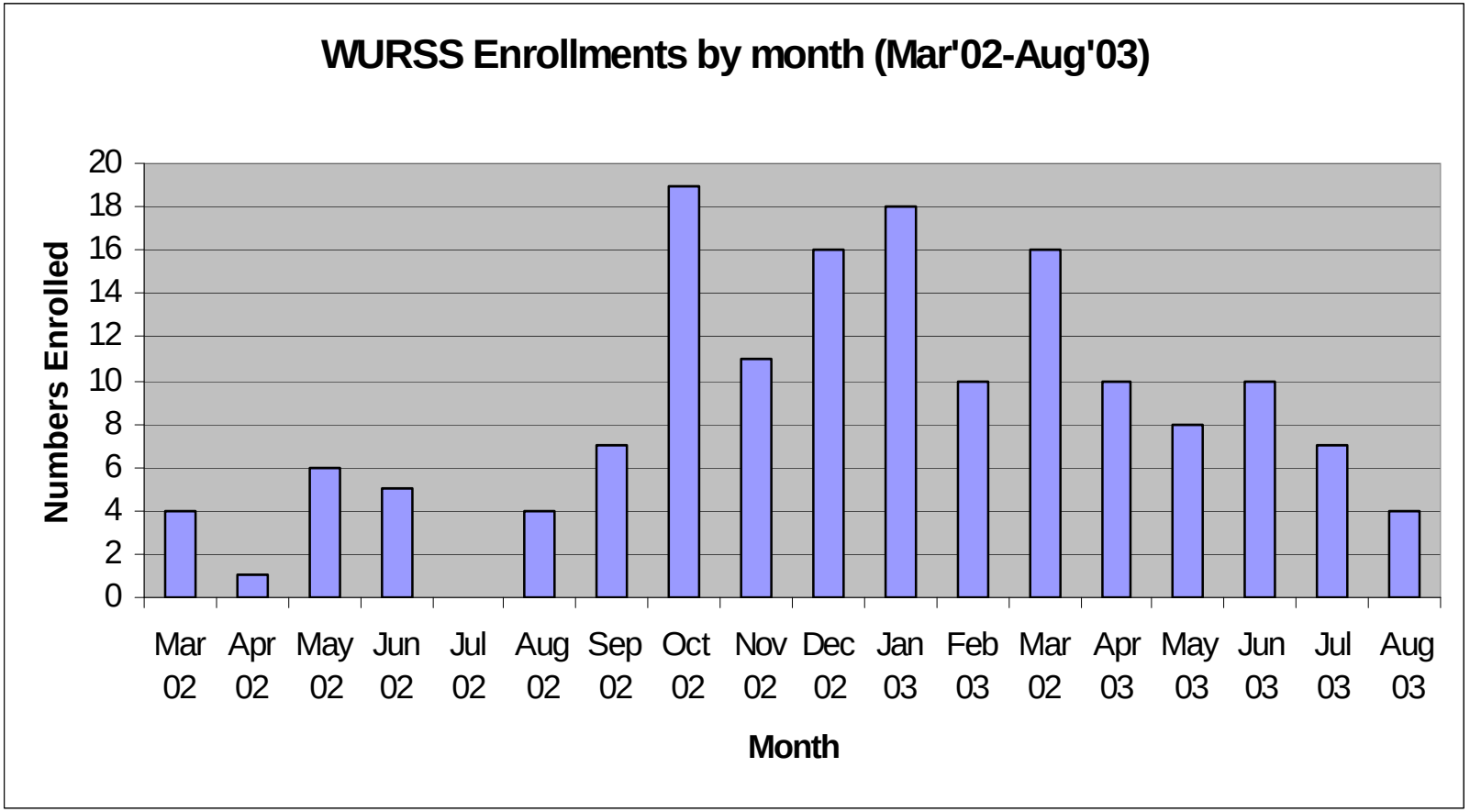
I. Marshall. Zinc for the common cold.
Cochrane Database of Systematic Reviews (Issue:4):4, 2002.

Tx scenario based on evidence regarding pleconaril, an antiviral

- "A $2 prescription-only pill must be taken 3 times daily for the first 3 days of the cold. Side effects are unknown. Severity of symptoms might be reduced by as much as 40%."

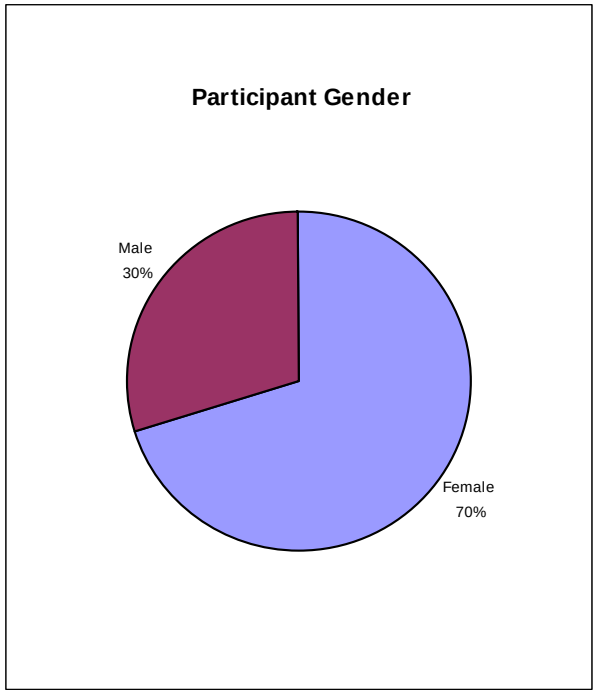
Hayden FG, Coats T, Kim K, Hassman HA, Blatter MM, Zhang B *et al*. Oral pleconaril treatment of picornavirus-associated viral respiratory illness in adults: efficacy and tolerability in phase II clinical trials. *Antiviral Therapy*. 2002;**7**:53-65.

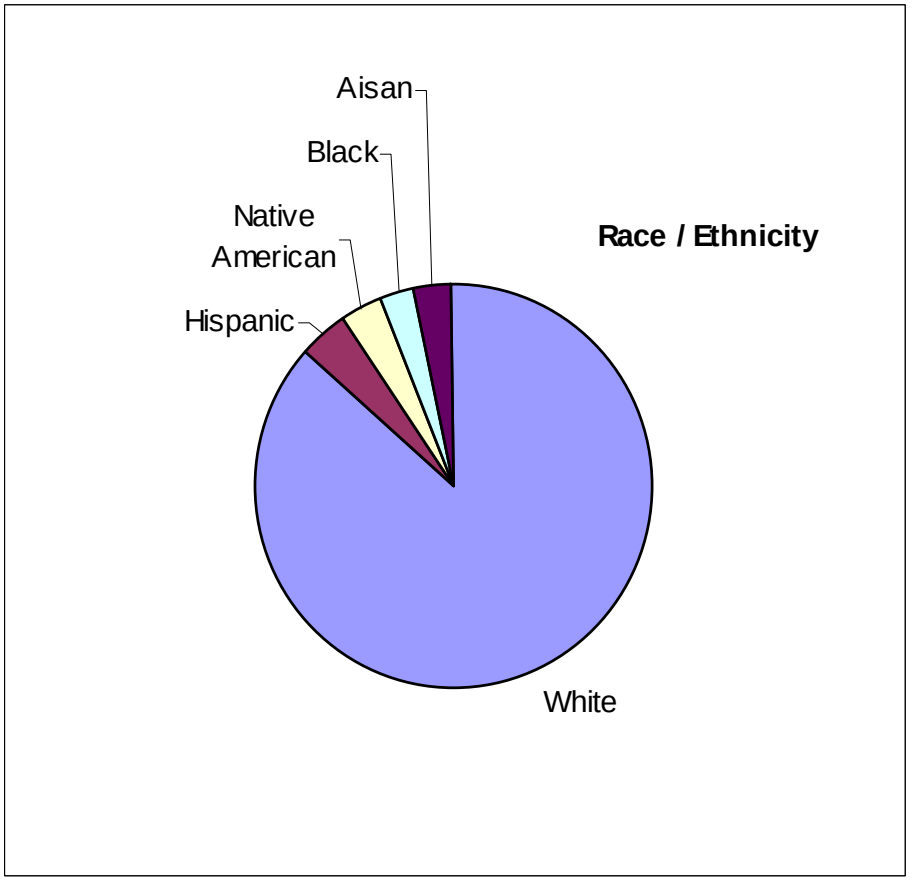
Phase 1 - March 2002 to August 2003



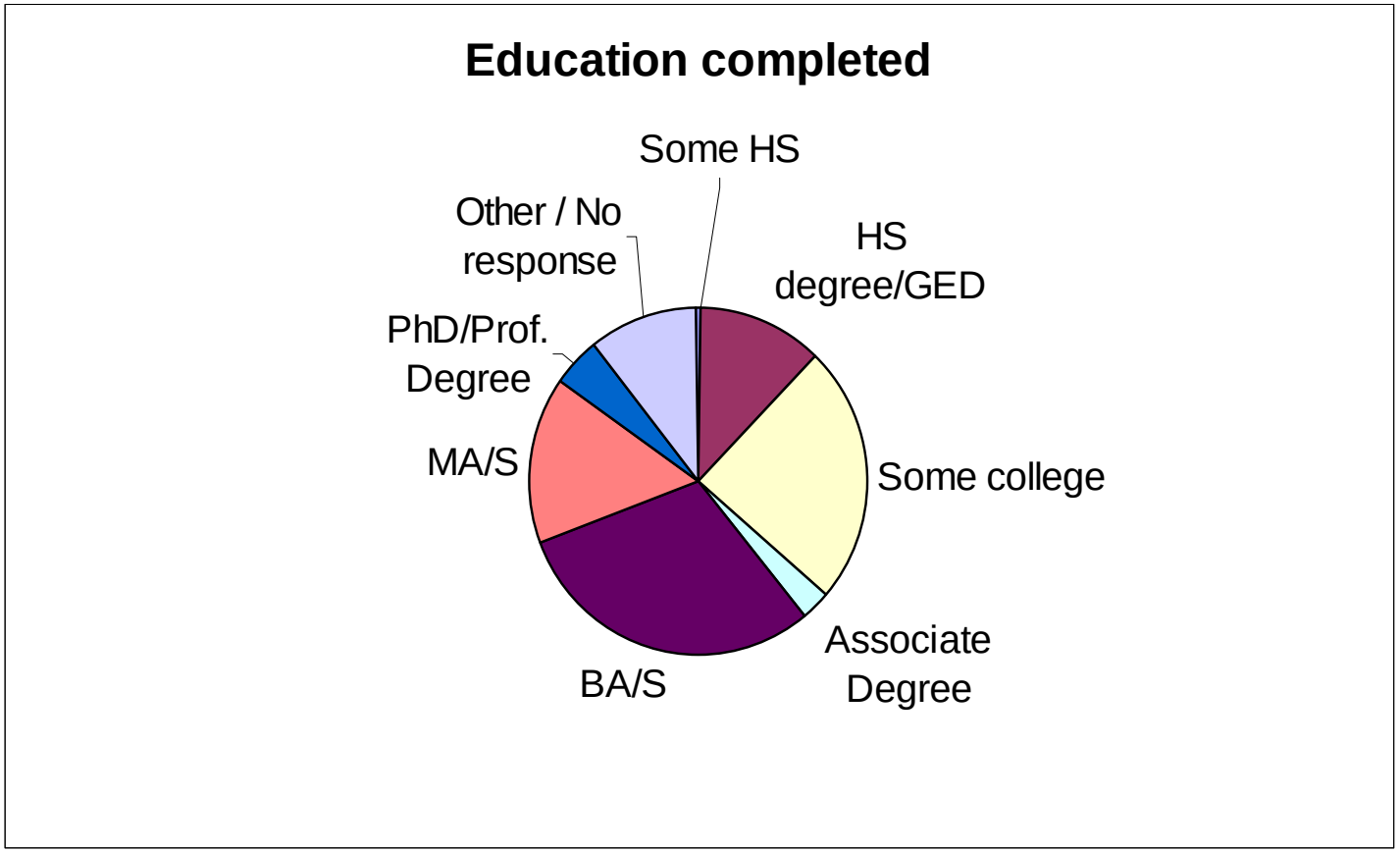
Age, gender & ethnicity of participants

Age range: 18 to 80 years
Mean = 35.5, SD = 14.7

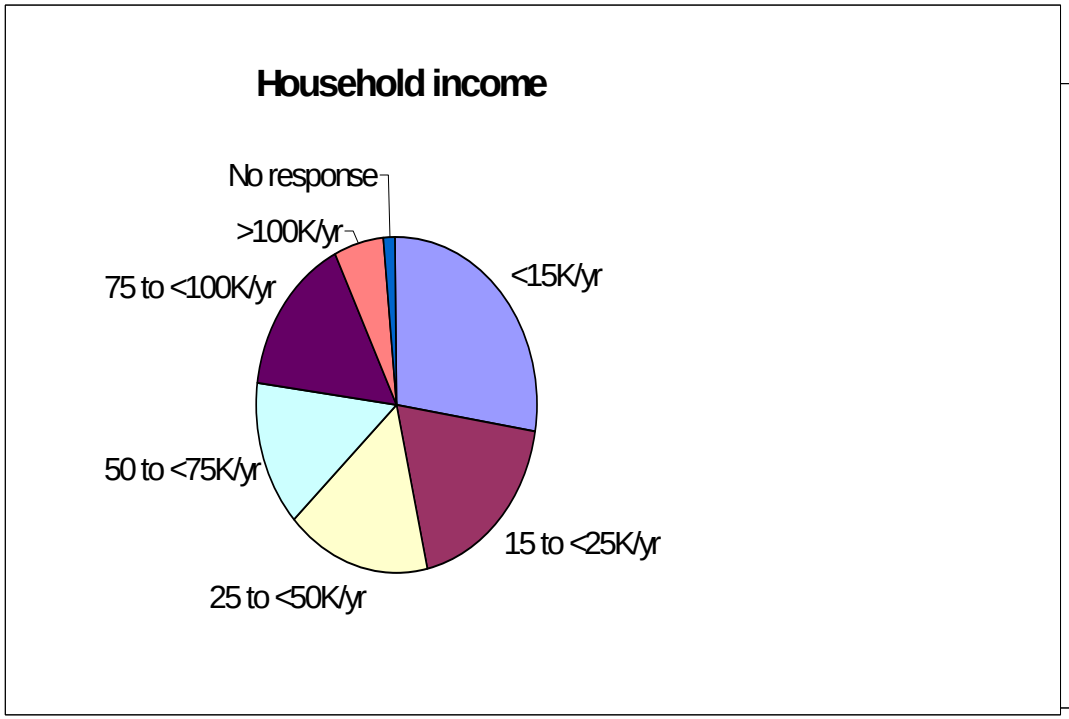


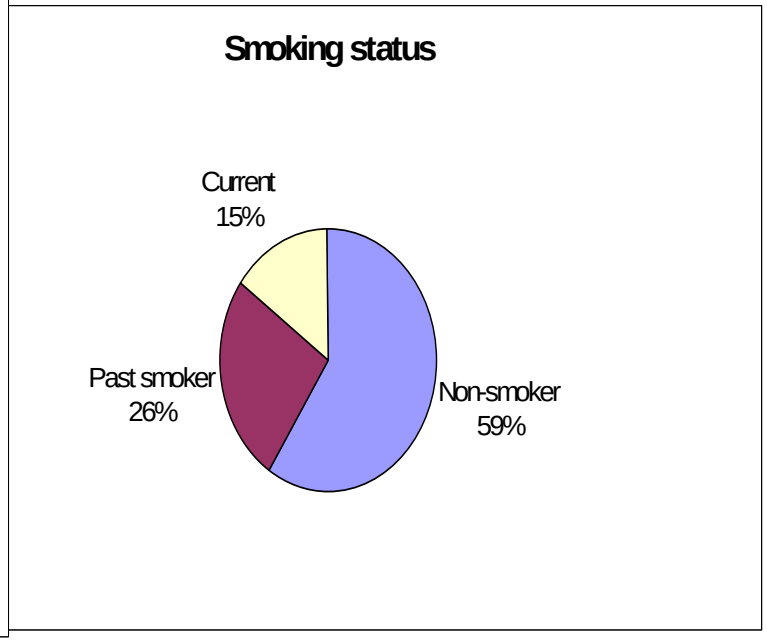


Educational level of participants



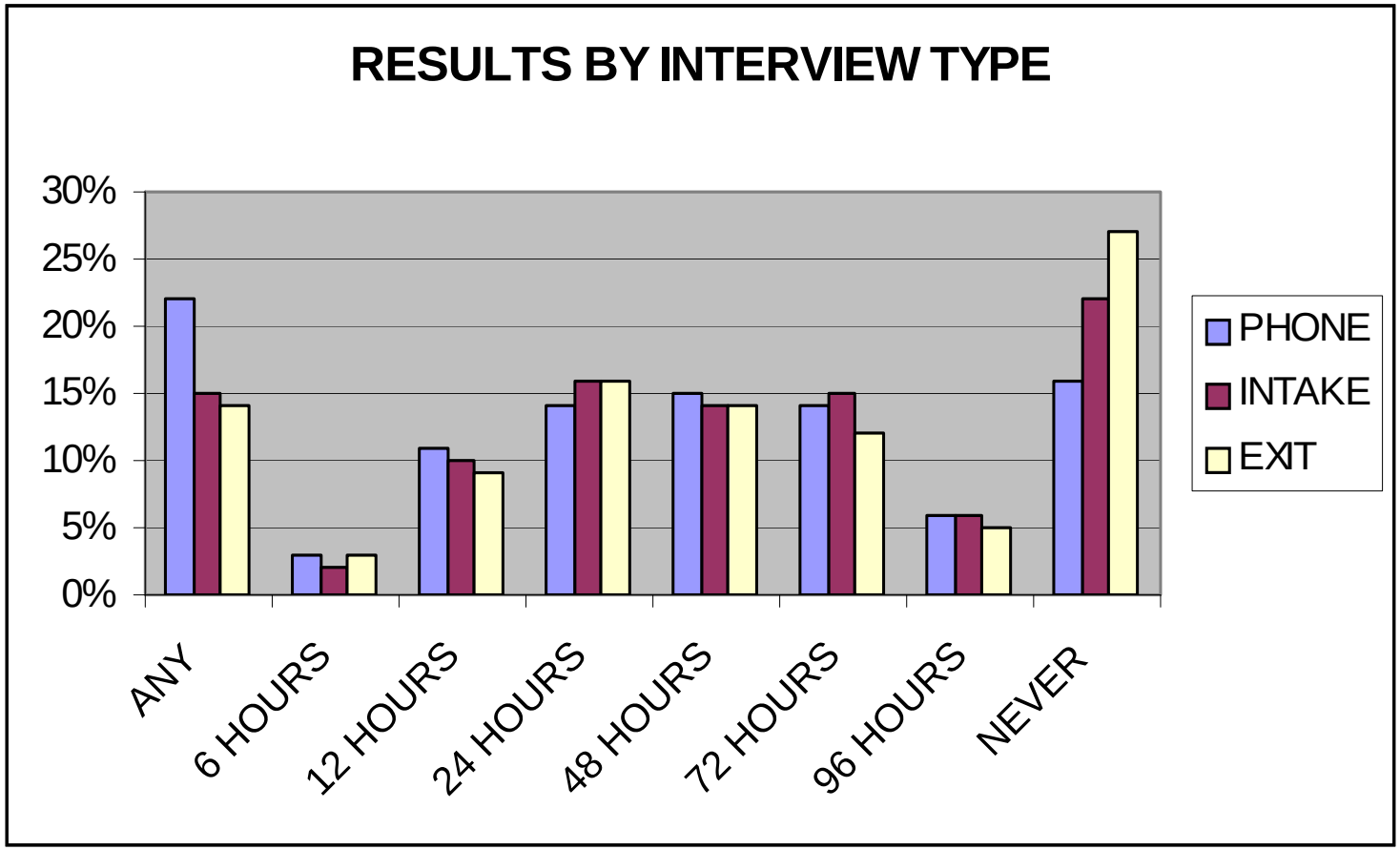
Income, smoking status



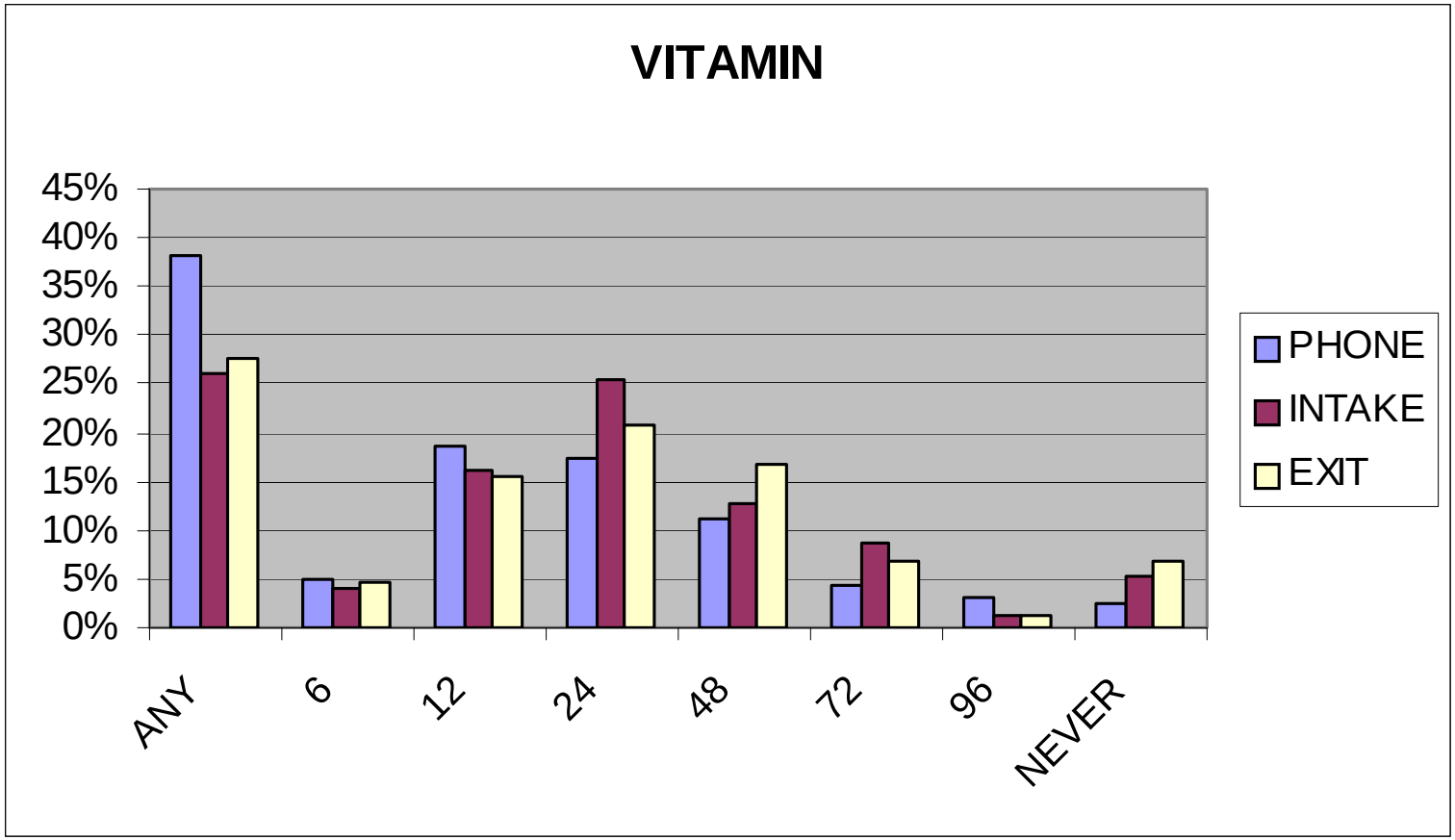


Overall distribution of responses

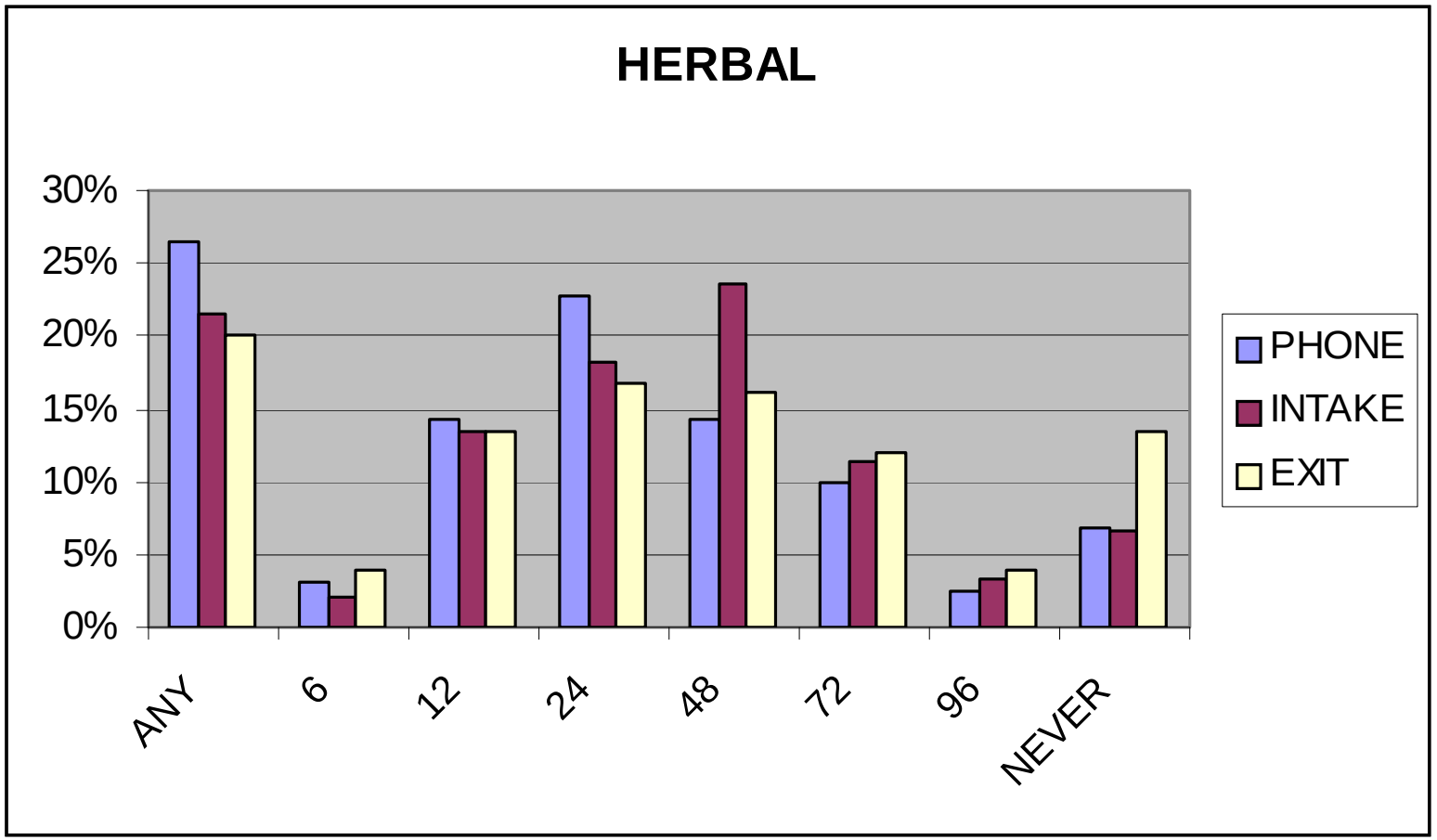
(460 interviews; 1,840 scenarios)



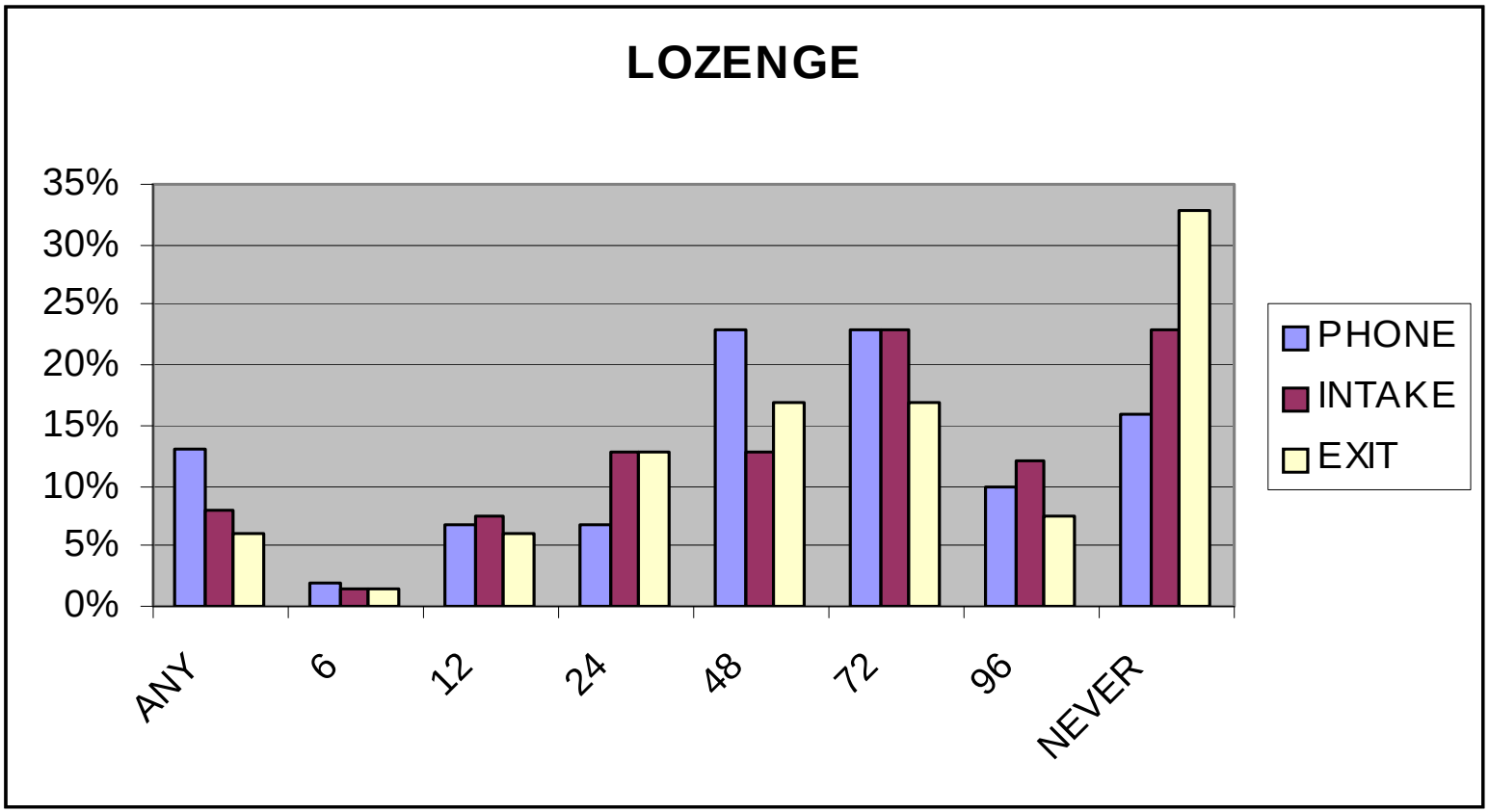
Responses to vitamin scenario



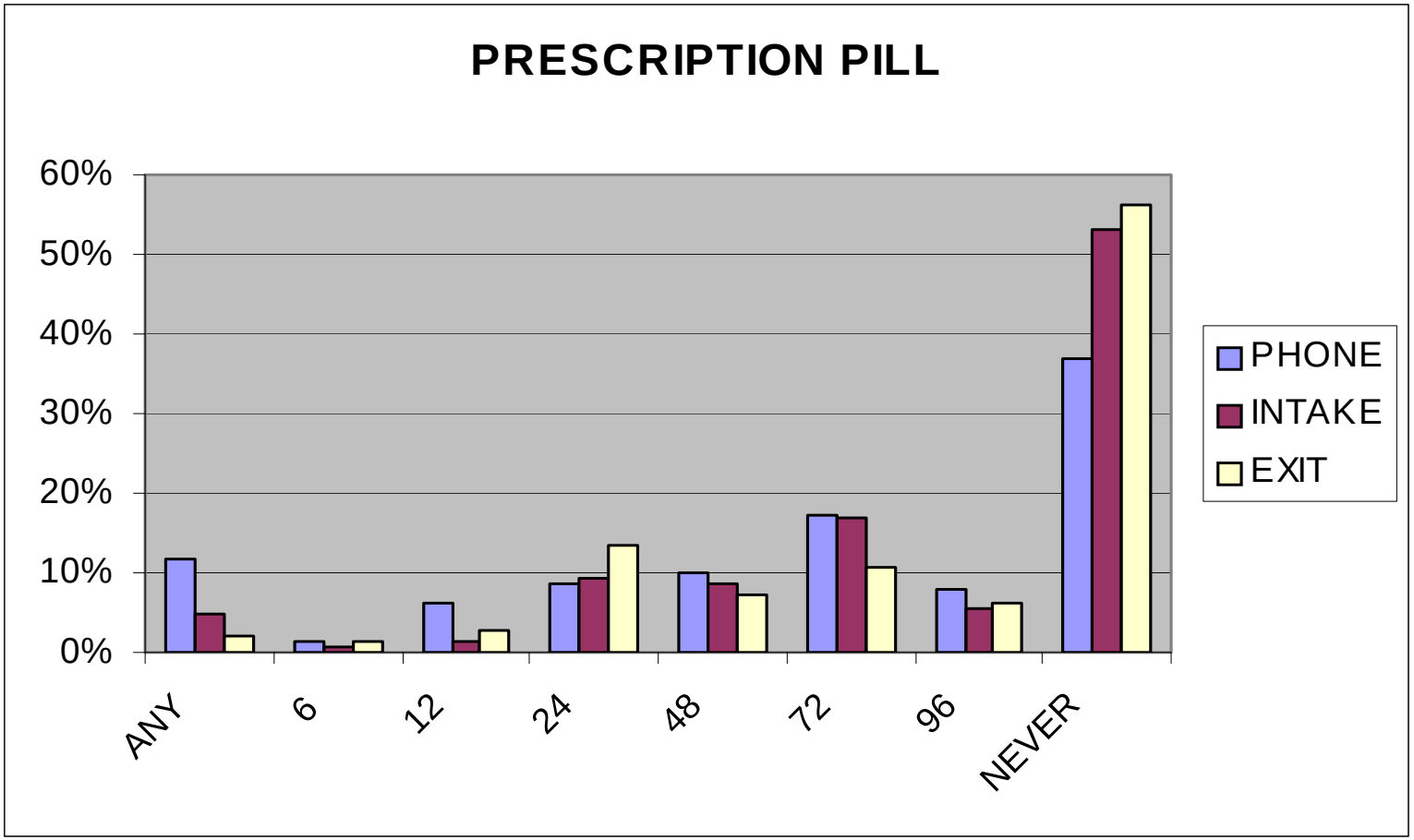
Responses to herbal scenario



Responses to lozenge scenario



Responses to antiviral scenario



These results were published as:

Using benefit harm tradeoffs to estimate sufficiently important difference: the case of the common cold.

Medical Decision Making 25 (1):47-55, 2005.

B.Barrett, R.Brown, M.Mundt, L.Dye, J.Alt, N.Safdar R.Maberry.

Next we repeat the model in another cohort, asking about severity reduction instead of duration reduction

SID severity cohort

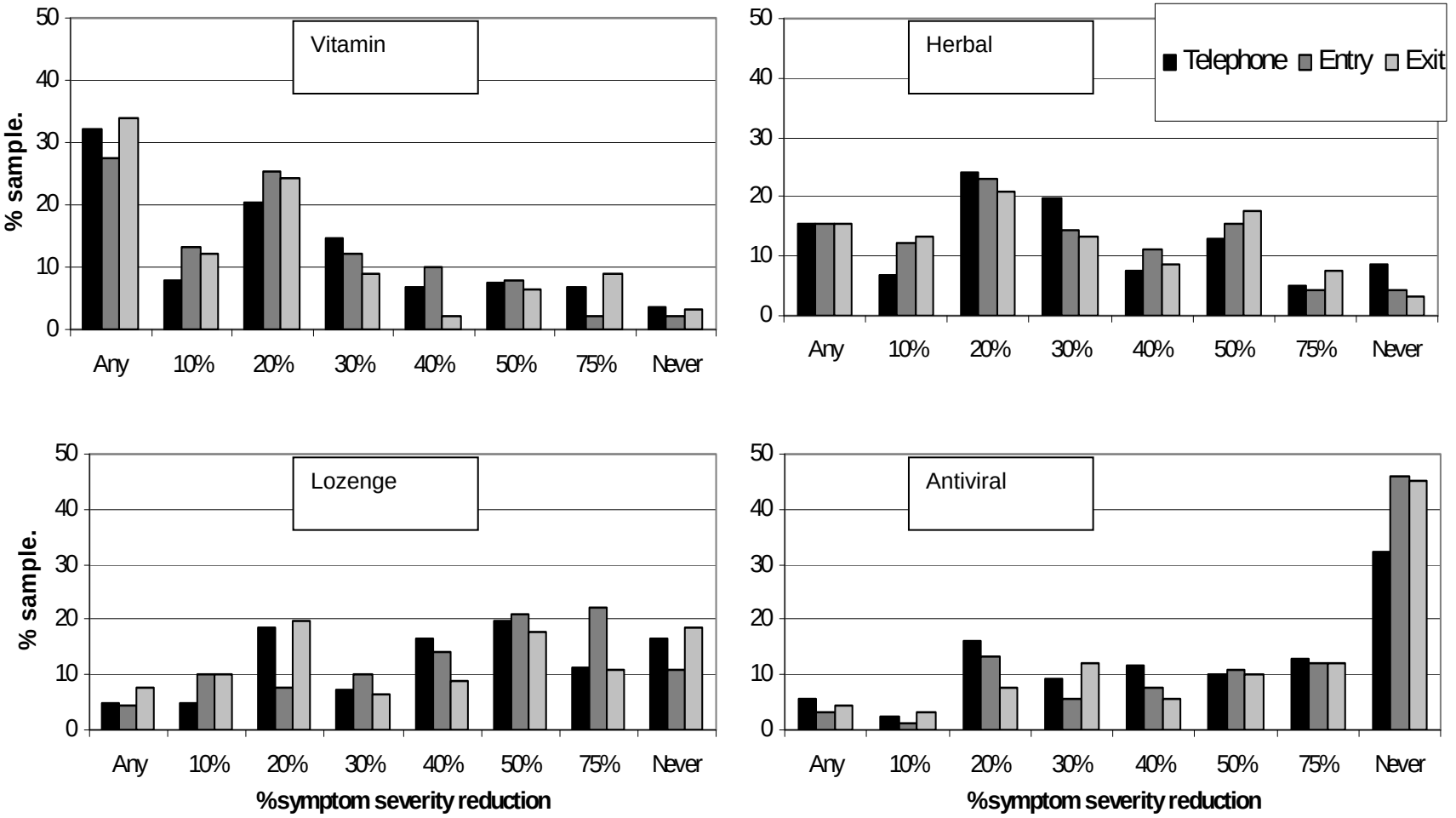
- 983 people screened; 253 participated (2004-05)
- N= 91 in person – at beginning & after cold ends (182 interviews)
- N= 162 telephone interviews; in first 5 days of cold
- N= 344 interviews; 1,376 scenarios & responses
- Mean age = 34.8 years (SD 13.3)
- 68% women, 68% white
- 46% reported household income < $25,000/year

G. G. Jackson, H. F. Dowling & R. L. Muldoon. Present concepts of the common cold. *Am J Public Health* 52 (6):940-945, 1962.

Bidding starts at 30% average overall severity reduction

- "Severity of symptoms might be reduced by *as much as 30%.*"
- Participants answering "Yes" are asked: "Would you take this [treatment] if it were able to reduce severity by 20%?"
- If still "Yes," hypothetical severity reduction is lowered to "10%," then if still "Yes" reduced to "5%," and, finally, "any?"
- If the first answer was "No," then severity reduction benefit was increased to "40%," then if still "No" to "50%," then "75%", then "any?"

SID Severity reduction by treatment scenario



Conclusions

- The SID framework seems valid
- Benefit harm trade-off interviews seem to work
- Response distribution is wide, but characteristic and reproducible
- Severity of illness and socioeconomic status have little impact on response distribution, at least for these people with common cold

Implications

- No currently available cold treatment provides the average benefit that most respondents say they would want in order to justify costs and risks
- If similar results are found for other interventions with small effect sizes

  (eg. anticholinergics for dementia, antihyperglycemics for type 2 diabetes, antidepressants for depression, statins for preventing heart attack and stroke, cancer screening, etc etc etc…)

  there may be profound policy implications

In terms of benefit harm trade-offs for medical decision-making, **one size does not fit all**!

Questions raised

Should trials be powered to detect SID effect sizes thought to be worthwhile by 50% of population? 75% ? 25% ? 10% ?

Should an intervention be approved by FDA or other regulators when 75%, 90% or 99% of potential users say the benefits don't justify the harms?

Should practice guidelines (and clinician report cards, bonuses, salaries) be based on effect sizes so small that most people wouldn't want them?

Should manufacturers be required to assess and report SID?

Should FDA/regulators or non-conflicted health researchers be empowered (funded) to assess and report SID?

Limitations

- Quality of evidence behind scenarios
- Our ability to summarize evidence without bias
- Participants' ability to weigh several hypothetical benefit and harm domains
- Potential bias arising from self-report of hypothetical intentions, not actual behaviors
- Loads of inter-person variability

Other related papers….

- B. Barrett, D. Brown, M. Mundt, and R. Brown. Sufficiently important difference: expanding the framework of clinical significance.
  Medical Decision Making 25 (3):250-261, 2005
- B. Barrett, R. Brown, R. Voland, R. Maberry, and R. Turner.
  Relations among questionnaire and laboratory measures of rhinovirus infection.
  European Respiratory Journal 28 (2):358-363, 2006
- B. Barrett, B. Harahan, D. Brown, Z. Zhang, and R. Brown.
  Sufficiently important difference for common cold: Severity reduction.
  Annals of Family Medicine 5 (4):216-223, 2007.
- B. Barrett, S. Endrizzi, P. Andreoli, S. Barlow, and Z. Zhang.
  Clinical significance of common cold treatment: professionals' opinions.
  Wisconsin Medical Journal 106 (8):473-480, 2007.
- B. Barrett, R. Brown, and M. Mundt. Comparison of anchor-based and distributional approaches in estimating important difference in common cold.
  Quality of Life Research 17 (1):75-85, 2008.