



Queensland University of Technology
Brisbane Australia

This is the author's version of a work that was submitted/accepted for publication in the following source:

Goldsmith, C.H., Duku, E., Brooks, P.M., Boers, M., Tugwell, P.S.L., & Baker, P.R. (1995) Interactive conference voting. *Journal of Rheumatology*, 22(7), pp. 1420-1430.

This file was downloaded from: <http://eprints.qut.edu.au/55749/>

© Copyright 1995 Journal of Rheumatology Publishing Co. Ltd.

Notice: *Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source:*

Request Details

http://docdel.librariesaustralia.nla.gov.au/zengine?VDXaction=IIIDe...

Request Details

HML email

Actions available: Shipped

Action



Mark as Unread

◀ Previous Next ▶

Request Details

QUT

Request ID

Our Number 9406115

Their Number 9406113

Status In Process

Authorisation Status Acknowledged

Service Details

Service 1

Service Type Copy non returnable

Media Type Printed Journal

Service 2

Service Type Loan

Media Type None

Item Details (Journal Article)

Journal Title The Journal of rheumatology.

Subtitle

Publisher Journal of Rheumatology Pub. Co.

ISSN 0315-162X 0315-162X

Volume / Issue 22 (7)n

Part Date 1995-07-01

Classmark

Call Number hmlp RC927 .J65 : 1(1974)-34(2007); 35(2008)-

Article Details

Title INTERACTIVE CONFERENCE VOTING

Article Author CH GOLDSMITH

Pages 1420-1430

Control Numbers

DNLM 7501984

Library of Congress cn 77318229

Local (AuCNLKIN) b11265231

Additional Nos ANBD: 000000909865;LCN: cn 77318229;CODEN JRHUA9;POSR: 737930
USPS

Ref. Source Libraries Australia/.biball-r20-db01

Additional Service Details

Requested By QUT Kelvin Grove

Service Level Core

Entry Date 07 Nov 2012 14:54

Interactive Conference Voting

CHARLIE H. GOLDSMITH, ERIC DUKU, PETER M. BROOKS, MAARTEN BOERS,
PETER S.L. TUGWELL, and PHILIP BAKER for the OMERACT II Committee

ABSTRACT. We describe and analyze opinion polling results from interactive voting procedures undertaken before and after presentations during the Outcome Measures in Rheumatoid Arthritis Clinical Trials Conference (OMERACT II) in Ottawa, Canada, June 30–July 2, 1994. The scoring procedure was a matched voting design; when a participant used the same keypad at the beginning and end of voting, change within a participant could be estimated. Participants, experienced in the rheumatic diseases included clinicians, researchers, methodologists, regulators, and representatives of the pharmaceutical industry. Patients under consideration were those with any rheumatic diseases. Questions were constructed to evaluate the change in voting behavior expected from the content of the presentation. Statistically significant and substantively important changes were evident in most questions. (*J Rheumatol* 1995;22:1420–30)

Key Indexing Terms:

REMOTE VOTING

CHANGE

EVALUATION

The OMERACT II Conference took place at the University of Ottawa, Ottawa, Canada, June 30 to July 2, 1994. At most conferences, data on the opinions of participants are not recorded along with the substantive written papers in the proceedings. This conference was different, in that an interactive voting system was used, and the results of the voting during the conference are summarized here.

MATERIALS AND METHODS

A set of 200 wireless keypads connected to a central computer and display screen was used by the participants. The keypads were placed on the arms of chairs in the conference auditorium. The wireless response system (RSVP) was provided by Photo Communications Corporation. Participants were asked to vote before and after a talk using the same keypad, so votes could be linked in the analysis. Participants seated in chairs without keypads did not have their votes counted. Also, if 2 different participants used the same keypad before and after a presentation, this could not be identified. During the voting period, participants were given about 6 seconds to respond, the vote was transmitted to the central computer, and the results were projected on a screen at the front of the auditorium. However, the connection

From the Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton; the Department of Physical Therapy, University of Western Ontario, London, Canada; the Department of Medicine, St. Vincent's Hospital, Sydney, Australia; the Department of Medicine/Rheumatology, University Hospital Maastricht, The Netherlands; and the Department of Medicine, University of Ottawa, Ottawa, Canada.

Analyses of these data were made possible by support provided by the Conn Smythe Foundation.

C.H. Goldsmith, PhD, Professor of Clinical Epidemiology and Biostatistics, Department of Clinical Epidemiology and Biostatistics, McMaster University and Honorary Professor of Physical Therapy, Department of Physical Therapy, University of Western Ontario; E. Duku, MSc, Research Associate, Department of Clinical Epidemiology and Biostatistics, McMaster University; P.M. Brooks, MD, FRACP, Professor of Medicine, Department of Medicine, St. Vincent's Hospital; M. Boers, MD, PhD, MSc, Associate Professor of Rheumatology, Department of Internal Medicine/Rheumatology, University Hospital Maastricht; P. Tugwell, MBBS, MD, MSc, Professor and Chair, Department of Medicine, University of Ottawa; P. Baker, BSc, Research Associate, Department of Medicine, University of Ottawa.

Address reprint requests to Dr. C.H. Goldsmith, St. Joseph's Hospital, Centre for Evaluation of Medicines, 50 Charlton Avenue East, Martha Wing, Hamilton, ON L8N 4A6, Canada.

between the before votes and the after votes was not presented during the conference. This is the first analysis of these connected votes.

RESULTS

Opening session. During the opening session Peter Tugwell welcomed participants and informed them about the interactive voting system. A series of questions were used to familiarize participants with the voting system, and to allow them to ask questions about how the system worked. The first 4 questions were designed to test participants' knowledge about Canadian and North American mnemonics, as follows.

Question: "RCMP stands for:"

Responses: 1. Really Cool Mountain Partyers. 2. Royal Canadian Mounted Police. 3. Royal Canadian Mashed Potatoes. 4. Royal Canadian Medical Practitioners.

	Response	Total	%1**	%2***
Time 1	1	12	13.5	6.0
	+ 2	45	50.6	22.5
	3	17	19.1	8.5
	4	15	16.9	7.5
	*	111		55.5
Total		200	100.1	100.0

* keypads with no votes recorded

** percentage of total responders

*** percentage of total possible keypads

* most likely answer

Each table of results shows the numbers for the response options (Responses) as well as a missing value label (*) for keypads with no one voting. The most appropriate answer is denoted with a plus sign (+). The column labelled %1 is the percentage of participants who responded in that response category, rounded to one decimal place. Similarly, the column labelled %2 is the percentage of the responses, including the missing data. From the percentage in last column [the missing value code (*)], one can compute the percentage of keypads used in that vote; i.e., $100 - 55.5 = 44.5\%$ (89/200). When a "most appropriate" response was designated, a 95% confidence interval (95% CI) using the exact method in the CIA software¹ was computed. In

this example, there were 89 respondents, 45 (50.6%) voted for Response 2, Royal Canadian Mounted Police, with a 95% CI of 39.8-61.3%.

Question: "Which of the following foods is not Canadian?"
 Responses: 1. Beaver tails. 2. Poutine. 3. Tourtiere. 4. Burger with the lot. 5. Bannock.

	Response	Total	%1	%2
Time 1	1	9	9.4	4.5
	2	7	7.3	3.5
	3	7	7.3	3.5
	+4	34	35.4	17.0
	5	39	40.6	19.5
	*	104		52.0
Total		200	100.0	100.0

There were 96 respondents with the most appropriate response, 4, having 34 (35.4%) votes with a 95% CI 25.9-45.8%.

Question: "Which of the following is not a Canadian beer?"
 Responses: 1. Kokanee. 2. XXXX. 3. Labatt's Blue. 4. Grizzly. 5. UC Rebellion.

	Response	Total	%1	%2
Time 1	1	12	12.5	6.0
	+2	44	45.8	22.0
	3	5	5.2	2.5
	4	8	8.3	4.0
	5	27	28.1	13.5
	*	104		52.0
Total		200	99.9	100.0

There were 96 respondents with the most appropriate response, 2, having 44 (45.8%) votes with a 95% CI 35.6-56.3%.

Question: "Which of the following is not a musketeer?"
 Responses: 1. Athos. 2. Porthos. 3. Aramis. 4. Dumas. 5. D'Artagnan.

	Response	Total	%1	%2
Time 1	1	3	3.3	1.5
	2	4	4.4	2.0
	+3	44	48.4	22.0
	4	30	33.0	15.0
	5	10	11.0	5.0
	*	109		54.5
Total		200	100.1	100.0

Question: "Which of the following factors do not increase the likelihood of an adverse reaction?"
 Responses: 1. Patient's age. 2. History of allergy. 3. Socioeconomic status. 4. Number of different drugs prescribed. 5. Drug dose.

Response	Time 2					Total	%1	%2	
	1	2	3	4	5				
Time 1	1	1		1		4	6.8	3.0	
	2		2	16		6	27.3	12.0	
	+3			47	1	3	58.0	25.5	
	4			2		1	5	5.7	2.5
	5			1			1	2.3	1.0
	*		8			104		56.0	
Total						200	100.1	100.0	
	Total	1	2	75	1	1	200		
	%1	1.3	2.5	93.8	1.3	1.3	100.2		
	%2	0.5	1.0	37.5	0.5	0.5	60.0	100.0	

There were 91 respondents with the most appropriate response, 3, having 44 (48.4%) votes with a 95% CI 37.7-59.1%.

M. Boers introduced a series of questions that focussed on the conference objectives.

Question: "In classic psychoanalysis, obsessive-compulsive behavior has the strongest relation with:"

Responses: 1. Unstable heart rate. 2. "Fight or flight" reaction. 3. Sphincter tone. 4. Fear of rotating restaurants.

	Response	Total	%1	%2
Time 1	1	5	8.9	2.5
	2	7	12.5	3.5
	+3	28	50.0	14.0
	4	16	28.6	8.0
	*	144		72.0
Total		200	100.0	100.0

There were 56 respondents with the most appropriate response, 3, having 28 (50.0%) votes with a 95% CI 36.3-63.7%.

Question: "The objectives of OMERACT II do not include:"

Responses: 1. Setting of a research agenda. 2. A solid consensus on the best measures in each of 3 modules: toxicity, health status, economics. 3. Sharing and acquiring knowledge in each of the modules. 4. Strengthening the network of people in measurement in rheumatology.

	Response	Total	%1	%2
Time 1	1	14	24.6	7.0
	2	38	66.7	19.0
	3	0	0.0	0.0
	4	5	8.8	2.5
	*	143		71.5
Total		200	100.1	100.0

Here, 57 participants responded; however, all answers could be sensible. Hence, no confidence intervals or most appropriate responses are highlighted.

Formal presentations. The first plenary speaker, R.O. Day, introduced 2 questions about adverse reactions. These questions were asked before the presentation (Time 1) and again after the presentation (Time 2).

The change in these tables is that the Total, %1, and %2 now correspond to the Response categories for the after voting; these are seen as rows at the bottom of the table. Now the most appropriate category can be estimated both before and after the presentation. There were 88 participants responding before with 51 (58.0%) voting most appropriately (95% CI: 47.0–68.4); and 80 responding after with 75 (93.8%) voting most appropriately (95% CI: 86.0–97.9). While this is an apparent increase in most appropriate responses, these estimates do not take into account the matched keypads responses. There were indeed 96 participants who responded at least once during Time 1 and Time 2 voting, and from these (deleting the 104 who were missing at both times) the matched estimates were before, 53.1% and after, 78.1%, with an increase of 25.0%; 95% CI: 14.0–31.0%. Using a McNemar correlated contingency table analysis with Chi-Square software², the chi square was $X^2(10) = 34.606$, $P2 = 0.0001$; and using the collapsed 2×2 table, $X^2(1) = 18.000$, $P2 < 0.0001$. Both the correlated confidence interval and chi squares show that there was an increase in the most appropriate response as a result of the presentation.

Question: "Which of the following features characterize a type B adverse reaction?"
 Responses: 1. Predictable. 2. Usually dose related. 3. Mortality more likely than type A adverse reaction. 4. Phenytoin induced ataxia an example.

Response	Time 2					Total	%1	%2
	1	2	3	4	*			
1	3	1	3		5	12	17.1	6.0
2		4	3	5	3	15	21.4	7.5
Time 1 +3	1		15	4	10	30	42.9	15.0
4		1	3	7	2	13	18.6	6.5
*		1	14	7	108	130		65.0
Total						200	100.0	100.0
Total	4	7	38	23	128	200		
%1	5.6	9.7	52.8	31.9		100.0		
%2	2.0	3.5	19.0	11.5	64.0	100.0		

There were 70 before respondents, with 30 (42.9%) voting response 3 (95% CI: 31.1–55.3) and 72 after respondents, with 38 (52.8%) voting response 3 (95% CI: 40.7–64.7). However, from the 92 that voted at least once, the correlated estimates were before, 32.6% and after, 41.3%, for an increase of 8.7% (95% CI: -5.5–21.4). The McNemar $X^2(9) = 17.254$, $P2 = 0.0449$; collapsed $X^2(1) = 1.684$, $P2 = 0.1944$. From these tests and the confidence interval, there was no change in the most appropriate response due to the presentation.

J. Fries posed 2 questions related to the toxicity of non-steroidal antiinflammatory drugs (NSAID).

Question: "Is it possible to definitely define the toxicity of a specific NSAID?"

Responses: 1. Yes. 2. No.

Response	Time 2			Total	%1	%2
	1	2	*			
1	19	7		26	31.0	13.0
Time 1 +2	11	42	5	58	69.0	29.0
*	3	5	108	116		58.0
Total				200	100.0	100.0
Total	33	54	113	200		
%1	37.9	62.1		100.0		
%2	16.5	27.0	56.5	100.0		

There were 84 before respondents, with 58 (69.0%) voting response 2 (95% CI: 58.0–78.7) and 87 after respondents, with 54 (62.1%) voting response 2 (95% CI: 51.0–72.3). However, from the 92 who responded at least once, the before rate was 63.0% and after, 58.7%, for a decrease of 4.3% (95% CI: -15.7–7.8). The McNemar $X^2(3) = 3.889$, $P2 = 0.2737$; the collapsed $X^2(1) = 0.571$, $P2 = 0.4497$. Hence, there was no change in the most appropriate responses due to the presentation.

Question: "When carefully recorded clinical experience is different from clinical trial experience, this most likely means:"

Responses: 1. The clinical trial is correct. 2. Clinical experience is correct. 3. Neither can be assumed correct. 4. Both may be correct, but asking different questions.

Response	Time 2					Total	%1	%2
	1	2	3	4	*			
1				1		1	1.3	0.5
2				1		1	1.3	0.5
Time 1 3			1	3		4	5.3	2.0
+4			1	65	4	70	92.1	35.0
*	1			15	108	124		62.0
Total						200	100.0	100.0
Total	1	0	2	85	112	200		
%1	1.1	0.0	2.3	96.6		100.0		
%2	0.5	0.0	1.0	42.5	56.0	100.0		

There were 76 before respondents, with 70 (92.1%) voting response 4 (95% CI: 83.6–97.1) and 88 after respondents, with 85 (96.6%) voting response 4 (95% CI: 90.4–99.3). However, from the 92 who responded at least once, the before rate was 76.1% and after, 92.4%, for an increase of 16.3% (95% CI: 5.0–23.5). Collapsing response 2 with *, the McNemar $X^2(4) = 10.200$, $P2 = 0.0372$; and the collapsed $X^2(1) = 12.960$, $P2 = 0.0032$. Hence the presentation clearly increased the most appropriate responses.

P. Peloso posed 2 questions that raised the issue of reporting toxicity in clinical trials. These questions used Likert scales, where there was no most appropriate response, since all opinions could be valid.

Question: "The assessment of patient toxicity in the rheumatology randomized clinical trial is well standardized."

Responses: 1. Strongly agree. 2. Agree. 3. Unsure. 4. Disagree. 5. Strongly disagree.

		Time 2					*	Total	%1	%2
Response		1	2	3	4	5				
Time 1	1					1	1	1.2	0.5	
	2		2		2		2	7.1	3.0	
	3		1	1	2	6		10	11.8	5.0
	4				6	23	4	33	38.8	16.5
	5			1	2	28	4	35	41.2	17.5
	*				2	7	106	115	57.5	
Total							200	100.1	100.0	
Total		0	3	2	14	65	116	200		
%1		0.0	3.6	2.4	16.7	77.4		100.1		
%2		0.0	1.5	1.0	7.0	32.5	58.0	100.0		

From the 85 before respondents the median was 4; while from the 84 after respondents the median was 5, an increase in the strength of the agreement after the presentation of 1 point on the 5 point scale. Collapsing response 1 with *, the

McNemar $X^2(8) = 30.211$, $P2 = 0.0002$. Clearly the presentation increased the disagreement with the standardization of toxicity statement.

Question: "The case report form for the assessment of toxicity is essentially equivalent across trials, with minor variations only."

Responses: 1. Strong agree. 2. Agree. 3. Unsure. 4. Disagree. 5. Strongly disagree.

		Time 2					*	Total	%1	%2
Response		1	2	3	4	5				
Time 1	1					1	1	2	2.2	1.0
	2	1	2		1	3		7	7.8	3.5
	3			1		4	1	6	6.7	3.0
	4	1	1	1	4	25	5	37	41.1	18.5
	5		1			32	5	38	42.2	19.0
	*		1	1		3	105	110	55.0	
Total							200	100.0	100.0	
Total		2	5	3	5	68	117	200		
%1		2.4	6.0	3.6	6.0	81.9		100.2		
%2		1.0	2.5	1.5	2.5	34.0	58.5	100.0		

From the 90 before respondents the median was 4; while from the 83 after respondents the median was 5, an increase on the strength of the agreement after the presentation of 1 point on the 5 point scale. The McNemar $X^2(13) = 41.500$, $P2 = 0.0001$. Clearly the presentation increased the disagreement with the standardization of case report form statement.

S. van der Linden discussed the role of sensitivity to change in the development of instruments to be used in the rheumatic diseases.

Question: "Sensitivity to clinically relevant change is most important for which one of the following instruments?"

Responses: 1. Remission. 2. Prognosis. 3. Classification. 4. Disease activity in clinical trials.

		Time 2					*	Total	%1	%2
Response		1	2	3	4	5				
Time 1	1				1	1	2	3.6	1.0	
	2		1		3	1	5	8.9	2.5	
	3				1		1	1.8	0.5	
	+4	1	1	41	5	48		85.7	24.0	
	*		2	2	15	125	144		72.0	
Total							200	100.0	100.0	
Total		0	4	3	61	132	200			
%1		0.0	5.9	4.4	89.7			100.0		
%2		0.0	2.0	1.5	30.5	66.0		100.0		

There were 56 before respondents, with 48 (85.7%) voting response 4 (95% CI: 73.8-93.6) and 68 after respondents, with 61 (89.7%) voting response 4 (95% CI: 79.9-95.8). However, from the 75 who responded at least once, the before rate was 64.0% and after, 81.3%, for an increase of 17.3% (95% CI: 2.7-28.0). Collapsing response 1 with *, the McNemar $X^2(5) = 9.095$, $P2 = 0.1053$; and the collapsed $X^2(1) = 6.259$, $P2 = 0.0124$. The presentation increased the most appropriate responses.

Question: "Sensitivity to clinically relevant change is least important for which one of the following instruments?"

Responses: 1. Remission. 2. Prognosis. 3. Classification. 4. Disease activity.

		Time 2					*	Total	%1	%2
Response		1	2	3	4	5				
Time 1	1	1		6			7	10.3	3.5	
	2		2	6			8	11.8	4.0	
	+3		3	41		3	47	69.1	23.5	
	4		1	4	1		6	8.8	3.0	
	*	1	1	11	1	118	132		66.0	
Total							200	100.0	100.0	
Total		2	7	68	2	121	200			
%1		2.5	8.9	86.1	2.5			100.0		
%2		1.0	3.5	34.0	1.0	60.5		100.0		

There were 68 before respondents, with 47 (69.1%) voting response 3 (95% CI: 56.7-79.8) and 79 after respondents, with 68 (86.1%) voting response 3 (95% CI: 76.4-92.8). However, from the 82 who responded at least once, the before rate was 57.3% and after, 82.9%, for an increase of 25.6% (95% CI: 11.7-34.6). The McNemar $X^2(8) = 19.571$, $P_2 = 0.0121$; and the collapsed $X^2(1) = 13.364$, $P_2 = 0.0003$. Clearly, the presentation increased the most appropriate responses.

G. Guyatt presented some issues relevant in the development of disease specific quality of life measures of outcome.

Question: "Generic health status measures share the following properties:"

Responses: 1. They can be aggregated into a single number. 2. They are ideal for economic analyses. 3. They are applicable to a wide variety of populations suffering different quality of life impairments. 4. All the above.

Response	Time 2					Total	%1	%2
	1	2	3	4	*			
1	1	1	1	1		3	3.9	1.5
2				1		1	1.3	0.5
3			13	14	2	29	37.7	14.5
+4			10	33	1	44	57.1	22.0
*			2	8	113	123		61.5
Total						200	100.0	100.0
Total	0	1	26	57	116	200		
%1	0.0	1.2	31.0	67.9			100.1	
%2	0.0	0.5	13.0	28.5	58.0			100.0

There were 77 before respondents, with 44 (57.1%) voting response 4 (95% CI: 45.4-68.4) and 84 after respondents, with 57 (67.9%) voting response 4 (95% CI: 56.8-68.4). However, from the 87 who responded at least once, the before rate was 50.6% and after, 65.5%, for an increase of 14.9% (95% CI: 0.6-26.7). Collapsing 1 with *, the McNemar $X^2(5) = 9.267$, $P_2 = 0.0989$; and the collapsed $X^2(1) = 4.829$, $P_2 = 0.0280$. The presentation increased the most appropriate responses.

Question: "Which of the following is the most true? Specific health status measures share the following limitations."

Responses: 1. They are less likely than generic measures to be responsive to small but important changes. 2. Quality of life impairment cannot be compared from one measure to another. 3. They tend to be too long and complicated for use in clinical trials. 4. They do not have an intuitive appeal for clinicians.

Response	Time 2					Total	%1	%2
	1	2	3	4	*			
1		3	3	3		9	12.5	4.5
+2		29	1	1	4	35	48.6	17.5
3	1	6	2	1	1	11	15.3	5.5
4		9	2	3	3	17	23.6	8.5
*	2	9	1		116	128		64.0
Total						200	100.0	100.0
Total	3	56	9	8	124	200		
%1	3.9	73.7	11.8	10.5			99.9	
%2	1.5	28.0	4.5	4.0	62.0			100.0

There were 72 before respondents, with 35 (48.6%) voting response 2 (95% CI: 36.7-60.7) and 76 after respondents, with 56 (73.7%) voting response 2 (95% CI: 62.3-83.1). However, from the 84 who responded at least once, the before rate was 41.7% and after, 66.7%, for an increase of 25.0% (95% CI: 11.4-33.8). The McNemar $X^2(10) = 24.228$, $P_2 = 0.0070$; and the collapsed $X^2(1) = 13.364$, $P_2 = 0.0003$. Clearly the presentation increased the most appropriate responses.

M. Drummond raised some issues relevant to using economic analyses with the results of clinical trials.

Question: "Economic analysis undertaken alongside clinical trials requires that we collect data on the following items:"

Responses: 1. Costs only. 2. Quality of life only. 3. Both costs and quality of life. 4. Always costs and sometimes quality of life.

Response	Time 2					Total	%1	%2
	1	2	3	4	*			
1			1			1	1.2	0.5
2					1	1	1.2	0.5
+3			33	8	7	48	57.8	24.0
4			14	19		33	39.8	16.5
*				1	116	117		58.5
Total						200	100.0	100.0
Total	0	0	49	27	124	200		
%1	0.0	0.0	64.5	35.5			100.0	
%2	0.0	0.0	24.5	13.5	62.0			100.0

There were 83 before respondents, with 48 (57.8%) voting response 3 (95% CI: 46.5-68.6) and 76 after respondents, with 49 (64.5%) voting response 3 (95% CI: 52.7-75.1). However, from the 84 who responded at least once, the before rate was 57.1% and after, 58.3%, for an increase of 1.2% (95% CI: -12.5-14.6). Collapsing responses 1 and 2 with *, the McNemar $X^2(2) = 4.142$, $P_2 = 0.1100$; and the collapsed $X^2(1) = 0.003$, $P_2 = 0.8575$. The presentation did not change the most appropriate responses.

Question: "The main concern we should have about undertaking economic analysis alongside clinical trials is that:"

Responses: 1. It takes considerable time and effort. 2. It requires an increase in sample size. 3. Many trials under typical conditions do not reflect normal practice.

Response	Time 2				Total	%1	%2
	1	2	3	*			
1	2		9		11	13.8	5.5
2	1	1	2		4	5.0	2.0
+3	6		53	6	65	81.2	32.5
*			2	118	120		60.0
Total					200	100.0	100.0
Total	9	1	66	124	200		
%1	11.8	1.3	86.8			99.9	
%2	4.5	0.5	33.0	62.0			100.0

There were 80 before respondents, with 65 (81.2%) voting response 3 (95% CI: 71.0-89.1) and 76 after respondents, with 66 (86.8%) voting response 3 (95% CI:

77.1-93.5). However, from the 82 who responded at least once, the before rate was 79.3% and after, 80.5%, for an increase of 1.2% (95% CI: -11.4-13.5). The McNemar $X^2(4) = 5.600$, $P2 = 0.2311$; and the collapsed $X^2(1) = 0.040$, $P2 = 0.8415$. The presentation did not change the most appropriate responses.

B. O'Brien first presented 2 questions on measuring cost effectiveness, and then showed a movie, where participants were asked for their opinions about a standard gamble and time tradeoff to estimate patient utilities. The latter 2 questions were not asked in the before/after mode.

Question: "If Drug A for RA costs \$5 per day and Drug B costs \$10 per day, which drug offers the best value for money?"
 Responses: 1. Drug A. 2. Drug B. 3. Can't tell.

Response	Time 2				Total	%1	%2	
	1	2	3	*				
Time 1	1	1	1	1	3	4.2	1.5	
	2				1	1.4	0.5	
	+3	1	1	62	4	68	94.4	34.0
	*			11	117	128		64.0
Total					200	100.0	100.0	
	Total	2	2	74	122	200		
	%1	2.6	2.6	94.9		100.1		
	%2	1.0	1.0	37.0	61.0	100.0		

There were 72 before respondents, with 68 (94.4%) voting response 3 (95% CI: 86.4-98.5) and 78 after respondents, with 74 (94.9%) voting response 3 (95% CI: 87.4-98.6). However, from the 83 who responded at least once, the before rate was 81.9% and after, 89.2%, for an increase of 7.2% (95% CI: -3.9-15.9). The McNemar $X^2(5) = 6.267$, $P2 = 0.2811$; and the collapsed $X^2(1) = 2.000$, $P2 = 0.1573$. The presentation did not change the most appropriate responses.

Question: "For a new therapy to be labelled cost effective, relative to some alternative, evidence must show that:"
 Responses: 1. It is more effective. 2. It is less costly. 3. It has the same or greater effect at lower cost. 4. It has a greater effect that is worth the additional cost.

Response	Time 2				Total	%1	%2	
	1	2	3	4				*
Time 1	1			2	1	3	4.2	1.5
	2			1		1	1.4	0.5
	3		16	19	3	38	53.5	19.0
	+4		5	24		29	40.8	14.5
	*		3	9	117	129		64.5
Total						200	99.9	100.0
	Total	0	0	24	55	121	200	
	%1	0.0	0.0	30.4	69.6		100.0	
	%2	0.0	0.0	12.0	27.5	60.5	100.0	

There were 71 before respondents, with 29 (40.8%) voting response 4 (95% CI: 29.3-53.2) and 79 after respondents, with 55 (69.6%) voting response 4 (95% CI: 58.2-79.5). However, from the 83 who responded at least

once, the before rate was 34.9% and after, 66.3%, for an increase of 31.3% (95% CI: 17.8-39.3). Collapsing responses 1 and 2 with *, the McNemar $X^2(3) = 20.167$, $P2 = 0.0002$; and the collapsed $X^2(1) = 18.778$, $P2 < 0.0001$. Clearly the presentation increased the most appropriate responses.

The movie questions were administered once, and thus provided individual estimates of the steps needed to conduct these interview exercises.

Question: "Standard Gamble 1. Which would you choose?"
 Responses: 1. Procedure with 80% chance of perfect health and 20% chance of immediate death. 2. Live the rest of your life in health status x. 3. 1 and 2 are equal.

Response	Total	%1	%2	
				1
Time 1	2	34	45.9	17.0
	3	4	5.4	2.0
	*	126		63.0
Total	200	99.9	100.0	

The 74 respondent estimates can be used in the utility estimation process.

Question: "Time tradeoff 1. Which would you choose?"
 Responses: 1. Procedure with perfect health for 25 years followed by immediate death. 2. Live for 30 years in health state x, followed by immediate death. 3. 1 and 2 are equal.

Response	Total	%1	%2	
				+1
Time 1	2	1	1.3	0.5
	3	1	1.3	0.5
	*	124		62.0
Total	200	100.0	100.0	

Since the 76 were expected to respond with response 1, the prevalence of this most appropriate response, some 74 (97.4%) did, with a 95% CI: 90.8-99.7%.

M. Drummond posed questions about ulcer hospitalization and NSAID use in 3 countries. The questions related to cost effectiveness and the probability of ulcer hospitalization.

Question: "Are relative prices of NSAID a good guide to their relative cost effectiveness?"
 Responses: 1. Yes. 2. No.

Response	Time 2			Total	%1	%2	
	1	2	*				
Time 1	1	6	10	16	21.6	8.0	
	+2	5	51	2	58	78.4	29.0
	*		6	120		63.0	
Total				200	100.0	100.0	
	Total	11	67	122	200		
	%1	14.1	85.9		100.0		
	%2	5.5	33.5	61.0	100.0		

There were 74 before respondents, with 58 (78.4%) voting response 2 (95% CI: 67.3-87.1) and 78 after respondents, with 67 (85.9%) voting response 2 (95% CI:

76.2-92.7). However, from the 80 who responded at least once, the before rate was 72.5% and after, 83.8%, for an increase of 11.2% (95% CI: -1.7-21.2). The McNemar $X^2(2) = 3.667$, $P2 = 0.1599$; and the collapsed $X^2(1) = 3.522$, $P2 = 0.0606$. The presentation did not change the most appropriate responses.

Question: "In which country would you have the highest probability of being admitted to a hospital if you had an ulcer?"
Responses: 1. United Kingdom. 2. USA. 3. Brazil.

	Response	Total	%1	%2
Time 1	1	24	32.9	12.0
	+2	37	50.7	18.5
	3	12	16.4	6.0
	*	127		63.5
Total		200	100.0	100.0

Since this question was asked before the presentation but not after, the prevalence of the most appropriate response from the 73 respondents was 37 (50.7%); 95% CI: 38.7-62.6%.

Question: "In which country would the costs of treating GI side effects have the biggest impact on the relative cost effectiveness of the 3 NSAID?"
Responses: 1. United Kingdom. 2. USA. 3. Brazil.

	Response	Total	%1	%2
Time 1	1	7	10.3	3.5
	+2	50	73.5	25.0
	3	11	16.2	5.5
	*	132		66.0
Total		200	100.0	100.0

Since this question was asked before the presentation but not after, the prevalence of the most appropriate response from the 68 respondents was 50 (73.5%); 95% CI: 61.4-83.5%.

S. Gabriel linked economic analysis with quality of life, and her questions related to these 2 issues.

Question: "Which of the following techniques is used to test the stability of the conclusions of an economic analysis?"
Responses: 1. Cost effectiveness. 2. Cost utility analysis. 3. Time tradeoff method. 4. Sensitivity analysis.

Response	Time 2					Total	%1	%2
	1	2	3	4	*			
Time 1	1	1	3	3	7	9.9	3.5	
	2			6	1	7	3.5	
	3	1		3	4	5.6	2.0	
	+4			46	7	53	74.6	26.5
	*	1		6	122	129	64.5	
Total					200	100.0	100.0	
	Total	1	2	3	64	130	200	
	%1	1.4	2.9	4.3	91.4	100.0		
	%2	0.5	1.0	1.5	32.0	65.0	100.0	

There were 71 before respondents, with 53 (74.6%) voting response 4 (95% CI: 62.9-84.2) and 70 after respondents, with 64 (91.4%) voting response 4 (95% CI:

82.3-96.8). However, from the 78 who responded at least once, the before rate was 67.9% and after, 82.1%, for an increase of 14.4% (95% CI: 0.4-24.3). The McNemar $X^2(9) = 19.077$, $P2 = 0.0245$; and the collapsed $X^2(1) = 4.840$, $P2 = 0.0278$. The presentation increased the most appropriate responses.

Question: "Which of the following techniques involves an analysis of costs relative to adjusted quality of life measurements?"
Responses: 1. Cost effectiveness analysis. 2. Cost utility analysis. 3. Time tradeoff method. 4. Sensitivity analysis.

Response	Time 2					Total	%1	%2
	1	2	3	4	*			
Time 1	1	1	8		1	10	13.5	5.0
	+2		49	2	5	56	75.7	28.0
	3		2		1	3	4.1	1.5
	4		4		1	5	6.8	2.5
	*	1	5		120	126		63.0
Total						200	100.1	100.0
	Total	2	68	2	1	127	200	
	%1	2.7	93.2	2.7	1.4	100.0		
	%2	1.0	34.0	1.0	0.5	63.5	100.0	

There were 74 before respondents, with 56 (75.7%) voting response 2 (95% CI: 64.3-84.9) and 73 after respondents, with 68 (93.2%) voting response 2 (95% CI: 84.7-97.7). However, from the 80 who responded at least once, the before rate was 70.0% and after, 85.0%, for an increase of 15.0% (95% CI: 1.4-25.0). The McNemar $X^2(6) = 13.000$, $P2 = 0.0430$; and the collapsed $X^2(1) = 5.538$, $P2 = 0.0186$. The presentation increased the most appropriate responses.

C. Bakker spoke on utility estimation and how utilities could be used to help make decisions about arthritis care delivery.

Question: "Which is true: 1. A utility expresses the value of a specific level of health between 0 and 1. 2. Utilities are generic, in that they are applied to all diseases and indications."
Responses: 1. 1 is true. 2. 2 is true. 3. Both are true. 4. Neither is true.

Response	Time 2					Total	%1	%2	
	1	2	3	4	*				
Time 1	1	2	2	4	4	12	17.1	6.0	
	2		3	2	1	6	8.6	3.0	
	+3	1	2	37	1	5	46	65.7	23.0
	4			6		6	8.6	3.0	
	*	1	2	6	121	130		65.0	
Total						200	100.0	100.0	
	Total	4	9	55	1	131	200		
	%1	5.8	13.0	79.7	1.4	99.9			
	%2	2.0	4.5	27.5	0.5	65.5	100.0		

There were 70 before respondents, with 46 (65.7%) voting response 3 (95% CI: 53.4-76.7) and 69 after respondents, with 55 (79.7%) voting response 3 (95% CI: 68.3-88.4). However, from the 79 who responded at least once, the before rate was 58.2% and after, 69.6%, for an increase of 11.4% (95% CI: -2.7-22.9). The McNemar $X^2(7) = 9.596$, $P2 = 0.2127$; and the collapsed $X^2(1) =$

3.000, $P_2 = 0.0833$. The presentation did not change the most appropriate responses.

Question: "Utilities are not measures of:"

Responses: 1. Satisfaction. 2. Patient preference. 3. Patient survivals. 4. Psychological factors.

	Response	Time 2					Total	%1	%2
		1	2	3	4	*			
Time 1	1			4		2	6	7.6	3.0
	2		1	1		1	3	3.8	1.5
	+3			47	2	9	58	73.4	29.0
	4		1	3	7	1	12	15.2	6.0
Total	*			4	1	116	121		60.5
	Total	0	2	59	10	129	200	100.0	100.0
	%1	0.0	2.8	83.1	14.1		100.0		
	%2	0.0	1.0	29.5	5.0	64.5	100.0		

There were 79 before respondents, with 58 (73.4%) voting response 3 (95% CI: 62.3-82.7) and 71 after respondents, with 59 (83.1%) voting response 3 (95% CI: 72.3-90.9). However, from the 84 who responded at least once, the before rate was 69.0% and after, 70.2%, for an increase of 1.2% (95% CI: -10.6-12.7). Collapsing response 1 with *, the McNemar $X^2(6) = 3.259$, $P_2 = 0.7757$; and the collapsed $X^2(1) = 0.043$, $P_2 = 0.8348$. The presentation did not change the most appropriate responses.

M. Liang discussed issues in the development of an instrument that is sensitive to change when real change in the patient has taken place.

Question: "What is the best statistical approach for evaluating the sensitivity of a questionnaire?"

Responses: 1. Standardized response mean. 2. Effect size. 3. Guyatt sensitivity statistic. 4. Receiver operating curves. 5. Option characteristics curves. 6. Potentially, any of the above.

	Response	Time						Total	%1	%2
		1	2	3	4	5	6			
Time 1	1	1	1					4	6.2	2.0
	2	1	3					18	27.7	9.0
	3							2	3.1	1.0
	4							4	9.2	3.0
	5							1	1.5	0.5
	+6	2	1					31	52.3	17.0
Total	*		3					11	121	67.5
	Total	4	8	0	0	0	63	125	200	100.0
	%1	5.3	10.7	0.0	0.0	0.0	84.0	100.0		
	%2	2.0	4.0	0.0	0.0	0.0	31.5	62.5	100.0	

There were 65 before respondents, with 34 (52.3%) voting response 6 (95% CI: 39.5-64.8) and 75 after respondents, with 63 (84.0%) voting response 6 (95% CI: 73.7-91.4). However, from the 79 who responded at least once, the before rate was 43.0% and after, 79.7%, for an

increase of 36.7 (95% CI: 23.9-42.7). Collapsing responses 3, 4, and 5 with *, the McNemar $X^2(5) = 27.508$, $P_2 < 0.0001$; and the collapsed $X^2(1) = 24.029$, $P_2 < 0.0001$. Clearly the presentation increased the most appropriate responses.

Question: "Which tactics can be used to improve sensitivity of a measure?"

Responses: 1. Increase number of response categories. 2. Eliminate unreliable questions. 3. Increase number of questions within a domain. 4. Use a visual analog response. 5. Increase size of questionnaire.

	Response	Time 2					Total	%1	%2
		1	2	3	4	5			
Time 1	1	7	3				11	14.9	5.5
	+2	9	26	4		1	44	59.5	22.0
	3	2	3				6	8.1	3.0
	4	1	3	1	1	1	8	10.8	4.0
	5	1	1			2	5	6.8	2.5
Total	*	1	2		1		122	126	63.0
	Total	21	38	5	2	4	130	200	100.1
	%1	30.0	54.3	7.1	2.9	5.7	100.0		
	%2	10.5	19.0	2.5	1.0	2.0	65.0	100.0	

There were 74 before respondents, with 44 (59.5%) voting response 2 (95% CI: 47.4–70.7) and 70 after respondents, with 38 (54.3%) voting response 2 (95% CI: 41.9–66.3). However, from the 78 who responded at least once, the before rate was 56.4% and after, 48.7%, for a decrease of 7.7% (95% CI: –21.0–7.2). The McNemar $X^2(14) = 14.810$, $P2 = 0.3913$; and the collapsed $X^2(1) = 1.200$, $P2 = 0.2733$. The presentation did not change the most appropriate responses.

Finally, G. Hawker presented the issues used to compare a disease specific quality of life measure with a generic quality of life measure.

Question: "How would a patient with multiple conditions in addition to a specific disease X be expected to score using a disease X specific instrument versus a generic health status instrument?"

Responses: 1. They would score "healthier" on the generic health status instrument. 2. They would score "healthier" on the disease X specific instrument. 3. They would score about the same on both.

Response	Time 2				Total	%1	%2	
	1	2	3	*				
Time 1	1	12	11	1	2	26	37.7	13.0
	+2	3	24			27	39.1	13.5
	3		8	6	2	16	23.2	8.0
	*	2	11	2	116	131		65.5
Total						200	100.0	100.0
	Total	17	54	9	120	200		
	%1	21.3	67.5	11.3			100.1	
	%2	8.5	27.0	4.5	60.0			100.0

There were 69 before respondents, with 27 (39.1%) voting response 2 (95% CI: 27.6–51.6) and 80 after respondents, with 54 (67.5%) voting response 2 (95% CI: 56.1–77.6). However, from the 84 who responded at least once, the before rate was 32.1% and after, 64.3%, for an increase of 32.1% (95% CI: 20.2–37.8). The McNemar $X^2(6) = 24.571$, $P2 = 0.0004$; and the collapsed $X^2(1) = 22.091$, $P2 < 0.0001$. Clearly the presentation increased the most appropriate responses.

Question: "Which health status instrument would be expected to discriminate best, if any, among individuals with varying degrees of severity for a specific disease?"

Responses: 1. The generic health status instrument would discriminate better. 2. The disease specific instrument would discriminate better. 3. They would discriminate equally well.

Response	Time 2				Total	%1	%2	
	1	2	3	*				
Time 1	1		3			3	3.4	1.5
	+2	4	70	1	6	81	92.0	40.5
	3		2	1	1	4	4.5	2.0
	*		2		110	112		56.0
Total						200	99.9	100.0
	Total	4	77	2	117	200		
	%1	4.8	92.8	2.4			100.0	
	%2	2.0	38.5	1.0	58.5			100.0

There were 88 before respondents, with 81 (92.0%) voting response 2 (95% CI: 84.3–96.7) and 83 after respondents, with 77 (92.8%) voting response 2 (95% CI: 84.9–97.3). However, from the 90 who responded at least once, the before rate was 90.0% and after, 85.6%, for a decrease of 4.4% (95% CI: –13.1–5.7). The McNemar $X^2(4) = 3.476$, $P2 = 0.4815$; and the collapsed $X^2(1) = 0.889$, $P2 = 0.3458$. The presentation did not change the most appropriate responses.

Closing session. During the closing session, Peter Tugwell raised 6 questions related to future research questions and possible expansion of research studies to incorporate economic instruments to gather research data. Each question had as responses: 1. Yes. 2. No.

Question: "Would you be prepared to include a toxicity index or supplemental questionnaire in addition to the usual case report form in your next rheumatology drug trial?"

	Response	Total	%1	%2
Time 1	1	48	80.0	24.0
	2	12	20.0	6.0
	*	140		70.0
Total		200	100.0	100.0

From the 60 respondents, 48 (80.0%) voted Yes; 95% CI: 67.7–89.2. Clearly the respondents were in favor of including some form of toxicity index in their next trial.

Question: "In your next rheumatology drug trial, in addition to the usual endpoints, would you be prepared to include one of the generic/utility questionnaires?"

	Response	Total	%1	%2
Time 1	1	56	87.5	28.0
	2	8	12.5	4.0
	*	136		68.0
Total		200	100.0	100.0

From the 64 respondents, 56 (87.5%) voted Yes; 95% CI: 76.8–94.4. Again, the respondents were clearly in favor of including a generic utility questionnaire in their next trial.

Question: "In your next rheumatology drug trial, in addition to the usual endpoints, would you be prepared to include a rating scale/feeling thermometer?"

	Response	Total	%1	%2
Time 1	1	46	71.9	23.0
	2	18	28.1	9.0
	*	136		68.0
Total		200	100.0	100.0

From the 64 respondents, 46 (71.9%) voted Yes; 95% CI: 59.2–82.4. Clearly, these respondents were prepared to use a feeling thermometer in their next trial.

Question: "In your next rheumatology drug trial, in addition to the usual endpoints, would you be prepared to include a standard gamble?"

	Response	Total	%1	%2
Time 1	1	11	17.2	5.5
	2	53	82.8	26.5
	*	136		68.0
Total		200	100.0	100.0

From the 64 respondents, 11 (17.2%) voted Yes; 95% CI: 8.9-28.7. Hence respondents were not in favor of using a standard gamble in their trials.

Question: "In your next rheumatology drug trial, in addition to the usual endpoints, would you be prepared to include a time tradeoff?"

	Response	Total	%1	%2
Time 1	1	18	28.1	9.0
	2	46	71.9	23.0
	*	136		68.0
Total		200	100.0	100.0

Again from the 64 respondents, 18 (28.1%) voted Yes; 95% CI: 17.6-40.8. Hence, respondents were not in favor of including a time tradeoff in their trials. From the discussion it was apparent these 2 instruments were perceived as too complicated to use with patients with RA.

Question: "In your next rheumatology drug trial, in addition to the usual endpoints, would you be prepared to include a cost component so that a concurrent cost effectiveness substudy can be included?"

	Response	Total	%1	%2
Time 1	1	56	87.5	28.0
	2	8	12.5	4.0
	*	136		68.0
Total		200	100.0	100.0

From the 64 respondents, 56 (87.5%) voted Yes; 95% CI: 76.8-94.4. Clearly the respondents were in favor of doing cost effectiveness substudies.

M. Boers raised 2 questions about future OMERACT conferences.

Question: "Future OMERACT conferences should be organized:"
Responses: 1. Never. 2. Every year. 3. Every 2 years. 4. Every 3 years. 5. Don't know.

	Response	Total	%1	%2
Time 1	1	3	4.2	1.5
	2	14	19.7	7.0
	+3	43	60.6	21.5
	4	6	8.5	3.0
	5	5	7.0	2.5
	*	129		64.5
Total		200	100.0	100.0

From the 71 respondents, 43 (60.6%); 95% CI: 48.2-72.0 thought OMERACT conferences should be held every 2 years. By dropping those who voted for response 5, Don't

know, of the 66 remaining respondents, 43 (65.2%); 95% CI: 52.4-76.5 voted every 2 years. The 2 year interval is the most popular choice.

Question: "The most important direction for the next OMERACT conference is:"

Responses: 1. Revalidate RA core set and improvement criteria based on prospective data. 2. Design core sets for other musculoskeletal diseases. 3. Evaluative research on efficacy/cost tradeoff on basis of research agenda proposed here. 4. More than one of items 1 to 3. 5. Other/Don't know.

	Response	Total	%1	%2
Time 1	1	9	13.0	4.5
	2	10	14.5	5.0
	3	19	27.5	9.5
	4	26	37.7	13.0
	5	5	7.2	2.5
Total	*	131		65.5
		200	99.9	100.0

From the 69 respondents, the clear winner is response category 4, for more than one of the first 3 categories; 26 (37.7%); 95% CI: 26.3-50.2 voted for this choice. However, when the 4th and 5th categories were eliminated in the next vote, the results were as shown in the next table.

Question: "The most important direction for the next OMERACT conference is:"

Responses: 1. Revalidate RA core set and improvement criteria based on prospective data. 2. Design core sets for other musculoskeletal diseases. 3. Evaluative research on efficacy/cost tradeoff on basis of research agenda proposed here.

	Response	Total	%1	%2
Time 1	1	13	18.6	6.5
	2	22	31.4	11.0
	3	35	50.0	17.5
	*	130		65.0
Total		200	100.0	100.0

Now the categories are rated 3 first, 2 second, and 1 third.

DISCUSSION

While some of the increases in the most appropriate responses may have been attributable to the presentation, setting α at 5% meant that the 95% confidence interval or the test statistic may not have eliminated the operation of chance. We have chosen to interpret these findings as if they were chance. No predetermined sample size was set to distinguish chance from the effect of the presentations.

The use of wireless keypads during a conference permits the acquisition of data directly from participants, permits the estimation of change, and allows for the opinions of the audience to be measured for future discussion. Other conferences should consider this type of equipment and the use of such a voting process to quantify the opinions of those who choose to vote, rather than usually those of the presenters alone.

ACKNOWLEDGMENT

Bonnie Geneslaw from MPE Communications, The Healthcare Resource, and Jerry Cropp from Photo Communications Corporation were instrumental in providing the keypads, running the software, and tabulating results during the conference. Jerry Cropp also provided a disk of the data from which these tables were constructed.

REFERENCES

1. Gardner SB, Winter PD, Gardner MJ: Confidence interval analysis. Version 1.1. British Medical Journal, London, 1991.
2. Chi-Square Analysis. Department of Clinical Epidemiology and Biostatistics, McMaster University, 1970.

Scanned by Uni. of Qld Library;
Herston Medical Library for supply
Under S.50 of Copyright Act 1968

9 - NOV 2012