How do different feedback strategies affect Delphi survey results? Results of a nested RCT

Steven MacLennan, Jamie Kirkham, Thomas Lam, Paula Williamson

Background

- Many core outcome set (COS) projects incorporate Delphi surveys with at least 2 stakeholder groups
- It is unclear which feedback strategy is best for between-round feedback in Delphi studies.
- We tested various feedback strategies whilst developing a COS for localised prostate cancer with patients and health professionals

Methods – Delphi overview

Generation of outcomes list - Systematic review:
1415 verbatim outcomes

17366 abstracts and 2080 full texts screened
325 studies included

Categorising outcomes:
76 outcomes

Generation of outcomes list: Patient interviews 3 *additional* outcomes

15 semi-structured interviews

List of 79 outcomes

Delphi Round 1: **79** outcomes scored
Scale: 1 (not important) – 9 (critically important)

(110 HCPs and 153 patients invited)
56 HPs and 118 patients completed

Delphi Round 2: **79** outcomes
Scale: 1 (not important) – 9 (critically important)

49 HPs and 109 patients completed

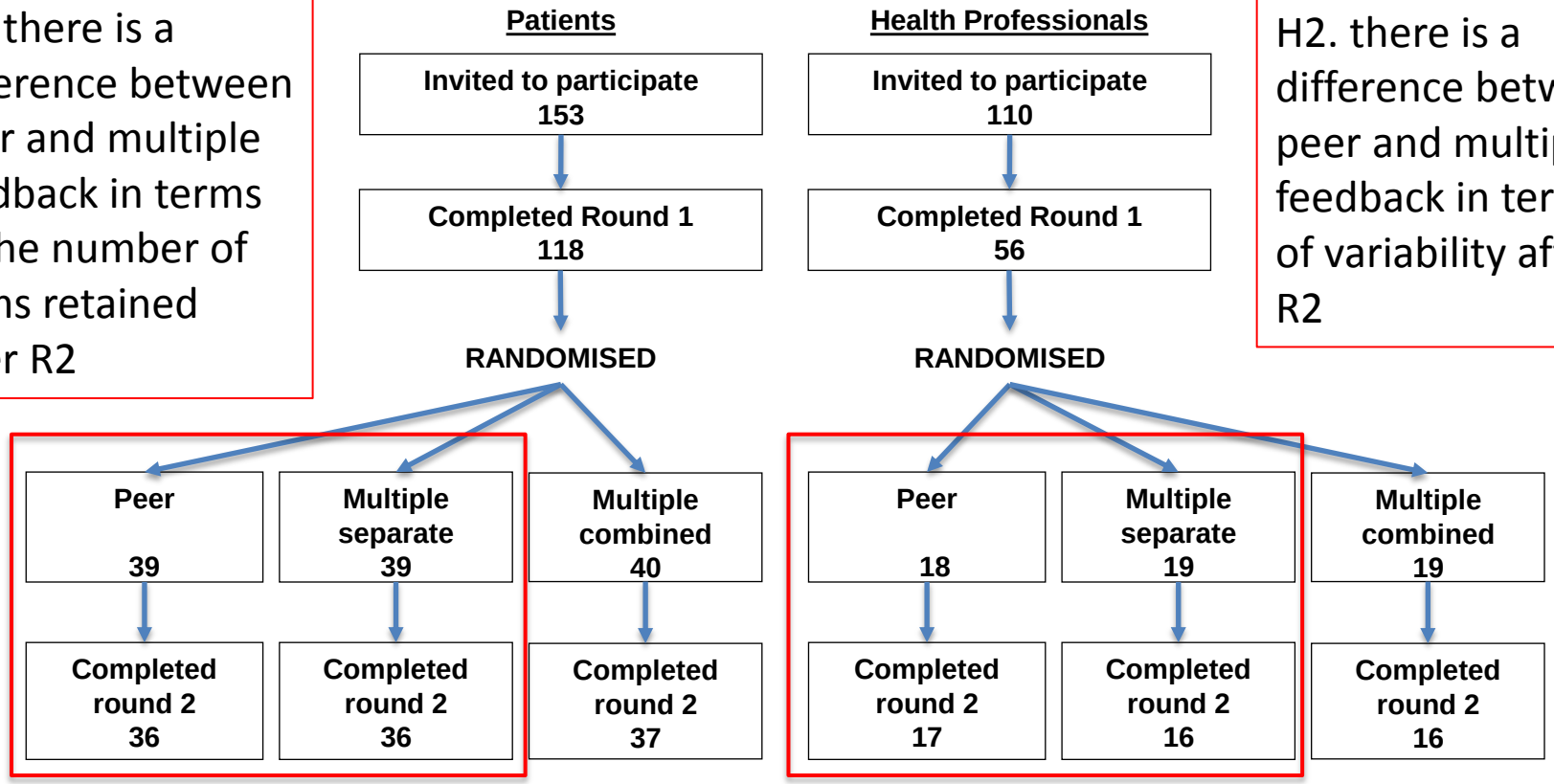
Delphi Round 3: **79** outcomes scored
Scale: 1 (not important) – 9 (critically important)

47 HPs and 105 patients completed

Methods – Delphi flow diagram

H1. there is a difference between peer and multiple feedback in terms of the number of items retained after R2

H2. there is a difference between peer and multiple feedback in terms of variability after R2



A. Cancer-specific outcomes and survival outcomes

Outcome	Total No of People Scoring 1 to 9	Importance									Una to Sco
		Not important			Important but not critical			Critical			
		1	2	3	4	5	6	7	8	9	
Applicable to all treatments											
1. Death from any cause *This outcome refers to the death of someone from any cause, including prostate cancer.*		○	○	○	○	○	○	○	○	○	○
Round 2 results for health professionals	15	0%	0%	0%	0%	0%	0%	7%	33%	60%	
Round 2 results for patients	35	6%	0%	3%	3%	6%	6%	6%	11%	60%	
2. Death from prostate cancer *This outcomes refers to the death of someone as a result of prostate cancer.*		○	○	○	○	○	○	○	○	○	
Round 2 results for health professionals	15	0%	0%	0%	0%	0%	7%	0%	7%	87%	
Round 2 results for patients	35	6%	0%	0%	0%	6%	6%	3%	9%	71%	
3. Death from causes other than prostate cancer *This outcome refers to the death of someone from any causes other than prostate cancer.*		○	○	○	○	○	○	○	○	○	
Round 2 results for health professionals	15	0%	0%	0%	0%	0%	7%	33%	40%	20%	
Round 2 results for patients	34	6%	0%	0%	3%	3%	12%	9%	3%	65%	

Yellow marker indicates score from previous round

Percentage of HPs and patients scoring the outcome at each score 1-9 shown in red

Methods – Analysis plan

H1. There is a difference between peer and multiple feedback in terms of the number of items retained after R2

Agreement analysis

- Items scored 7-9 ≥70% *and* 1-3 by <15% after round 2 = 'retained'
- Calculated number of outcomes retained for a) both stakeholder groups or b) neither group (i.e. agreement) c) each stakeholder group independently (i.e. discordant): stratified by *peer* and *multiple separate* feedback groups
- Calculated number (%) of items where there was 'agreement' or 'discordance'

H2. There is a difference in between peer and multiple feedback in terms of the reduction in variability after R2

Variability analysis

- Calculated mean (SD) score across all 79 items for the *peer* and *multiple separate* groups in round 1 (ignoring stakeholder group)
- Calculate mean of SDs
- Repeat for round 2
- Calculate difference in mean SDs (SD) between R1 and R2 for 2 randomised groups
- Paired t-test comparing difference in mean differences

Results – Agreement

Core set	Feedback	Number of participants		Items retained at end of round 2, no. (%)					
		Patients	HCP	Retained by both stakeholder groups	Retained by patients only	Retained by health professionals only	Retained by neither stakeholder group	% discordant	% agreement
Prostate	Peer	36	17	16/79	4/79	2/79	57/79	8%	92%
	Multiple separate	36	16	14/79	6/79	2/79	57/79	10%	90%
Breast*	Peer	95	36	16/34	4/34	2/34	12/34	18 %	82 %
	Multiple separate	95	33	17/34	2/34	0/34	15/34	6 %	94 %
Colorectal*	Peer	42	41	10/46	5/46	11/46	20/46	35 %	65 %
	Multiple separate	44	37	10/46	3/46	9/46	24/46	26 %	74 %
Oesophageal*	Peer	76	52	11/29	6/29	4/29	8/29	35 %	65 %
	Multiple separate	69	55	14/51	4/51	7/51	26/51	22 %	78 %

*Brookes, S. et al Trials (2016) **17**(1): 1.

- Not much difference, but *peer* feedback group retained more items
- 3 Bristol Delphi's: *multiple* feedback group had consistently better agreement

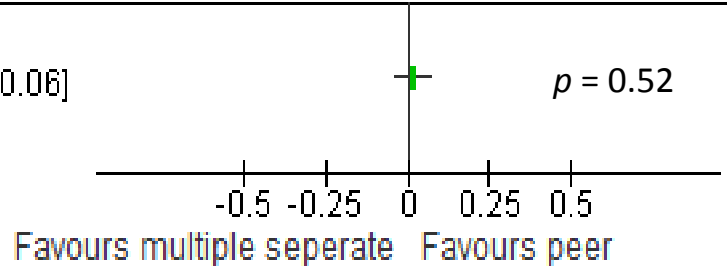
Results – Agreement

Core set	Feedback	Number of participants		Items retained at end of round 2, no. (%)					
		Patients	HCP	Retained by both stakeholder groups	Retained by patients only	Retained by health professionals only	Retained by neither stakeholder group	% discordant	% agreement
Prostate	Peer	36	17	16/79	4/79	2/79	57/79	8%	92%
	Multiple separate	36	16	14/79	6/79	2/79	57/79	10%	90%
Prostate R1	Peer	36	17	9/79	3/79	4/79	63/79	9%	91%
	Multiple separate	36	16	10/79	7/79	1/79	61/79	10%	90%

- Agreement was already good in round 1

Results – Variability

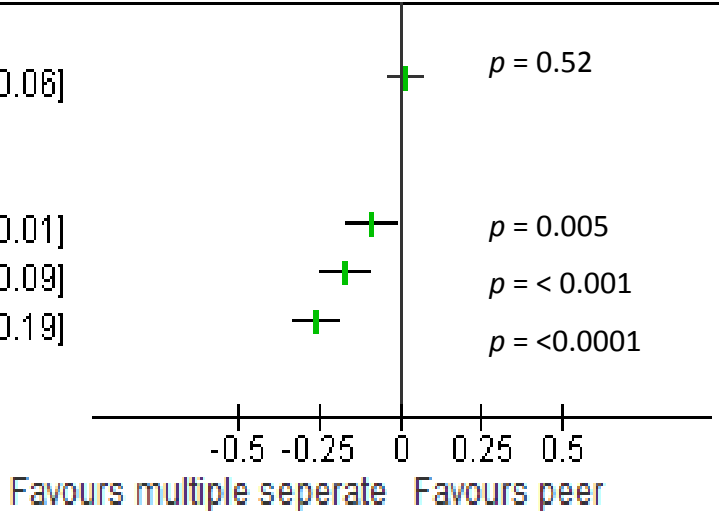
Study or Subgroup	Peer Mean	SD	Total	Multiple seperate Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
1.1.1 Prostate Delphi							
Prostate	0.217	0.19	79	0.201	0.115	79	0.02 [-0.03, 0.06]



- Prostate Delphi: no evidence of a difference in variability between *peer* and *multiple separate* feedback

Results – Variability

Study or Subgroup	Peer Mean	SD	Total	Multiple seperate Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
1.1.2 Prostate Delphi							
Prostate	0.217	0.19	79	0.201	0.115	79	0.02 [-0.03, 0.06]
1.1.3 Previous Delphis *							
Breast	-0.04	0.15	34	0.05	0.17	34	-0.09 [-0.17, -0.01]
Colorectal	0.34	0.21	46	0.51	0.16	46	-0.17 [-0.25, -0.09]
Oesophageal	0.19	0.18	29	0.45	0.11	51	-0.26 [-0.33, -0.19]



- Prostate Delphi: no evidence of a difference in variability between *peer* and *multiple separate* feedback
- 3 previous Delphis: *multiple separate* feedback consistently and significantly reduced variability

*Brookes, S. et al Trials (2016) **17**(1): 1.

Discussion and conclusions

- 95% CI in prostate Delphi suggest differences may not be as large as observed in other studies
- Good R1 agreement and the retention of all items after R1 may explain differences between prostate and 3 other Delphis
- Does dropping items after R1 bias results in favour of reduced variability?

<u>Conclusions</u>

- There is no evidence that *peer only* feedback is beneficial
- There some evidence that *multiple separate* feedback benefits agreement and reduces variability – particularly when there are heterogeneous views to start with

Thanks

Reference:
Brookes, S. T., R. C. Macefield, P. R. Williamson, A. G. McNair, S. Potter, N. S. Blencowe, S. Strong and J. M. Blazeby (2016). "Three nested randomized controlled trials of peer-only or multiple stakeholder group feedback within Delphi surveys during core outcome and information set development." Trials **17**(1): 1.

Any questions