

Is there a consensus when physicians evaluate the relevance of retrieved systematic reviews? Online appendices

Appendix 1. Sample screen from the web-based survey application

Evaluate Evidence Retrieved for Vignette #6

Full Description

An almost-3-year-old girl is sent in from a walk-in clinic by car. She was diagnosed with asthma at age 1 year and has been assessed in the Chest Clinic for recurrent wheezing episodes. Her last exacerbation was 7 months ago, despite compliant use of daily Flovent. She has allergy to nuts and fish. There is a family history of asthma. 30 hours ago, she developed URTI symptoms but has not had a fever. The parents have given Ventolin 5 times in the last 24 hrs. She is afebrile, saturations 93%, HR 124, RR 32. She has moderate distress with air entry diminished to the bases, suprasternal and scalene retractions, and both inspiratory and expiratory wheezing.

Diagnosis

Severe exacerbation (verified in follow-up)

Treatment

Oxygen, beta agonists, corticosteroids, anticholinergics, magnesium sulfate, methylxanthines (according to the CAEP paediatric asthma guideline)

Retrieved Systematic Reviews

For the systematic reviews listed, please rank their relevance to the clinical scenario, from 1 (*most relevant*) to 3 (*least relevant*), or X (*not relevant at all*).

Review	Relevance
3 Combined inhaled anticholinergics and beta2-agonists for initial treatment of acute asthma in children	1 <input type="radio"/> 2 <input checked="" type="radio"/> 3 <input type="radio"/> X <input type="radio"/>
1 Oral xanthines as maintenance treatment for asthma in children	1 <input checked="" type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> X <input type="radio"/>
2 Intravenous aminophylline for acute severe asthma in children over two years receiving inhaled bronchodilators	1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> X <input checked="" type="radio"/>

Comments

If you have any comments on the retrieved reviews or other issues related to the current encounter, please provide them below.

The last review is irrelevant for this vignette. I would recommend replacing it by review #34.

Retrieval ranks, not presented to the participant

Evaluation ranks provided by the participant

Appendix 2. Group value function

Code	Marginal value		
	Position 1	Position 2	Position 3
Y	0.52	0.32	0.16
N	0.26	0.26	0.11
X	0.00	0.00	0.00

Group value function offers a thorough assessment of a triple that is richer than precision at 3. For example, the two coded triples *NXY* and *YXN* have the same value of precision at 3 (it is $2/3 = 0.66$), while the group value function gives a better insight into ranking as it indicates that the latter triple is “more preferred” than the former ($0.26 + 0.00 + 0.16 = 0.42$ for *NXY* vs. $0.52 + 0.00 + 0.11 = 0.63$ for *YXN*).

Appendix 3. Coded triples representing relevance evaluations by physicians

Participant	Vignette													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
sp2	XXX	XXX	YXX	NXN	YNN	NYN	YXX	NYN	XXN	NYN	YXX	NYN	YXX	YXX
sp3	XNN	XXX	XNN	NXN	YNN	NYN	YXX	NYN	XYX	NYN	YNN	NYN	NNY	YXN
sp4	NNY	XXX	NXN	XXN	YXN	XYN	YXX	XYN	YXN	XNX	YXN	XYN	XXN	NNN
sp5	XXX	XXX	XXX	XXX	NXX	XNN	YXX	XNN	NNX	XNN	NXX	XNN	XXX	NXX
sp6	XNX	YXN	XXN	XXN	NXN	NYN	NNN	NYN	XXX	NXN	YNN	NYN	YYY	YXN
sp7	XXN	XXX	XXY	XXX	YXY	NYN	YNX	NNN	XXY	XYN	YXX	XYN	XXY	YXX
sp8	XXX	NXN	XYN	NXN	YNN	XYN	YNN	XYN	NXN	XXN	YNN	XXY	XYN	YXN
sp9	NNY	YXX	XXX	XXN	YXX	XYN	YXX	XYN	XXX	XNN	YXX	XYN	YYY	YNN
sp10	NNY	YXN	NXN	NXN	YNN	XYN	YYY	XYN	XXY	XXN	YYY	XYN	YYY	YNN
sp11	NNY	YXN	NXN	NXN	YNN	NYN	YXX	NYN	XXX	XYN	YNN	NYN	NNN	YNN
sp12	NNN	XXX	NXN	NXN	YXN	NYN	YYX	NYN	YXX	NXN	YYY	XYN	YXN	YXN
sp14	NYN	YXX	NNN	NNN	YNN	NNN	YYX	NYN	YXX	NXN	YYY	NYN	YYY	YXN
sp16	NNN	YXX	XXN	NXN	NXN	NNN	YYX	NYN	YXX	NXN	NNX	NNN	YXN	YXN
sp17	XXN	NXX	XXN	XXX	YXX	XNN	YXX	XNN	XXY	XNN	YXY	XNN	XNN	YXY
sp18	NYN	XXX	NNN	YXN	YNN	NYN	YXX	XYN	YNX	NXN	YXN	XXN	NXN	YXN
sp19	XYN	XXX	YXN	XXX	YYY	NNN	YXX	NYN	YXX	NYN	YXX	XYN	NXN	NXN
sp21	NXN	YXX	XXN	NXN	NXN	NYN	YXX	NYN	XXN	NYN	YYY	NYN	YNN	YXN
sp22	NNN	XNN	NNN	YNN	YXX	NYN	YYX	XYN	YYY	NYN	YXX	XYN	XXN	YXX
sp25	NNX	XXX	NXN	XXX	XXN	NNX	XNX	NNX	XXX	NXX	XXN	XNX	XNX	XXN
sp26	YYY	YNN	NNN	NNN	YNN	NNN	YYY	XYN	YYY	XNN	YNN	NNN	NNY	YNN
sp28	YNN	YNN	NNN	NXN	NXN	NYN	YNX	NYN	YNX	XYN	YXN	XYN	XNN	YNN
sp29	NNY	NXN	NXN	NXN	YNN	NYN	YNX	NYN	NYX	XYN	YYY	NYN	NYN	YXN

There are major differences between evaluations provided by different physicians for the same vignette. For example, physician sp25 evaluated reviews retrieved for vignette 9 as *XXX* (considered all of them to be irrelevant), while physician sp26 evaluated the same reviews as *YYY* (considered all of them not only as relevant but also correctly ranked by the retrieval algorithm).

Appendix 4. Precision at 3 and group value function averaged over vignettes (95% CI)

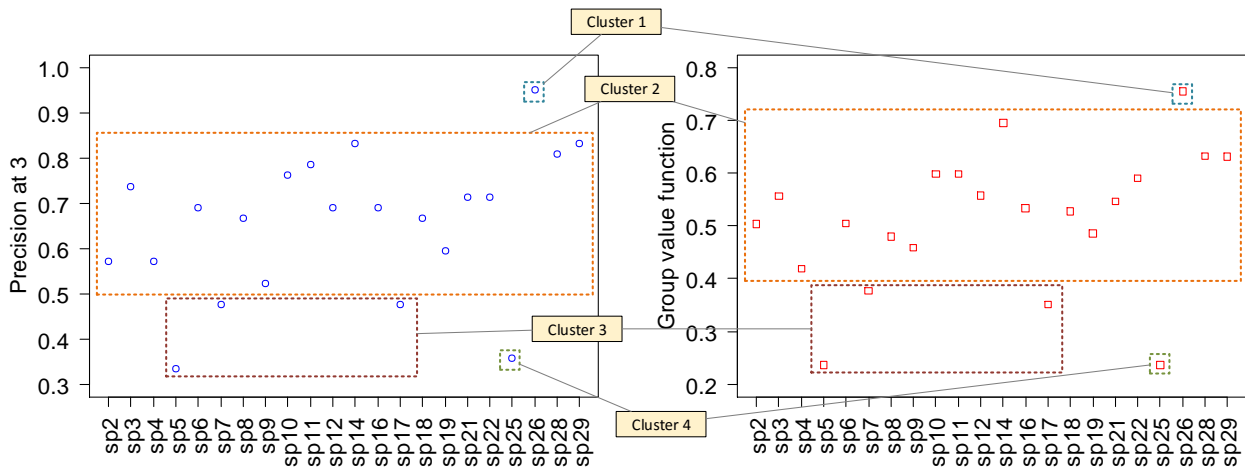
Participant	Precision at 3	Group value function
sp2	0.57 [0.37, 0.77]	0.50 [0.35, 0.65]
sp3	0.74 [0.57, 0.91]	0.56 [0.43, 0.69]
sp4	0.57 [0.43, 0.72]	0.42 [0.30, 0.54]
sp5	0.33 [0.18, 0.49]	0.24 [0.13, 0.34]
sp6	0.69 [0.52, 0.86]	0.51 [0.35, 0.66]
sp7	0.48 [0.31, 0.64]	0.38 [0.23, 0.52]
sp8	0.67 [0.53, 0.80]	0.48 [0.34, 0.62]
sp9	0.52 [0.35, 0.70]	0.46 [0.31, 0.61]
sp10	0.76 [0.64, 0.89]	0.60 [0.43, 0.76]
sp11	0.79 [0.62, 0.95]	0.60 [0.47, 0.73]
sp12	0.69 [0.55, 0.84]	0.56 [0.43, 0.68]
sp14	0.83 [0.70, 0.97]	0.70 [0.60, 0.79]
sp16	0.69 [0.56, 0.82]	0.53 [0.44, 0.63]
sp17	0.48 [0.36, 0.59]	0.35 [0.24, 0.46]
sp18	0.67 [0.51, 0.82]	0.53 [0.40, 0.66]
sp19	0.60 [0.41, 0.78]	0.49 [0.35, 0.62]
sp21	0.71 [0.56, 0.87]	0.55 [0.41, 0.68]
sp22	0.71 [0.56, 0.87]	0.59 [0.47, 0.71]
sp25	0.36 [0.23, 0.48]	0.24 [0.14, 0.34]
sp26	0.95 [0.89, 1.02]	0.76 [0.64, 0.87]
sp28	0.81 [0.72, 0.90]	0.63 [0.52, 0.74]
sp29	0.83 [0.74, 0.92]	0.63 [0.53, 0.73]

Appendix 5. Values of Cohen's kappa for pairs of participants

Participant	Participant																					
	sp2	sp3	sp4	sp5	sp6	sp7	sp8	sp9	sp10	sp11	sp12	sp14	sp16	sp17	sp18	sp19	sp21	sp22	sp25	sp26	sp28	sp29
sp2		0.50	0.27	0.18	0.32	0.44	0.24	0.34	0.27	0.44	0.43	0.34	0.32	0.14	0.32	0.42	0.54	0.26	-0.05	0.08	0.19	0.34
sp3	0.50		0.23	0.03	0.47	0.38	0.30	0.34	0.28	0.61	0.49	0.47	0.38	0.22	0.51	0.43	0.55	0.30	0.09	0.29	0.37	0.50
sp4	0.27	0.23		0.20	0.15	0.26	0.30	0.56	0.44	0.45	0.49	0.19	0.21	0.32	0.45	0.38	0.31	0.35	0.11	0.20	0.38	0.32
sp5	0.18	0.03	0.20		0.00	0.34	0.07	0.29	0.01	-0.05	0.06	-0.01	0.17	0.44	0.10	0.27	0.07	0.12	-0.19	0.02	0.07	0.04
sp6	0.32	0.47	0.15	0.00		0.30	0.41	0.40	0.40	0.43	0.41	0.37	0.46	0.09	0.18	0.21	0.48	0.09	0.17	0.09	0.31	0.36
sp7	0.44	0.38	0.26	0.34	0.30		0.22	0.35	0.21	0.26	0.37	0.20	0.23	0.50	0.22	0.49	0.37	0.37	0.07	0.05	0.30	0.27
sp8	0.24	0.30	0.30	0.07	0.41	0.22		0.26	0.43	0.33	0.27	0.20	0.13	0.27	0.26	0.10	0.27	0.13	-0.09	0.15	0.25	0.41
sp9	0.34	0.34	0.56	0.29	0.40	0.35	0.26		0.58	0.49	0.40	0.32	0.34	0.39	0.29	0.27	0.41	0.26	-0.06	0.22	0.32	0.36
sp10	0.27	0.28	0.44	0.01	0.40	0.21	0.43	0.58		0.57	0.54	0.48	0.29	0.25	0.29	0.09	0.39	0.24	0.04	0.40	0.31	0.52
sp11	0.44	0.61	0.45	-0.05	0.43	0.26	0.33	0.49	0.57		0.48	0.38	0.37	0.19	0.40	0.32	0.55	0.22	0.11	0.38	0.52	0.68
sp12	0.43	0.49	0.49	0.06	0.41	0.37	0.27	0.40	0.54	0.48		0.67	0.67	0.25	0.60	0.50	0.64	0.49	0.24	0.11	0.44	0.56
sp14	0.34	0.47	0.19	-0.01	0.37	0.20	0.20	0.32	0.48	0.38	0.67		0.55	0.16	0.49	0.37	0.52	0.30	0.14	0.34	0.26	0.46
sp16	0.32	0.38	0.21	0.17	0.46	0.23	0.13	0.34	0.29	0.37	0.67	0.55		0.19	0.35	0.38	0.58	0.27	0.15	0.15	0.29	0.31
sp17	0.14	0.22	0.32	0.44	0.09	0.50	0.27	0.39	0.25	0.19	0.25	0.16	0.19		0.19	0.33	0.36	0.22	0.01	0.15	0.16	0.18
sp18	0.32	0.51	0.45	0.10	0.18	0.22	0.26	0.29	0.29	0.40	0.60	0.49	0.35	0.19		0.49	0.33	0.49	0.08	0.25	0.39	0.37
sp19	0.42	0.43	0.38	0.27	0.21	0.49	0.10	0.27	0.09	0.32	0.50	0.37	0.38	0.33	0.49		0.36	0.32	0.12	0.07	0.25	0.30
sp21	0.54	0.55	0.31	0.07	0.48	0.37	0.27	0.41	0.39	0.55	0.64	0.52	0.58	0.36	0.33	0.36		0.24	0.09	0.05	0.39	0.47
sp22	0.26	0.30	0.35	0.12	0.09	0.37	0.13	0.26	0.24	0.22	0.49	0.30	0.27	0.22	0.49	0.32	0.24		-0.03	0.25	0.44	0.30
sp25	-0.05	0.09	0.11	-0.19	0.17	0.07	-0.09	-0.06	0.04	0.11	0.24	0.14	0.15	0.01	0.08	0.12	0.09	-0.03		-0.04	0.12	0.09
sp26	0.08	0.29	0.20	0.02	0.09	0.05	0.15	0.22	0.40	0.38	0.11	0.34	0.15	0.15	0.25	0.07	0.05	0.25	-0.04		0.28	0.21
sp28	0.19	0.37	0.38	0.07	0.31	0.30	0.25	0.32	0.31	0.52	0.44	0.26	0.29	0.16	0.39	0.25	0.39	0.44	0.12	0.28		0.40
sp29	0.34	0.50	0.32	0.04	0.36	0.27	0.41	0.36	0.52	0.68	0.56	0.46	0.31	0.18	0.37	0.30	0.47	0.30	0.09	0.21	0.40	

For each pair of participants kappa value was computed using coded triples obtained for these participants (see Appendix 3). Specifically, for each participant, we concatenated all 14 associated triples (one triple per vignette) into a single vector of relevancy evaluations. Triples were concatenated in the same order (corresponding to the sequence of vignettes) and the resulting vector contained 52 entries corresponding to coded relevancy evaluations of individual systematic reviews. Then, these vectors were exported to the R system, where we calculated kappa values.

Appendix 6. Selected clustering of physicians in the context of precision at 3 and group value function



Physician sp5 from cluster 3 and sp25 from cluster 4 are very similar in terms of these two measures. However, a closer look at their coded triples (Appendix 3) reveals differences in evaluations across vignettes. For example, reviews retrieved for vignette 1 were evaluated as *XXX* by sp5 and as *NNX* by sp25, while for vignette 10 the evaluations were *NNX* for sp5 and *XXX* for sp25. While these differences were compensated after averaging values of both measures over all vignettes, they were captured by the kappa coefficient (that indicated the lack of agreement between sp5 and sp25) and resulted in placing these two physicians in two different clusters.