# Supplementary Material S1. Symptom diary

4 Symptoms Each day, please complet boxes using the number w shows how your child h last 24 hours	ch you think best 1 = Very little problem 5 = Very bad	uld be Your child's study number 999999
This column is for the day your child saw the doctor or nurse	aw kaw ktor Day Day Day Day Day Day Day Day Day Day	Day Day Day Day Day Day Day Day
Fill in the day of the week (for example M, T, W, T, F, S, S)	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
How was your child's cough?	_	
How short of breath (breathing faster) was your child?	_	
How well did your child sleep (last night)?	=	
How well did your child cope with normal activities?	_	
How unwell was your child?	_	
How was your child's temperature? Medicines		
If your child has been taking any medicines plea the number of times, per day, your child took ea		
Write the medicine name here	isaw Week 1 (days 1 to 7) Week 2 (days 8 to 14) Week 3 (days 15 to 7) Jay Day Day Day Day Day Day Day Day Day D	
	_ <u> </u>	
	_	
	_	
	_	

# Supplementary Material S2. Case report form

	CASE REP		RM	Q	TARG	ET
ID Today's date	Background informati	on Y Y Y Y	Mother's age		hildren in hom (inc. unwell child	
		hnicity	mother s	bes the <mark>□</mark> moke?	No <sub>0</sub> Yes <sub>1</sub>	Don't know <sub>88</sub>
Informed consent for study obtained 🗖 1	Male <sub>1 PTO i</sub> ethnicity, pleas	or codes, if othe e describe belov	feeding of	breast 🗖 child at onths?	No <sub>0</sub> Yes <sub>1</sub>	Don't know <sub>88</sub>
Carer reported symptoms	How unwell does the pare consider the child to be?	ent 0 Well 🔲 [		5 6 7	8 9 10	y unwell
Duration of days	s illness got a lot ☐ No₀ [ worse recently?	Yes <sub>1</sub>	f Yes, how ma did it start t			5
Symptoms present	During illne Noo Ye	ess? Last: s <sub>1</sub> If yes No <sub>0</sub>	24 hours? Yes <sub>1</sub> If yes	Severity in Mild <sub>1</sub>	last 24 hours (t Moderate <sub>2</sub>	<i>ick</i> one) Severe <sub>3</sub>
Dry cough						
Productive/ wet cough		] ≻ □				
Barking/ croupy cough						
Blocked or runny nose		i → Π	$\overline{n} >$	Ē	Ē	<b>n</b>
Change in cry			П	H	- H	Н
Breathing faster than normal (sho			H	H	H	H
Wheeze or whistling in the chest			H	H	H	H
Fever			H		H	H
			H	H	H	H
Chills/ shivering			H	H	H	H
Diarrhoea			H	H	H	님
Vomiting (including after cough)				H	H	님
Taking fewer fluids/ milk feeds			L >	님	님	님
Eating less		<u>i</u> > []		H	H	님
Low energy/ fatigue/ lethargy				<u> </u>	닏	브
Disturbed sleep						
Passing urine less often/ dryer na	ppies					
Please tick NA if the child is too y	oung/ uncommunicative for	the parent to	o know about	the followin	ig > NA 🗖 1	
Chest/ shoulder pain		] > 🗆				
Headache		] > 🗆				
Muscle aches all over		] > 🗖				
Confusion/ disorientation						
Clinician examination and m	anagement Present <sub>1</sub> Temperature	₽	°C Pulse	bp	om	
Pallor			о <b>Г</b>		Unable to	take O <sub>2</sub>
Grunting	Respiratory rate	e bpm	O <sub>2</sub> sat	%	sat/no eq	
Nasal flaring	Consci	ousness leve	l 🔲 normal₀	irritab	e <sub>1</sub> drows	Y2
Stridor			_		_	
Inter/ subcostal recession	Capil	lary refill time	2 second	s or less	3 seconds	or more <sub>1</sub>
Inflamed pharynx/ tonsils		ow unwell do	you consider	the child to	be?	
Absent	Unilateral <sub>1</sub> Bilateral <sub>2</sub>		3 4 5 6			العن
Wheeze					Very unv	ven
Crackles/ crepitations			Т	hroat swab t	aken? 🚺 No	Yes1
Bronchial breathing			If	No, reason:		
Main working respiratory tract dia	gnosis			Child refus		
My gut feeling is 'something is wr	ong' 🔲 No <sub>0</sub> 🔛 Yes <sub>1</sub>		Ē	Other (spe		
Antibiotics prescribed?	Yes, immediate, 🔲 Yes,	delayed <sub>2</sub> by	days			
Referral for acute admission		· L	┷┛┊╞			
			L			

2011 04 08 TARGET CRF Bristol centre version 1.0

# **Supplementary material S3.** LLCA models – model fit and comparisons. Table. Fit statistics for "cough" models (n = 1,408)

Timing	# classes	# parameters	aBIC	Entropy	Smallest class	LL	ΔLL	LMR pvalue	BLRT pvalue
	1	30	26845.6	NA	100.0%	13361.7		-	-
	2	61	21064.5	0.942	25.9%	10408.0	2953.6	< 0.001	< 0.001
	3	92	19142.9	0.915	16.6%	9384.1	1024.0	< 0.001	< 0.001
1-15 days (skip 0)	4	123	18513.5	0.932	7.2%	9006.3	377.8	< 0.001	< 0.001
	5	154	17920.7	0.930	6.8%	8646.7	359.5	< 0.001	< 0.001
	6	185	17570.4	0.932	5.7%	8408.4	238.3	< 0.001	< 0.001
	7	216	17290.6	0.932	5.1%	8205.4	203.0	0.0035	0.0040
	1	16	14247.7	NA	100.0%	7091.3		-	-
	2	33	11896.6	0.903	22.8%	5881.1	1210.2	< 0.001	< 0.001
	3	50	11288.3	0.843	14.8%	5542.3	338.8	< 0.001	< 0.001
1-15 days (skip 1)	4	67	11070.3	0.867	6.6%	5398.7	143.6	< 0.001	< 0.001
	5	84	10925.4	0.900	6.0%	5291.6	107.1	< 0.001	< 0.001
	6	101	10875.9	0.900	2.5%	5232.2	59.4	0.0115	< 0.001
	7	118	10826.6	0.920	2.2%	5173.0	59.2	1.0000	< 0.001
	1	10	8765.9	NA	100.0%	4362.6		-	-
	2	21	7928.3	0.820	24.0%	3921.4	441.2	< 0.001	< 0.001
1-15 days	3	32	7766.7	0.742	13.0%	3818.2	103.2	< 0.001	< 0.001
(skip 2)	4	43	7711.0	0.805	9.4%	3767.9	50.3	0.0045	< 0.001
	5	54	7695.1	0.817	2.0%	3737.6	30.4	0.0513	< 0.001
	6	65	7700.1	0.833	1.5%	3717.7	19.9	0.0166	0.3960

LL= log-likelihood, ΔLL= change in log-likelihood, aBIC = Sample-size adjusted BIC; BLRT = Bootstrap Likelihood Ratio Test; LMR = Lo-Mendell Rubin test

#### Model fit statistics considered

#### aBIC - sample-size adjusted Bayesian Information Criterion.

The Bayesian Information Criterion (BIC: Schwartz, G., 1978) is the most commonly-used fit statistic for comparing mixture models. A function of both the likelihood and the number of estimated parameters, the BIC penalises model complexity. We opted for the sample-size adjusted version which incorporates the sample-size as an additional term. BIC will typically decrease and then increase following the incremental additional of classes. Using this statistic the model with the lowest BIC (or other models with BIC values in the vicinity) would be deemed satisfactory however in some instances this statistic did not reach a minimum within the range of models considered..

#### Bootstrap tests for nested models.

The Bootstrap Likelihood Ratio Test (BLRT) and the Lo-Mendell-Rubin (LMR) test statistics (Nylund et al., 2007) both assess change in model fit when adding an additional class. Here a high p-value for a k-class model indicates no substantial improvement in fit compared to the k-1 class solution. Unlike the LMR, The BLRT makes no distributional assumptions and simulation work has so far shown this measure to be superior (Nylund et al., 2007) however in our experience the BLRT can be extremely conservative and may reject all the models considered.

# Entropy.

Mixture modelling output consists primarily of class-assignment probabilities which describe the confidence with which each participant can be assigned to each latent class. Entropy, also referred to as classification accuracy, summarises this information as a single measure which can take values form zero to one, with one indicating no assignment uncertainty. Entropy is of little use in determining the optimal model (Tein et al, 2013) and can be poor in simulation studies even when the correct model is estimated (Heron et al., 2015). Whilst LCA has been promoted as a method to facilitate targeted interventions (Lanza and Rhoades, 2013) we propose that such a strategy is dependent on clearly defined and well-separated groups of individuals. Consequently, we regard entropy as an indicator of model *utility* since if entropy is low and individuals can only be poorly classified then the resulting classification is of little use as a targeting tool.

In addition, entropy has been shown to be important when it comes to the level of bias resulting from a standard three-step analysis (Vermunt, 2010; Bakk, Tekle, & Vermunt, 2013; Bakk, Oberski, & Vermunt, 2014). Whilst not directly related to the issue of model utility, entropy will influence the analytical approach employed when assessing covariate and outcome associations.

#### Smallest class size.

As a latent class analysis is usually the initial stage of a project with the intention of deriving a number of groups for further research, analysts often place a limit on the size of the results classes. This pragmatic decision is to facilitate the planned further study since there is little one could reasonably do with a class of ten participants other than drop them from the sample. Here we only considered models where all classes contained at least 5% of the participants.

#### Bivariate residuals

In addition to the figures shown in the previous table we also examined the bivariate residuals from each model. Mixture modelling, as with continuous trait modelling, is based on the assumption of conditional independence, namely that the class indicators should be independent conditional on the latent variable. The pattern and magnitude of these residuals is examined over the following few pages.

Bakk, Z., Tekle, F. T., & Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology*, 43, 272-311.

Bakk, Z., Oberski, D. L., & Vermunt, J. K. (2014). Relating Latent Class Assignments to External Variables: Standard Errors for Correct Inference. *Political Analysis*, 22, 520-540.

Heron, J. E., Croudace, T. J., Barker, E. D., & Tilling, K. (2015). A comparison of approaches for assessing covariate effects in latent class analysis. *Longitudinal and Life Course Studies*; Vol 6, No 4.

Lanza ST and Rhoades BL. Latent Class Analysis: An Alternative Perspective on Subgroup Analysis in Prevention and Treatment. *Prevention Science*. 2013, Volume 14, Issue 2, pp 157-168

Nylund KL, Asparouhov T, Muthen BO. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modelling: A Monte Carlo Simulation Study. *Structural Equation Modelling: A Multidisciplinary Journal* 2007;14(4):535-569.

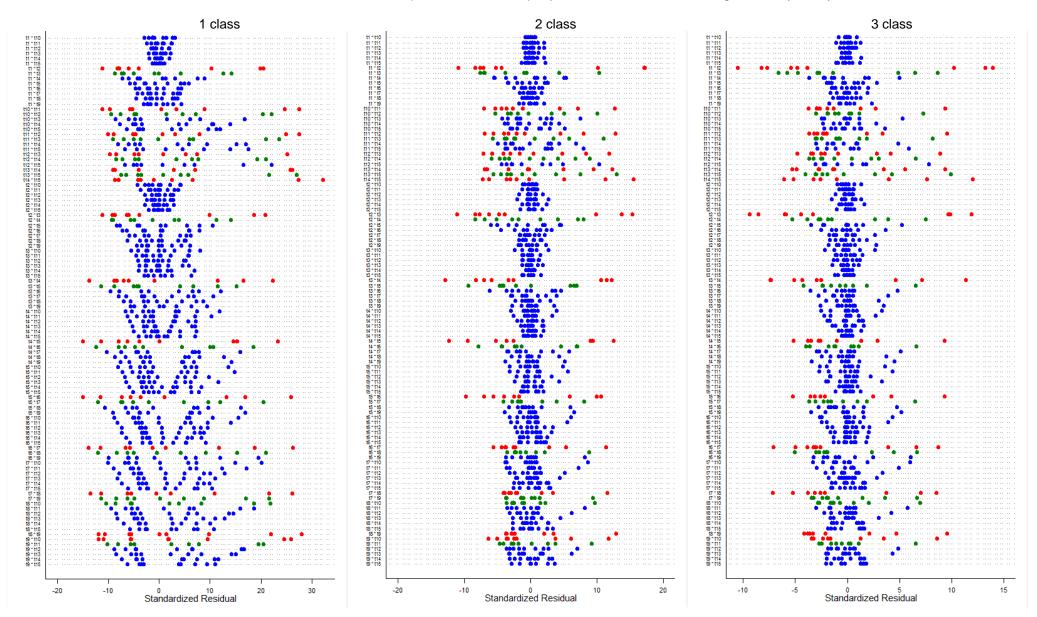
Schwarz G. Estimating the dimension of a model. Annals of Statistics 1978;6:461-464

Tein J-Y, Coxe S, Cham H. Statistical Power to Detect the Correct Number of Classes in Latent Profile Analysis. *Structural Equation Modeling: A Multidisciplinary Journal.* Vol. 20, Iss. 4, 2013.

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18, 450-469.

## Bivariate residuals from models of 15 consecutive days' data on COUGH.

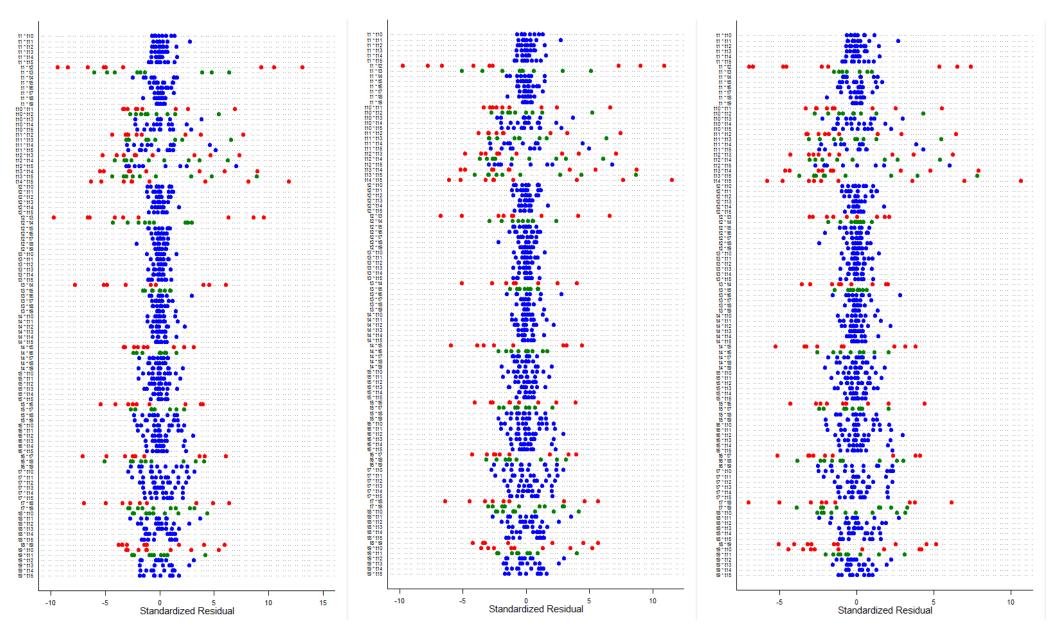
Graphs over next two pages show residuals for 1 class, 2 class, ... 6 class models. As each class indicator has three categories, there are nine residuals for each pair of measures (3x3 cells in the contingency table). Red indicates residuals between adjacent time points, green indicates residuals between measures two days apart and blue the remaining residuals. It's apparent that these models are failing to capture two aspects of the data (i) the strong association between measures taken very close together and (ii) the association between measures taken towards the start and end of the two-week period when the majority of children exhibit little change from day to day.



4 class

5 class

6 class



Not surprisingly, we see marked improvement in the magnitude of the residuals when adjacent measurements are dropped from the model. These figures show pairwise residuals from 1 through 6-class models for the cough data using alternate measures. Once again, but to a lesser extent, we see that the model is less able to model the data at the two extremes of the two-week measurement window.

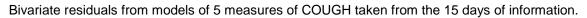
I	1 class	2 class	1	3 class		
t1 * t11	· · · · · · · · · · · · · · · · · · ·	t1 * t11		t1 * t11		
t1 * t13	• • • • • • • • • • • • • • • • • • •	t1 * t13		t1 * t13	• • • • • • • • • • • • • • • • • • • •	
t1 * t15		t1 * t15		t1 * t15		
t1 * t3		t1 * t3		t1 * t3		
t1 * t5		t1 * t5	• • • • • • • • • • • • • • • • • • • •	t1 * t5		
t1 * t7		t1 * t7		t1 * t7		
t1 * t9	· · · · · · · · · · · · · · · · · · ·	t1 * t9		t1 * t9		
11 * t13		t11 * t13		t11 * t13		
t11 * t15		t11 * t15		t11 * t15		
13 * t15		t13 * t15		t13 * t15		
t3 * t11		t3 * t11		t3 * t11		
t3 * t13		t3 * t13		t3 * t13		
t3 * t15		t3 * t15		t3 * t15	•••••	
t3 * t5	• • • • • • • • • • • • • • • •	t3 * t5		t3 * t5		
t3 * t7		t3 * t7		t3 * t7	• • • • • • • • • • • • • • • • • • • •	
t3 * t9	· · · · · · · · · · · · · · · · · · ·	t3 * t9		t3 * t9		
t5 * t11		t5 * t11		t5 * t11		
t5 * t13		t5 * t13		t5 * t13		
t5 * t15	· · · · · · · • • • • • • • • • • • • •	t5 * t15		t5 * t15		
t5 * t7		t5 * t7		t5 * t7		
t5 * t9		t5 * t9		t5 * t9		
t7 * t11		t7 * t11		t7 * t11		
t7 * t13		t7 * t13		t7 * t13		
t7 * t15		t7 * t15		t7 * t15		
t7 * t9		t7 * t9		t7 * t9		
t9 * t11		t9 * t11		t9 * t11		
t9 * t13		t9 * t13		t9 * t13		
t9 * t15		t9 * t15		t9 * t15		
	-10 0 10 20 30 Standardized Residual		-10 -5 0 5 10 Standardized Residual		-10 -5 0 5 10	

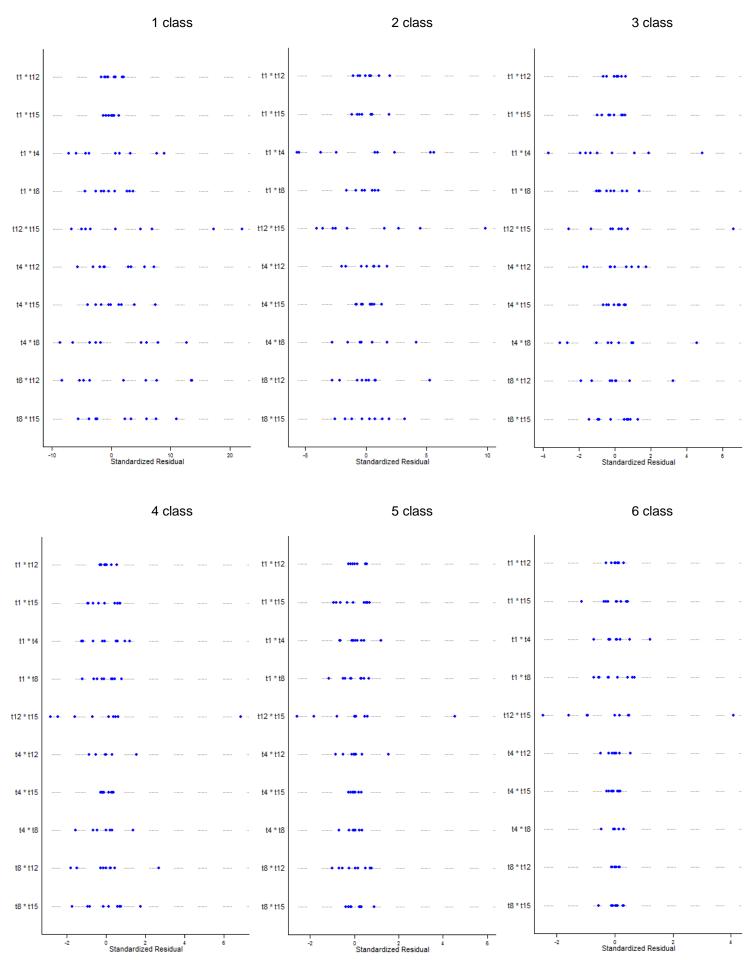
1		1			
t1 * t11		t1 * t11		•••••• t1 * t11	
t1 * t13		t1 * t13		······ t1 * t13	
t1 * t15		t1 * t15		•••••• t1 * t15	
t1 * t3		t1 * t3		t1 * t3	
t1 * t5	· · · · · · · · · · · · · · · · · · ·	t1 * t5		t1 * t5	
t1 * t7		t1 * t7		t1 * t7	
t1 * t9		t1 * t9		t1 * t9	
t11 * t13		t11 * t13		t11 * t13	
t11 * t15		t11 * t15		t11 * t15	
t13 * t15		t13 * t15		t13 * t15	
t3 * t11	• • • • • • • • • • • • • • • • • • •	t3 * t11		t3 * t11	
t3 * t13		t3 * t13		t3 * t13	
t3 * t15		t3 * t15		t3 * t15	
t3 * t5		t3 * t5		t3 * t5	
t3 * t7		t3 * t7		t3 * t7	
t3 * t9		t3 * t9		t3 * t9	
t5 * t11		t5 * t11		t5 * t11	
t5 * t13		t5 * t13		t5 * t13	
t5 * t15		t5 * t15		t5 * t15	
t5 * t7		t5 * t7		t5 * t7	
t5 * t9		t5 * t9		t5 * t9	
t7 * t11		t7 * t11		t7 * t11	
t7 * t13		t7 * t13		t7 * t13	
t7 * t15		t7 * t15		t7 * t15	
t7 * t9		t7 * t9	· · · · · · · • • • · · • • • • • • • •	t7 * t9	
t9 * t11		t9 * t11		t9 * t11	
t9 * t13		t9 * t13	· · · · · · · · · • • • • • • • • • • · · · • ·	t9 * t13	
t9 * t15		t9 * t15		t9 * t15	
l		l	1 1	- <del></del>	
	-5 0 5 10 Standardized Residual		-5 0 5 Standardized Residual	10	-5 0 5 Standardized Residual

4 class

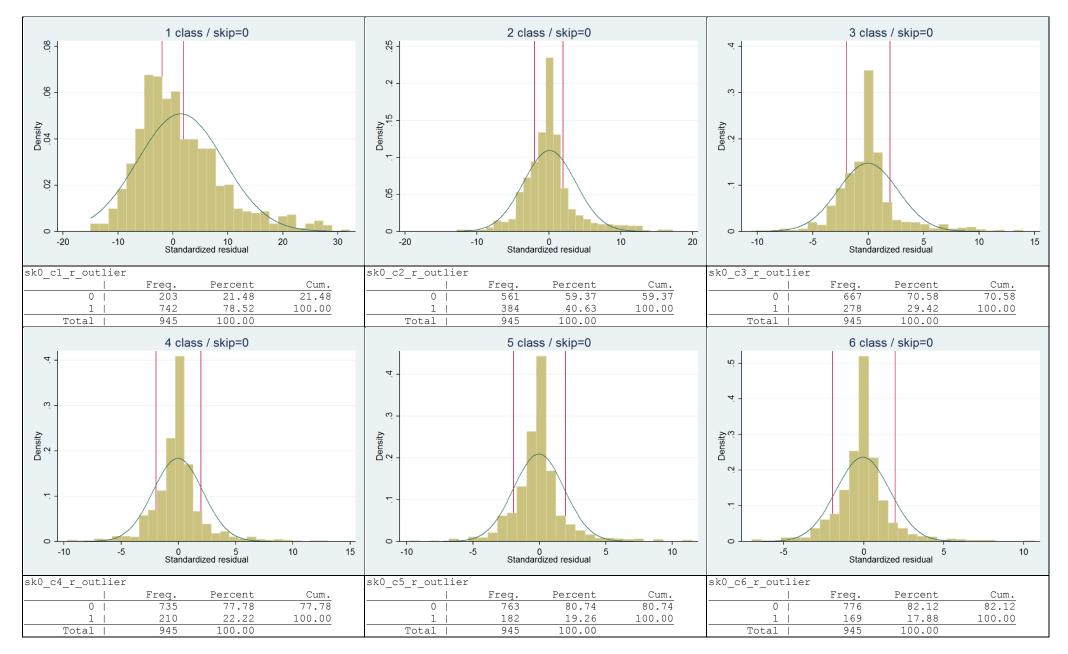
5 class

6 class

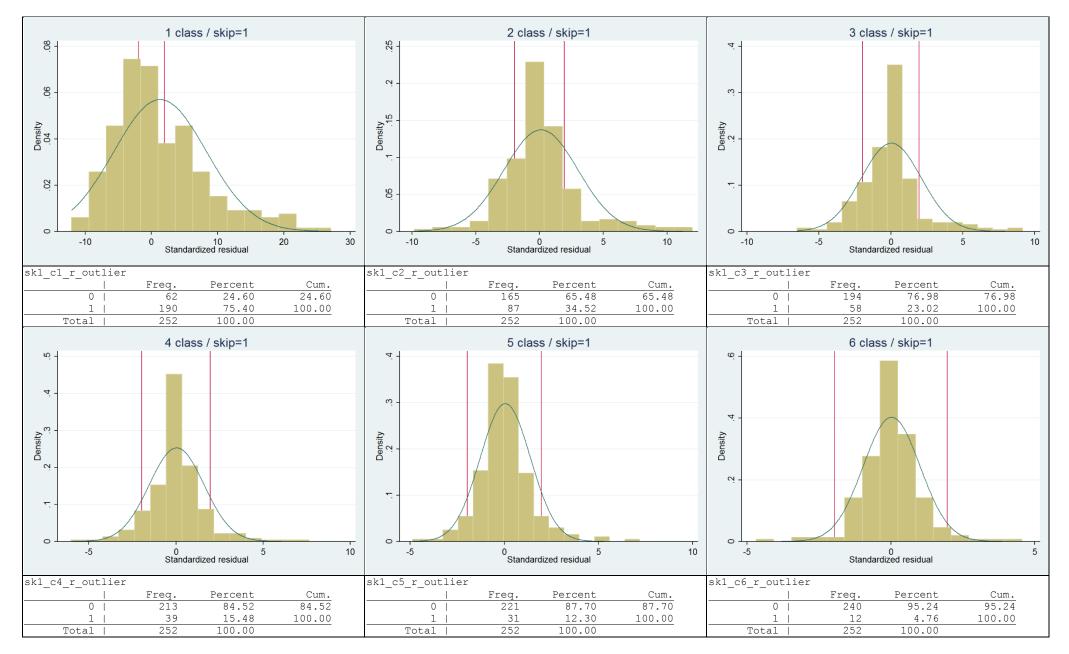




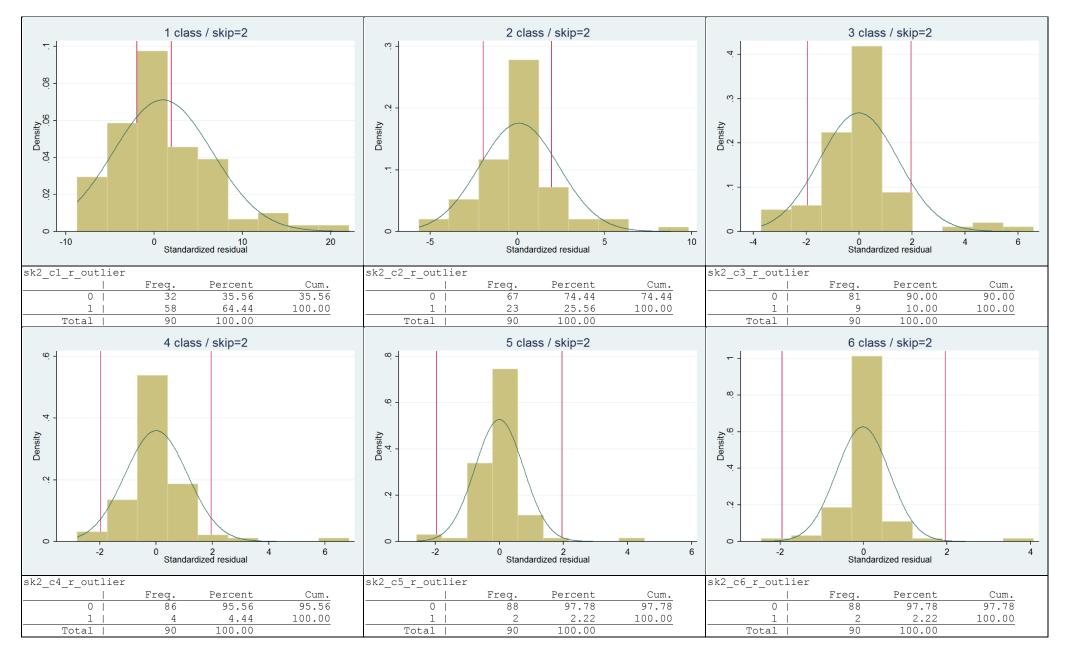
# Distribution of bivariate residuals with red-lines indicating +/- 1.96 (Skip 0: All 15 measures)



# Distribution of bivariate residuals with red-lines indicating +/- 1.96 (Skip 1: Alternate measures)



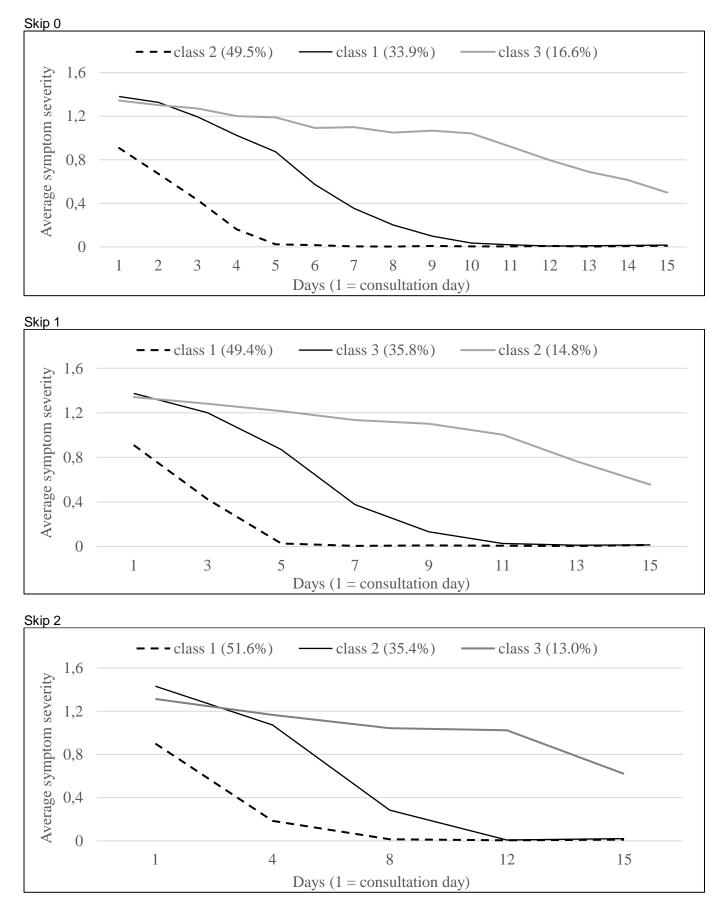
# Distribution of bivariate residuals with red-lines indicating +/- 1.96 (Skip 2: Every third measure)



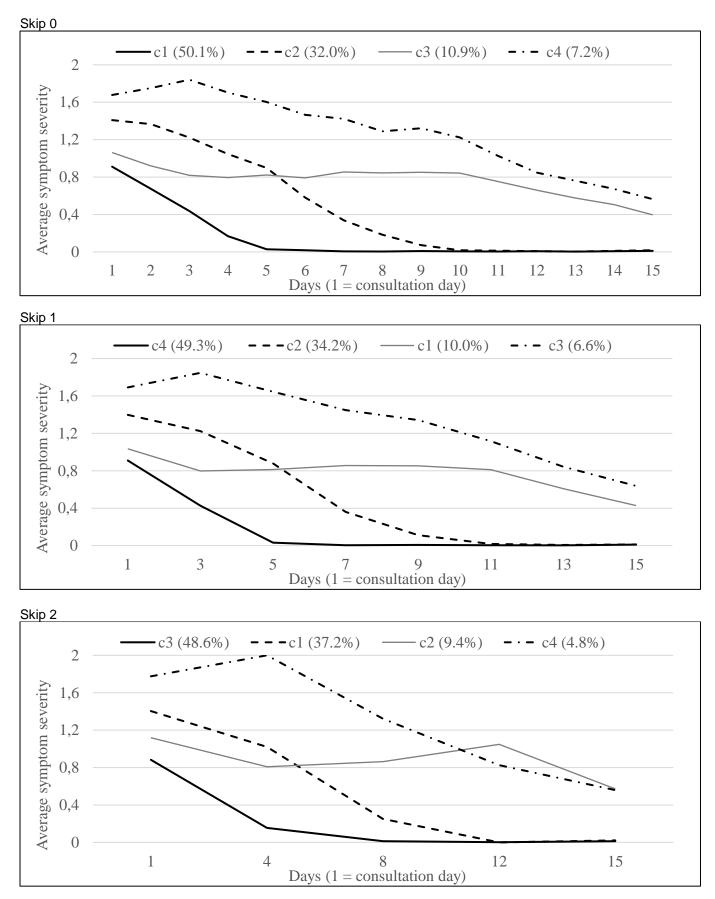
# Estimated trajectories

Modelling three-category ordinal measures poses a slight problem when it comes to plotting the results as all three categories describe a trajectory through time meaning two plots are necessary. To avoid this we have collapsed the within-class profiles by turning the probabilities into a single measure of severity (severity = 0\*P(Category 0) + 1\*P(Category 1) + 2\*P(Category 2)).

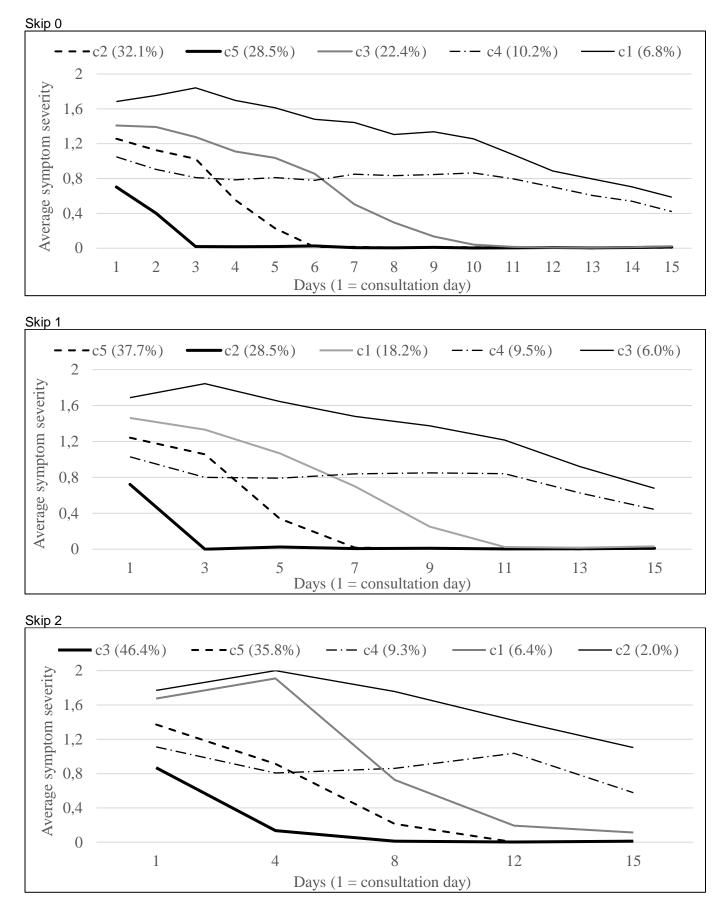
# 3-class models



4-class models



## 5-class models



Supplementary Table S1. Univariable analyses of change† in cough severity from day of recruitment to the following day (day 2) in 1,385 children presenting
to primary care with respiratory tract infection between July 2011 and May 2013 <sup>\$</sup> .

/ariable		"Exposed"	Improved	Unchanged	Worse	p-value
Nac of obild	group	group	OR [95% CI]	OR	OR [95% CI]	
Age of child	2 – 15 years	0 – 1 year	1.20 [0.94, 1.53]	1.00 ref	1.65 [1.09, 2.50]	0.040
Gender	Female	Male	1.13 [0.90, 1.43]	1.00 ref	1.14 [0.76, 1.71]	0.532
Ethnicity	White	Non-white	1.16 [0.80, 1.67]	1.00 ref	1.03 [0.53, 2.01]	0.742
Mother smoking	No	Yes	1.20 [0.87, 1.64]	1.00 ref	0.81 [0.44, 1.49]	0.354
iving in 20% most deprived	No	Yes	0.74 [0.53, 1.03]	1.00 ref	0.81 [0.46, 1.45]	0.176
	0 1	× ۲	• • •			
No of consultations last year	0 – 4 consultations	≥ 5 consultations	0.96 [0.62, 1.48]	1.00 ref	1.13 [0.54, 2.33]	0.918
Current asthma	No	Yes	1.31 [0.91, 1.91]	1.00 ref	1.65 [0.91, 2.98]	0.153
Previous asthma	No	Yes	1.21 [0.71, 2.06]	1.00 ref	1.08 [0.42, 2.81]	0.785
Pre-consultation illness duration	>3 days	≤3 days	1.21 [0.93, 1.57]	1.00 ref	1.42 [0.91, 2.21]	0.163
Severe dry cough <sup>b</sup>	No	Yes	1.27 [0.82, 1.99]	1.00 ref	0.90 [0.38, 2.16]	0.520
Severe productive cough <sup>b</sup>	No	Yes	1.41 [0.95, 2.07]	1.00 ref	1.06 [0.51, 2.18]	0.229
Severe barking cough <sup>b</sup>	No	Yes	1.10 [0.68, 1.79]	1.00 ref	0.83 [0.32, 2.12]	0.827
Any severe cough <sup>b</sup>	No	Yes	1.35 [1.00, 1.82]	1.00 ref	1.02 [0.59, 1.77]	0.139
Severe blocked or runny nose <sup>b</sup>	No	Yes	1.03 [0.66, 1.59]	1.00 ref	1.60 [0.83, 3.08]	0.396
Moderate-to-severe change in cry <sup>b</sup>	No	Yes	1.03 [0.72, 1.46]	1.00 ref	0.76 [0.38, 1.51]	0.686
Moderate-to-severe shortness of preath <sup>b</sup>	No	Yes	1.42 [1.09, 1.84]	1.00 ref	1.16 [0.73, 1.86]	0.034
Noderate-to-severe wheezeb	No	Yes	1.25 [0.97, 1.62]	1.00 ref	1.04 [0.66, 1.66]	0.236
Moderate-to-severe diarrhoeab	No	Yes	0.81 [0.45, 1.46]	1.00 ref	1.97 [0.93, 4.20]	0.141
Moderate-to-severe vomiting <sup>b</sup>	No	Yes	0.88 [0.61, 1.28]	1.00 ref	0.89 [0.46, 1.72]	0.774
Moderate-to-severe reduced fluid	No	Yes	1.06 [0.77, 1.46]	1.00 ref	1.04 [0.59, 1.82]	0. 944
Severe reduced food intake <sup>b</sup>	No	Yes	1.09 [0.68, 1.76]	1.00 ref	0.63 [0.22, 1.79]	0.564
Severe reduced energy <sup>b</sup>	No	Yes	0.73 [0.42, 1.28]	1.00 ref	0.67 [0.24, 1.91]	0.450
Severe disturbed sleep <sup>b</sup>	No	Yes	1.41 [1.05, 1.89]	1.00 ref	1.55 [0.95, 2.53]	0.036
Moderate-to-severe reduction in urine passed <sup>b</sup>	No	Yes	1.05 [0.66, 1.67]	1.00 ref	0.43 [0.13, 1.40]	0.255
Severe fever <sup>b</sup>	No	Yes	1.04 [0.67, 1.64]	1.00 ref	0.68 [0.27, 1.73]	0.651
Temperature ≥37·8°C at	No	Yes	. , .			
consultation			0.95 [0.67, 1.34]	1.00 ref	0.94 [0.51, 1.74]	0.941
Pallor	No	Yes	0.93 [0.71, 1.22]	1.00 ref	0.96 [0.60, 1.54]	0.873

Inter-/subcostal recession	No	Yes	2.05 [1.20, 3.50]	1.00 ref	0.87 [0.26, 2.92]	0.026
Inflamed pharynx/tonsils	No	Yes	0.92 [0.70, 1.20]	1.00 ref	0.96 [0.60, 1.54]	0.832
Wheeze assessed by clinician	No	Yes	1.29 [0.97, 1.72]	1.00 ref	1.11 [0.66, 1.87]	0.231
Crackles assessed by clinician	No	Yes	1.17 [0.89, 1.54]	1.00 ref	0.71 [0.41, 1.22]	0.168
Severity by parent	Quartiles 1-3	Top quartile	0.99 [0.70, 1.41]	1.00 ref	0.98 [0.53, 1.81]	0.997
Severity by clinician	Quartiles 1-3	Top quartile	1.11 [0.82, 1.50]	1.00 ref	0.87 [0.50, 1.53]	0.646

† 'Improved' and 'worse' is defined by change of ≥1 point on a 7 level Likert scale.
<sup>\$</sup> Number of children less than 1,408 due to missing data.
<sup>a</sup> Pearson's chi<sup>2</sup>.
<sup>b</sup> In the 24 hours prior to consultation.