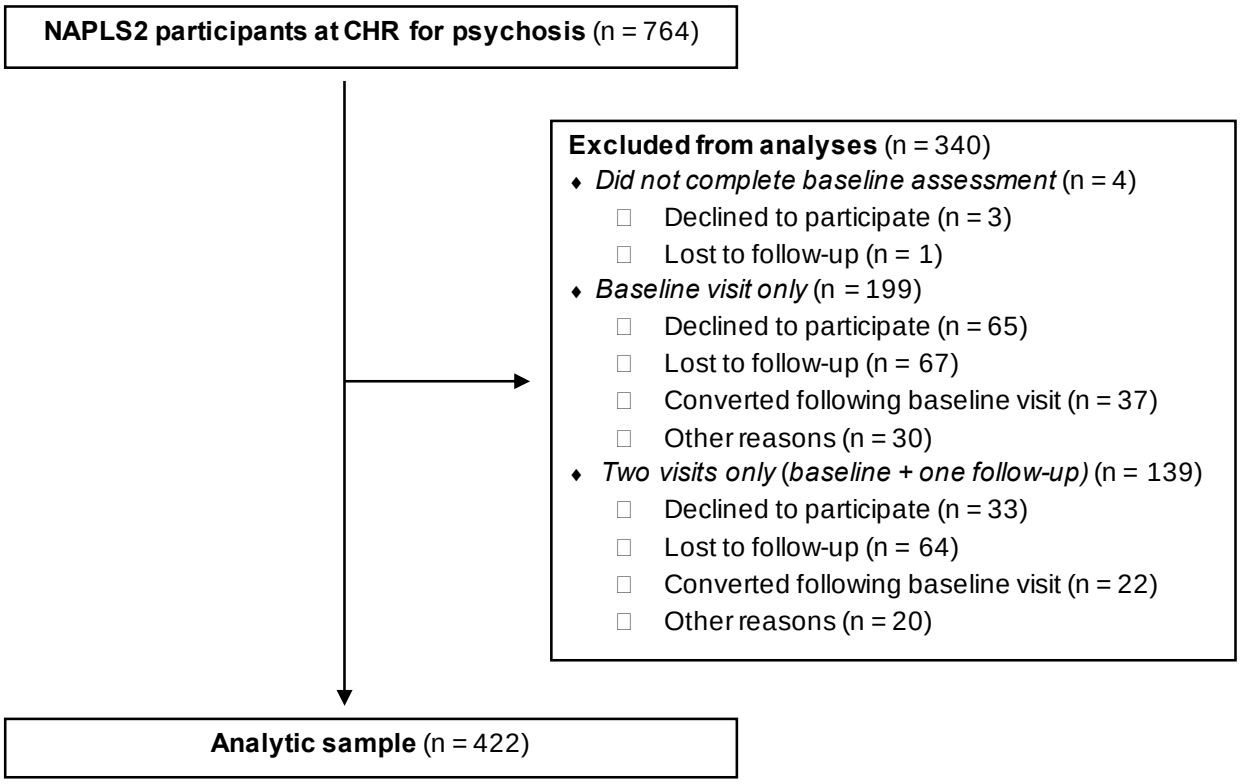
Online supplement for Allswede et al., Characterizing Covariant Trajectories of Individuals at Clinical High Risk for Psychosis Across Symptomatic and Functional Domains. Am J Psychiatry (doi: 10.1176/appi.ajp.2019.18111290)

Supplemental Information

Methods

Participants. ***NAPLS2.*** Of the initial 764 CHR participants, four individuals were excluded due to incomplete data from the baseline visit and a further 336 (44.47%) were excluded due to participation in only one or two visits. Figure S1 presents a flow chart of participant inclusion and exclusion for the present analyses and details reasons for exclusion. See (1) for more information about NAPLS2 study design and recruitment. To assess whether the analytic sample differed from the excluded sample on covariates of interest, we examined differences between these groups using chi square tests, ANOVAs, and logistic regression.

Figure S1. Participant flow chart for NAPLS2 analyses.

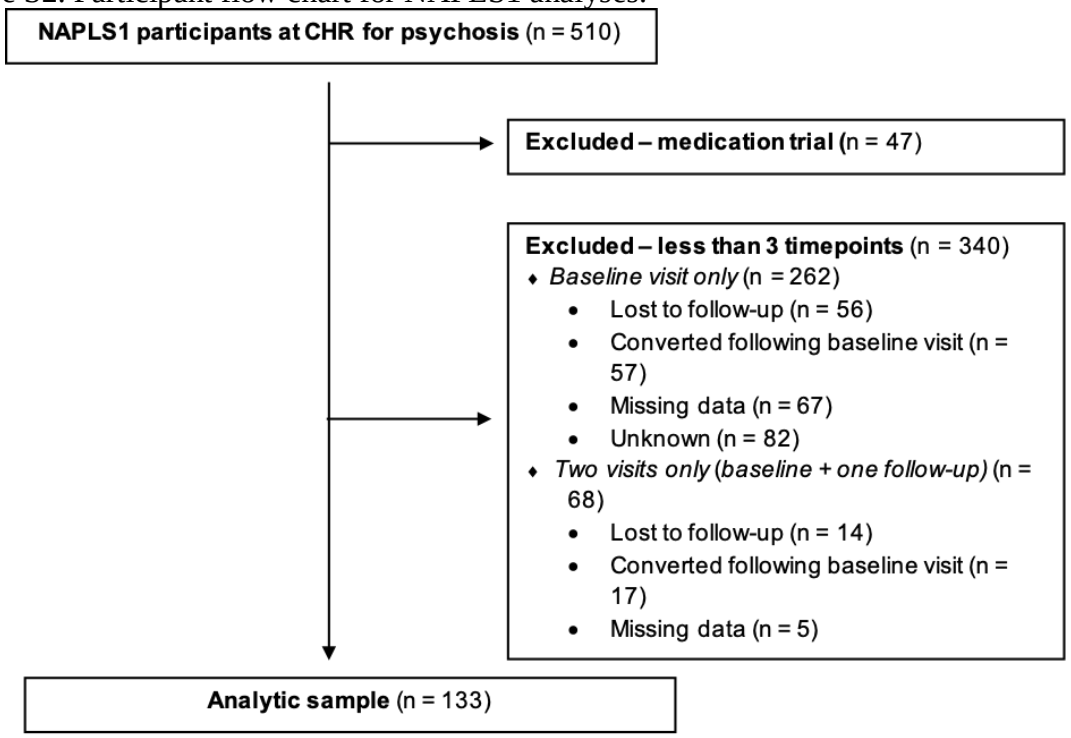


Individuals with one or two visits did not differ significantly from those with three or more visits on the following variables examined: age at baseline assessment, sex, race, highest parent education level, baseline rates of anxiety, mood, or non-cluster A disorders, or mean negative symptom severity at baseline. Individuals with one or two visits were significantly more likely to ultimately convert to psychosis (24.85%, n = 84) than individuals with three or more visits (4.50%, n = 19), X^2 (1,757) = 66.34, $p < .001$. Those with one or two visits also exhibited significantly higher baseline symptom severity sum scores on positive (M=12.33, SD=3.63), disorganized (M=5.43, SD=3.27), (F(1,757) = 4.18, p = .041, and general domains (M=9.54, SD=4.18) compared to those with three or more visits (positive: M=11.56, SD=3.92, F(1,757) = 7.71, p = .006) (disorganized: M=4.95, SD=3.06, F(1,757) = 4.18, p = .041) (general: M=8.86, SD=4.32, F(1,757) = 4.70, p = .031). Those with two or less visits also exhibited a significantly

lower mean GAF at baseline (M=47.42, SD=10.18) relative to those with three or more visits (M=49.16, SD=11.14), $F(1,757) = 4.90$, $p = .027$. When conversion, symptom severity sum scores at baseline, and GAF at baseline were assessed concurrently in a logistic regression model predicting inclusion versus exclusion in sample membership, only conversion ($X^2[1]=53.10$, $p < .001$) remained significant, suggesting that the greater proportion of converters in the excluded sample accounts for the differences observed on baseline symptom severity sum scores and GAF. The higher rates of converters among those who completed only one or two visits was expected based on prior work with this sample (2), which identified that a substantial portion of individuals at clinical high risk who ultimately convert to psychosis do so within six to twelve months of seeking treatment. We acknowledge this limitation in the Discussion section of the manuscript.

NAPLS1. Figure S2 presents a flow chart of participant inclusion and exclusion for the present analyses, and details reasons for exclusion. Of the 510 participants who attended the baseline visit, 133 (26.1%) attended at least two additional visits (prior to conversion) during the course of the study and were not simultaneously enrolled in a medication trial. As the quadratic longitudinal modeling approach utilized required at least three data points for each individual, the 330 (64.7%) participants who did not data available from at least 3 visits were excluded from the analytic sample. As no individuals in NAPLS2 were simultaneously enrolled in medication trials, NAPLS1 participants who concurrently participated in medication trials were also excluded and were assessed separately (see Assessing the Influence of Medication Trial Participants section below). See (3) for more information about NAPLS1 study design and recruitment. To assess whether the analytic sample differed from the excluded sample on covariates of interest, we examined differences between these groups using t-tests and ANOVAs.

Figure S2. Participant flow chart for NAPLS1 analyses.



Individuals with one or two visits versus three or more visits did not differ significantly on the following variables examined: age at baseline assessment, sex, race, rates of lifetime anxiety or non-cluster A personality disorders, mean GAF score at baseline, or positive or general symptom severity sum scores at baseline. Individuals with one or two visits were significantly more likely to convert to psychosis (21.8%) relative to those with two or more visits (3.8%), $\chi^2(1) = 22.30$, $p < .001$. Significantly higher symptom sum scores at baseline on negative (M=12.25, SD=6.43) and disorganized (M=6.57, SD=3.55) domains were observed for those with one or two visits relative to negative (M=10.78, SD=6.52, $F(1,354) = 4.26$, $p = .040$) and disorganized (M=5.48, SD=3.42, $F(1,354) = 8.03$, $p = .005$) sum scores for those with three or more visits. Those with one or two visits were significantly more likely to have a highest parent education level of high school or less (27.8%) relative to those with three or more visits (12.8%), $\chi^2(1) = 10.90$, $p = .001$. Additionally, individuals with one or two visits exhibited lower rates of lifetime depressive disorders (38.63%) compared to those with three or more visits (52.27%), $\chi^2(1) = 5.87$, $p = .019$.

Logistic regression models predicting one visit and accounting for conversion, parental education, race, lifetime depressive disorder, and baseline positive disorganized symptom severity sum scores simultaneously found that conversion, parental education, and lifetime depressive disorder remained significant predictors whereas positive and disorganized symptoms did not. These findings suggest that differences in rates of conversion, parental education, and lifetime depressive disorders likely account for the differences observed in positive and disorganized symptoms observed across samples. As in the NAPLS2 sample, previous work in the NAPLS1 sample aligns with the observation of higher rates of conversion within the first year after assessment (2,4). In interpreting the results of analyses in this paper and relating them to the help-seeking population more broadly, it will be important to consider that the NAPLS1 analytic sample exhibited a slightly higher parental education and higher rates of baseline depressive disorders on average relative to the full help-seeking CHR population.

Results

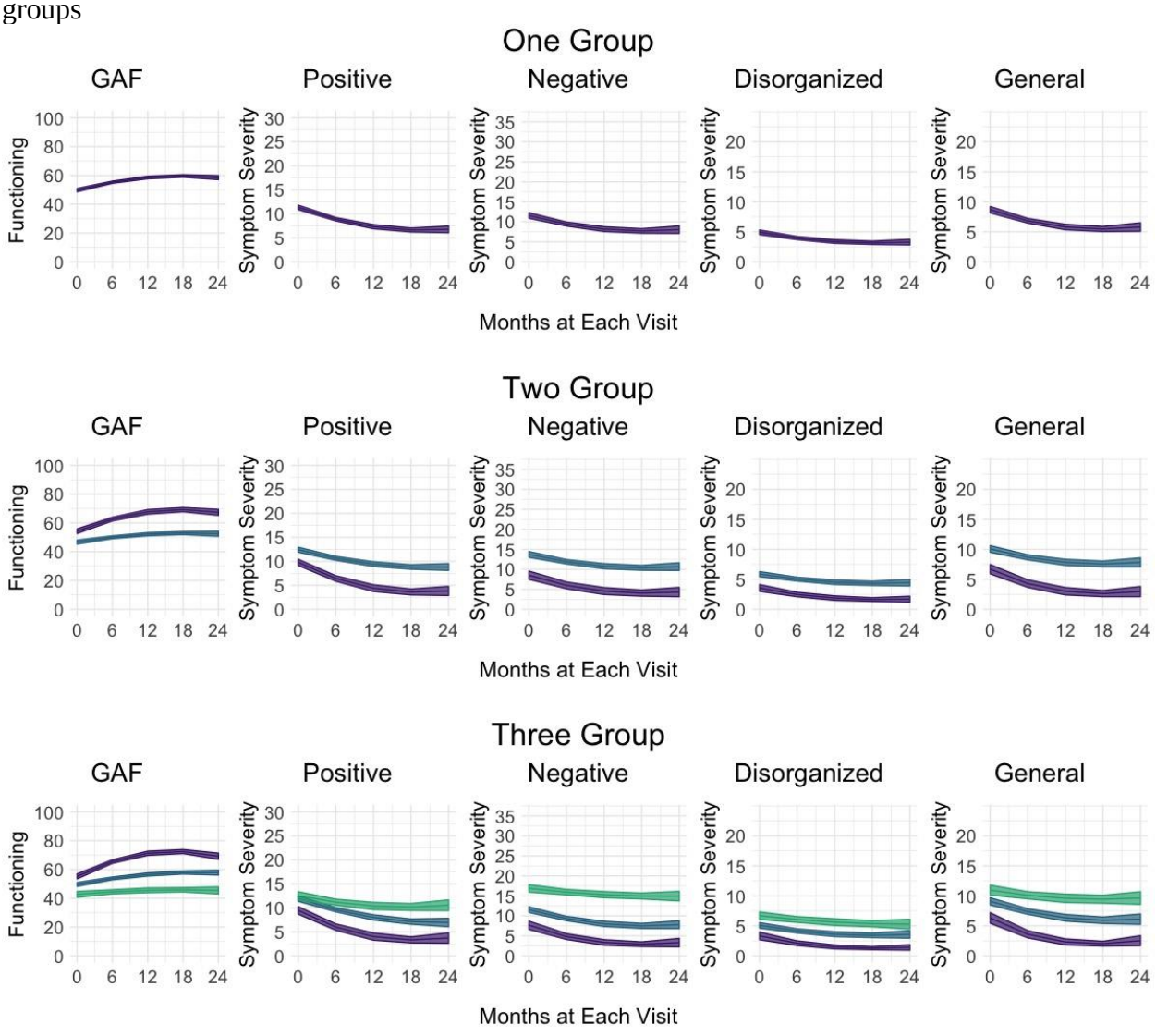
NAPLS2. ***Multivariate model search and assessment of final model.*** Univariate group-based trajectory models derived for GAF and positive, negative, disorganized, and general symptom severity suggested that three to four distinct and meaningful trajectories were likely to be represented within each domain. With this in mind, we followed the same procedure for deriving multi-group trajectory models as was followed with univariate models (Table S1, Figure S3). As measures of model adequacy continued to be fulfilled as the number of groups increased, the final model was selected based on identification of unique and theoretically meaningful combinations of trajectories that reflected those observed in the final univariate models. The three-group model was selected and is described in detail in the main manuscript. Model adequacy parameters are listed in Table S2. The quadratic, slope, and intercept parameters for each of the three trajectory groups within each functioning and symptom domain are reported and compared using Wald X^2 tests in Table S3. Table S4 compares the three trajectory groups on basic demographic information.

As the final multivariate model would not converge when dropout parameters would add to the model, the mean number of visits was compared across groups. The groups differed significantly from one another (Table S4), with the highest mean number of visits observed for Group 1 and the lowest observed for Group 3.

Table S1. NAPLS2 group-based multi-trajectory model search

Number of groups	BIC (*N*=8788)	BIC (*N*=422)	Log likelihood	Average posterior probability	Odds of correct classification	Percent assigned to group
1	-26709.98	-26679.62	-26619.17	..	..	100
2	-25752.83	-25698.18	-25589.37	.97 .96	42 19	41 59
3	-25487.52	-25408.58	-25251.41	.97 .96 .96	94 25 77	29 49 22
4	-25386.63	-25283.40	-25077.87	.93 .97 .91 .98	42 81 22 213	24 29 30 17
5	-25316.54	-25189.02	-24935.13	.89 .95 .94 .90 .94	36 74 37 46 109	19 21 31 17 12

Figure S3. NAPLS2 group-based multi-trajectory model search graphs for one to six derived groups



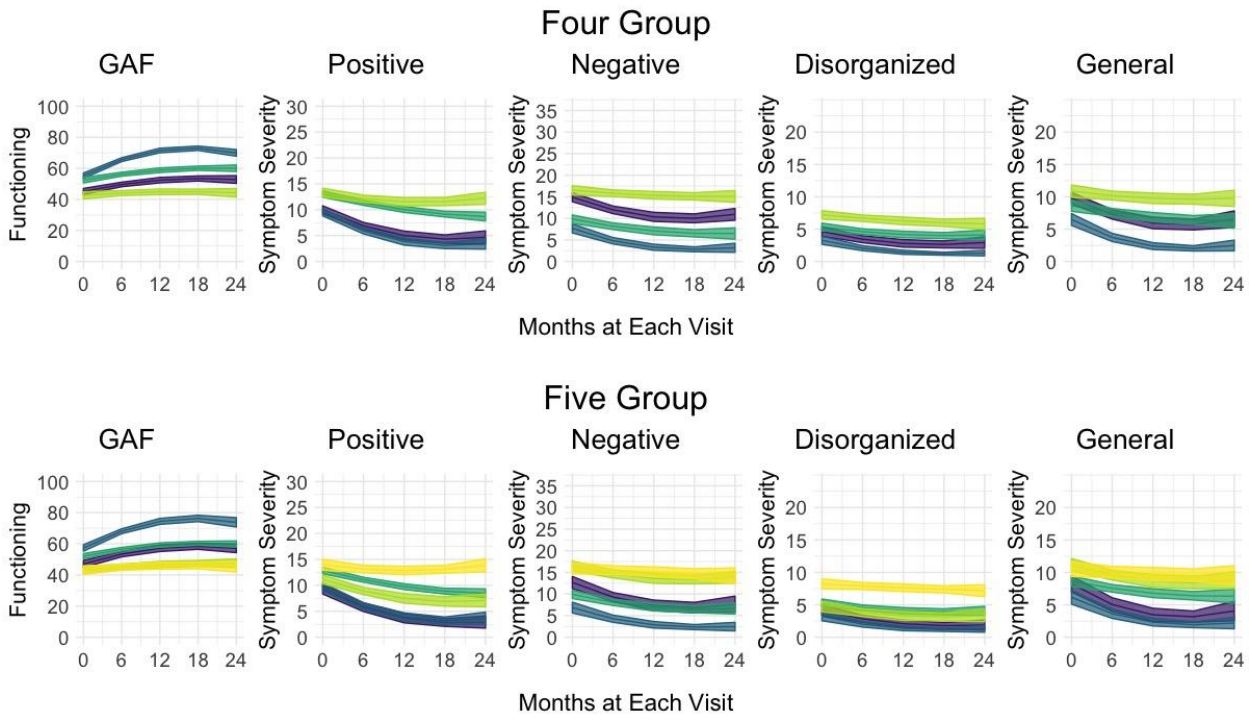


Table S2. Diagnostics of group-based model adequacy for the preferred three-group multi-trajectory model for NAPLS2.

Group	Probability of group membership	95% CI	Proportion assigned to group	Average posterior probability	Odds of correct classification*
1	.296	[.249, .343]	.294	.974	92.8
2	.489	[.436, .542]	.491	.959	24.4
3	.215	[.172, .258]	.216	.950	69.4

**improvement in odds above chance classification*
BIC (*N* = 8788): -25467.45 *BIC* (*N* = 422): -25400.66 *L*: -25267.67

Table S3. Parameters from the NAPLS2 final group-based multi-trajectory model, including Wald X^2 tests assessing differences in quadratic, linear, and intercept estimates between groups.

Group	**Parameter**	**Slope/SE**	**X^2 (2) or X^2 (1)**	**Group comparisons**
GAF				
Group 1	Quadratic	-2.26 (.27)	20.09**	1 < 2
Group 2	Quadratic	-0.72 (.21)		
Group 3	Quadratic	*NA*		

Group 1	Linear	17.03 (1.60)	26.82**	1 > 2
Group 2	Linear	6.40 (1.26)		
Group 3	Linear	*NA*		
Group 1	Intercept	40.44 (2.05)	2.02	1 = 2 = 3
Group 2	Intercept	43.77 (1.60)		
Group 3	Intercept	44.63 (.56)		
Positive				
Group 1	Quadratic	0.69 (.11)	3.74*	1 > 2 = 3
Group 2	Quadratic	0.36 (.08)		
Group 3	Quadratic	0.28 (.13)		
Group 1	Linear	-5.67 (.65)	13.53**	1 > 2 = 3
Group 2	Linear	-3.44 (.50)		
Group 3	Linear	-2.13 (.77)		
Group 1	Intercept	14.46 (.82)	.39	1 = 2 = 3
Group 2	Intercept	15.01 (.64)		
Group 3	Intercept	14.41 (.98)		
Negative				
Group 1	Quadratic	0.61 (.12)	1.27	1 = 2
Group 2	Quadratic	0.44 (.10)		
Group 3	Quadratic	*NA*		
Group 1	Linear	-4.95 (.74)	2.04	1 = 2
Group 2	Linear	-3.60 (.57)		
Group 3	Linear	*NA*		
Group 1	Intercept	11.88 (.94)	16.30**	1 < 2 = 3
Group 2	Intercept	14.87 (.74)		
Group 3	Intercept	15.76 (.28)		
Disorganized				
Group 1	Quadratic	0.29 (.08)	.97	1 = 2
Group 2	Quadratic	0.19 (.06)		
Group 3	Quadratic	*NA*		
Group 1	Linear	-2.38 (.45)	1.94	1 = 2

Group 2	Linear	-1.58 (.34)		
Group 3	Linear	*NA*		
Group 1	Intercept	5.26 (.57)	2.91	1 = 2 = 3
Group 2	Intercept	6.48 (.44)		
Group 3	Intercept	5.91 (.18)		
General				
Group 1	Quadratic	0.64 (.11)	5.20*	1 > 2
Group 2	Quadratic	0.33 (.08)		
Group 3	Quadratic	*NA*		
Group 1	Linear	-5.00 (.64)	7.33**	1 < 2
Group 2	Linear	-2.78 (.49)		
Group 3	Linear	*NA*		
Group 1	Intercept	10.57 (.81)	5.64	1 = 2 = 3
Group 2	Intercept	11.62 (.63)		
Group 3	Intercept	10.06 (.22)		

***p*<.001 **p* < .05

Table S4. Demographic differences by group assignment for NAPLS2.

Group	1 (*n* = 124)	2 (*n* = 207)	3 (*n* = 91)	F/χ^2	*p*
Age (M/SD)	18.50 (4.19)	18.25 (4.35)	19.24 (4.38)	1.69	.186
Sex (% male)	61.29%	56.04%	63.74%	1.86	.395
Parent Ed (completed HS or less)	25.20%	11.27%	20.88%	11.27	.004
Race (%)				4.55	.337
White	58.87%	54.11%	56.04%		
African-American	11.29%	20.29%	17.58%		
Other	29.84%	25.60%	26.37%		
Visits (M/SD)	4.24 (.80)	4.20 (.86)	4.02 (.86)	2.01	.135

NAPLS2. *Outcomes statistics.* Rates of outcomes for positive symptom severity and functional impairment for the full sample are listed in Table S5 (see Methods in main manuscript for definitions of each outcome). To ensure sufficient cell sizes for comparison of outcomes

across the identified trajectory groups, outcomes were grouped as "favorable" (i.e., remission, recovery) or "unfavorable" (i.e., persistent, recurrent, relapse). Statistics assessing differences in the rates of favorable outcomes on positive symptom severity, functioning, and both criteria are listed in Table S6 below.

Table S5. Rates of favorable outcomes on positive symptom severity and functioning in the NAPLS2 sample.

NAPLS2 (*N* = 422)						
	Persistent	*Remission*	*Recurrent*	*Recovery*	*Relapse*	***Favorable Outcomes***
Positive Symptom Severity	51.90% (*n* = 219)	9.95% (*n* = 42)	8.29% (*n* =35)	27.01% (*n* = 114)	2.84% (*n* = 12)	36.95% (*n* = 156)
GAF	46.45% (*n* =196)	11.37% (*n* = 48)	6.16% (*n* = 26)	26.54% (*n* = 112)	9.48% (*n* = 40)	37.91% (*n* = 160)

Table S6. Comparison of rates of favorable outcomes on positive symptoms and level of functioning by group using Pearson's X^2 tests.

NAPLS2			
Positive		64.93**	1 > 2 > 3
Group 1	64.23%		
Group 2	31.55%		
Group 3	12.90%		
GAF		110.320**	1 > 2 > 3
Group 1	73.17%		
Group 2	31.55%		
Group 3	5.38%		
Both		80.840**	1 > 2 > 3
Group 1	47.15%		
Group 2	13.59%		
Group 3	1.08%		

**p<.001 *p < .05

Alternative outcomes criteria. We used a GAF threshold of 61 for our main favorable outcomes analyses to promote consistency with the CHR literature (5–7). However, we also calculated

rates of favorable outcomes with a more stringent GAF threshold of 71 ("a person with both mild symptoms and slight impairment in social, work, and school functioning") as a secondary approach to examining improvements in functioning. With the more stringent GAF threshold, rates of favorable functional outcomes across the sample dropped to 17% (Table S7). Within groups, rates of favorable functional outcomes in Group 1 were still relatively high (51%) but drop dramatically for Group 2 (12%) and Group 3 (0%) (Figure S4, Table S8). This approach suggests that Group 1 particularly differs from Groups 2 and 3 on rates of functional outcomes when functional thresholds are increased.

Table S7. Rates of favorable outcomes on positive symptom severity and functioning (GAF threshold of 71) in the NAPLS2 sample.

NAPLS2 (*N* = 422)						
	Persistent	*Remission*	*Recurrent*	*Recovery*	*Relapse*	***Favorable Outcomes***
Positive Symptom Severity	51.90% (*n* = 219)	9.95% (*n* = 42)	8.29% (*n* = 35)	27.01% (*n* = 114)	2.84% (*n* = 12)	36.95% (*n* = 156)
GAF	72.75% (*n* =307)	6.40% (*n* = 27)	6.40% (*n* = 27)	10.90% (*n* = 46)	3.55% (*n* = 15)	17.30% (*n* = 73)

Figure S4. Percent of individuals in NAPLS2 who exhibited favorable outcomes (i.e., remitted or recovered) on positive symptom severity, functional impairment (GAF threshold of 71), or both within each group.

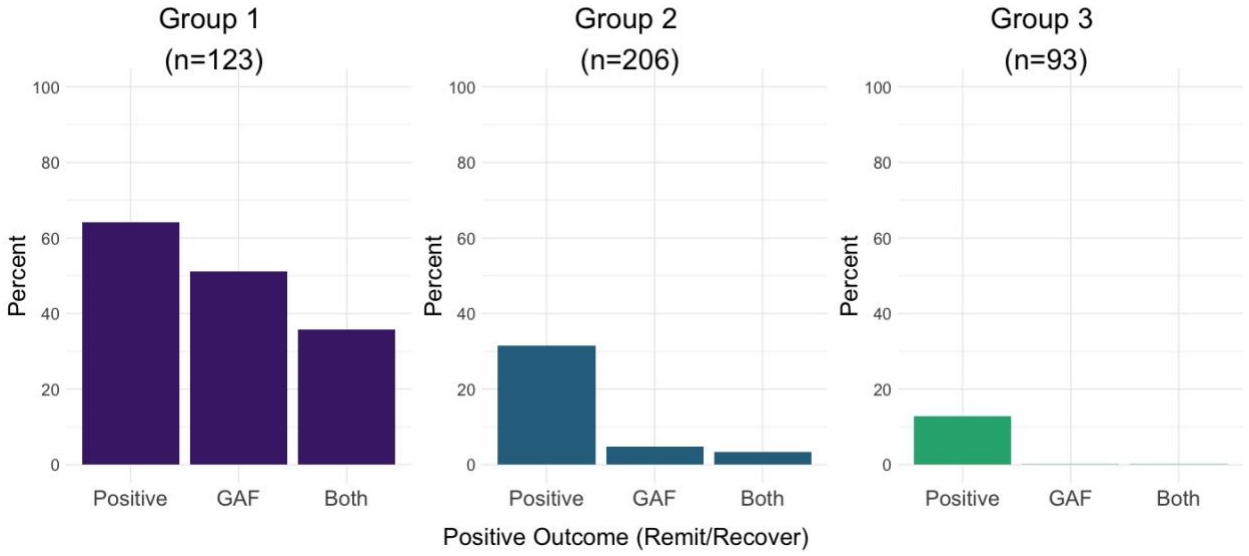


Table S8. Comparison of rates of favorable outcomes on positive symptoms and level of functioning (GAF threshold of 71) by group in NAPLS2 using Pearson's X^2 tests.

NAPLS2			
Positive		64.93**	1 > 2 > 3
Group 1	64.23%		
Group 2	31.55%		
Group 3	12.90%		
GAF		140.68**	1 > 2 > 3
Group 1	51.22%		
Group 2	4.85%		
Group 3	0%		
Both		92.37**	1 > 2 = 3
Group 1	35.77%		
Group 2	3.40%		
Group 3	0%		

**p<.001 *p < .05

Assessing Influence of Converters

To maximize the generalizability of the models for use with baseline data at prodromal clinics, all individuals who participated in at least three visits prior to conversion to psychosis were included in analyses. As only about ~5% of the analytic sample was comprised of individuals who ultimately converted (n = 19), we did not expect the inclusion of these individuals to influence the derived trajectory groups. To assess the potential influence of converters on model results, we characterized converters relative to non-converters within the analytic sample using demographic and clinical characteristics (Table S9). Converters attended significantly fewer visits than non-converters, but otherwise did not differ significantly from non-converters on examined variables.

Table S9. Demographic, clinical, and study participation differences between members of the NAPLS2 analytic sample who converted to psychosis during the course of the study versus those who did not convert to psychosis.

	Converters (n = 19)	Non-converters (n = 403)	$F/t/\chi^2$	p
Demographics				
Age (years)	17.35 (3.79)	18.59 (4.34)	1.490	.224
Sex (% male)	63.16%	59.06%	.126	.722

Parent Education (completed HS or less)	21.05%	17.29%	.178	.673
Race			.211	.900
White	52.63%	56.08%		
African-American	15.79%	17.12%		
Other	31.58%	26.80%		
Baseline symptom severity				
GAF	47.79 (13.88)	49.22 (11.01)	.300	.584
Positive Sum	11.11 (4.27)	11.58 (3.92)	.270	.607
Negative Sum	11.37 (6.80)	11.82 (5.98)	.100	.748
Disorganized Sum	5.37 (3.62)	4.94 (3.04)	.360	.548
General Sum	8.47 (4.43)	8.88 (4.32)	.160	.689
Study Completion				
Mean number of visits	3.58 (.77)	4.20 (.84)	10.18	.002

We then re-ran final univariate models with converters excluded and compared trajectory shapes and model adequacy parameters to the original final model for each variable. Excluding converters from the final NAPLS2 multivariate model did not lead to appreciable change in final model estimates or adequacy parameters. Estimated trajectories appear the same across models that include and exclude converters (Figure S5), and group size estimates changed by less than 0.5%. Model adequacy measures remained similar (Table S10). Thus, the inclusion of converters (n = 19) is not expected to have an appreciable effect on the NAPLS2 multi-trajectory model parameters.

Figure S5. Individual trajectories and group trajectory estimates for functioning (GAF) and symptom severity across SOPS domains for the three derived groups in NAPLS2 (distinguished by color), excluding converters (n = 19).

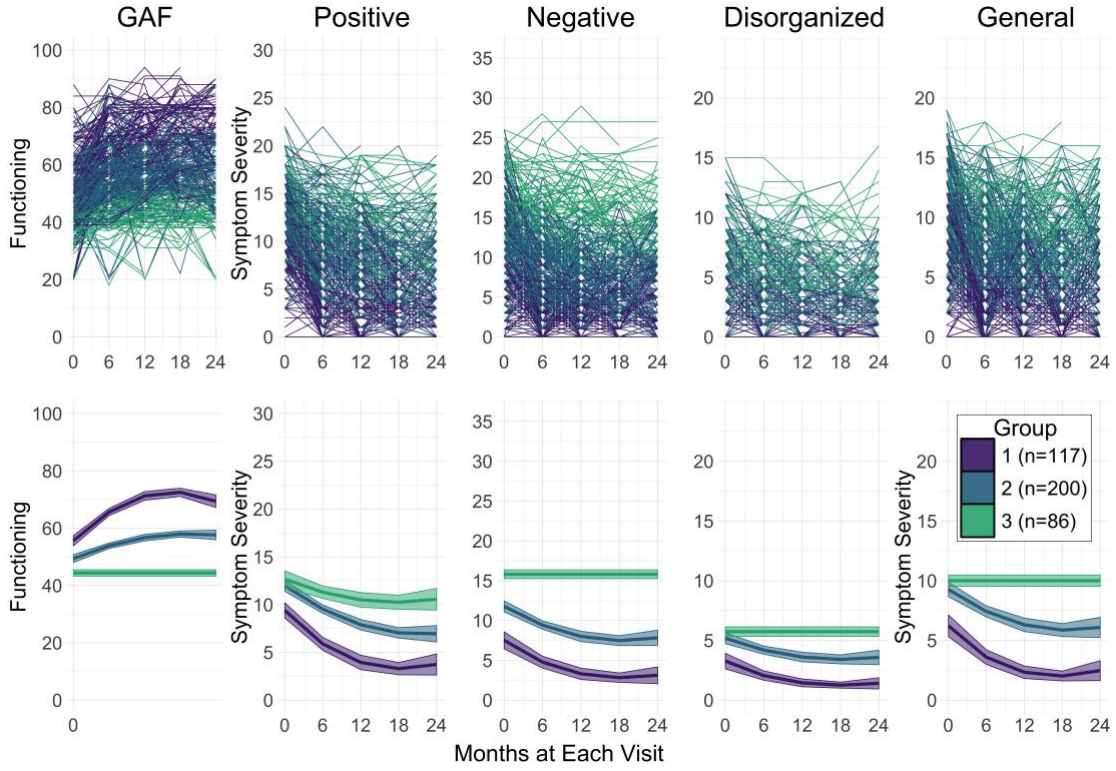


Table S10. Diagnostics of group-based model adequacy for the preferred three-group multi-trajectory model for NAPLS2, with converters excluded (n = 403).

Class	Probability of group membership	95% CI	Proportion assigned to group	Average posterior probability	Odds of correct classification*
1	.293	[.246, .340]	.290	.976	98.1
2	.494	[.441, .547]	.496	.962	25.9
3	.213	[.170, .256]	.213	.954	76.6

NAPLS1. ***Assessment of multivariate model.*** Model adequacy parameters of the three-group multivariate model are listed in Table S11. The quadratic, linear, and intercept parameters for each of the three trajectory groups within each functioning and symptom domain are reported and compared using Wald X^2 tests in Table S12. Table S13 compares the three trajectory groups on basic demographic information.

As the final multivariate model would not converge when dropout parameters would add to the model, the mean number of visits was compared across groups. The groups did not differ significantly, suggesting that there are no differences between the groups in the extent to which final estimates may be inflated due to dropout rates. Based on assessments of dropout rates

across groups in univariate models (see GRoLTs NAPLS1 section below), we expect that the estimated size of Groups 1 and 2 may be approximately 3-5% inflated and deflated, respectively, which we note in our interpretation of the model in the manuscript.

Table S11. Diagnostics of group-based model adequacy for the preferred three-group multi-trajectory model for NAPLS1.

Class	Probability of group membership	95% CI	Proportion assigned to group	Average posterior probability	Odds of correct classification*
1	.285	[.195, .375]	.271	.968	75.9
2	.513	[.417, .609]	.526	.945	16.3
3	.202	[.129, .275]	.203	.967	115.8

**improvement in odds above chance classification*
BIC (*N* = 2393): -6938.57 *BIC* (*N* = 133): -6877.88 *L*: -6775.18

Table S12. Parameters from the NAPL1 final group-based multi-trajectory model, including Wald X^2 tests assessing differences in quadratic, linear, and intercept estimates between groups.

Group	**Parameter**	**Slope/SE**	**X^2 (2) or X^2 (1)**	**Group comparisons**
GAF				
Group 1	Quadratic	-2.00 (.52)	1.66	1 = 2
Group 2	Quadratic	-1.10 (.42)		
Group 3	Quadratic	*NA*		
Group 1	Linear	15.55 (3.01)	2.77	1 = 2
Group 2	Linear	8.95 (2.37)		
Group 3	Linear	*NA*		
Group 1	Intercept	46.58 (3.72)	1.25	1 = 2 = 3
Group 2	Intercept	41.23 (2.90)		
Group 3	Intercept	42.93 (1.02)		
Positive				
Group 1	Quadratic	.94 (.21)	.80	1 = 2
Group 2	Quadratic	.70 (.16)		
Group 3	Quadratic	*NA*		
Group 1	Linear	-7.47 (1.24)	1.12	1 = 2
Group 2	Linear	-5.79 (.91)		

Group 3	Linear	*NA*		
Group 1	Intercept	15.89 (1.51)	46.61**	1 = 2 > 3
Group 2	Intercept	15.82 (1.11)		
Group 3	Intercept	9.14 (.43)		
Negative				
Group 1	Quadratic	.60 (.27)	.12	1 = 2
Group 2	Quadratic	.72 (.21)		
Group 3	Quadratic	*NA*		
Group 1	Linear	-5.25 (1.56)	.001	1 = 2
Group 2	Linear	-5.33 (1.22)		
Group 3	Linear	*NA*		
Group 1	Intercept	10.51 (1.96)	12.83*	1 = 2 > 3
Group 2	Intercept	14.75 (1.49)		
Group 3	Intercept	17.26 (.56)		
Disorganized				
Group 1	Quadratic	.42 (.17)	.01	1 = 2
Group 2	Quadratic	.39 (.14)		
Group 3	Quadratic	*NA*		
Group 1	Linear	-3.19 (1.00)	.03	1 = 2
Group 2	Linear	-2.98 (.77)		
Group 3	Linear	*NA*		
Group 1	Intercept	5.88 (1.26)	1.95	1 = 2 = 3
Group 2	Intercept	8.01 (.94)		
Group 3	Intercept	7.60 (.36)		
General				
Group 1	Quadratic	.80 (.23)	.54	1 = 2
Group 2	Quadratic	.58 (.19)		
Group 3	Quadratic	*NA*		
Group 1	Linear	-5.78 (1.34)	.46	1 = 2
Group 2	Linear	-4.62 (1.05)		
Group 3	Linear	*NA*		

Group 1	Intercept	10.08 (1.64)	7.18*	1 = 2 > 3
Group 2	Intercept	12.14 (1.24)		
Group 3	Intercept	8.70 (.45)		

**p<.001 *p < .05

Table S13. Demographic differences by group assignment for NAPLS1.

Group	1 (n = 36)	2 (n = 70)	3 (n = 27)	F/χ^2	p
Age (M/SD)	20.01 (5.77)	17.54 (4.13)	16.55 (2.85)	5.51	.005
Sex (% male)	61.11%	57.14%	74.07%	2.37	.306
Parent Ed (completed HS or less)	9.38%	11.76%	20.00%	1.56	.458
Race (%)				4.07	.396
White	86.11%	77.14%	74.04%		
African-American	8.33%	7.14%	3.70%		
Other	5.56%	15.71%	22.22%		
Visits (M/SD)	4.00 (.89)	3.94 (.80)	3.70 (.78)	1.13	.327

NAPLS1. *Outcomes statistics.* Rates of outcomes for positive symptom severity and functional impairment for the full sample are listed in Table S14 (see Methods in main manuscript for definitions of each outcome). To ensure sufficient cell sizes for comparison of outcomes across the identified trajectory groups, outcomes were grouped as "favorable" (i.e., remission, recovery) or "unfavorable" (i.e., persistent, recurrent, relapse). Statistics assessing differences in the rates of favorable outcomes on positive symptom severity, functioning, and both criteria are listed in Table S15 below.

Table S14. Rates of favorable outcomes on positive symptom severity and functioning in the NAPLS1 sample.

NAPLS1 (N = 133)						
	Persistent	*Remission*	*Recurrent*	*Recovery*	*Relapse*	***Favorable Outcomes***
Positive Symptom Severity	30.83% (n = 41)	12.78% (n = 17)	9.02% (n =12)	42.11% (n = 56)	5.26% (n = 7)	54.89% (n = 73)
GAF	44.36% (n = 59)	13.53% (n = 18)	9.02% (n = 12)	29.32% (n = 39)	3.76% (n = 5)	42.86% (n = 57)

Table S15. Comparison of rates of favorable outcomes on positive symptoms and level of functioning by group in NAPLS1 using Pearson's X^2 tests.

NAPLS1			
Positive		15.73**	1 > 2 > 3
Group 1	77.14%		
Group 2	55.07%		
Group 3	27.97%		
GAF		41.65**	1 > 2 > 3
Group 1	82.86%		
Group 2	39.13%		
Group 3	3.45%		
Both		29.96**	1 > 2 > 3
Group 1	60.00%		
Group 2	23.19%		
Group 3	0%		

**p<.001 *p < .05

Alternative outcomes criteria. As in NAPLS2, we used a GAF threshold of 61 for our main favorable outcomes analyses to promote consistency with the clinical high-risk literature (5–7). However, we also calculated rates of favorable outcomes with a more stringent GAF threshold of 71 ("a person with both mild symptoms and slight impairment in social, work, and school functioning") as a secondary approach to examining improvements in functioning. With the more stringent GAF threshold, overall rates of positive GAF outcomes decreased to 20% across the sample (Table S16). Within groups, rates of positive functional outcomes in Group 1 were still relatively high (51%) but dropped dramatically for Group 2 (12%) and Group 3 (0%) (Figure S6, Table S17). This approach suggests that Group 1 particularly differs from Groups 2 and 3 on rates of functional outcomes when functional thresholds are increased.

Table S16. Comparison of favorable outcomes on positive symptoms and level of functioning (using GAF threshold of 71) by group in NAPLS1 using Pearson's X^2 tests.

NAPLS1 (*N* = 133)						
	Persistent	*Remission*	*Recurrent*	*Recovery*	*Relapse*	***Favorable Outcomes***
Positive Symptom Severity	30.83% (*n* = 41)	12.78% (*n* = 17)	9.02% (*n* = 12)	42.11% (*n* = 56)	5.26% (*n* = 7)	54.89% (*n* = 73)
GAF	69.17% (*n* = 92)	9.01% (*n* = 12)	6.77% (*n* = 9)	10.53% (*n* = 14)	4.51% (*n* = 6)	19.55% (*n* = 26)

Figure S6. Percent of individuals in NAPLS1 who exhibited favorable outcomes (i.e., remitted or recovered) on positive symptom severity, functional impairment (GAF threshold of 71), or both within each group.

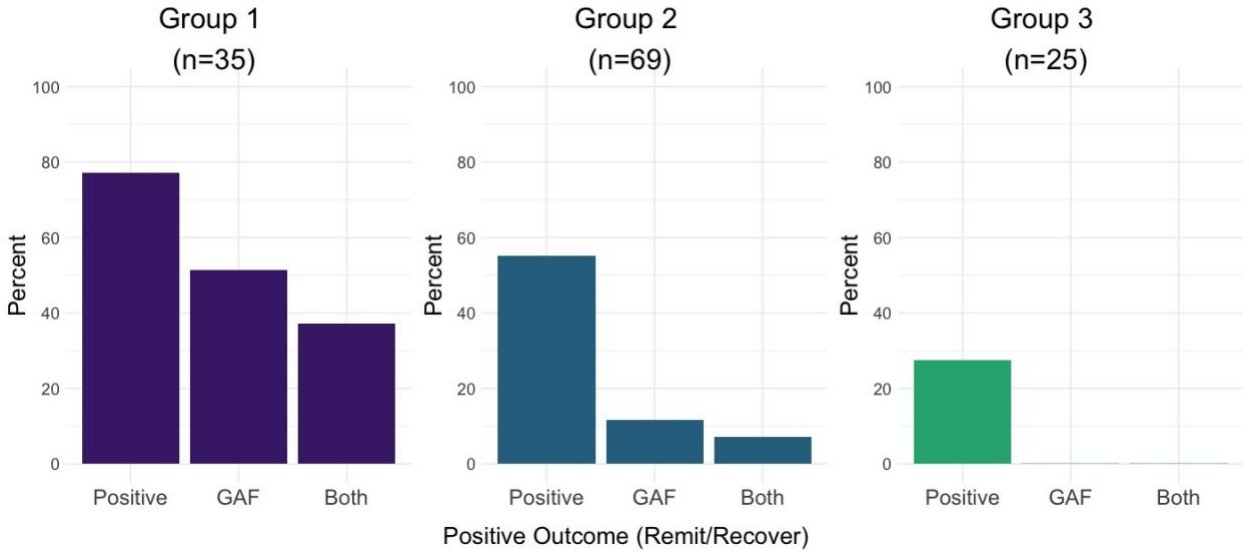


Table S17. Comparison of favorable on positive symptoms and level of functioning (GAF threshold of 71) by group in NAPLS1 using Pearson's X^2 tests.

NAPLS1			
Positive		15.73**	1 > 2 > 3
Group 1	77.14%		
Group 2	55.07%		
Group 3	27.97%		
GAF		32.44**	1 > 2 = 3
Group 1	51.43%		
Group 2	11.59%		
Group 3	0%		
Both		23.54**	1 > 2 = 3
Group 1	37.14%		
Group 2	7.25%		
Group 3	0%		

**p<.001 *p < .05

Assessing Influence of Converters

As in NAPLS2, all individuals in NAPLS1 who participated in at least three visits prior to conversion to psychosis were included in analyses. As only about ~4% of the analytic sample was comprised of individuals who ultimately converted (n = 5), we did not expect the inclusion of these individuals to influence the derived trajectory groups. To assess the potential influence of converters on model results, we characterized converters relative to non-converters within the analytic sample using demographic and clinical characteristics (Table S18). Converters exhibited significantly lower GAF scores and higher negative symptom severity scores at baseline relative to non-converters; however, due to the large discrepancy in group sizes (n = 5, n = 128), these statistics should be interpreted with caution.

Table S18. Demographic, clinical, and study participation differences between members of the NAPLS1 analytic sample who converted to psychosis during the course of the study versus those who did not convert to psychosis.

	Converters (n = 5)	Non-converters (n = 128)	$F/t/\chi^2$	p
Demographics				
Age (years)	16.80 (2.06)	18.06 (4.65)	.36	.549
Sex (% male)	100%	60.2%	3.23	.072

Parent Education (completed HS or less)	20.0%	12.5%	.24	.623
Race			4.74	.093
White	40.0%	80.5%		
African-American	20.0%	6.3%		
Other	40.0%	13.3%		
Baseline symptom severity				
GAF	37.80 (4.92)	50.36 (12.09)	5.33	.023
Positive Sum	13.80 (4.32)	10.65 (3.52)	3.80	.053
Negative Sum	20.80 (7.16)	10.39 (6.22)	12.84	<.001
Disorganized Sum	7.60 (3.71)	5.40 (3.39)	2.01	.158
General Sum	9.20 (5.40)	7.75 (4.44)	.51	.478
Study Completion				
Mean number of visits	3.40 (.55)	3.93 (.82)	2.02	.158

We then re-ran final univariate models with converters excluded and compared trajectory shapes and model adequacy parameters to the original final model for each variable. Excluding converters from the final NAPLS1 multivariate model did not lead to appreciable change in final model estimates or adequacy parameters. Estimated trajectories appear the same across models that include and exclude converters (Figure S7), and group size estimates changed by less than 2%. Model adequacy measures remained similar (Table S19). Thus, the inclusion of converters (N=5) is not expected to have an appreciable effect on the NAPLS1 multi-trajectory model parameters.

Unfortunately, we could not assess functional and symptomatic patterns of converters prior to and following conversion due to lack of data for these domains at conversion visits. Prior to conversion, most (n = 4) converters followed the chronic Group 3 trajectory, though one converter was a member of the rapidly improving Group 1 prior to conversion.

Figure S7. Individual trajectories and group trajectory estimates for functioning (GAF) and symptom severity across SOPS domains for the three derived groups in NAPLS1 (distinguished by color), excluding converters (*n* = 5).

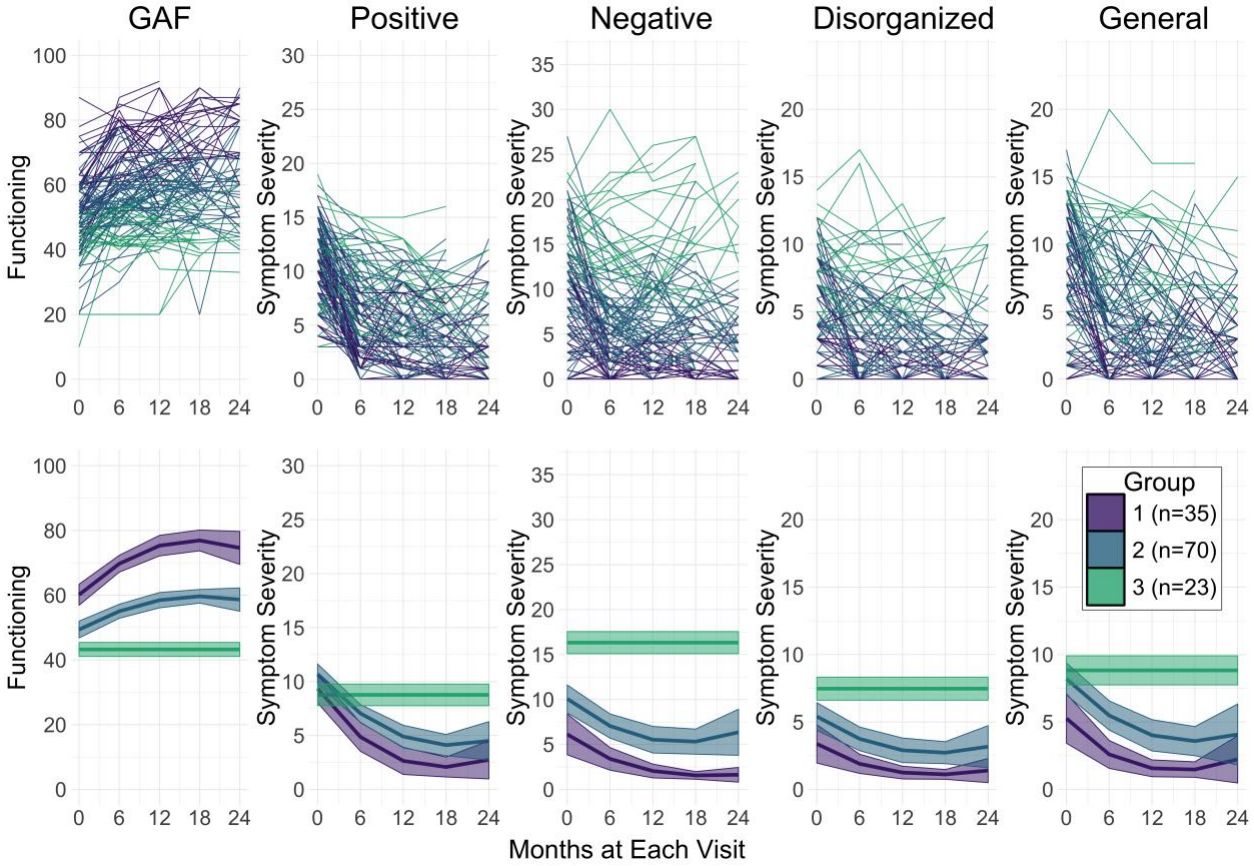


Table S19. Diagnostics of group-based model adequacy for the preferred three-group multi-trajectory model for NAPLS1, with converters excluded (*n* = 128).

Class	Probability of group membership	95% CI	Proportion assigned to group	Average posterior probability	Odds of correct classification*
1	.290	[.196, .384]	.273	.975	95.5
2	.524	[.424, .624]	.547	.936	13.3
3	.186	[.113, .259]	.180	.973	157.7

**improvement in odds above chance classification*
BIC (*N* = 2308): -6669.48 *BIC* (*N* = 128): -6608.75 *L*: -6506.85

Assessing Influence of Medication Trial Participants

In NAPLS1, 27 participants who completed at least 3 visits were simultaneously enrolled in medication trials and received either active or placebo treatment with olanzapine (*n* = 20), Abilify (*n* = 3), glycine (*n* = 2), and omega 3 (*n* = 2) (8–11). As no NAPLS2 participants were simultaneously enrolled in medication trials, we excluded individuals in NAPLS1 who were

concurrently participating in medication trials to maximize the consistency of the study population to NAPLS2 and to prodromal clinics broadly. To assess the potential influence of excluding medication trial participants on model results, we characterized differences between excluded medication trial participants with three or more visits and the analytic sample using demographic and clinical characteristics (Table S20). Medication trial participants exhibited significantly lower GAF scores and higher positive and disorganized symptom severity scores at baseline relative to the analytic sample. Medication trial participants also seemed to have higher rates of parent education of high school or less relative to the analytic sample, though these statistics should be interpreted with caution due to small cell size (e.g., 5 medication trial participants with parental education of high school or less).

Table S20. Demographic, clinical, and study participation differences between the NAPLS1 analytic sample and excluded individuals that participated in medication trials.

	Medication trial (n = 27)	Analytic sample (n = 133)	$F/t/\chi^2$	p
Demographics				
Age (years)	16.65 (5.18)	18.01 (4.58)	1.89	.171
Sex (% male)	62.96%	61.65%	.02	.898
Parent Education (completed HS or less)[#]	33.33%	12.80%	4.43	.035
Race			2.68	.261
White	85.19%	78.95%		
African-American	11.11%	6.77%		
Other	3.70%	14.29%		
Baseline symptom severity				
GAF	41.74 (10.21)	49.89 (12.13)	10.63	.001
Positive Sum	12.70 (3.43)	10.77 (3.58)	6.65	.011
Negative Sum	12.67 (7.11)	10.78 (6.52)	1.82	.180
Disorganized Sum	7.15 (5.00)	5.48 (3.42)	4.48	.036
General Sum	8.22 (5.68)	7.80 (4.46)	.18	.673
Study Completion				
Mean number of visits	4.19 (.92)	3.91 (.82)	2.42	.122

[#]20 individuals missing parent education data

To assess the influence of excluding medication trial individuals on derivation of trajectory models, we re-ran final univariate models including medication trial participants with three or more visits (n = 27) and compared trajectory shapes and model adequacy parameters to the original final model for each variable. Including medication trial participants did not lead to substantial change in final model estimates or adequacy parameters. Estimated trajectories appear the same across models that included and excluded medication trial participants (Figure S8) and group size estimates changed by less than 3.5%. Model adequacy measures remained

similar to parameters observed for the original model (Table S21). Aside from a slight increase in the size of Group 2, the exclusion of medication trial participants is not expected to have an appreciable effect on multi-trajectory models.

Figure S8. Individual trajectories and group trajectory estimates for functioning (GAF) and symptom severity across SOPS domains for the three derived groups in NAPLS1 (distinguished by color), including medication trial participations (n = 27).

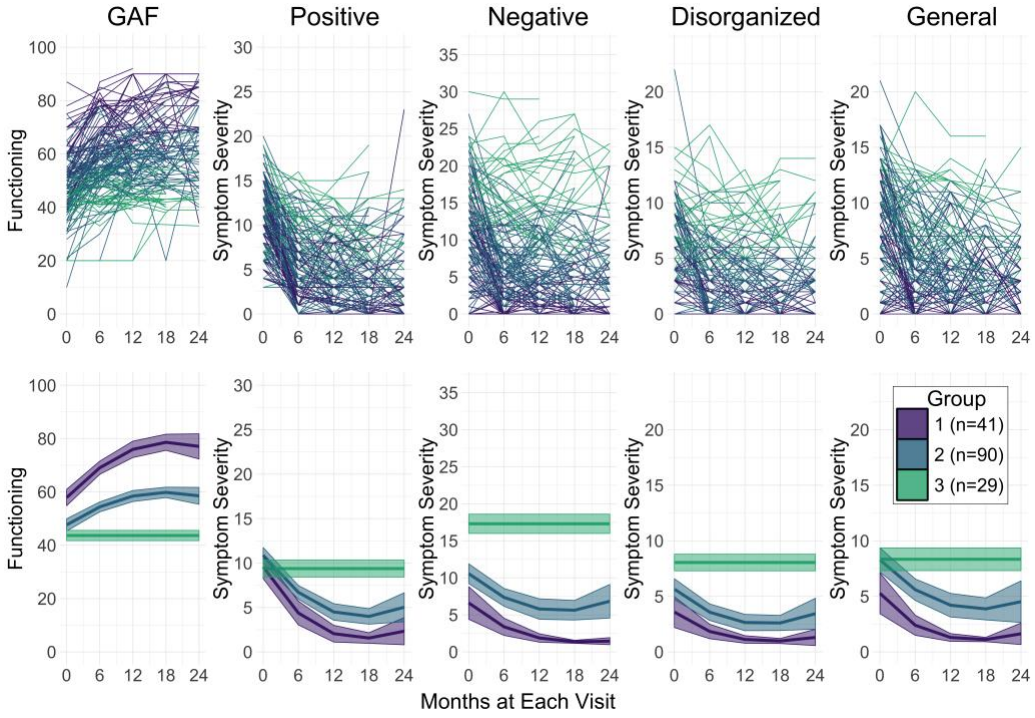


Table S21. Diagnostics of group-based model adequacy for the preferred three-group multi-trajectory model for NAPLS1, with medication trial participants included (n = 27).

Class	Probability of group membership	95% CI	Proportion assigned to group	Average posterior probability	Odds of correct classification*
1	.267	[.187, .347]	.256	.957	61.1
2	.549	[.461, .637]	.563	.947	14.7
3	.184	[.117, .251]	.181	.971	148.5

**improvement in odds above chance classification*
BIC (N = 2952): -8525.26 *BIC* (N = 160): -8464.05 *L*: -8357.47

Assessing Influence of Comorbidity

At baseline, 49.05% of the NAPLS2 sample met SCID criteria for an anxiety disorder, 42.89% qualified for a mood disorder, and 15.46% met criteria for a non-cluster A personality disorder. Similar rates of comorbidity at baseline were observed in the NAPLS1 sample, with 40.91%

meeting criteria for an anxiety disorder, 52.27% qualifying for a depressive disorder, and 28.57% meeting criteria for a non-cluster A personality disorder. Of note, the NAPLS1 baseline comorbidity rates include lifetime as well as current SCID diagnoses (4). To assess whether differences in rates of baseline and persistent affective and non-cluster A personality disorders could explain the observed trajectory groups, we compared comorbidity rates across the three groups (Table S22). Due to small cell sizes in NAPLS1, we collapsed NAPLS2 and NAPLS1 for these analyses. Across baseline and persistent disorders, we observed lower rates of comorbid disorders in Group 1 relative to Groups 2 and 3, and no differences in rates between Group 2 and Group 3. As rates of incidence of non-psychotic disorders were low in these samples (4,12), differences in emergence of comorbid disorders likely do not distinguish Group 2 and 3. These analyses suggest that comorbid affective and non-cluster A personality disorders cannot fully account for the observed differences in group trajectories and rates of positive functional and symptomatic outcomes.

Table S22. Rates of baseline and persistent anxiety, depressive, and non-cluster A personality disorder by group membership in NAPLS2 and NAPLS1 (combined).

Combined samples	*Group 1 (n=160)*	*Group 2 (n=277)*	*Group 3 (n=117)*	X^2	*Group Comparison*
BL anxiety disorder	31.88%	50.54%	59.83%	23.81**	1 < 2 = 3
BL depressive disorder	30.00%	50.90%	52.14%	20.84**	1 < 2 = 3
BL Non-Cluster A personality disorder[#]	11.11%	19.69%	25.47%	9.25*	1 < 2 = 3
Persistent anxiety disorder[#]	23.87%	52.42%	56.36%	39.68**	1 < 2 = 3
Persistent depressive disorder[#]	16.77%	28.25%	35.45%	12.51*	1 < 2 = 3
Persistent Non-Cluster A[#]	4.73%	15.45%	19.00%	13.47*	1 < 2 = 3

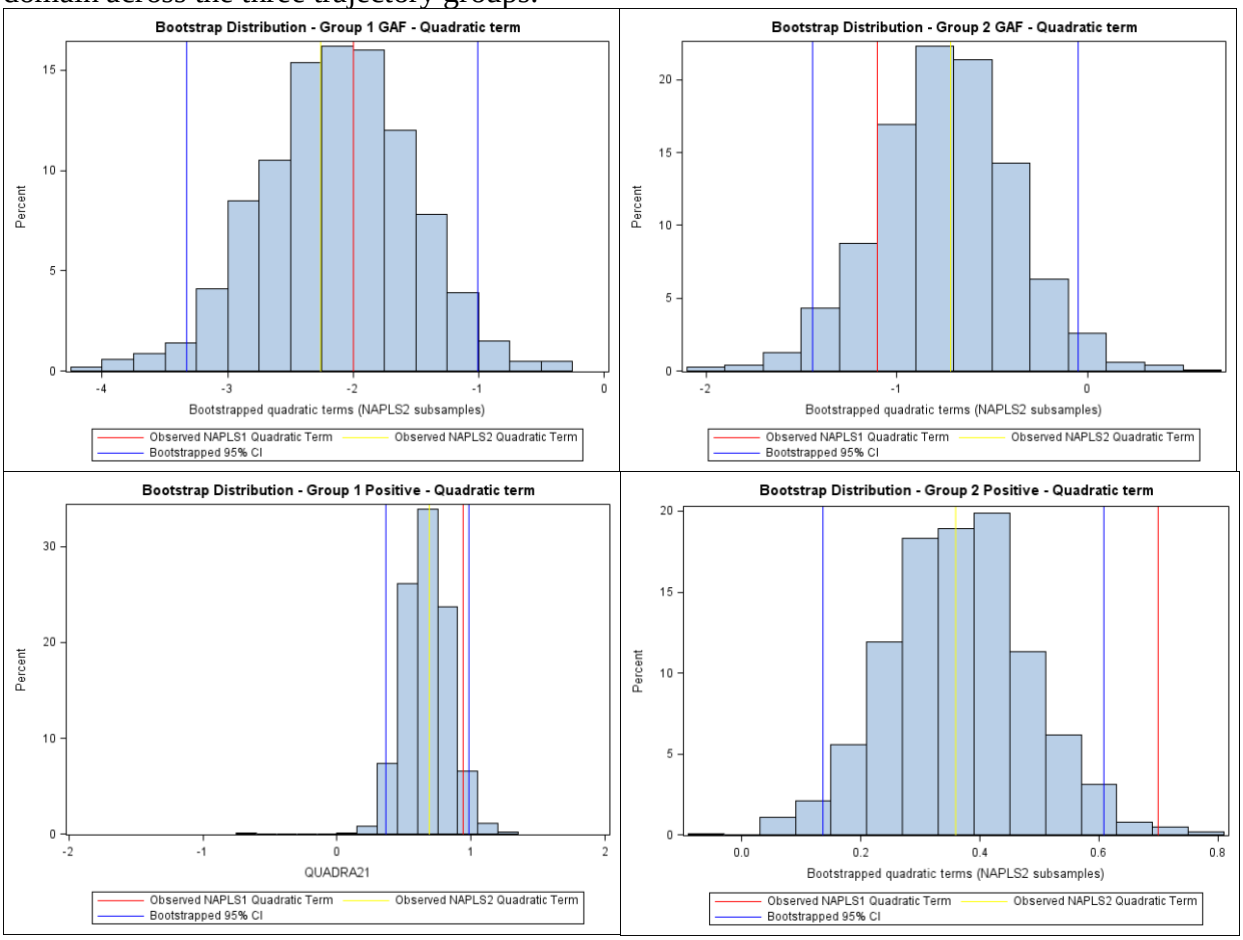
*****p <.05; **p < .001

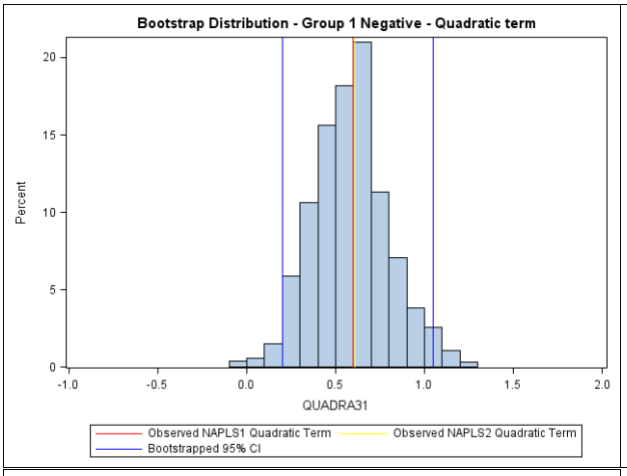
[#]42 individuals missing personality disorder data at baseline, 21 individuals missing follow-up anxiety and depressive disorder data

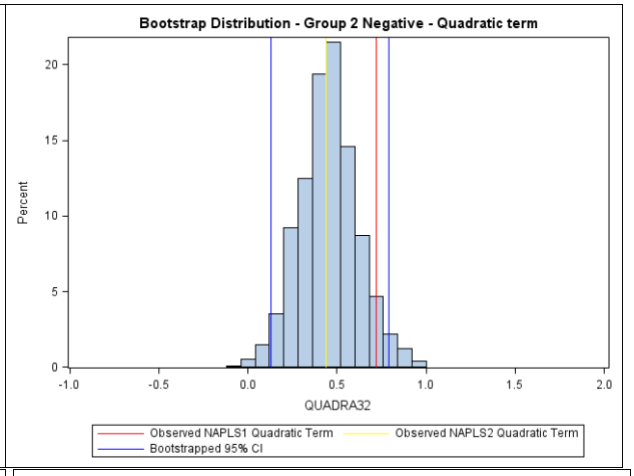
Comparison of NAPLS2 and NAPLS1. ***Bootstrapped confidence intervals for parameter estimates.*** We assessed the replicability of parameter estimates across NAPLS2 and NAPLS1 using bootstrapped confidence intervals for parameter estimates. We derived 1000 samples of 133 participants (sampled with replacement) using the NAPLS2 sample, which corresponded in size to the NAPLS1 sample (N = 133). Models were run in each subsample, and parameter estimates were plotted for each domain by group (Figures S9-S11). 95% confidence intervals for

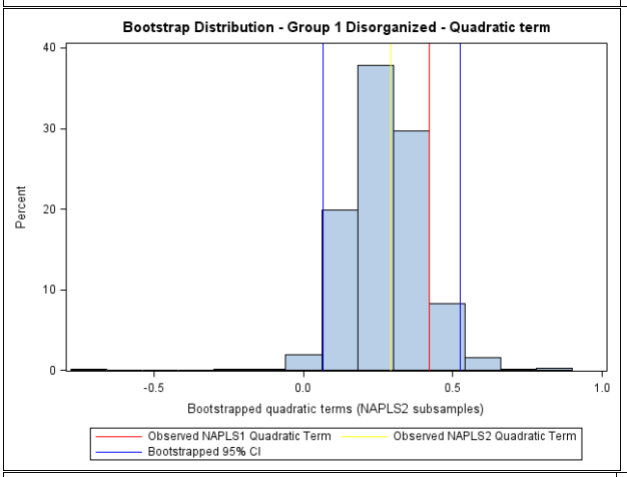
the true NAPLS2 parameter were identified using the distribution, and the NAPLS1 estimate was plotted to assess if it fell within this confidence interval. If the NAPLS1 parameter fell within the confidence interval, parameters were not considered significantly different across the NAPLS2 and NAPLS1 sample. In contrast, parameters were considered to be significantly different across samples if the NAPLS1 estimate fell outside of the bootstrapped 95% confidence interval for the NAPSL2 parameter. Most NAPLS1 parameters fell within the NAPLS2 bootstrapped 95% confidence interval, indicating that the parameters could not be considered significantly different. Exceptions included quadratic and slope terms for Group 2 Positive, Disorganized, and General symptoms (Figures S9, S10) and intercepts for Group 2 Positive (Figure S11).

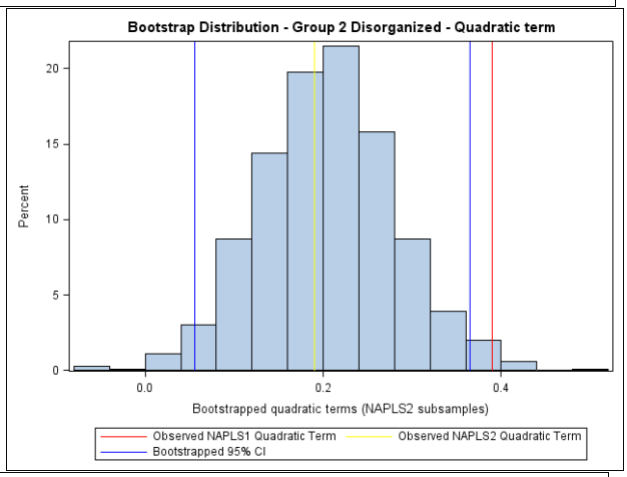
Figure S9. Bootstrapped distributions for quadratic estimates for each symptom and functioning domain across the three trajectory groups.

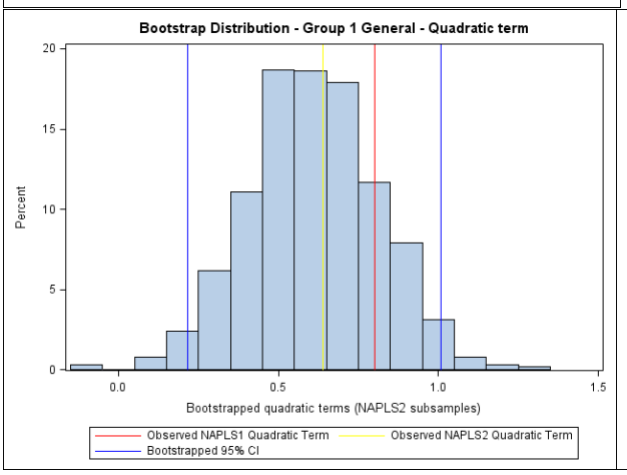



Bootstrap Distribution - Group 1 Negative - Quadratic term
Percent
QUADRA31
Observed NAPLS1 Quadratic Term
Observed NAPLS2 Quadratic Term
Bootstrapped 95% CI


Bootstrap Distribution - Group 2 Negative - Quadratic term
Percent
QUADRA32
Observed NAPLS1 Quadratic Term
Observed NAPLS2 Quadratic Term
Bootstrapped 95% CI


Bootstrap Distribution - Group 1 Disorganized - Quadratic term
Percent
Bootstrapped quadratic terms (NAPLS2 subsamples)
Observed NAPLS1 Quadratic Term
Observed NAPLS2 Quadratic Term
Bootstrapped 95% CI


Bootstrap Distribution - Group 2 Disorganized - Quadratic term
Percent
Bootstrapped quadratic terms (NAPLS2 subsamples)
Observed NAPLS1 Quadratic Term
Observed NAPLS2 Quadratic Term
Bootstrapped 95% CI


Bootstrap Distribution - Group 1 General - Quadratic term
Percent
Bootstrapped quadratic terms (NAPLS2 subsamples)
Observed NAPLS1 Quadratic Term
Observed NAPLS2 Quadratic Term
Bootstrapped 95% CI

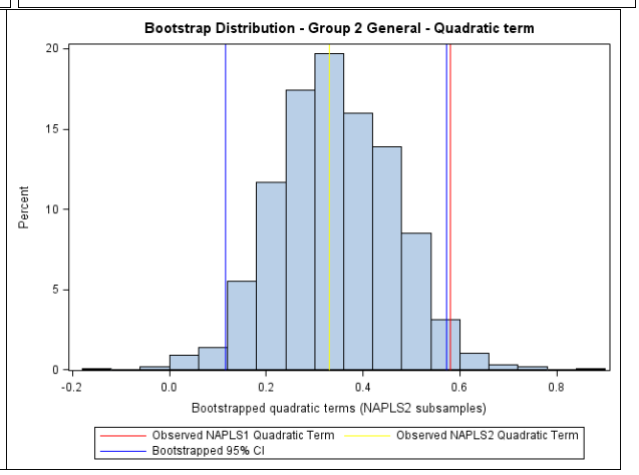
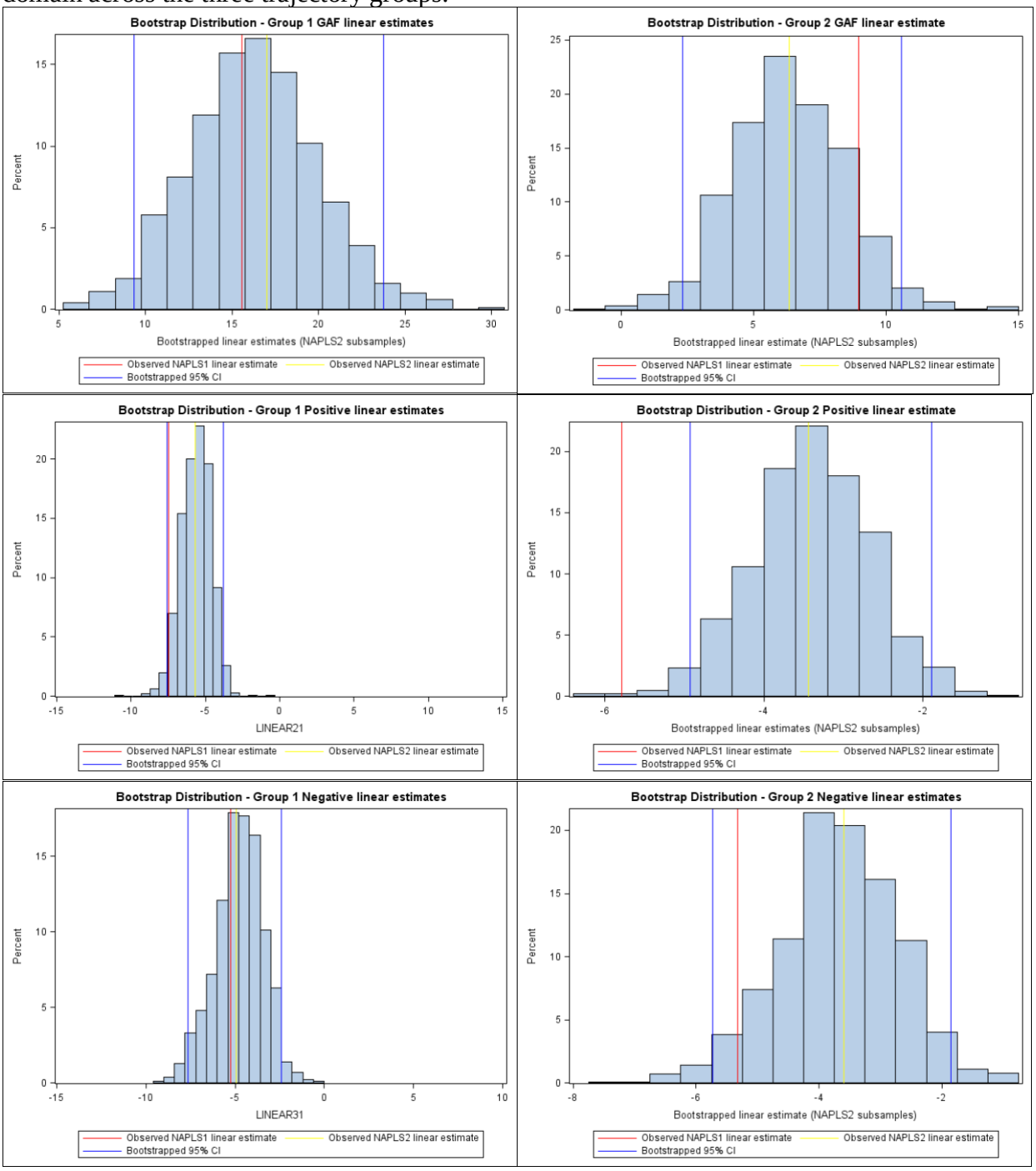

Bootstrap Distribution - Group 2 General - Quadratic term
Percent
Bootstrapped quadratic terms (NAPLS2 subsamples)
Observed NAPLS1 Quadratic Term
Observed NAPLS2 Quadratic Term
Bootstrapped 95% CI

Figure S10. Bootstrapped distributions for linear estimates for each symptom and functioning domain across the three trajectory groups.



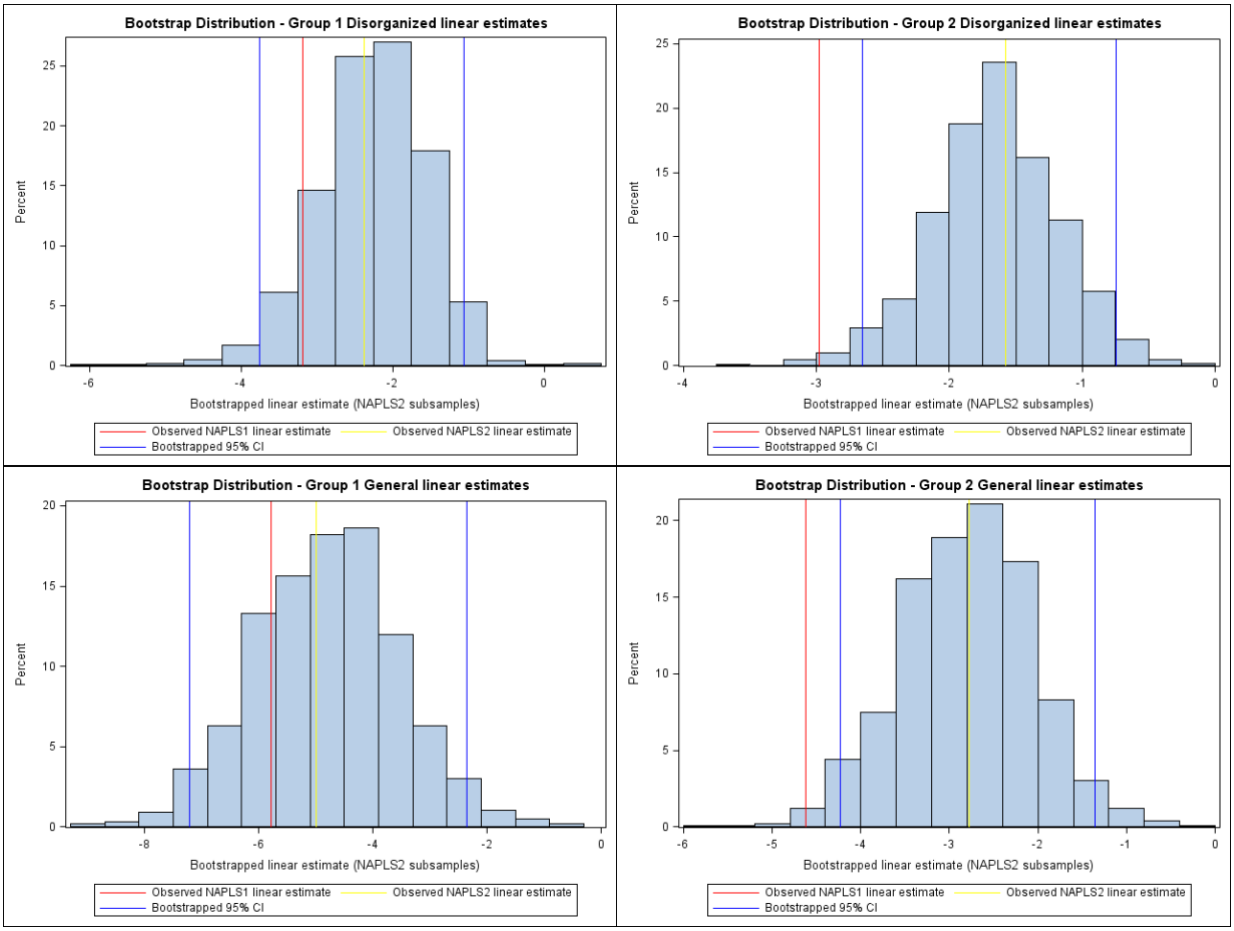
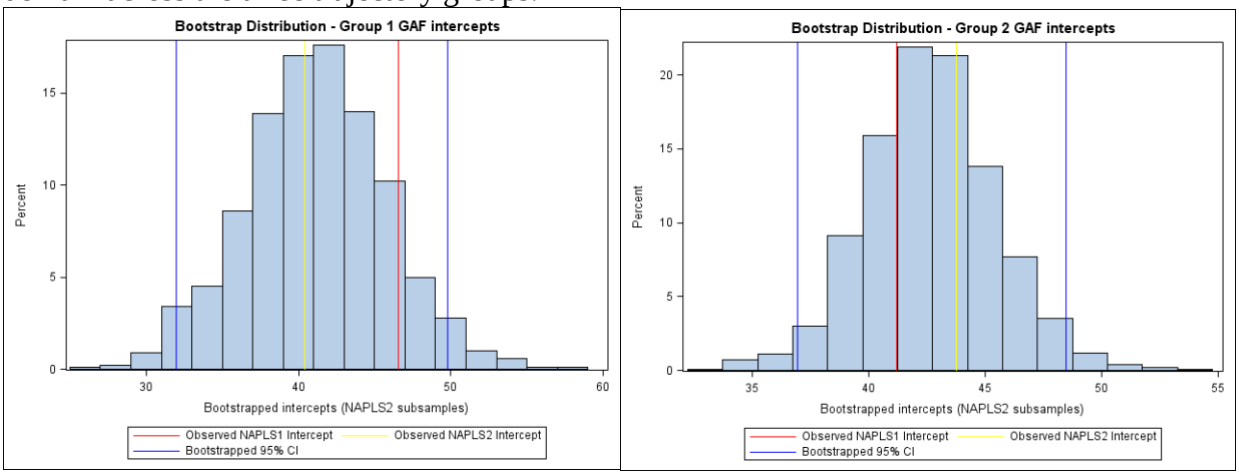
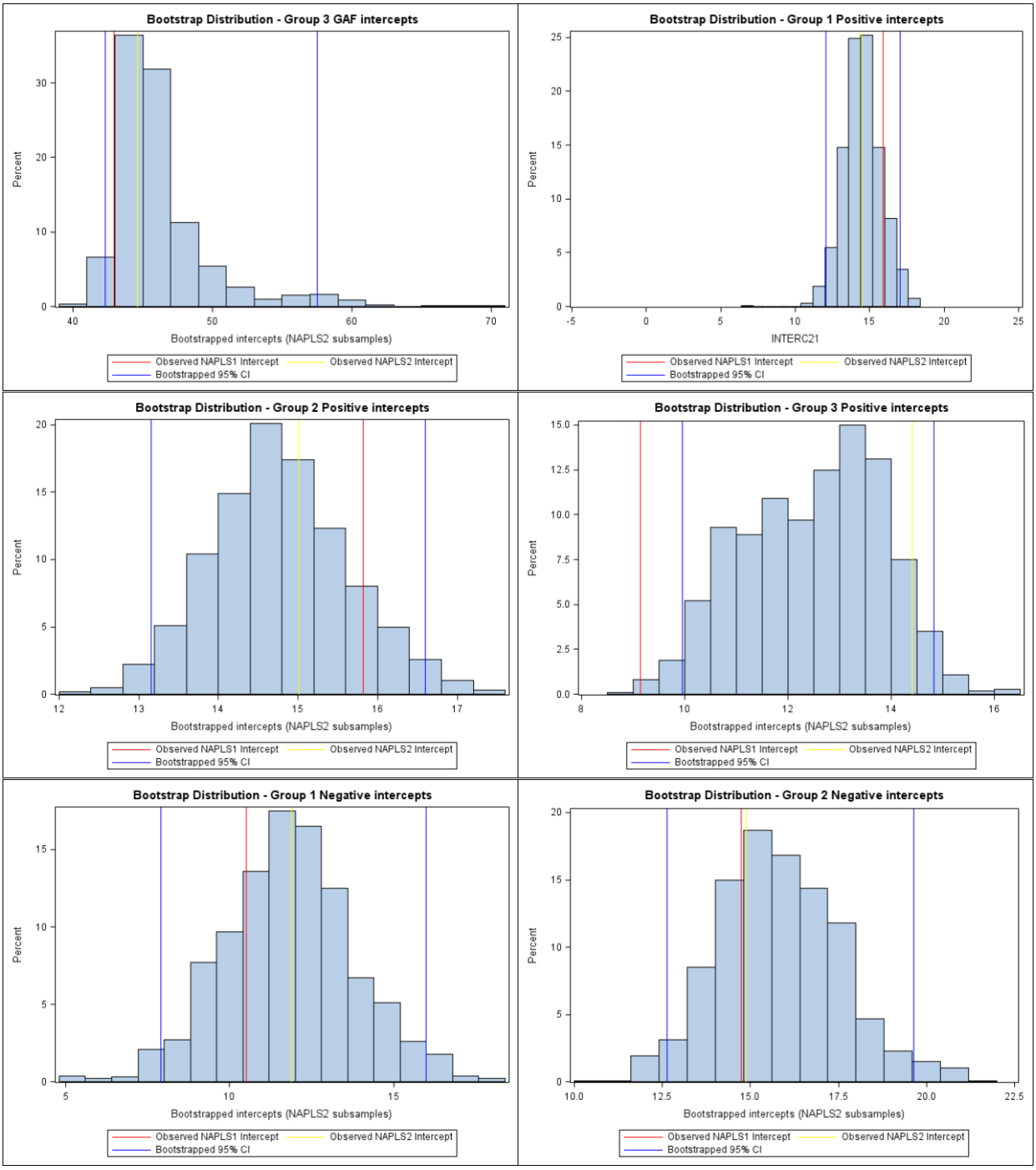
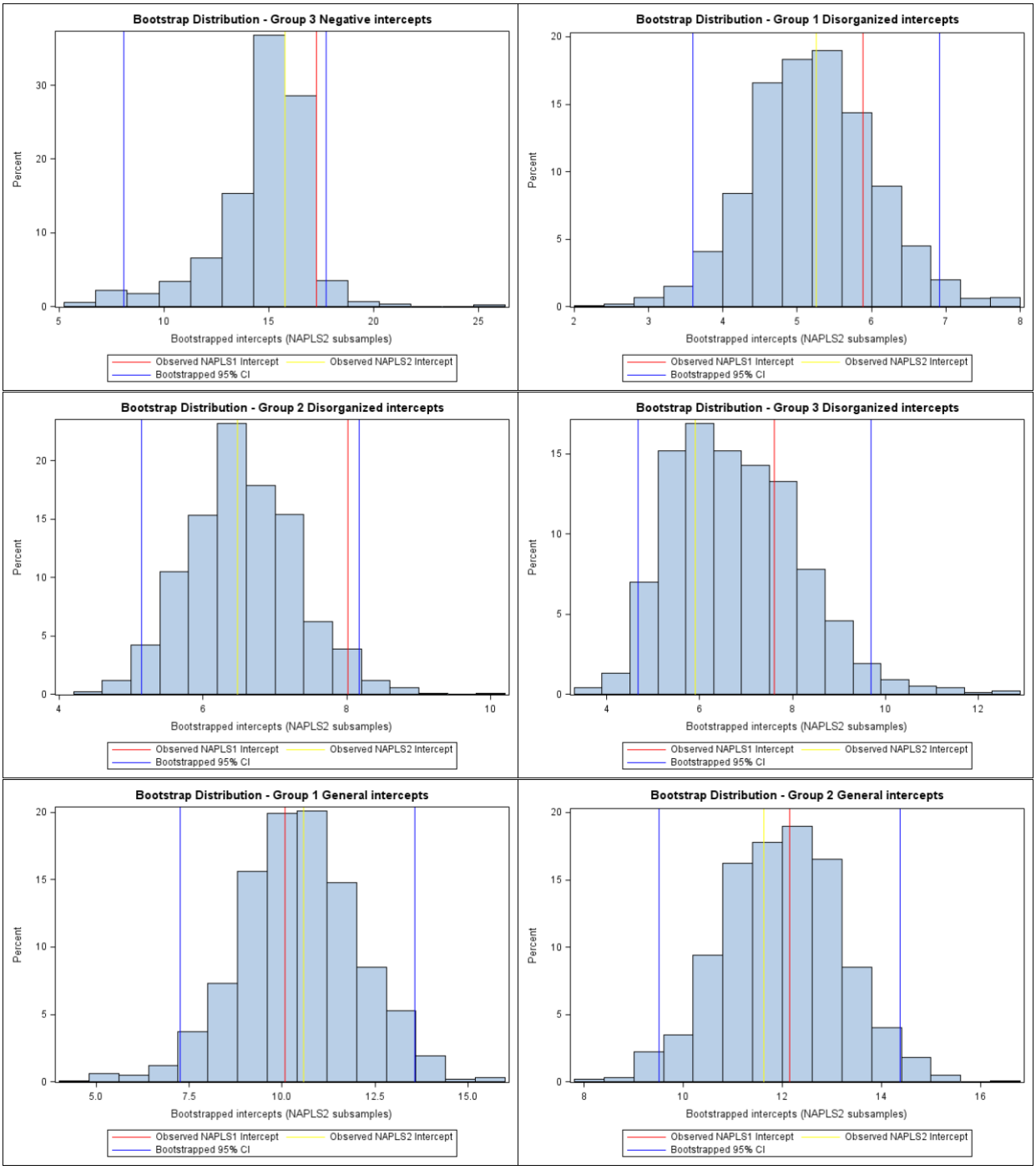
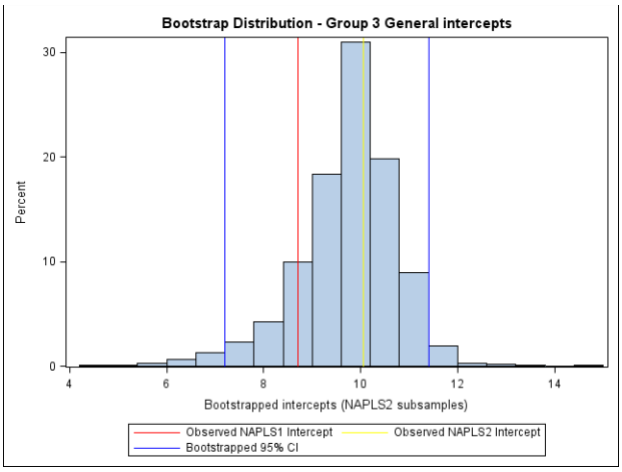


Figure S11. Bootstrapped distributions for intercept estimates for each symptom and functioning domain across the three trajectory groups.




Bootstrap Distribution - Group 3 GAF intercepts
Percent
Bootstrapped intercepts (NAPLS2 subsamples)
Observed NAPLS1 Intercept
Observed NAPLS2 Intercept
Bootstrapped 95% CI
Bootstrap Distribution - Group 1 Positive intercepts
Percent
INTERC21
Observed NAPLS1 Intercept
Observed NAPLS2 Intercept
Bootstrapped 95% CI
Bootstrap Distribution - Group 2 Positive intercepts
Percent
Bootstrapped intercepts (NAPLS2 subsamples)
Observed NAPLS1 Intercept
Observed NAPLS2 Intercept
Bootstrapped 95% CI
Bootstrap Distribution - Group 3 Positive intercepts
Percent
Bootstrapped intercepts (NAPLS2 subsamples)
Observed NAPLS1 Intercept
Observed NAPLS2 Intercept
Bootstrapped 95% CI
Bootstrap Distribution - Group 1 Negative intercepts
Percent
Bootstrapped intercepts (NAPLS2 subsamples)
Observed NAPLS1 Intercept
Observed NAPLS2 Intercept
Bootstrapped 95% CI
Bootstrap Distribution - Group 2 Negative intercepts
Percent
Bootstrapped intercepts (NAPLS2 subsamples)
Observed NAPLS1 Intercept
Observed NAPLS2 Intercept
Bootstrapped 95% CI


Bootstrap Distribution - Group 3 Negative intercepts
Percent
Bootstrapped intercepts (NAPLS2 subsamples)
Observed NAPLS1 Intercept
Observed NAPLS2 Intercept
Bootstrapped 95% CI
Bootstrap Distribution - Group 1 Disorganized intercepts
Percent
Bootstrapped intercepts (NAPLS2 subsamples)
Observed NAPLS1 Intercept
Observed NAPLS2 Intercept
Bootstrapped 95% CI
Bootstrap Distribution - Group 2 Disorganized intercepts
Percent
Bootstrapped intercepts (NAPLS2 subsamples)
Observed NAPLS1 Intercept
Observed NAPLS2 Intercept
Bootstrapped 95% CI
Bootstrap Distribution - Group 3 Disorganized intercepts
Percent
Bootstrapped intercepts (NAPLS2 subsamples)
Observed NAPLS1 Intercept
Observed NAPLS2 Intercept
Bootstrapped 95% CI
Bootstrap Distribution - Group 1 General intercepts
Percent
Bootstrapped intercepts (NAPLS2 subsamples)
Observed NAPLS1 Intercept
Observed NAPLS2 Intercept
Bootstrapped 95% CI
Bootstrap Distribution - Group 2 General intercepts
Percent
Bootstrapped intercepts (NAPLS2 subsamples)
Observed NAPLS1 Intercept
Observed NAPLS2 Intercept
Bootstrapped 95% CI



Guidelines for Reporting on Latent Trajectory Studies (GRoLTS) Checklist (13)
NAPLS2

1) *Metric of time used in the statistical model*

Timepoints used in the model corresponded to the number of months since the baseline assessment. The study was designed to consist of five assessments, which were intended to take place every 6 months. The five assessments timepoints were labeled: baseline, 6 months, 12 months, 18 months, 24 months.

2) *Information presented about the mean and variance of time within a wave*

Across the assessment waves, the average date of assessment ranged from 6 months and four days (12-month assessment) to 6 months and 20 days (6-month assessment) following the previous assessment (Table S23).

Table S23. Mean number of days elapsed since previous assessment for each wave. Expected number of days between assessments is 182.5.

Assessment	N	Mean (days)	SD
6 months	395	201.92	39.18
12 months	367	186.76	49.45
18 months	252	187.32	45.72
24 months	215	194.44	53.83

3a) *Missing data mechanism reported*

In PROC TRAJ, missing data is assumed to be missing at random (MAR). This assumes that attrition and missing data are independent of unobserved outcomes, conditional on observed outcome values and any covariates included in the model (14).

3b) *Description of what variables are related to attrition*

Of the 422 participants in the analytic sample, 54.29% (*n* = 229) missed at least one of the five assessment visits over the course of the two-year study. 26.07% (*n* = 110) missed one visit, and 28.20% (*n* = 119) missed two visits. Of the individuals who missed at least one visit, 64.19% (*n* = 147) dropped out of the study, whereas the remaining 35.81 returned for the final study assessment. To assess the potential for selective drop-out effects to bias results, we compared individuals who completed the study to those who did not complete the study on demographics and clinical characteristics available at baseline (Table S24). Significantly more males had missing data than complete data, but otherwise the groups did not differ.

Table S24. Demographic and baseline clinical differences between members of the NAPLS2 analytic sample with versus without at least one missing study visit.

	Completed study (*n* = 193)	Dropout (*n* = 229)	t/χ^2	*p*
Demographics				
Age (years)	18.71 (4.37)	18.38 (4.28)	.60	.438
Sex (% male)	53.89%	63.76%	4.23	.040
Parent Education (completed HS or less)	20.53%	14.91%	2.27	.132
Race (%)			3.01	.222
White	55.44%	56.33%		
African-American	20.21%	14.41%		
Other	24.35%	29.26%		
Lifetime anxiety disorder (%)	52.85%	54.14%	.07	.790
Lifetime mood disorder (%)	63.73%	59.35%	.83	.36
Lifetime non-cluster A disorder (%)	19.23%	15.07%	1.22	.269
Mean number of comorbid disorders across domains	1.35 (.88)	1.28 (.88)	.62	.431
Baseline symptom severity				
GAF	49.36 (11.07)	48.99 (11.22)	.11	.737
Positive Sum	11.48 (4.01)	11.62 (3.86)	.14	.711
Negative Sum	12.03 (6.10)	11.61 (5.94)	.51	.474
Disorganized Sum	5.21 (3.29)	4.74 (2.85)	2.42	.120
General Sum	8.93 (4.37)	8.81 (4.28)	.08	.777

3c) *Description of how missing data in the analyses were dealt with*

Common reasons for missing data in NAPLS2 included: participant dropout, loss to follow up, no-show, not booked, or rater forgot. To support the MAR assumption, higher rates of

missing data among males needed to be addressed to support accurate estimates of model parameters. Additionally, differential rates of missing data across derived trajectory groups could also lead to bias in estimation of model parameters. In standard group-based trajectory models, the probability of trajectory group membership is expected to remain consistent across the length of the study and estimated trajectory probabilities do not account for missing data or attrition (14). If dropout rates are relatively consistent among trajectory groups, model estimation is relatively robust. However, differential rates of dropout can lead to mild underestimation of group size probabilities and overestimation of trajectory estimate values at later visits within those groups (14).

The potential influence of sex and differential rates of missing data across trajectory groups were addressed by incorporating dropout and sex parameters into univariate trajectory models. The addition of dropout parameters allowed for estimation of rates of missing data within each identified trajectory group, and adjusted accordingly for potential inflation of the expected size of the trajectory groups and parameter estimates at later visits (14). Including sex as a covariate in the model accounted for higher rates of missing data among males in dropout parameter estimation as well (14).

Due to the complexity of modeling longitudinal outcomes for variables (GAF and severity of symptoms scores), inclusion of dropout and covariate parameters in multi-trajectory models led to model instability and could not be estimated. To address this concern, we re-ran final univariate models with the addition of parameters to assess dropout and sex, and assessed similarities and differences with the original final model (see Assessing the Influence of Attrition section below). In sum, we found that accounting for dropout did not substantially affect models, addressing this concern.

4) *Information about the distribution of observed variables (DV)*

The dependent variables investigated (GAF, sum of symptom severity scores at each visit) were coded as continuous variables. Non-normal distributions of dependent variables can affect model estimation and increase bias towards over-extraction of trajectory groups. To minimize the impact of potential non-normal distributions due to clustering of values at the minimum value of the symptom severity scale (in this case, 0), a censored normal (also known as tobit) distribution was utilized for all dependent variables in all models. Utilization of the censored normal distribution modification in this scenario supports the consistent estimation of parameters of trajectory groups across all outcome variables (15). To aid with visualization of the distribution of observed variables relative to parameter estimates at each timepoint within classes, spaghetti plots were also included in both univariate models (SI) as well as final multivariate models (manuscript).

5) *Software mentioned*

Analyses were conducted in SAS 9.4 using PROC TRAJ (16). Graphs were made in R using ggplot2 (17) and gridExtra (18). Start values were tested in STATA 15 using a macro created by Bobby Jones (personal communication, 4.12.18).

6a) *Alternative specifications of within-class heterogeneity considered (e.g., LGCA vs LGMM), or if not, sufficient justification provided as to eliminate certain specifications from consideration*

Group-based multi-trajectory modeling is a form of latent class growth analysis (LCGA) in which it is assumed that within-class individual growth trajectories are consistent and variance and covariance estimates for trajectory estimates within each class are set to zero (13,15). Another variant of longitudinal modeling, latent growth mixture modeling (LGMM), permits flexibility of variance around trajectory estimates within each class (19). Though LGMM is better suited to fitting variation in trajectories within groups, the computational demands of modeling are also higher and can lead to issues with model convergence (13). Due to computational limitations, we chose to apply the LGCA-based approach in our dataset. However, the assumption of consistent variance around trajectory estimates within groups is pragmatically motivated. Theoretically, the trajectories identified are considered to be clusters of individuals who exhibit *approximately* similar longitudinal courses in the domains examined, rather than truly homogenous subgroups within a population (15). To aid in assessment of individual trajectories within classes, we include spaghetti plots of all models along with trajectory parameter estimates. Broadly, these figures suggest that there is individual variation in the magnitude of change on measures within classes, which is also indicated by large confidence intervals around trajectory parameters. This variability suggests that our ability to draw conclusions about individual outcomes within a class is limited, and results should be interpreted cautiously (20).

6b) *Alternative specifications of the between-class differences in variance-covariance matrix structure considered and clearly documented, or if not, sufficient justification provided as to eliminate certain specifications from consideration*

Group-based multi-trajectory modeling also makes the assumption of homogeneity of variance and covariance matrices across trajectory groups (i.e., the extent to which individual trajectories vary from the trajectory estimates is consistent across groups) (13). Again, this assumption is made to support model convergence, and likely oversimplifies the relationship between groups. To support assessment of the variance of individual trajectories across groups, we provide spaghetti plots of all of our models. Generally, there appears to be some variability in variance from the estimated trajectory parameters across groups, particularly for symptoms with a relatively higher prevalence of the minimum scale value (zero) in certain groups. Our use of censored normal distributions for these variables in the model (see #4 above) likely supports accurate parameter estimation in these instances of heterogeneity. In other instances, model estimates may be slightly influenced by differences in variance across trajectory groups. Overall, our approach is to cautiously interpret the trajectory group results under this assumption and acknowledge that the final model could change if the variance and covariance matrices were freely estimated. Larger studies with sufficient power to support convergence in models with flexible variance and covariance matrices could assess the influence of this assumption on model derivation in this scenario.

7) *Alternative shape/functional forms of the trajectories described*

Trajectories were initially estimated with quadratic parameters, and the statistical significance of these parameters were assessed. Trajectories were trimmed to linear or intercept-only shapes if statistical support for higher-order parameters was not strong.

8) *Confirmation that analyses can still be replicate if covariates have been used*

In the present analyses, the effects of covariates were only examined as predictors of class membership (i.e., unconditional models) (13). This use of covariates contributes to the identification of which group an individual is likely to belong to, and does not affect the derivation of groups, shapes of trajectories, or calculation of posterior probabilities of group membership (15). The association between covariates and assigned trajectory group membership were analyzed using linear regression, separately from the trajectory model derivation (i.e., standard three-step approach) (13). Though this approach ignores the uncertainty inherent in class membership, the high posterior probabilities of class membership suggest that there are few classification errors and minimal bias in prediction of class membership (21). However, it should be noted that this assumption may lead to an underestimation of the true relationship between covariates and groups (13).

9) *Information reported about the number of random start values and final iterations*

In GBTM and GBMTM, model parameters are derived using quasi-Newtonian maximum likelihood estimation. As this approach can be vulnerable to identifying local rather than global maxima, the selection of starting values can influence the final model parameters (22). A common approach to addressing this concern is to test a variety of start values to ensure that the global maximum is identified. Guidelines for identifying start values for univariate trajectory models to ensure robust assessment of the full parameter space have been identified (23) and involve the use of theory to set ranges for the start values to be tested for each model parameter. Though full control over these parameters for start value testing is not currently available in the PROC TRAJ software package, we were able to use a macro developed by Bobby Jones (personal communication, 4.12.18) to test a variety of start values for final univariate GBTM models for each symptom as well as the final multivariate GBMTM model. For univariate models, default start values were initially used to identify the preferred number of groups. In PROC TRAJ, default start values are created by determining group intercepts based on the range or standard deviation of the outcome variable, with the assumption that group probabilities were equal. Following the initial estimation of the univariate trajectories, group parameter estimates were inputted into the macro and used to generate 100 sets of start values that randomly varied by a sigma value of 10. BIC values of the models generated by each set of start values were tabulated and assessed to determine if the best-fitting model was generated by the initial starting values or one of the random variants. The start values macro was used in the same way as above for the final multivariate models.

10) *Model comparison (and selection) tools described from a statistical perspective*

Guidelines for model comparison parameters put forth by the developers of group-based trajectory modeling (15) were followed. Parameters used in model comparison include: two BICs (calculated for both the number of participants and the number of observations included), model log-likelihood, average posterior probability, odds of correct classification, and trajectory group size (percent size). Final models were selected based on model parameter guidelines (e.g., larger BIC compared to previous model, average posterior probabilities above .7 for all groups, odds of correct classification above 5 for all groups, and trajectory group sizes above 5% of the total sample) (15). In cases in which model parameters were above these thresholds and BICs continued to increase as the number of groups derived approached 10, the model with the lowest number of trajectory groups that were theoretically distinct and meaningful was selected (13).

The parameters listed above were reported for final models, along with the probability of group membership and associated 95% confidence intervals (to facilitate comparison with the actual trajectory group size).

11) *Total number of fitted models reported, including a one-class solution*

As recommended by van de Schoot et al (2017), we followed a forward modeling approach and started with a one class solution. The total number of fitted models, as well as model fit statistics used in model selection, are reported below for all univariate GBTM models as well as the multivariate GBMTM models.

12) *Number of cases per class reported for each model (absolute sample size or proportion)*

The percentage of cases allocated to each trajectory group is reported for all models tested (see below). Models with groups smaller than 5% were excluded from consideration of final models due to concerns about estimation reliability and accuracy (13,15).

13) Entropy reported if classification of cases in a trajectory is the goal

As classification of individuals into trajectory groups is a primary goal of these analyses, average posterior probability of group membership was reported for all models. This parameter is conceptually similar to entropy, which indexes the difference between posterior probabilities across classes (13). Consistency among posterior probabilities across groups was examined throughout model comparison to assess the ability of the model to parse data reliably (13,15).

14) *Plot with estimated mean trajectories included for all considered models, and plots with observed individual trajectories split out for each latent class included for each final model*

Plots of estimated mean trajectories for all univariate and multivariate solutions considered during model selection in (see below). Spaghetti plots are included along with estimated mean trajectories for all final models (see below).

15) *Characteristics of the final class solution numerically described (i.e., means, SD/SE, n, CI, etc.)*

tables of results, including slope estimates, standard deviations, *p* values, confidence intervals, and sample size, were included for all final models (see below).

16) *Syntax files available (either in the appendix, supplementary materials, or from the authors)*

Syntax files in SAS and R will be shared with anyone who inquires with the corresponding author.

Univariate Model Derivation

Multi-group trajectory modeling is a powerful tool because each trajectory group encapsulates information about change over time of multiple outcomes. However, this complexity makes model searching for multi-trajectory models more challenging. Estimating trajectory models for each of the outcomes separately first can facilitate multi-group modeling by identifying key trajectories for each outcome that ought to be represented in the multi-group model. Importantly, the goal for univariate model search is not to optimize the preferred model for each outcome, but to clarify which specific trajectories are expected to be included in the multi-group models for each outcome (16).

To achieve this aim, we derived univariate group-based trajectory models for each of our outcomes (GAF, sum of positive/negative/disorganized/general symptoms), beginning with a one-group model and continuing until model fit parameter thresholds were crossed (e.g., percent assigned to group below 5%, negligible improvement in BIC above lowest absolute value identified thus far, average posterior probability below 0.7, odds of correct classification below 5) or additional trajectory groups derived were not of theoretical interest (15). We used the best supported univariate models to assess assumptions of group-based trajectory modeling, including homogeneity of variance between and within trajectory groups (which, if invalid, could potentially lead to over-extraction of groups).

Group-based trajectory modeling assumes that data are missing at random, accounting for observed variables included in the model (14). Additionally, differential rates of dropout between groups can also differential rates of dropout can lead to mild underestimation of group size probabilities and overestimation of trajectory estimate values at later visits within those groups (14). Unfortunately, dropout and baseline GAF could not be included in final multivariate models due to issues with model convergence. To assess the potential influence of lower baseline GAF scores among those with more missing data and any differences in rates of dropout across groups on model parameters, we re-ran final univariate models that included baseline GAF and dropout parameters as well.

Global Assessment of Functioning

Group-based trajectory models of GAF were supported up to three trajectory groups (Figure S12, Table S25). The four-group model was rejected due to a very small increase BIC and a group size of 4% for one of the derived groups (Table S25). The selected three-group model suggests that common GAF trajectories include: poor and relatively stable functioning across time (Group 1), moderate and slightly improving across time (Group 2), and good functioning that improves most rapidly within the first six months of the study (Group 3) (Figure S13, Table S26). Further assessment of model fit parameters suggest that the selected three-group model adequately fits the underlying data (Table S27). Estimated spaghetti plots of each trajectory group (Figure S14) suggests that the three groups are relatively well separated for one another and that variance is approximately homogeneous within and across groups. Rapid improvement in GAF is particularly pronounced between visit 1 (baseline) and visit 2 (6-months) for Groups 2 and 3, suggesting that rapid change in functioning may take place within the first six months of seeking treatment for these individuals with relatively better functioning at baseline compared to Group 1 individuals. Variants of this study design with more frequent assessments within the first year following help-seeking would further explore this phenomenon and hopefully yield further insight into changes that might be occurring during this period.

Including dropout and sex parameters in the final model of GAF did not lead to substantial change in final model estimates or adequacy parameters. Though dropout parameters were significant, estimated trajectories appear very similar across models that exclude and include dropout parameters (Figure S15), model parameters remained consistent (Table S28), and group sizes changed by 1.3% at most (Table S29). Thus, the effects of attrition are not expected to have an appreciable effect on GAF trajectory modeling in multi-trajectory models.

Table S25. Diagnostics of group-based model adequacy for GAF model comparison.

Number of groups	BIC (*N*=1758)	BIC (*N*=422)	Log likelihood	Average posterior probability	Odds of correct classification	Percent assigned to group
1	-6987.29	-6984.36	-6972.27	..	..	100
2	-6771.67	-6765.96	-6741.78	.94 .91	10 15	60 40
3	-6724.84	-6716.28	-6680.01	.88 .85 .89	16 7 38	33 49 18
4	-6723.97	-6712.55	-6664.19	.88 .85 .89 .77	16 6 41 86	32 47 17 4

Figure S12. Model comparison plots of estimated trajectories (dashed) and mean GAF values at each visit (solid) for derived trajectory groups in each model.

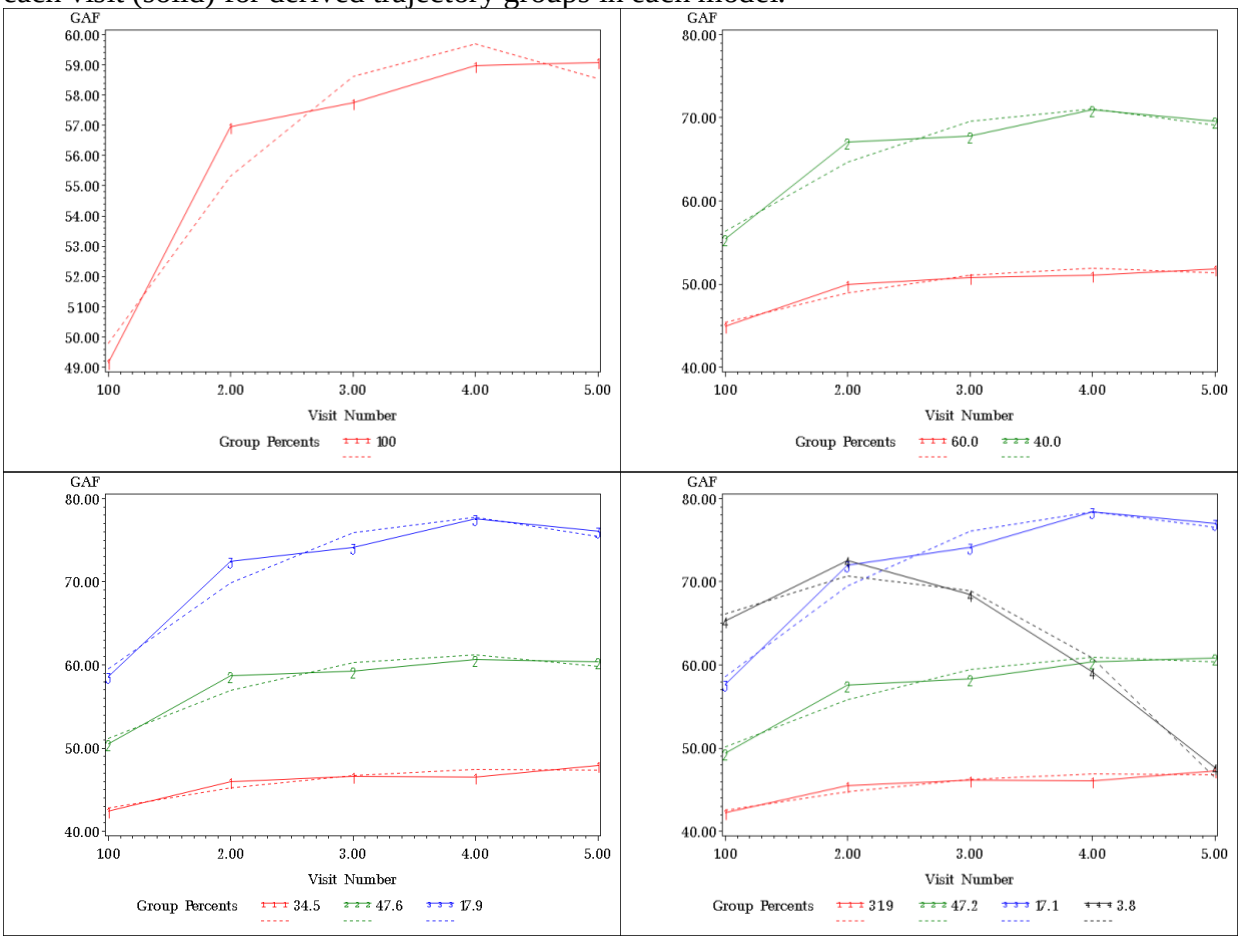


Figure S13. Plot of estimated trajectories (dashed) and mean GAF values at each visit (solid) for the preferred three-group GAF model.

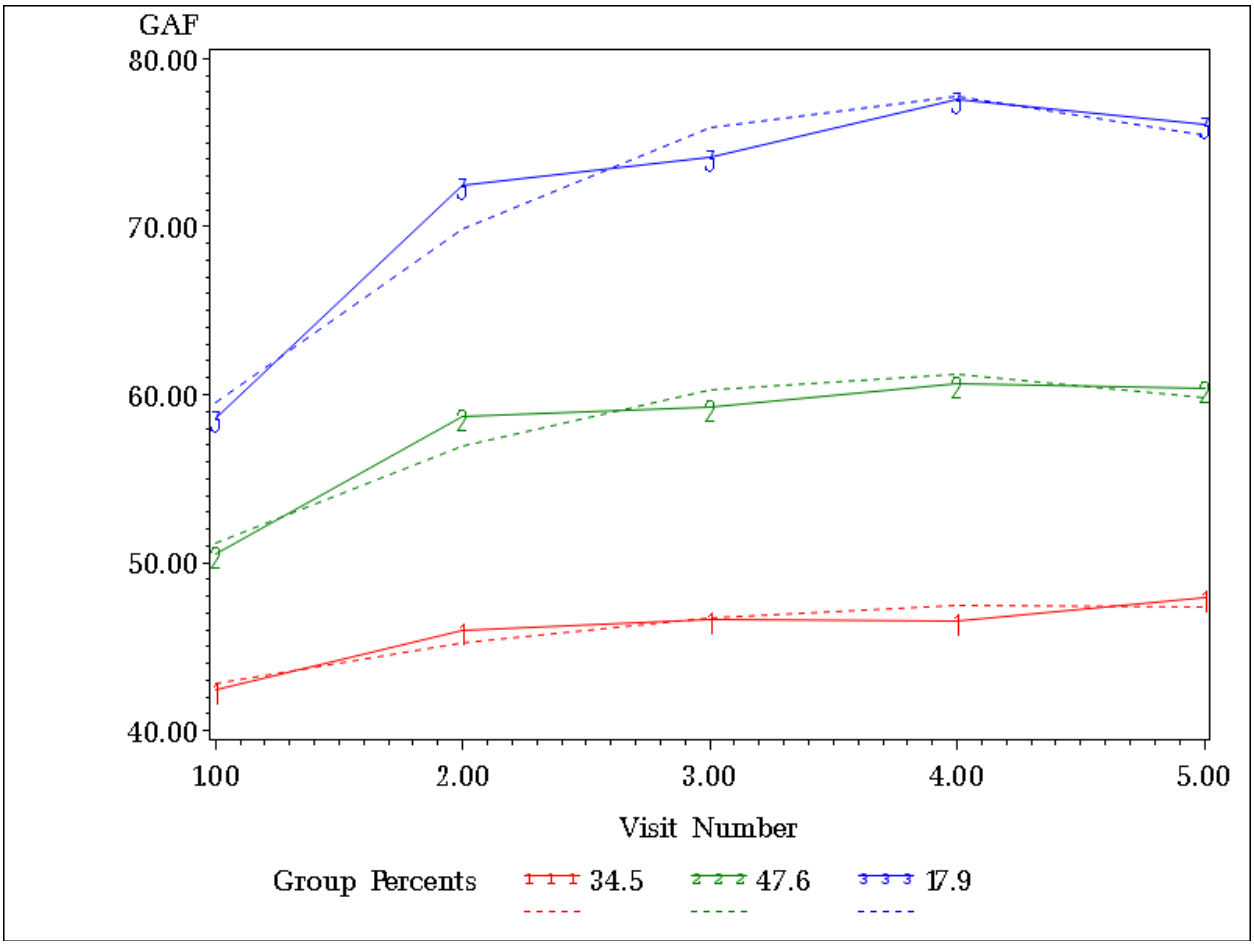


Table S26. Model parameters for the preferred three-group GAF model.

Group	Parameter	Estimate	SE	T	P
1	Intercept	39.59	1.91	20.744	<.001
	Linear	3.45	1.52	2.263	.024
	Quadratic	-.38	.26	-1.504	.133
2	Intercept	42.96	1.67	25.741	<.001
	Linear	9.20	1.38	6.664	<.001
	Quadratic	-1.17	.23	-5.054	<.001
3	Intercept	44.99	2.86	15.724	<.001
	Linear	16.64	2.21	7.517	<.001
	Quadratic	-2.12	.37	-5.697	<.001

	Sigma	9.09	.17	54.743	<.001
Group Membership					
1	(%)	33.18	4.31	7.694	<.001
2	(%)	48.68	3.90	12.493	<.001
3	(%)	18.14	2.76	6.576	<.001

Table S27. Diagnostics of group-based model adequacy for the preferred three-group GAF mode

Class	Probability of group membership	95% CI	Proportion assigned to group	Average posterior probability	Odds of correct classification
1	.332	[.248, .416]	.353	.885	15.5
2	.471	[.395, .547]	.499	.855	6.6
3	.181	[.126, .236]	.148	.893	37.8

Figure S14. Spaghetti plots of individual longitudinal data and estimated trajectories for the preferred three-group GAF model.

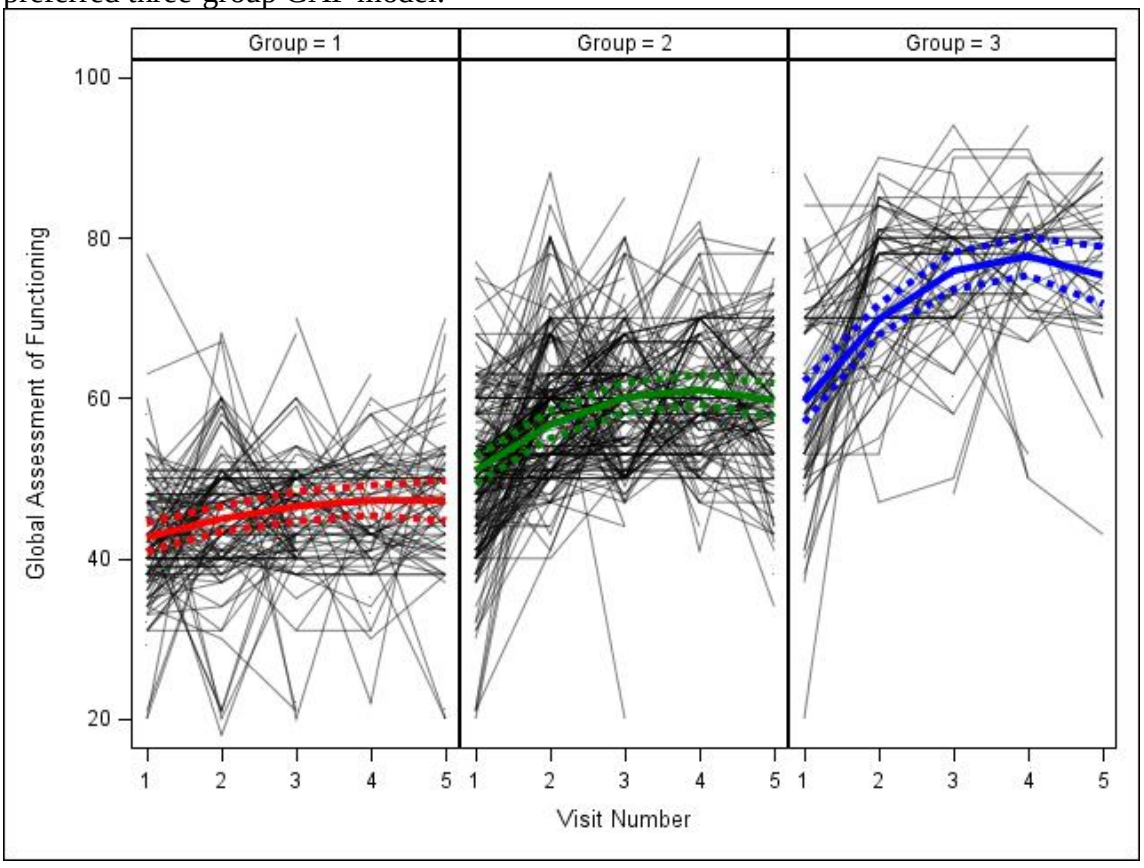


Figure S15. Comparison of GAF final model (panel 1) with final model when sex and dropout at each visit are included in the model (panel 2)

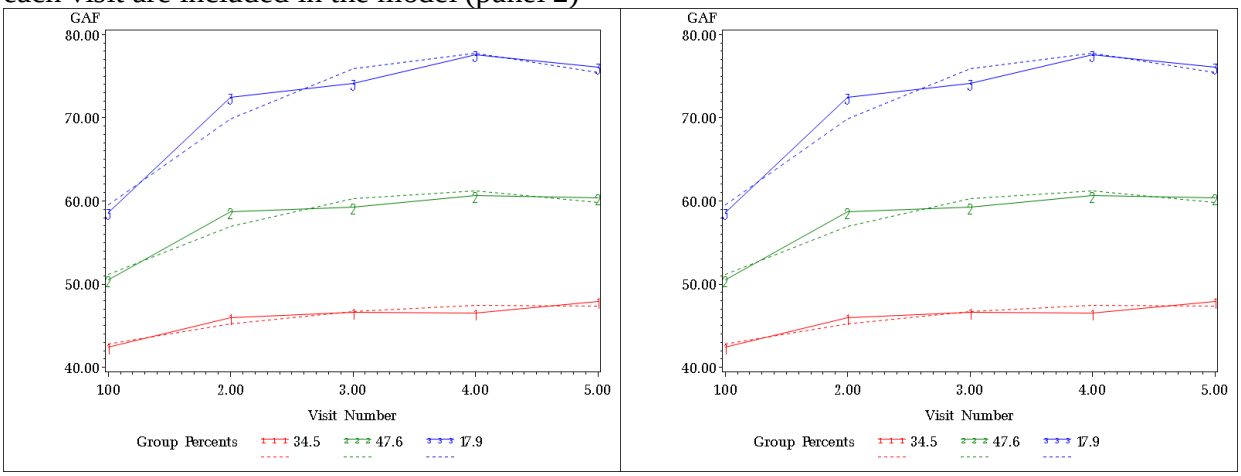


Table S28. Model parameters for the preferred three-group GAF model, including dropout and sex parameters.

Group	Parameter	Estimate	SE	T	P
1	Intercept	39.56	1.88	21.069	<.001
	Linear	3.64	1.51	2.408	.016
	Quadratic	-.42	.25	-1.639	.101
2	Intercept	43.05	1.71	25.210	<.001
	Linear	9.30	1.41	6.619	<.001
	Quadratic	-1.19	.24	-5.047	<.001
3	Intercept	45.01	2.90	15.524	<.001
	Linear	16.65	2.25	7.413	<.001
	Quadratic	-2.11	.38	-5.604	<.001
1	Drop0	-2.18	.16	-13.782	<.001
2	Drop0	-2.35	.15	-15.884	<.001
3	Drop0	-2.31	.23	-10.149	<.001
	Sigma	9.09	.17	54.609	<.001
Group Membership					
1	Constant	.	.	.	.
2	Constant	-.18	.48	-.381	.703
	Sex	.36	.28	1.280	.200

3	Constant	-1.06	.54	-1.968	.049
	Sex	.29	.32	.901	.368

BIC (N=1758): -7228.85 *BIC (N=422):* -7218.14 *L:* -7172.81

Table S29. Diagnostics of group-based model adequacy for the three-group GAF model with dropout and sex parameters included compared to the base three-group GAF model.

	Base model		*Model with dropout and sex*	
Class	Probability of group membership	Average posterior probability	Probability of group membership	Average posterior probability
1	.332	.885	.345	.888
2	.471	.855	.476	.847
3	.181	.893	.179	.897

Sum of Positive Symptom Severity

Group-based trajectory models of the sum of positive symptom severity were supported up to four trajectory groups (Table S30, Figure S16). The five-group model was rejected due to a negligible increase in BIC and lack of uniqueness of the fifth derived trajectory group (Table S30). The selected four-group model suggests that common trajectories for the sum of positive symptoms include: low and rapidly improving symptom severity across time that almost reaches zero severity (Group 1), moderate severity that improves especially rapidly during the first 6 months of the study (Group 2), high severity that somewhat improves at a consistent rate across time (Group 3), and extreme severity that remains stable across the visits (Group 4) (Figure S17, Table S31). Further assessment of model fit parameters suggest that the selected four-group model adequately fits the underlying data (Table S32). Estimated spaghetti plots of each trajectory group (Figure S18) suggests that the four groups are relatively well separated for one another, and that variance is approximately homogeneous within and across groups. Rapid improvement in the sum of positive symptom severity between baseline and the 6-month visit was observed for all groups except the stable chronic group (Group 4), which we hope to further explore in future studies with more frequent assessments during this period. A floor effect is observed for Group 1 and may be relevant for Group 2 as well, which could theoretically influence the assumptions of homogeneity. However, all outcomes were modeled using the censored normal (also known as tobit) option in the group-based trajectory modeling code, which accounts for clustering of observation at the minimum or maximum values of a scale and prevents over-extraction of groups or misestimation of trajectory parameters due to associated non-normality (15).

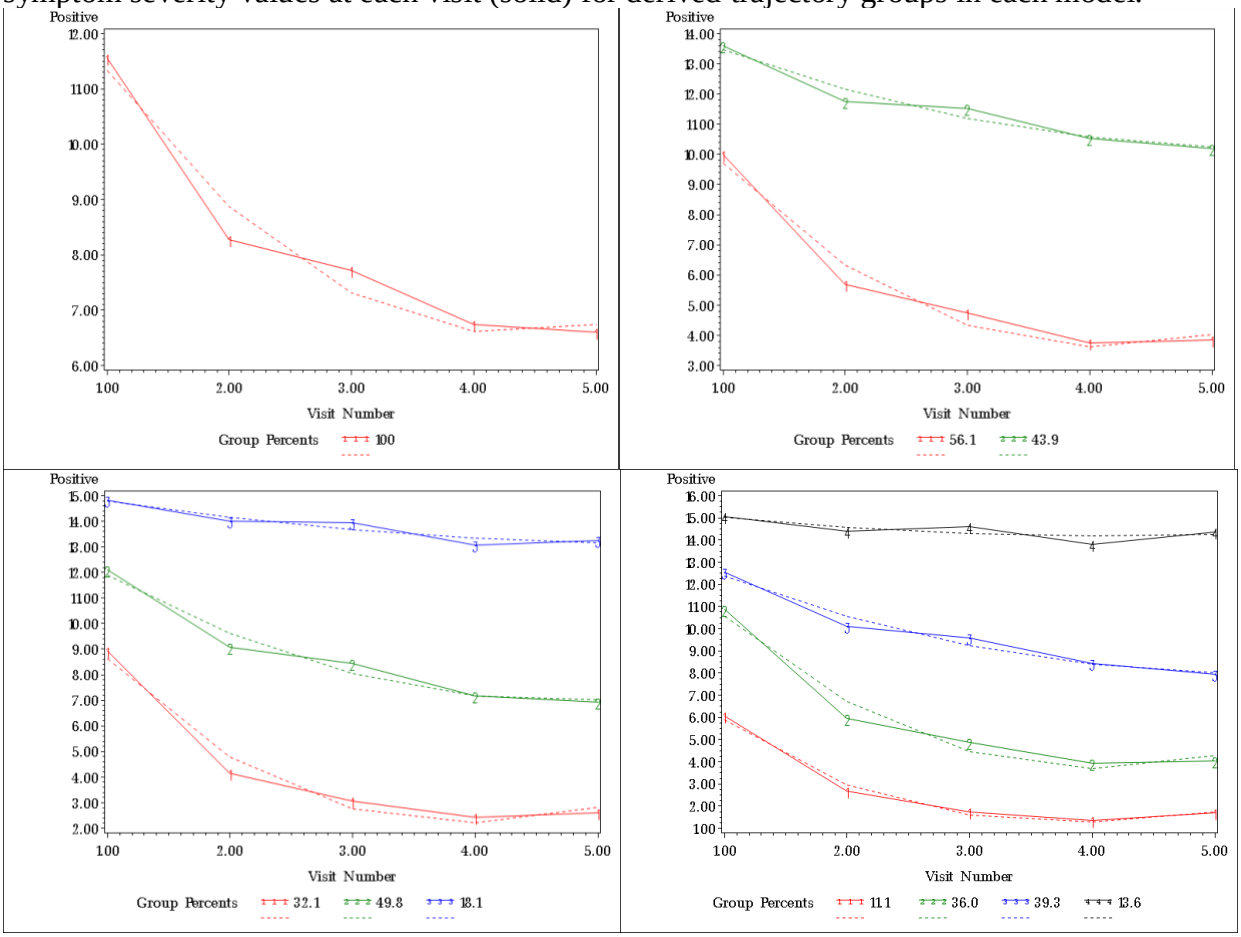
Including dropout and sex parameters in the final model of sum of positive symptom severity did not lead to substantial change in final model estimates or adequacy parameters. Though dropout parameters were significant, estimated trajectories appear very similar across models that exclude and include dropout parameters (Figure S19), model parameters remained consistent (Table S33), and group sizes changed by less than 1% across groups (Table S34).

Thus, the effects of attrition are not expected to have an appreciable effect on positive symptom trajectory modeling in multi-trajectory models.

Table S30. Diagnostics of group-based model adequacy for sum of positive symptom severity model comparison.

Number of groups	BIC (*N*=1758)	BIC (*N*=422)	Log likelihood	Average posterior probability	Odds of correct classification	Percent assigned to group
1	-5060.08	-5057.22	-5045.13	..	..	100
2	-4800.64	-4794.43	-4770.75	.93 .92	11 15	56 44
3	-4717.76	-4709.20	-4672.93	.91 .90 .91	20 9 45	32 50 18
4	-4705.50	-4694.08	-4645.72	.86 .84 .85 .90	50 9 9 57	11 36 39 14
5	-4702.89	-4688.82	-4628.17	.81 .83 .82 .81 .89	54 14 8 15 5	7 26 38 23 6

Figure S16. Model comparison plots of estimated trajectories (dashed) and mean sum of positive symptom severity values at each visit (solid) for derived trajectory groups in each model.



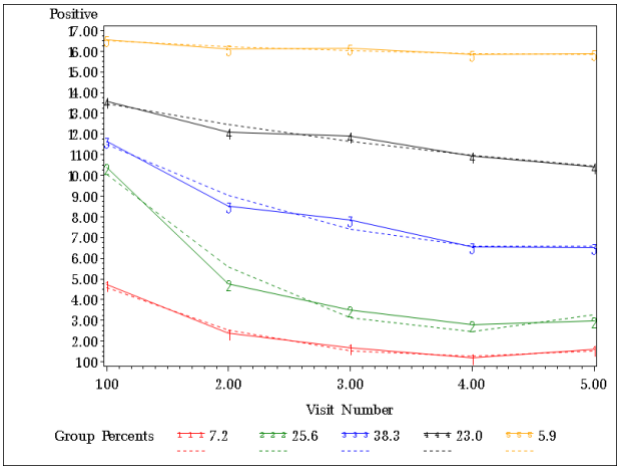


Figure S17. Plot of estimated trajectories (dashed) and mean value of sum of positive symptom severity values at each visit (solid) for the preferred four-group positive symptom severity sum model.

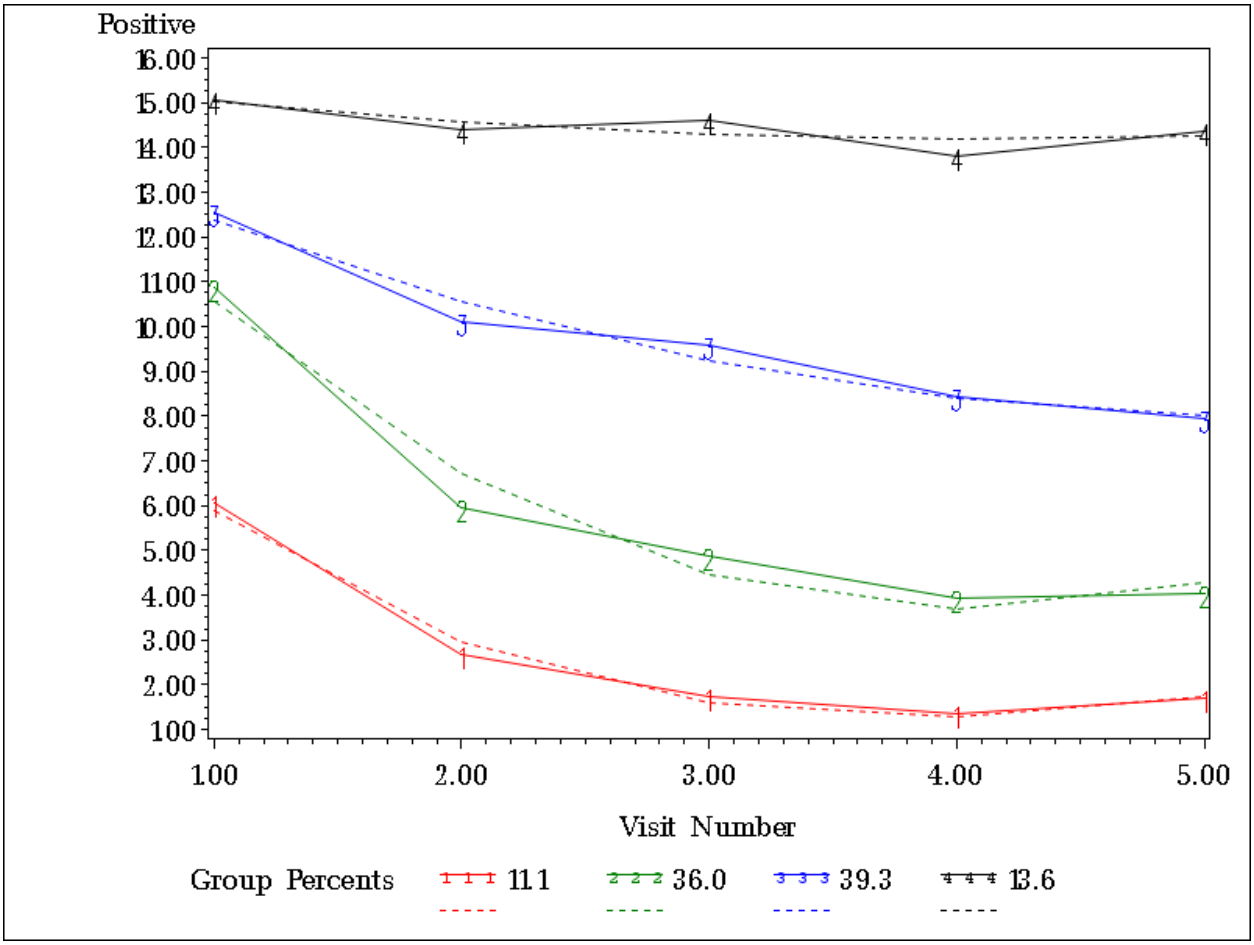


Table S31. Model parameters for the preferred four-group sum of positive symptom severity model.

Group	Parameter	Estimate	SE	T	P
1	Intercept	10.40	1.44	7.209	< .001
	Linear	-5.20	.94	-5.552	< .001
	Quadratic	.66	.15	4.441	< .001
2	Intercept	15.91	.66	23.997	< .001
	Linear	-6.10	.59	-10.418	< .001
	Quadratic	.75	.09	8.061	< .001
3	Intercept	14.66	.60	24.360	< .001
	Linear	-2.52	.56	-4.501	< .001
	Quadratic	.24	.09	2.689	.007
4	Intercept	15.69	.96	16.289	< .001
	Linear	-.74	.75	-.988	.323
	Quadratic	.09	.12	.729	.466
	Sigma	2.95	.06	51.861	< .001
Group Membership					
1	(%)	11.12	3.29	3.38	< .001
2	(%)	35.99	4.76	7.56	< .001
3	(%)	39.28	5.24	7.49	< .001
4	(%)	13.61	2.55	5.33	< .001

Table S32. Diagnostics of group-based model adequacy for the preferred four-group sum of positive symptom severity model.

Class	Probability of group membership	95% CI	Proportion assigned to group	Average posterior probability	Odds of correct classification
1	.111	[.046, .176]	.107	.860	49.2
2	.360	[.266, .454]	.360	.837	9.1
3	.392	[.291, .495]	.400	.849	8.7
4	.136	[.085, .187]	.133	.900	57.2

Figure S18. Spaghetti plots of individual longitudinal data and estimated trajectories for the preferred four-group sum of positive symptom severity model.

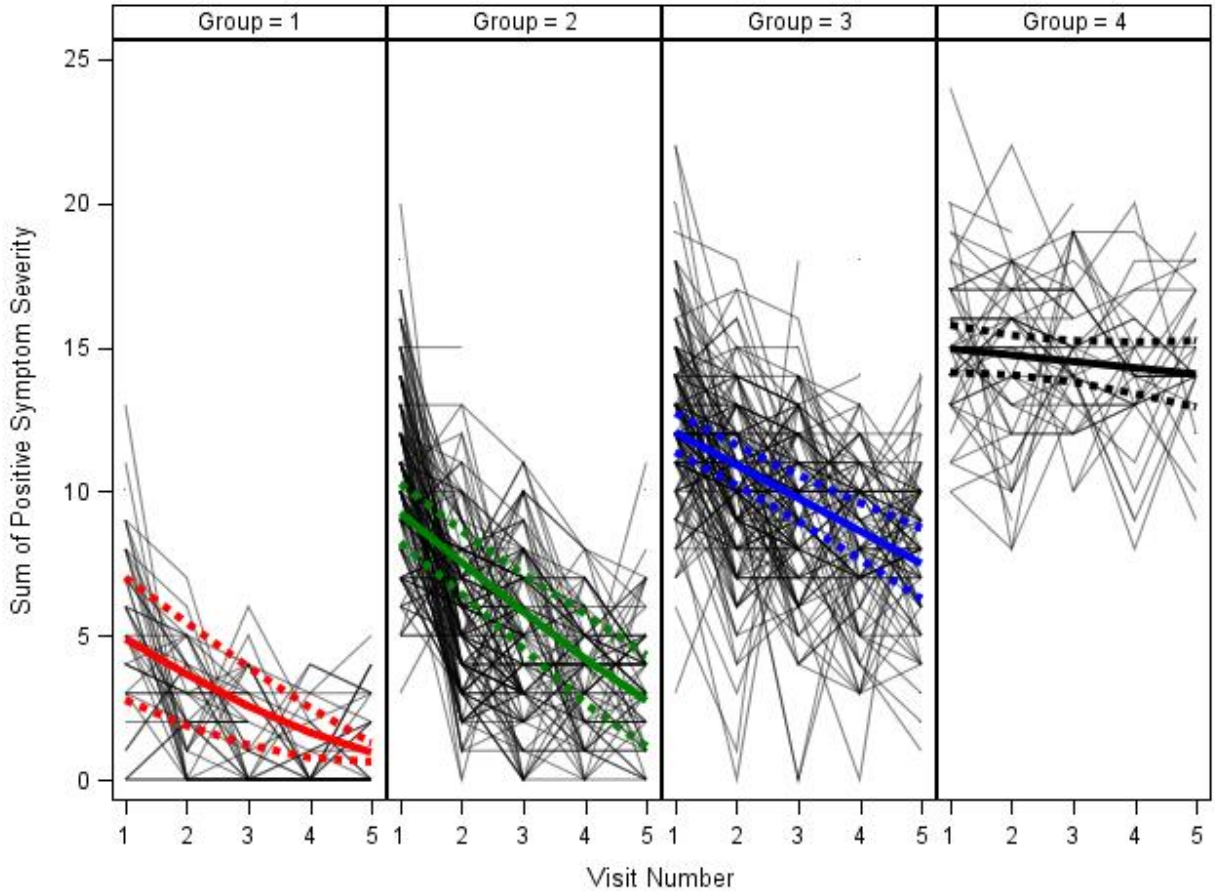


Figure S19. Comparison of sum of positive symptom severity final model (panel 1) with final positive model when sex and dropout at each visit is included in the model (panel 2).

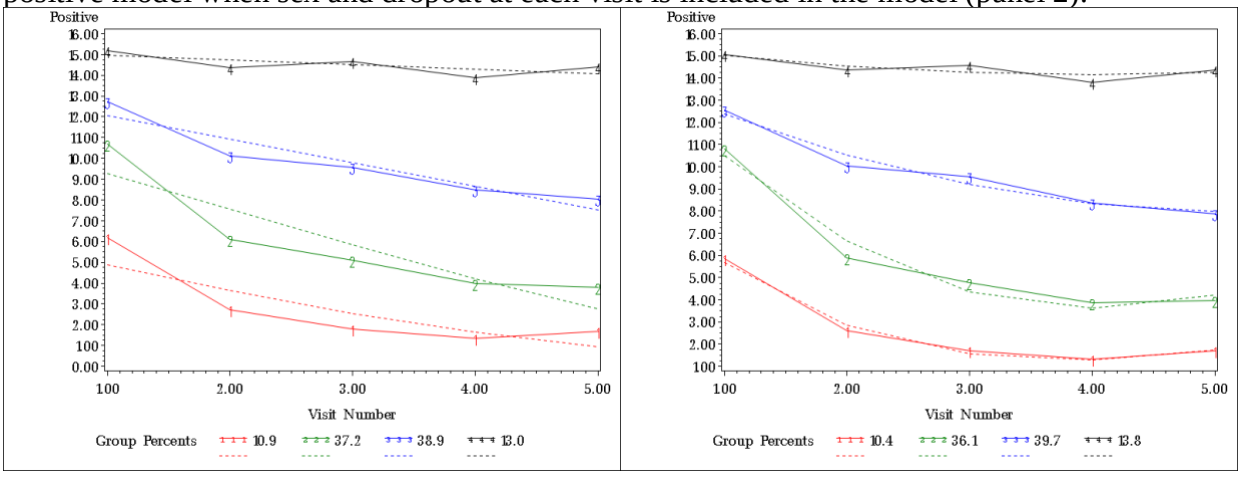


Table S33. Model parameters for the preferred four-group sum of positive symptom severity model, including dropout and sex parameters.

Group	Parameter	Estimate	SE	T	P
1	Intercept	10.10	1.41	7.185	< .001
	Linear	-5.11	.96	-5.326	< .001
	Quadratic	.66	.15	4.286	< .001
2	Intercept	15.89	.66	24.097	< .001
	Linear	-6.15	.57	-10.845	< .001
	Quadratic	.76	.09	8.320	< .001
3	Intercept	14.69	.59	24.717	< .001
	Linear	-2.57	.53	-4.884	< .001
	Quadratic	.24	.08	2.898	.003
4	Intercept	15.68	.96	16.388	< .001
	Linear	-.75	.75	-1.012	.312
	Quadratic	.09	.12	.753	.451
1	Drop0	-3.01	.46	-6.603	< .001
2	Drop0	-2.14	.16	-13.589	< .001
3	Drop0	-2.32	.16	-14.290	< .001
4	Drop0	-2.19	.25	-8.891	< .001
	Sigma	2.95	.06	51.897	< .001
Group Membership					
1	Constant	.	.	.	.
	(ref group)				
2	Constant	.64	.66	.967	.334
	Sex	.44	.47	.942	.346
3	Constant	.89	.61	1.461	.144
	Sex	.33	.43	.775	.438
4	Constant	.14	.70	.199	.842
	Sex	.11	.50	.217	.828

BIC (N=1758): -5219.57 *BIC (N=422):* -5203.16 *L:* -5133.64

Table S34. Diagnostics of group-based model adequacy for the four-group sum of positive symptom severity model with dropout and sex parameters included compared to the base four-group positive symptoms model.

	Base model		*Model with dropout and sex*	
Class	Probability of group membership	Average posterior probability	Probability of group membership	Average posterior probability
1	.111	.860	.104	.869
2	.360	.837	.361	.843
3	.392	.849	.397	.852
4	.136	.900	.138	.908

Sum of Negative Symptom Severity

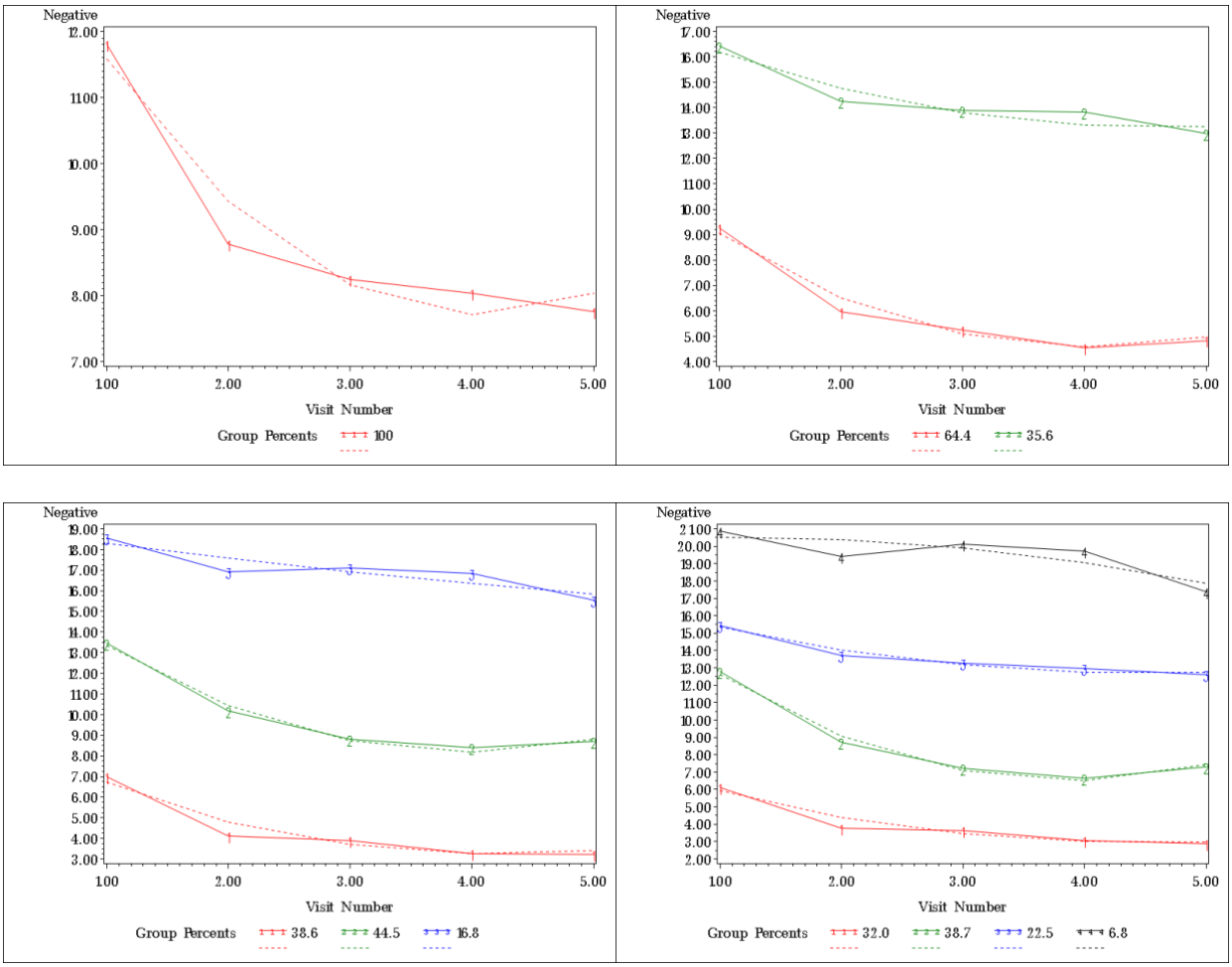
Group-based trajectory models of the sum of negative symptom severity were supported up to four trajectory groups (Table S35, Figure S20). The five-group model was rejected due to a negligible increase in BIC (Table S35). The selected four-group model suggests that common trajectories for the sum of negative symptoms include: low and steadily improving symptom severity across time that almost reaches zero severity (Group 1), moderate severity that improves most rapidly during the first six months (Group 2), higher severity that remains consistent (Group 3), and extreme severity that remains stable across the visits (Group 4) (Figure S21, Table S36). Further assessment of model fit parameters suggest that the selected four-group model adequately fits the underlying data (Table S37). Estimated spaghetti plots of each trajectory group (Figure S22) suggests that the four groups are relatively well separated for one another and that variance is approximately homogeneous within and across groups. Similar to the sum of positive symptom severity model, a floor effect is observed for Group 1 and is accounted for by the use of the censored normal option within models.

Including dropout and sex parameters in the final model of sum of negative symptom severity did not lead to substantial change in final model estimates or adequacy parameters. Though dropout parameters were significant, estimated trajectories appear very similar across models that exclude and include dropout parameters (Figure S23), model parameters remained consistent (Table S38), and group sizes changed by less than .1% across groups (Table S39). Thus, the effects of attrition are not expected to have a noticeable effect on negative symptom trajectory modeling in multi-trajectory models.

Table S35. Diagnostics of group-based model adequacy for sum of negative symptom severity model comparison.

Number of groups	BIC (*N*=1757)	BIC (*N*=422)	Log likelihood	Average posterior probability	Odds of correct classification	Percent assigned to group
1	-5509.43	-5506.57	-5494.48	.	.	100
2	-5240.19	-5234.48	-5210.30	.96 .92	13 21	64 36
3	-5156.87	-5148.36	-5112.04	.91 .88 .90	16 9 46	39 44 17
4	-5144.67	-5133.26	-5084.90	.90 .83 .84 .89	20 8 18 112	32 39 22 7
5	-5143.39	-5129.13	-5068.68	.84 .71 .76 .84 .93	24 8 10 15 147	18 23 24 27 8

Figure S20. Model comparison plots of estimated linear trajectories (dashed) and mean value of sum of negative symptom severity values at each visit (solid) for derived trajectory groups in each model.



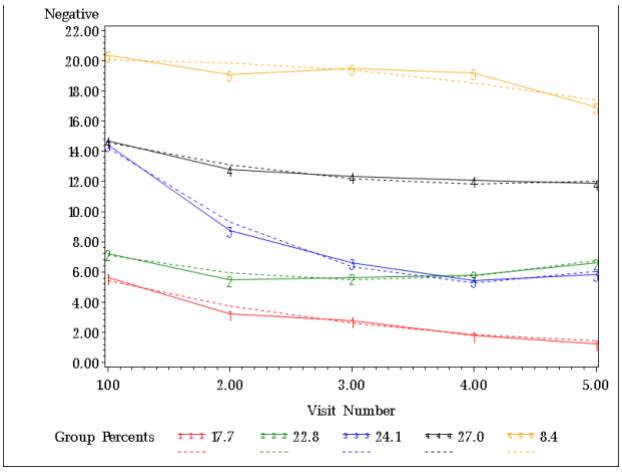


Figure S21. Plot of estimated linear trajectories (dashed) and mean negative symptom severity sum values at each visit (solid) for the preferred four-group sum of negative symptom severity model.

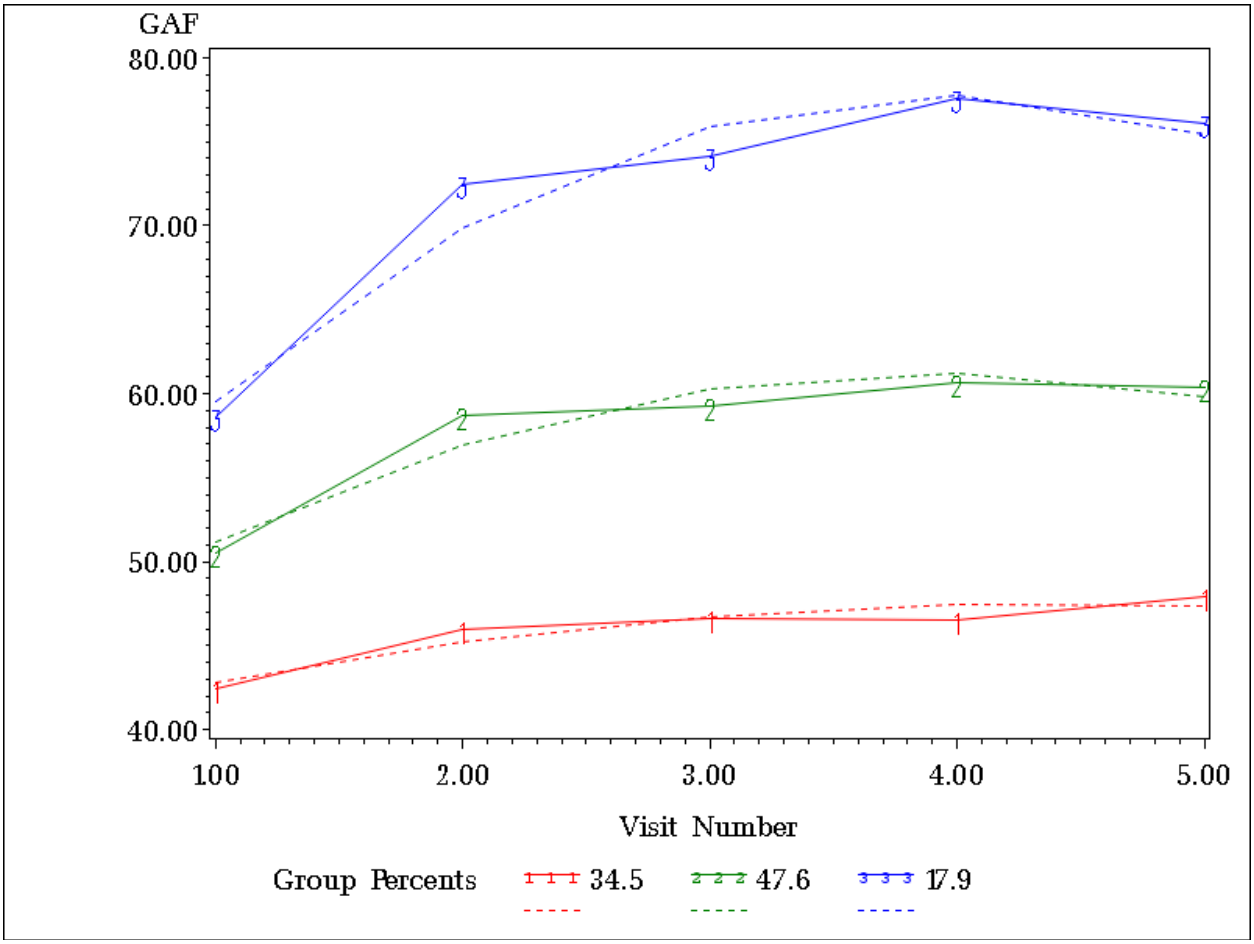


Table S36. Model parameters for the preferred four-group sum of negative symptom severity model.

Group	Parameter	Estimate	SE	T	P
1	Intercept	8.08	.95	8.534	< .001
	Linear	-2.55	.69	-3.685	< .001
	Quadratic	.28	.11	2.439	.015
2	Intercept	17.65	.95	18.530	< .001
	Linear	-5.78	.70	-8.292	< .001
	Quadratic	.75	.11	6.739	< .001
3	Intercept	17.07	1.09	15.715	< .001
	Linear	-1.96	.85	-2.298	.022
	Quadratic	.75	.14	1.550	.123
4	Intercept	20.27	1.77	11.448	< .001
	Linear	.42	1.35	.308	.758
	Quadratic	-.18	.23	-.792	.428
	Sigma	3.87	.08	51.125	< .001
Group Membership					< .001
1	(%)	32.0	3.64	8.78	< .001
2	(%)	38.7	4.09	9.46	< .001
3	(%)	22.5	3.35	6.72	< .001
4	(%)	6.8	1.78	3.84	< .001

Table S37. Diagnostics of group-based model adequacy for the preferred four-group sum of negative symptom severity model.

Class	Probability of group membership	95% CI	Proportion assigned to group	Average posterior probability	Odds of correct classification
1	.320	[.249, .391]	.313	.904	20.0
2	.387	[.307, .467]	.398	.829	7.7
3	.225	[.160, .290]	.223	.838	17.8
4	.068	[.033, .103]	.066	.891	112.0

Figure S22. Spaghetti plots of individual longitudinal data and estimated trajectories for the preferred four-group sum of negative symptom severity model.

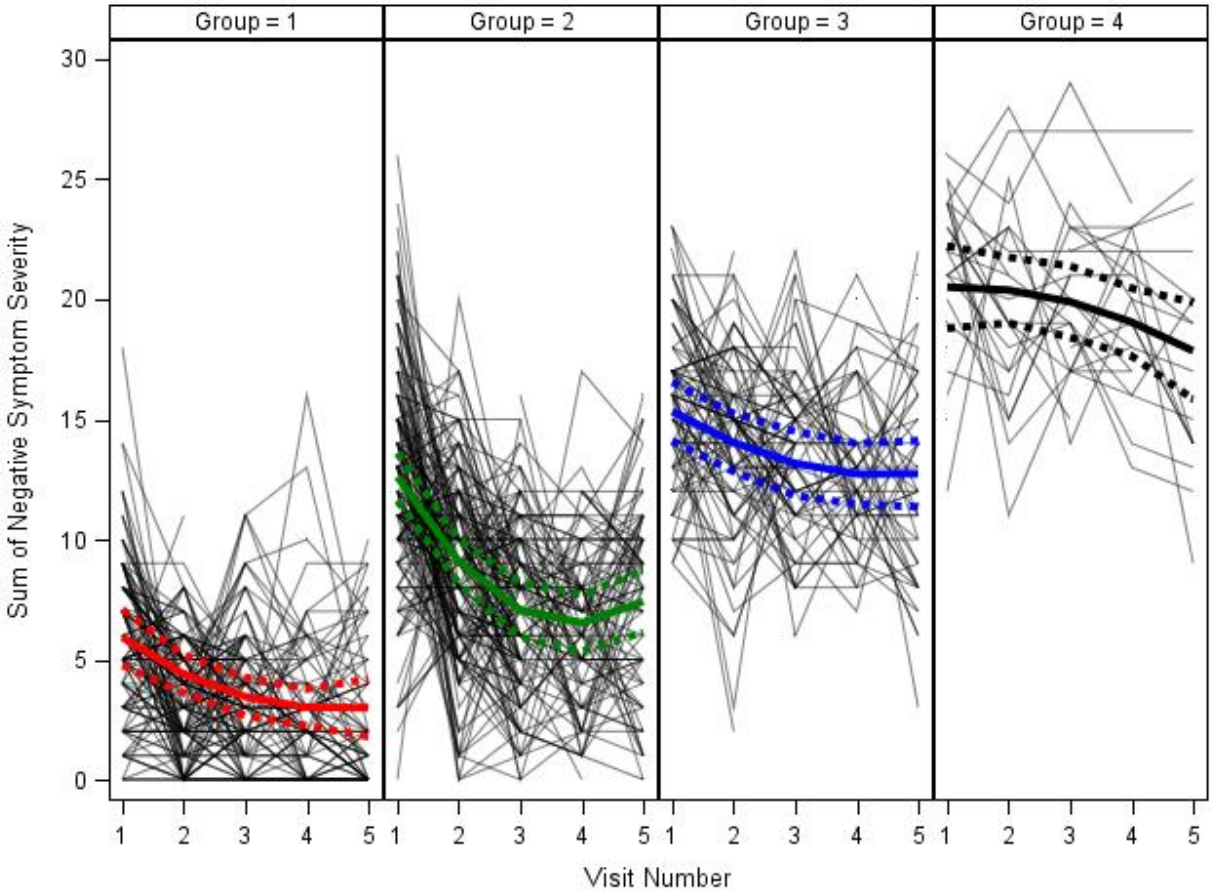


Figure S23. Comparison of sum of negative symptom severity final model (panel 1) with final negative model when sex and dropout at each visit is included in the model (panel 2).

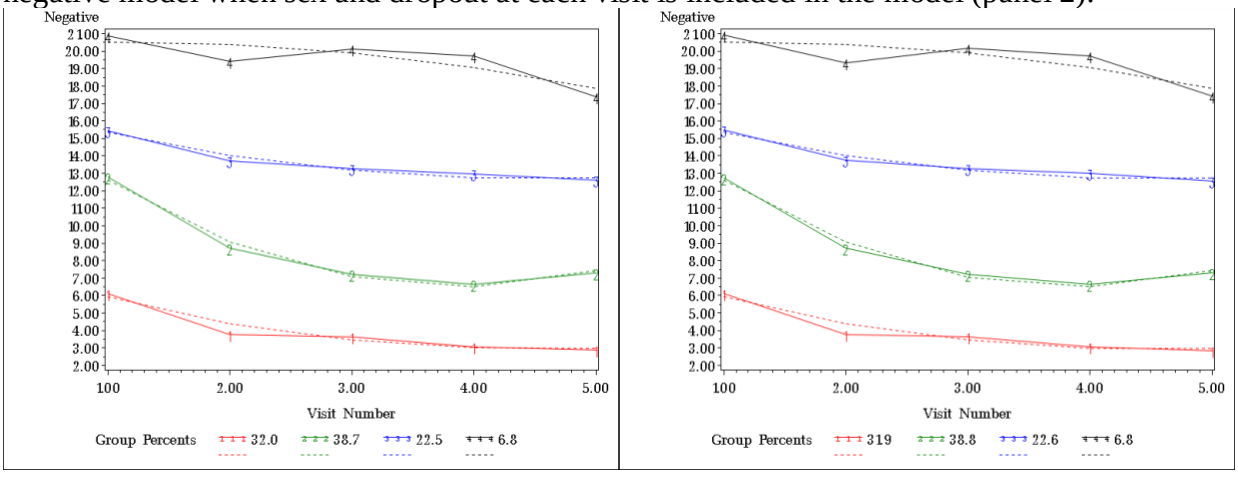


Table S38. Model parameters for the preferred four-group sum of negative symptom severity model, including dropout and sex parameters.

Group	Parameter	Estimate	SE	T	P
1	Intercept	8.07	.95	8.485	< .001
	Linear	-2.53	.69	-3.653	< .001
	Quadratic	.28	.12	2.395	.017
2	Intercept	17.63	.95	18.564	< .001
	Linear	-5.79	.69	-8.349	< .001
	Quadratic	.75	.11	6.797	< .001
3	Intercept	17.06	1.08	15.766	< .001
	Linear	-1.94	.85	-2.291	.022
	Quadratic	.21	.14	1.529	.127
4	Intercept	20.35	1.78	11.444	< .001
	Linear	.35	1.36	.260	.795
	Quadratic	-.17	.23	-.741	.459
1	Drop0	-2.25	.17	-13.424	< .001
2	Drop0	-2.35	.17	-13.740	< .001
3	Drop0	-2.25	.21	-10.728	< .001
4	Drop0	-2.28	.36	-6.285	< .001
	Sigma	3.87	.08	51.088	< .001
Group Membership					
1	Constant (ref group)	.	.	.	.
2	Constant	.46	.48	.949	.343
	Sex	-.18	.29	-.622	.534
3	Constant	.37	.49	.758	.448
	Sex	-.51	.32	-1.615	.107
4	Constant	-.08	.74	-.115	.908
	Sex	-1.09	.54	-2.007	.045

BIC (N=1757): -5657.97 *BIC (N=422):* -5641.57 *L:* -5572.05

Table S39. Diagnostics of group-based model adequacy for the four-group sum of negative symptom severity model with dropout and sex parameters included compared to the base four-group negative symptoms model.

	Base model		*Model with dropout and sex*	
Class	Probability of group membership	Average posterior probability	Probability of group membership	Average posterior probability
1	.320	.904	.319	.903
2	.387	.829	.388	.828
3	.225	.838	.226	.840
4	.068	.891	.068	.905

Sum of Disorganized Symptom Severity

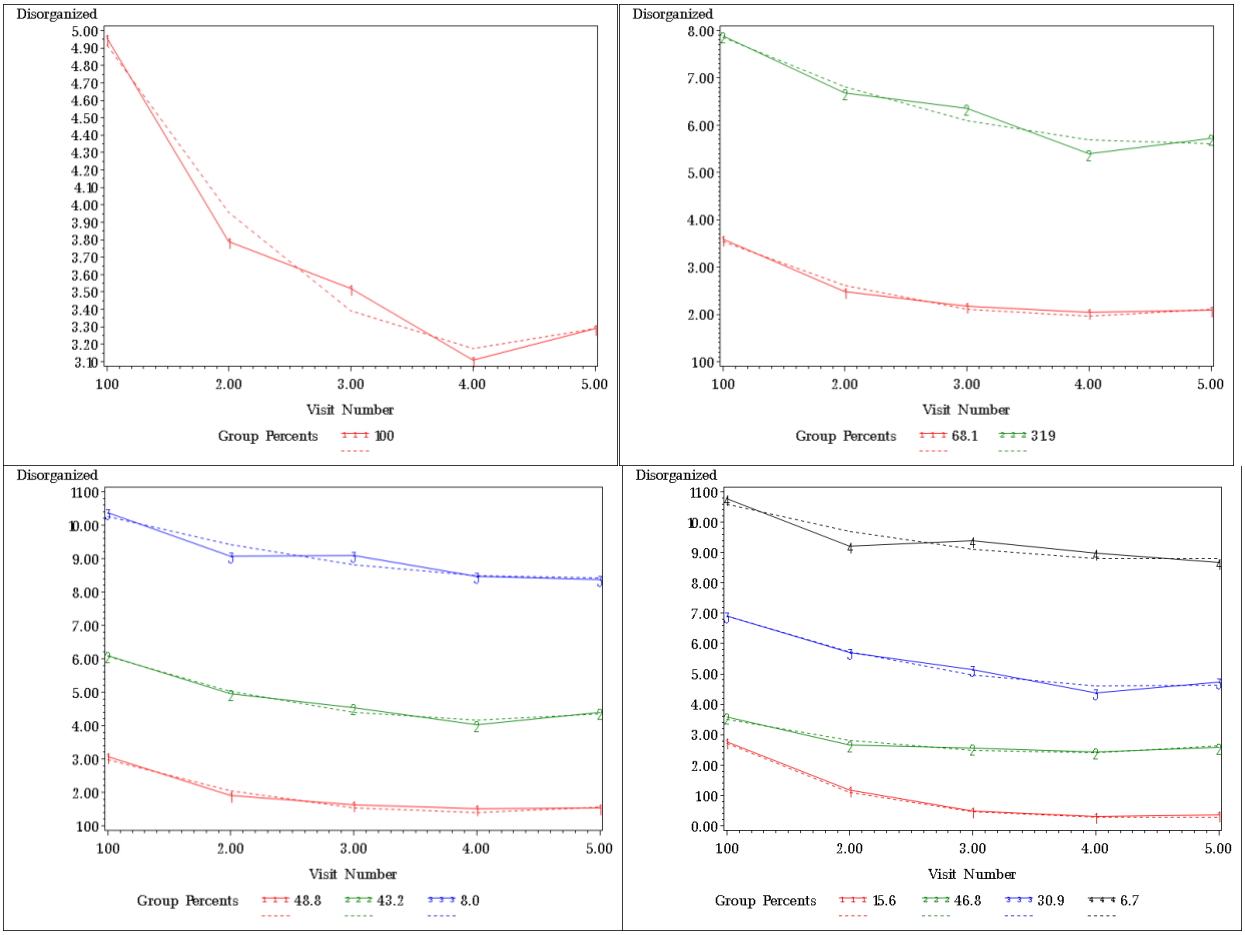
Group-based trajectory models of the sum of disorganized symptom severity were supported up to four trajectory groups (Table S40, Figure S24). The five-group model was rejected due to a decrease in BIC (Table S40). The selected four-group model suggests that common trajectories for the sum of disorganized symptoms include: low and improving symptom severity across time that decreases to zero severity (Group 1), low severity that remains consistent (Group 2), moderate severity that somewhat decreases across visits (Group 3), and extreme severity that slightly improves with time (Group 4) (Figure S25, Table S41). Further assessment of model fit parameters suggest that the selected four-group model adequately fits the underlying data (Table S42). Estimated spaghetti plots of each trajectory group (Figure S26) suggests that the four groups are relatively well separated from one another and that variance is approximately homogeneous within and across groups. Again, the floor effect observed for Groups 1 and 2 is accounted for by use of the censored normal option.

Including dropout and sex parameters in the final model of sum of disorganized symptom severity did not lead to substantial change in final model estimates or adequacy parameters. Though dropout parameters were significant, estimated trajectories appear very similar across models that exclude and include dropout parameters (Figure S27), model parameters remained consistent (Table S43), and group sizes changed by less than 1% across groups (Table S44). Thus, the effects of attrition are not expected to have a noticeable effect on disorganized symptom trajectory modeling in multi-trajectory models.

Table S40. Diagnostics of group-based model adequacy for sum of disorganized symptom severity model comparison.

Number of groups	BIC (*N*=1758)	BIC (*N*=422)	Log likelihood	Average posterior probability	Odds of correct classification	Percent assigned to group
1	-4209.90	-4207.04	-4194.95	.	.	100
2	-3977.49	-3971.78	-3947.60	.95 .92	9 24	68 32
3	-3894.15	-3885.58	-3849.31	.92 .90 .95	12 11 227	49 43 8
4	-3865.61	-3854.91	-3805.83	.85 .87 .88 .93	30 8 16 197	15 47 31 7
5	-3872.54	-3858.27	-3797.82	.75 .78 .87 .88 .94	26 59 8 17 219	10 6 46 31 7

Figure S24. Model comparison plots of estimated trajectories (dashed) and mean sum of disorganized symptom severity values at each visit (solid) for derived trajectory groups in each model.



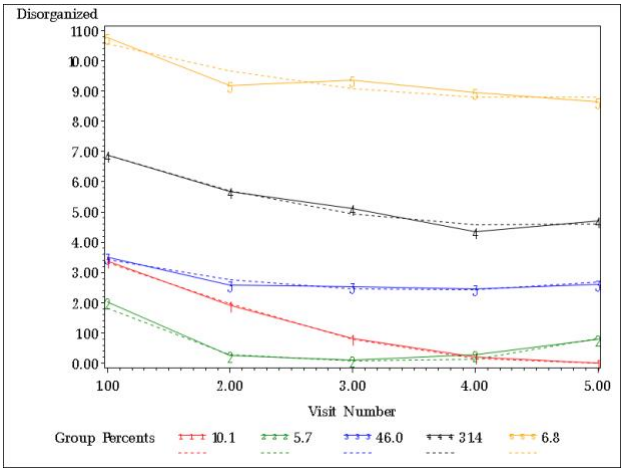


Figure S25. Plot of estimated trajectories (dashed) and mean disorganized symptom severity sum values at each visit (solid) for the preferred four-group sum of disorganized symptom severity model.

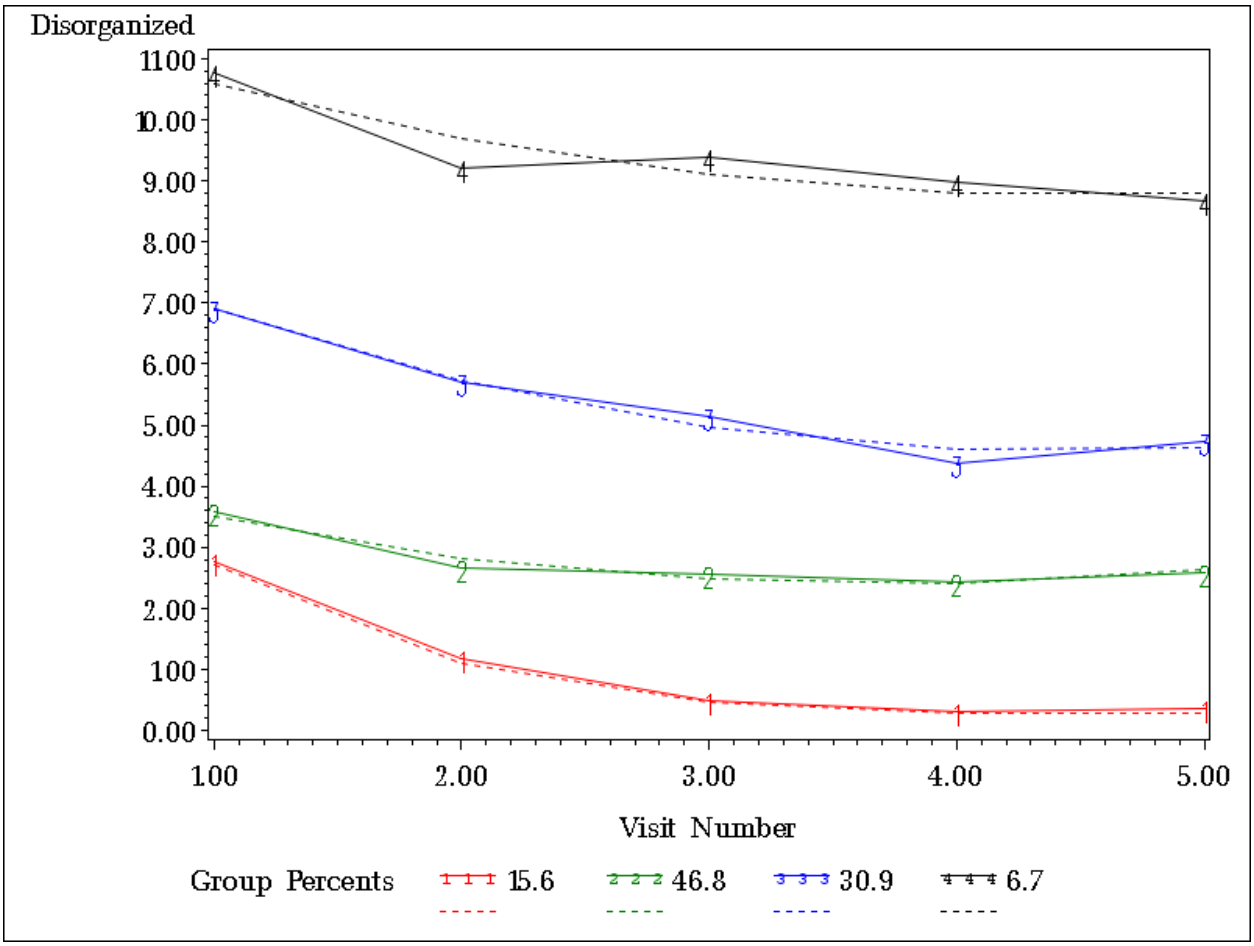


Table S41. Model parameters for the preferred four-group sum of disorganized symptom severity model.

Group	Parameter	Estimate	SE	T	P
1	Intercept	5.40	.77	7.054	< .001
	Linear	-3.15	.69	-4.544	< .001
	Quadratic	.38	.12	3.028	.003
2	Intercept	4.50	.39	11.518	< .001
	Linear	-1.20	.29	-4.183	< .001
	Quadratic	.16	.05	3.395	< .001
3	Intercept	8.48	.47	18.103	< .001
	Linear	-1.77	.35	-5.108	< .001
	Quadratic	.20	.06	3.487	< .001
4	Intercept	11.78	.96	12.304	< .001
	Linear	-1.34	.71	-1.887	.059
	Quadratic	.15	.17	1.282	.200
	Sigma	2.02	.05	49.713	< .001
Group Membership					
1	(%)	15.56	2.97	5.247	< .001
2	(%)	46.78	3.73	12.551	< .001
3	(%)	30.94	3.30	9.375	< .001
4	(%)	6.71	1.40	4.808	< .001

Table S42. Diagnostics of group-based model adequacy for the preferred four-group sum of disorganized symptom severity model.

Class	Probability of group membership	95% CI	Proportion assigned to group	Average posterior probability	Odds of correct classification
1	.156	[.097, .215]	.149	.848	30.2
2	.468	[.395, .541]	.474	.869	7.5
3	.309	[.244, .374]	.310	.876	15.8
4	.067	[.040, .094]	.066	.934	197.1

Figure S26. Spaghetti plots of individual longitudinal data and estimated trajectories for the preferred four-group sum of disorganized symptom severity model.

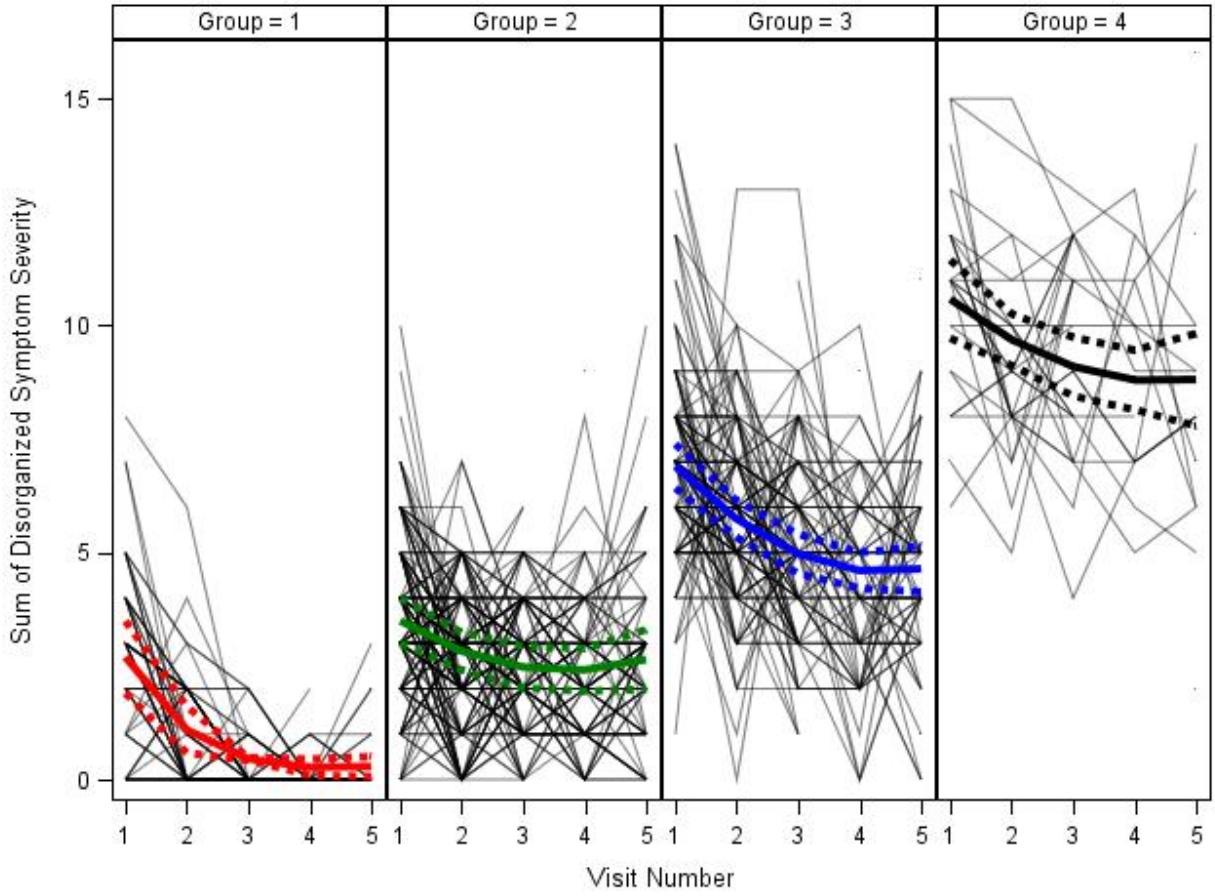


Figure S27. Comparison of sum of disorganized symptom severity final model (panel 1) with final disorganized model when sex and dropout at each visit is included in the model (panel 2).

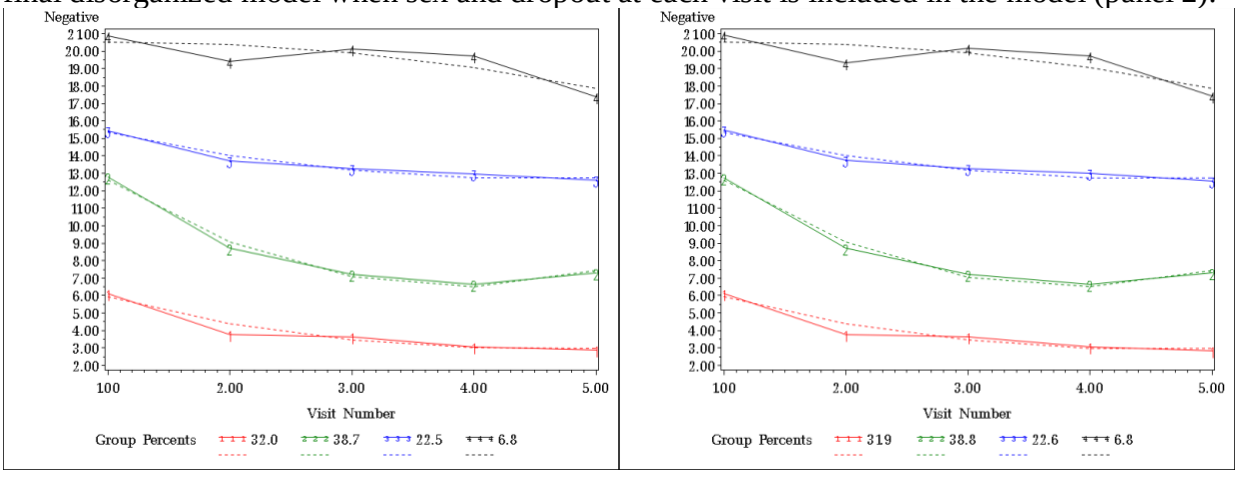


Table S43. Model parameters for the preferred four-group sum of disorganized symptom severity model, including dropout and sex parameters.

Group	Parameter	Estimate	SE	T	P
1	Intercept	5.45	.78	7.007	< .001
	Linear	-3.24	.71	-4.558	< .001
	Quadratic	.38	.12	3.129	.002
2	Intercept	4.53	.39	11.600	< .001
	Linear	-1.21	.29	-4.216	< .001
	Quadratic	.16	.05	3.398	< .001
3	Intercept	8.49	.47	17.959	< .001
	Linear	-1.77	.35	-5.064	< .001
	Quadratic	.20	.06	3.444	< .001
4	Intercept	11.77	.95	12.325	< .001
	Linear	-1.34	.71	-1.883	.060
	Quadratic	.15	.12	1.286	.199
1	Drop0	-2.54	.30	-8.483	< .001
2	Drop0	-2.19	.14	-16.060	< .001
3	Drop0	-2.35	.18	-12.870	< .001
4	Drop0	-2.24	.35	-6.320	< .001
	Sigma	2.02	.04	49.632	< .001
Group Membership					
1	Constant (ref group)	.	.	.	.
2	Constant	.73	.57	1.296	.195
	Sex	.28	.38	.756	.450
3	Constant	.41	.55	.747	.455
	Sex	.21	.37	.561	.575
4	Constant	-.65	.76	-.854	.393
	Sex	-.13	.53	-.246	.805

BIC (N=1758): -4381.22 *BIC (N=422):* -4364.81 *L:* -4295.29

Table S44. Diagnostics of group-based model adequacy for the four-group sum of disorganized symptom severity model with dropout and sex parameters included compared to the base four-group disorganized symptoms model.

	Base model		*Model with dropout and sex*	
Class	Probability of group membership	Average posterior probability	Probability of group membership	Average posterior probability
1	.156	.848	.153	.836
2	.468	.869	.473	.876
3	.309	.876	.306	.875
4	.067	.934	.067	.935

Sum of General Symptom Severity

Group-based trajectory models of the sum of general symptom severity were supported up to three trajectory groups (Table S45, Figure S28). The four-group model was rejected due to a decrease in BIC (Table S45). The selected three-group model suggests that common trajectories for the sum of general symptoms include: low and improving symptom severity across time that almost reaches zero severity (Group 1), moderate severity that improves most rapidly at the beginning of the study (Group 2), and higher severity that slightly improves in the first half of the study (Group 3) (Figure S29, Table S46). Further assessment of model fit parameters suggest that the selected three-group model adequately fits the underlying data (Table S47). Estimated spaghetti plots of each trajectory group (Figure S30) suggests that the three groups are relatively well separated for one another, and that variance is approximately homogeneous within and across groups. As in other models, the floor effect observed for Group 1 is accounted for by the use of the censored normal option within models.

Including dropout and sex parameters in the final model of sum of general symptom severity did not lead to substantial change in final model estimates or adequacy parameters. Though dropout parameters were significant, estimated trajectories appear very similar across models that exclude and include dropout parameters (Figure S31), model parameters remained consistent (Table S48), and group sizes changed by less than 0.5% across groups (Table S49). Thus, the effects of attrition are not expected to have a noticeable effect on general symptom trajectory modeling in multi-trajectory models.

Table S45. Diagnostics of group-based model adequacy for sum of general symptom severity model comparison.

Number of groups	BIC (*N*=1756)	BIC (*N*=422)	Log likelihood	Average posterior probability	Odds of correct classification	Percent assigned to group
1	-4927.27	-4924.42	-4912.33	.	.	100
2	-4718.78	-4713.07	-4688.89	.93 .92	12 13	53 47
3	-4696.39	-4687.83	-4651.56	.88 .82 .89	22 6 17	24 44 32
4	-4702.53	-4691.13	-4642.77	.76 .86 .72 .89	6 19 25 18	35 24 10 31

Figure S28. Model comparison plots of estimated trajectories (dashed) and mean score for sum of general symptom severity at each visit (solid) for derived trajectory groups in each model.

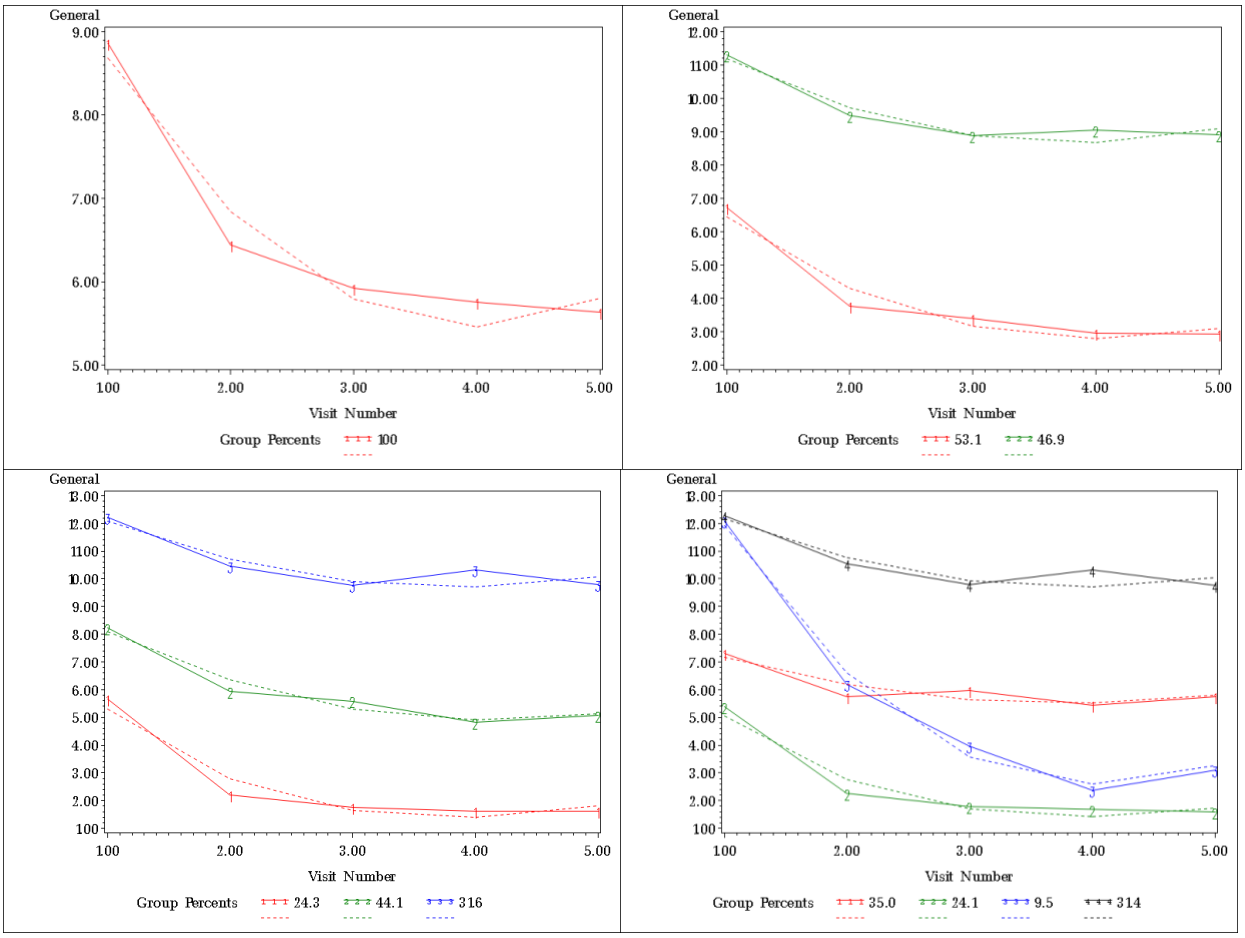


Figure S29. Plot of estimated trajectories (dashed) and mean general symptom severity sum values at each visit (solid) for the preferred three-group sum of general symptom severity model.

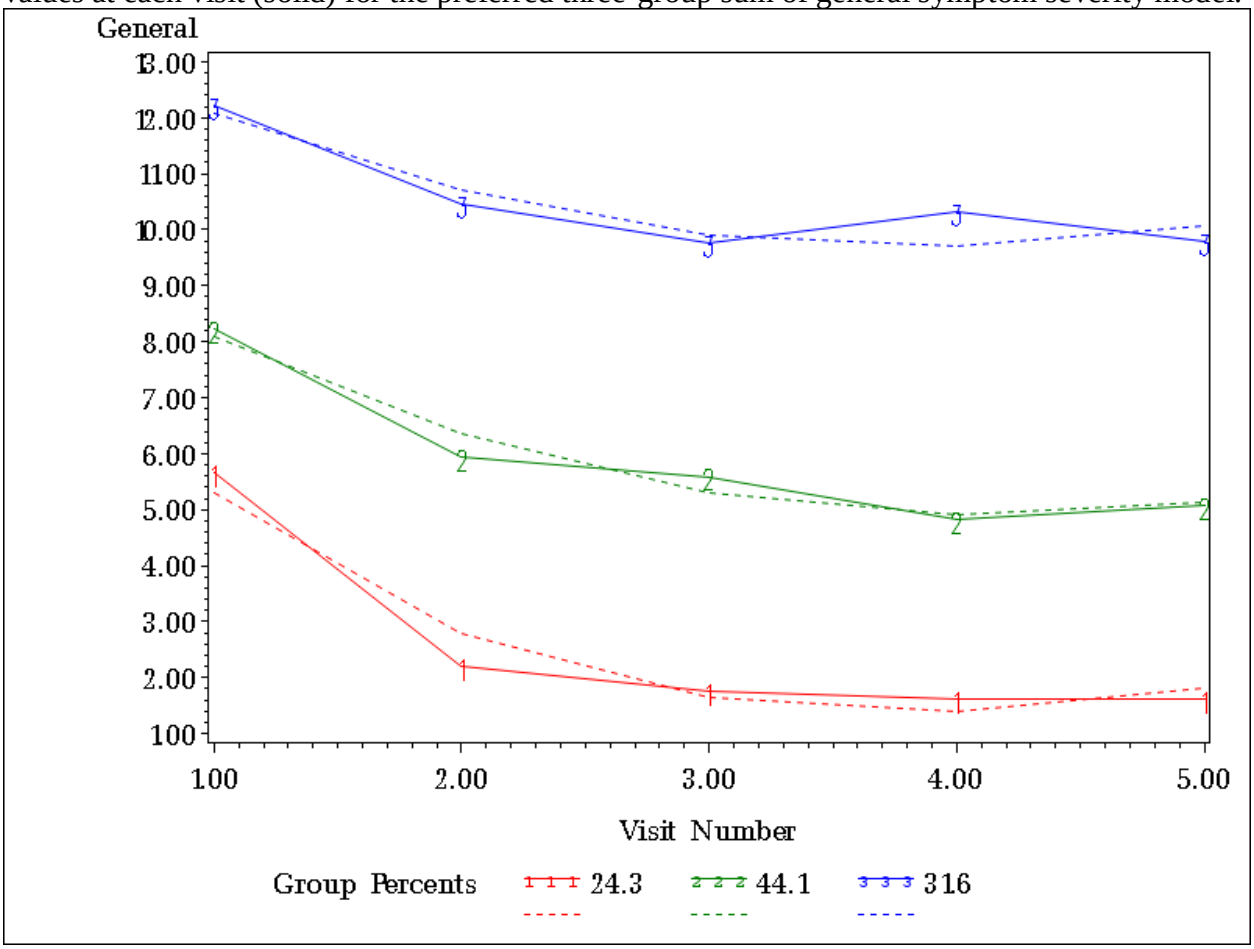


Table S46. Model parameters for the preferred three-group sum of general symptom severity model.

Group	Parameter	Estimate	SE	T	P
1	Intercept	9.39	.95	9.895	< .001
	Linear	-4.78	.78	-6.127	< .001
	Quadratic	.61	.13	4.810	< .001
2	Intercept	10.52	.71	14.819	< .001
	Linear	-2.78	.52	-5.327	< .001
	Quadratic	.34	.09	3.918	< .001
3	Intercept	14.04	.74	19.057	< .001
	Linear	-2.25	.57	-3.972	< .001

	Quadratic	.29	.09	3.084	.002
	Sigma	3.37	.07	50.656	< .001
Group Membership					
1	(%)	24.31	4.09	5.938	< .001
2	(%)	44.09	3.91	11.286	< .001
3	(%)	31.61	3.61	8.754	< .001

Table S47. Diagnostics of group-based model adequacy for the preferred three-group sum of general symptom severity model.

Class	Probability of group membership	95% CI	Proportion assigned to group	Average posterior probability	Odds of correct classification
1	.243	[.163, .323]	.225	.876	22.0
2	.441	[.365, .517]	.460	.822	5.9
3	.316	[.245, .387]	.315	.889	17.3

Figure S30. Spaghetti plots of individual longitudinal data and estimated trajectories for the preferred three-group sum of general symptom severity model.

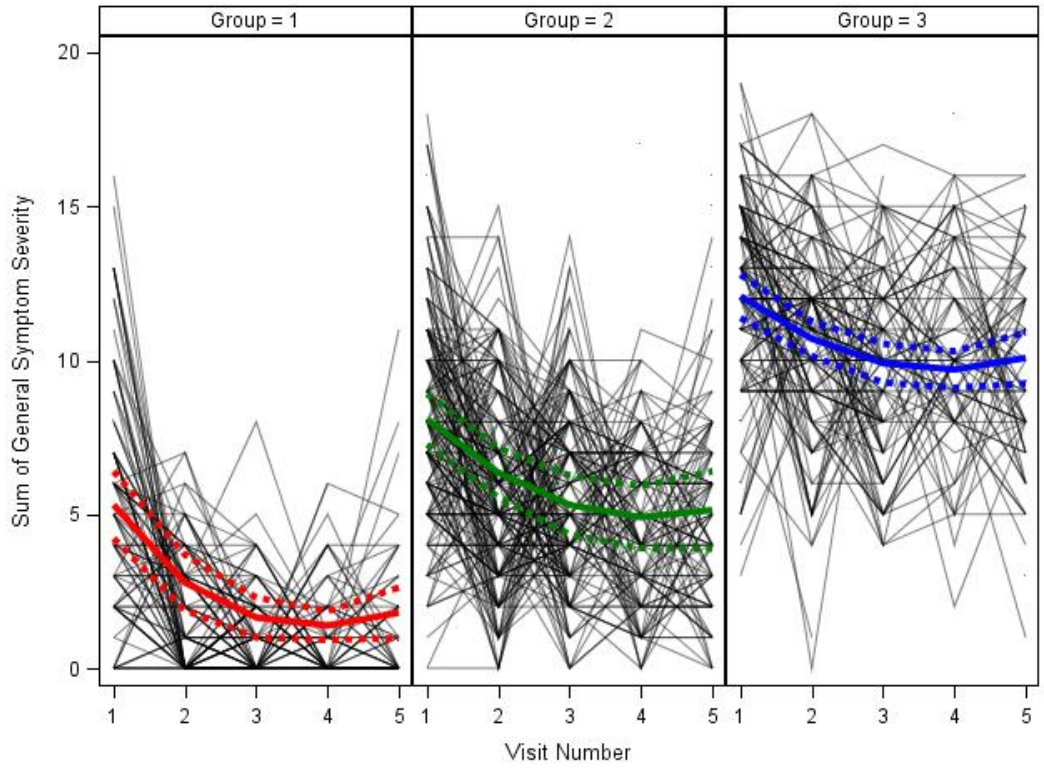


Figure S31. Comparison of the sum of general symptom severity final model (panel 1) with the final general model when sex and dropout at each visit is included in the model (panel 2).

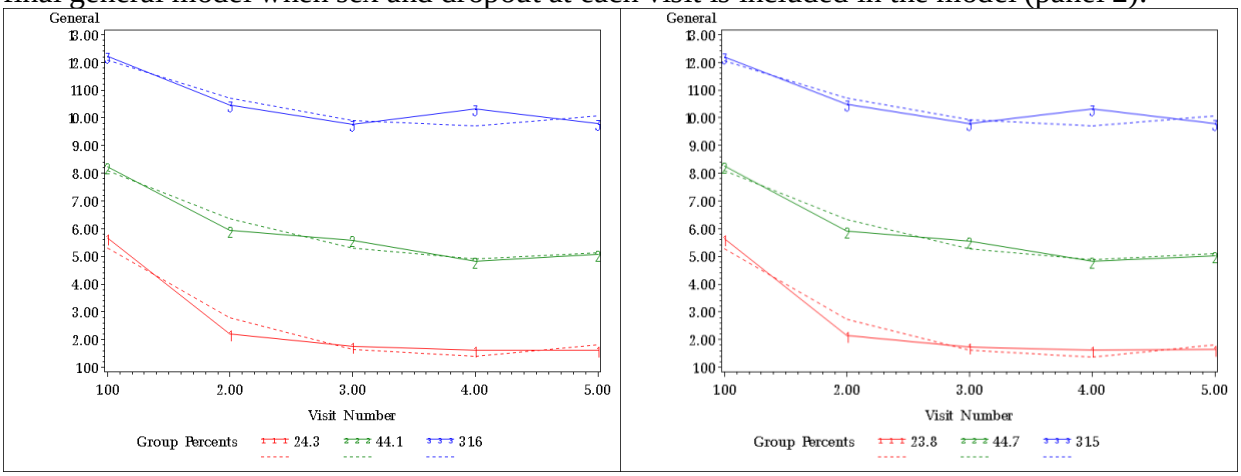


Table S48. Model parameters for the preferred three-group sum of general symptom severity model, including dropout and sex parameters.

Group	Parameter	Estimate	SE	T	P
1	Intercept	9.39	.96	9.777	< .001
	Linear	-4.84	.80	-6.089	< .001
	Quadratic	.63	.13	4.819	< .001
2	Intercept	10.57	.70	15.030	< .001
	Linear	-2.82	.52	-5.452	< .001
	Quadratic	.34	.09	4.000	< .001
3	Intercept	13.98	.74	18.993	< .001
	Linear	-2.21	.57	-3.871	< .001
	Quadratic	.28	.10	2.984	.003
1	Drop0	-2.43	.22	-10.990	< .001
2	Drop0	-2.26	.15	-15.092	< .001
3	Drop0	-2.23	.17	-13.343	< .001
	Sigma	3.38	.07	50.541	< .001
Group Membership					
1	Constant (ref group)	.	.	.	.
2	Constant	-0.01	.51	-.018	.986
	Sex	.48	.35	1.353	.176

3	Constant	-1.17	.50	-2.343	.019
	Sex	1.04	.32	3.190	.001

BIC (N=1756): -5199.33 *BIC (N=422):* -5187.2 *L:* -5135.83

Table S49. Diagnostics of group-based model adequacy for the three-group sum of general symptom severity model with dropout and sex parameters included, compared to the base three-group general symptoms model.

	Base model		*Model with dropout and sex*	
Class	Probability of group membership	Average posterior probability	Probability of group membership	Average posterior probability
1	.243	.876	.238	.871
2	.441	.822	.447	.827
3	.316	.889	.315	.892

Summary

Overall, preferred univariate models of GAF and the four sum of symptom severity measures suggest that there are 3-4 distinct trajectory groups for each outcome that should be represented in multi-trajectory models. In general, trajectory groups within each outcome ranged from low severity that approached remission to high severity that remained chronically poor across the two-year assessment period, with one or two moderate groups between these two extremes. Overall, assumptions of heterogeneity within and between trajectory groups appeared to be met. Continuing to use the censored normal option within modeling will be important to minimize the influence of floor effects, particularly for Groups 1 and 2 across outcomes. Large improvements between the baseline and six-month visits were observed across most outcomes for all but the most severe groups, suggesting that substantial change may occur during this period for many prodromal individuals. Future studies that include more assessments during this period and are adequately powered to estimate quadratic trajectories may further clarify common patterns of change in symptom severity and functioning that occur soon after seeking treatment among prodromes.

Models that included dropout parameters and accounted for higher rates of missing data among males in univariate models of GAF and symptom severity sums did not result in appreciable change in trajectory estimates or probabilities of group membership. Though it is unfortunate that dropout and sex cannot be accounted for in the final multivariate model due to challenges with model conversion, these results suggest that the estimates identified in the multivariate model are not substantially influenced by missing data and can be interpreted with reasonable confidence.

Guidelines for Reporting on Latent Trajectory Studies (GRoLTS) Checklist
NAPLS1

1) *Metric of time used in the statistical model*

Timepoints used in the model corresponded to the number of months since the baseline assessment. Assessments took place approximately every 6 months. The five assessments timepoints used in analyses were labeled: baseline, 6 months, 12 months, 18 months, 24 months. A sixth visit (30+ months) was included for some individuals in NAPLS1; however, to facilitate comparison with the NAPLS2 dataset, data from the 6th visit were excluded from all analyses.

2) *Information presented about the mean and variance of time within a wave*

Unfortunately, specific dates of assessment were not available for the NAPLS1 dataset. The allowable range for dates of visit assessment (in months since baseline assessment) for each visit were as follows: *6-month visit*: 3-8 months, *12-month visit:* 9-14 months, *18-month visit*: 15-20 months, *24-month visit*: 21-26 months, *30(+) month visit*: 27-30+ months. As the design of NAPLS1 is very similar to the design of NAPLS2, and the NAPLS2 distribution of assessment dates around each visit was relatively similar across visits, we expect that the mean and variance of assessment dates do not vary substantially across waves in NAPLS1.

3a) *Missing data mechanism reported*

In PROC TRAJ, missing data is assumed to be missing at random (MAR). This assumes that attrition and missing data are independent of unobserved outcomes, conditional on observed outcome values and any covariates included in the model (14).

3b) *Description of what variables are related to attrition/missing data*

Over the course of the two-year study, 70.68% (*n* = 94) of the analytic sample (*n* = 133) missed at least one of the five assessment sessions. 32.33% (*n* = 43) missed only one visit, and 38.35% (*n* = 51) missed two visits. Of those with missing data, 71.28% (*n* = 67) dropped out of the study, whereas the remaining 28.72% missed an earlier visit but returned for the final visit. To assess the potential for selective missing data effects to bias results, we compared individuals with missing data to those with no missing data on demographics and clinical characteristics available at baseline (Table S50). The participants with missing data were significantly more likely to have lower GAF scores at baseline relative to individuals who completed all five visits. Unfortunately, detailed information regarding loss to follow-up was not available in the NAPLS1 dataset.

Table S50. Demographic, baseline clinical, and study participation differences between members of the NAPLS1 analytic sample who completed all five assessments relative to individuals with missing data from at least one assessment.

	No missing sessions (*n* = 29)	At least 1 missing session (*n* = 94)	F/χ^2	*p*
Demographics				
Age (years)	18.88 (5.55)	17.65 (4.10)	2.01	.159

Sex (% male)	74.36%	56.38%	3.767	.052
Parent Education (completed HS or less)	16.67%	11.24%	.677	.411
Race			2.372	.306
White	87.18%	75.53%		
African-American	5.13%	7.45%		
Other	7.69%	17.02%		
Lifetime anxiety disorder	28.21%	46.24	3.696	.055
Lifetime depressive disorder	53.85%	51.61%	.055	.815
Lifetime non-Cluster A disorder	25.64%	30.14%	.252	.616
Mean number of disorders across the above domains	1.08 (1.06)	1.22 (.93)	.560	.457
Baseline symptom severity				
GAF	53.08 (11.96)	48.55 (12.01)	3.92	.050
Positive Sum	10.49 (3.52)	10.88 (3.62)	.33	.564
Negative Sum	10.64 (5.59)	10.84 (6.91)	.030	.875
Disorganized Sum	5.53 (3.67)	5.46 (3.31)	.010	.923
General Sum	8.13 (4.35)	7.67 (4.52)	.290	.590

3c) *Description of how missing data in the analyses were dealt with*

Detailed information about reasons for missing data was not available in the NAPLS1 dataset. To support the MAR assumption, lower GAF scores among individuals with missing data relative to those without missing data needed to be addressed to support accurate estimates of model parameters. Additionally, differential rates of missing data across derived trajectory groups could also lead to bias in estimation of model parameters. In standard group-based trajectory models, the probability of trajectory group membership is expected to remain consistent across the length of the study and estimated trajectory probabilities do not account for missing data or attrition (14). If dropout rates are relatively consistent among trajectory groups, model estimation is relatively robust. However, differential rates of dropout can lead to mild underestimation of group size probabilities and overestimation of trajectory estimate values at later visits within those groups (14).

The potential influence of baseline GAF and differential rates of missing data across trajectory groups were addressed by incorporating dropout and baseline GAF parameters into univariate trajectory models. The addition of dropout parameters allowed for estimation of rates of missing data within each identified trajectory group, and adjusted accordingly for potential inflation of the expected size of the trajectory groups and parameter estimates at later visits (14). Including baseline GAF as a covariate in the model accounted for lower mean GAF scores for those with missing data in dropout parameter estimation as well (14).

Due to the complexity of multivariate models, inclusion of dropout and covariate parameters in multi-trajectory models led to model instability and could not be estimated. To address this concern, we re-ran final univariate models with the addition of parameters to assess

dropout and sex, and assessed similarities and differences with the original final model (see Assessing the Influence of Attrition section below). In sum, we found that accounting for dropout did not substantially affect models, addressing this concern. We used knowledge obtained from the univariate models and assessed the mean number of visits among the multi-trajectory groups to inform the interpretation of trajectory group sizes and estimates. For trajectory groups in which higher rates of dropout were suspected, group estimates were assumed to be slightly deflated and later trajectory estimates were expected to be slightly inflated (24).

4-16) same approach taken as in NAPLS2 analyses (see above)

Univariate Model Derivation

As in the NAPLS2 dataset, we first derived univariate group-based trajectory models for each of our outcomes (GAF, sum of positive/negative/disorganized/general symptoms), beginning with a one-group model and continuing until model fit parameter thresholds were crossed (e.g., percent assigned to group below 5%, negligible change in BIC, average posterior probability below 0.7, odds of correct classification below 5) or additional trajectory groups derived were not of theoretical interest (15). We used the best supported univariate models to assess assumptions of group-based trajectory modeling, including homogeneity of variance between and within trajectory groups (which, if invalid, could lead to over-extraction of groups). As the intention of the NAPLS1 sample was to assess pattern replication of the four-group multi-trajectory model derived in the NAPLS2 sample, factors that could influence the likelihood of replicability (e.g., similarity of number of trajectory groups supported, range of values observed) were also considered.

Group-based trajectory modeling assumes that data are missing at random, accounting for observed variables included in the model (14). Additionally, differential rates of dropout between groups can also differential rates of dropout can lead to mild underestimation of group size probabilities and overestimation of trajectory estimate values at later visits within those groups (14). Unfortunately, dropout and baseline GAF could not be included in final multivariate models due to issues with model convergence. To assess the potential influence of lower baseline GAF scores among those with more missing data and any differences in rates of dropout across groups on model parameters, we re-ran final univariate models that included baseline GAF and dropout parameters as well.

Global Assessment of Functioning

Group-based trajectory models of GAF were supported up to three trajectory groups (Table S51, Figure S32). The four-group model was rejected due to a group size of 2% (Table S51). The derived trajectories in NAPLS1 appeared very similar in intercepts and slopes to the final three-group model selected in NAPLS2 upon visual inspection (Figure S33, Table S52).

The selected three-group model for NAPLS1 suggests that common GAF trajectories include: poor functioning with slight improvement across time (Group 1), moderate functioning that improves most rapidly during the first portion of the study, and high functioning that improves most rapidly during the beginning of the study (Figure S34). Further assessment of model fit parameters suggest that the selected three-group model adequately fits the underlying data (Table S53). Estimated spaghetti plots of each trajectory group (Figure S35) suggests that the three groups are relatively well separated from one another, and that variance is approximately homogeneous within and across groups. As discussed in the NAPLS2 section,

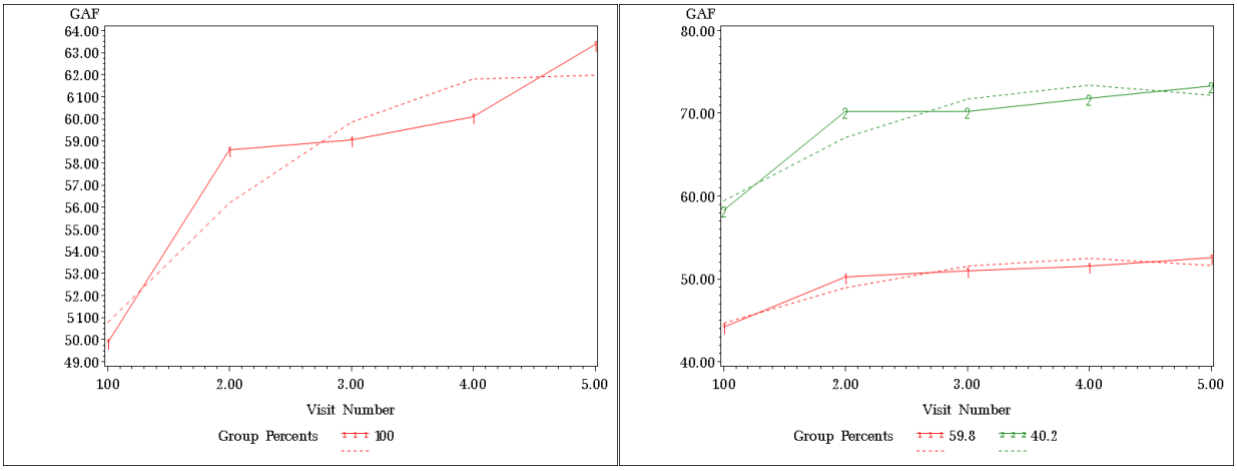
future studies that include more frequent assessments within the first year following help-seeking can further explore patterns of change within Groups 1 and 2 and will hopefully yield further insight into changes that might be occurring during this period.

Baseline GAF could not be included as a risk factor due to lack of model convergence. Including dropout parameters in the final model of GAF did not lead to substantial change in final model estimates or adequacy parameters. Though dropout parameters were significant, estimated trajectories appear very similar across models that exclude and include dropout parameters (Figure S36), model parameters remained consistent (Table S54), and group sizes changed by less than 1% at most (Table S55). Thus, the effects of attrition are not expected to have an appreciable effect on GAF trajectory modeling in multi-trajectory models.

Table S51. Diagnostics of group-based model adequacy for Global Assessment of Functioning model comparison.

Number of groups	BIC (*N*=496)	BIC (*N*=133)	Log likelihood	Average posterior probability	Odds of correct classification	Percent assigned to group
1	-2014.63	-2012.00	-2002.22	..	..	100
2	-1949.26	-1944.00	-1924.44	.95 .93	13 20	60 40
3	-1945.39	-1937.49	-1908.15	.82 .87 .90	15 7 21	23 48 29
4	-1941.11	-1930.58	-1891.46	.99 .86 .81 .91	42,436 14 6 30	2 30 41 26

Figure S32. Model comparison plots of estimated linear trajectories (dashed) and mean GAF score at each visit (solid) for derived trajectory groups in each model.



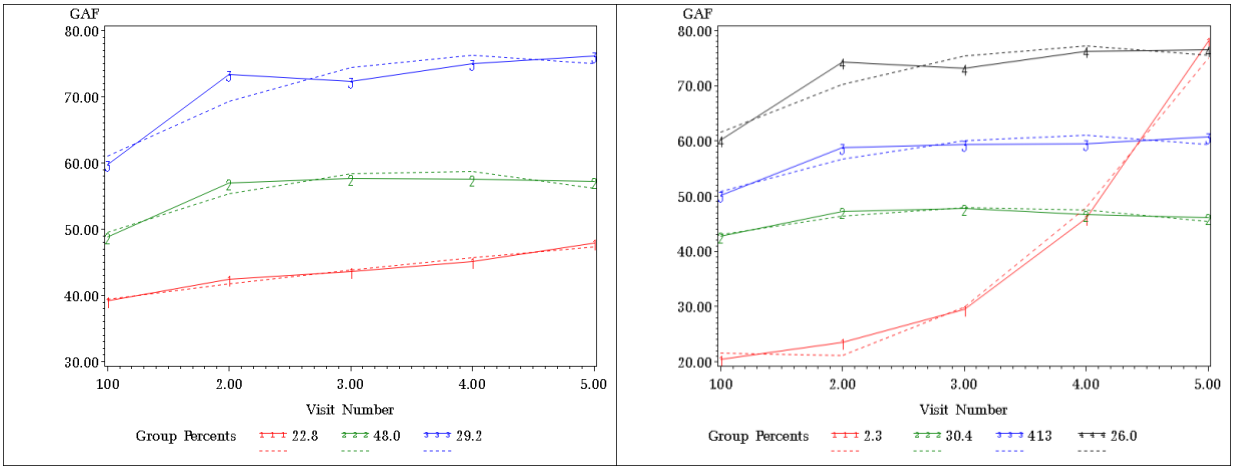


Figure S33. Model comparison plots of the selected three-group model for GAF in NAPLS2 (left) and the corresponding three-group model in NAPLS1 (right).

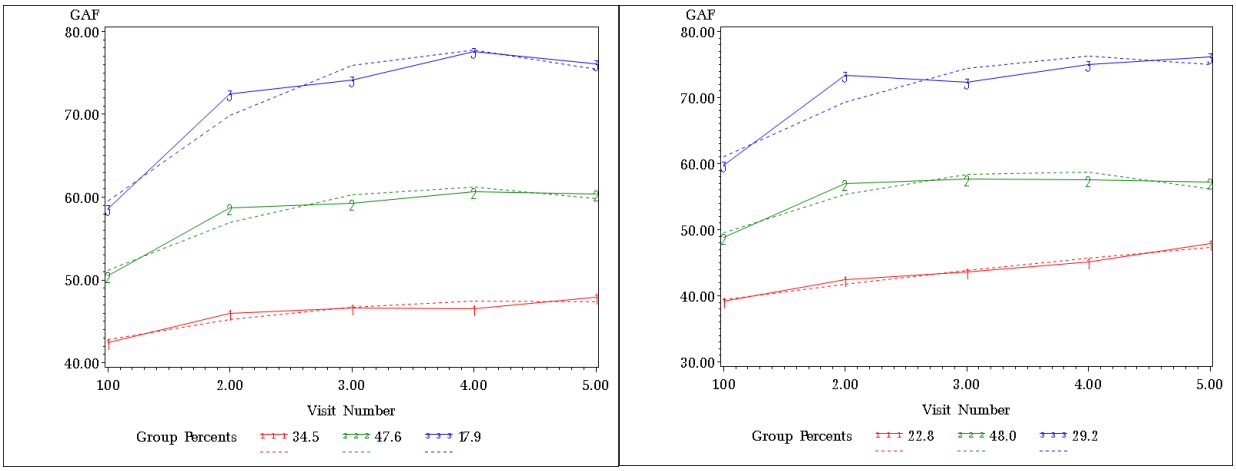


Figure S34. Plot of estimated linear trajectories (dashed) and mean GAF values at each visit (solid) for the preferred three-group GAF model.

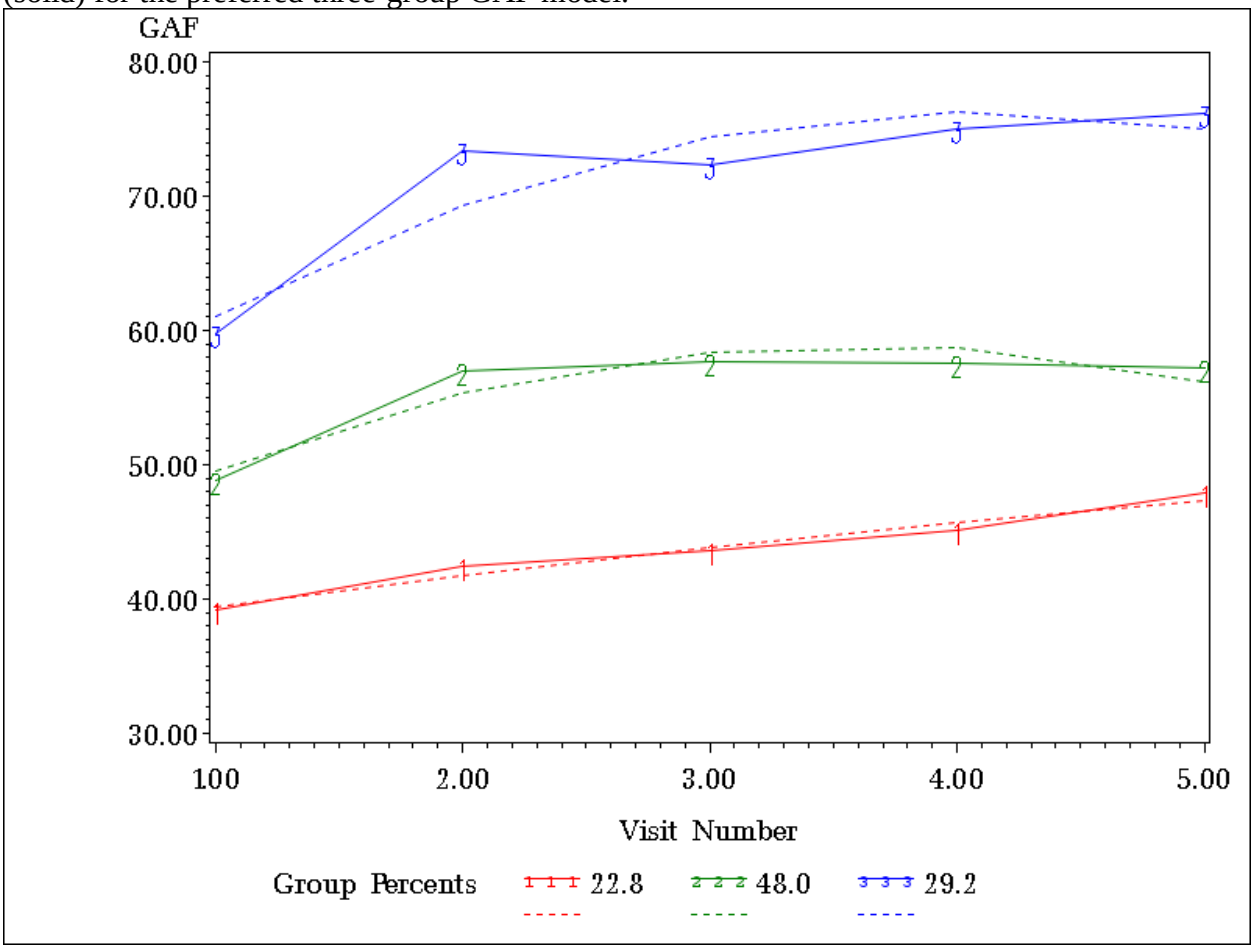


Table S52. Model parameters for the preferred GAF model in NAPLS1.

Group	Parameter	Estimate	SE	T	P
1	Intercept	36.97	4.59	8.059	< .001
	Linear	2.57	3.61	.712	.477
	Quadratic	-.10	.63	-.157	.875
2	Intercept	40.88	3.15	12.970	< .001
	Linear	9.99	2.55	3.918	< .001
	Quadratic	-1.38	.45	-3.077	.002
3	Intercept	49.66	3.63	13.691	< .001
	Linear	13.00	2.95	4.408	< .001
	Quadratic	-1.59	.50	-3.194	.002

	Sigma	9.30	.33	28.410	< .001
Group Membership					
1	(%)	22.77	6.73	3.382	< .001
2	(%)	48.02	6.54	7.336	< .001
3	(%)	29.21	5.49	5.317	< .001

Table S53. Diagnostics of group-based model adequacy for the preferred three-group GAF model.

Class	Probability of group membership	95% CI	Proportion assigned to group	Average posterior probability	Odds of correct classification
1	.228	[.097, .359]	.233	.819	15.3
2	.480	[.353, .607]	.466	.873	7.4
3	.291	[.184, .400]	.301	.897	21.1

Figure S35. Spaghetti plots of individual longitudinal data and estimated trajectories for the preferred three-group GAF model.

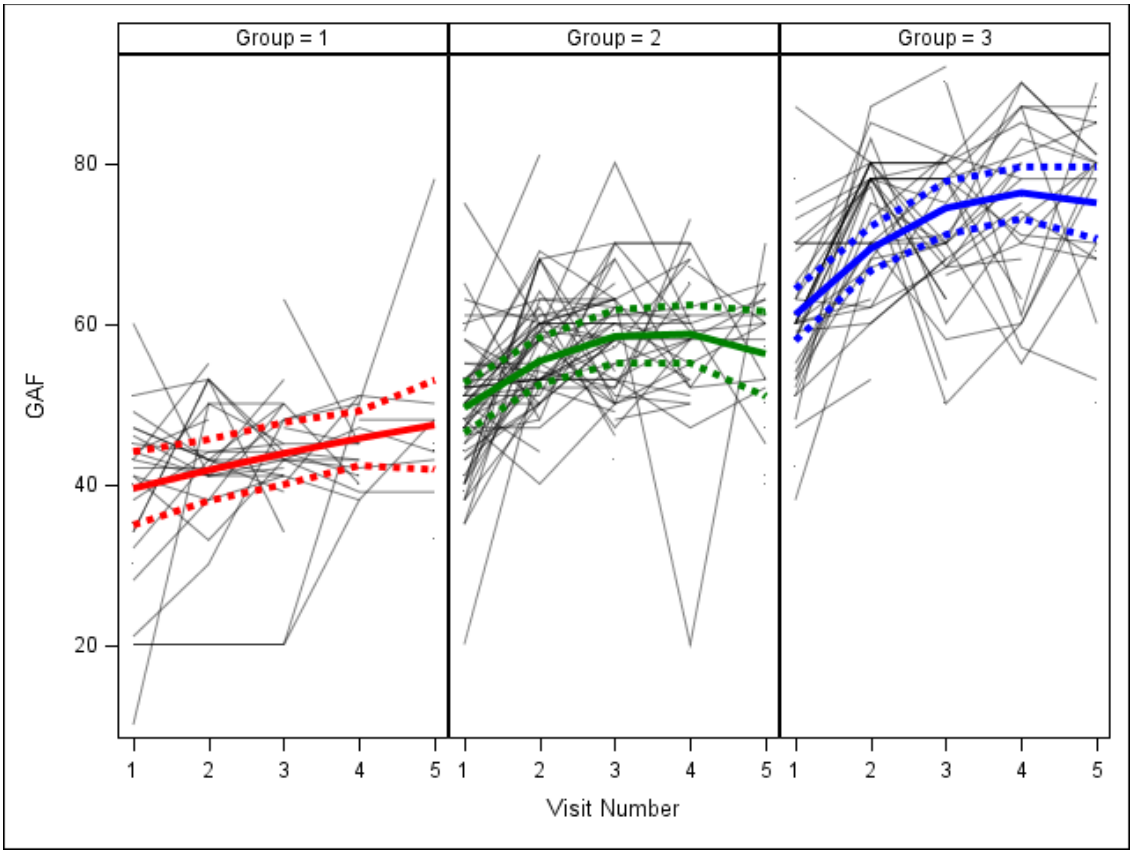


Figure S36. Comparison of GAF final model (panel 1) with final GAF model when dropout at each visit is included in the model (panel 2).

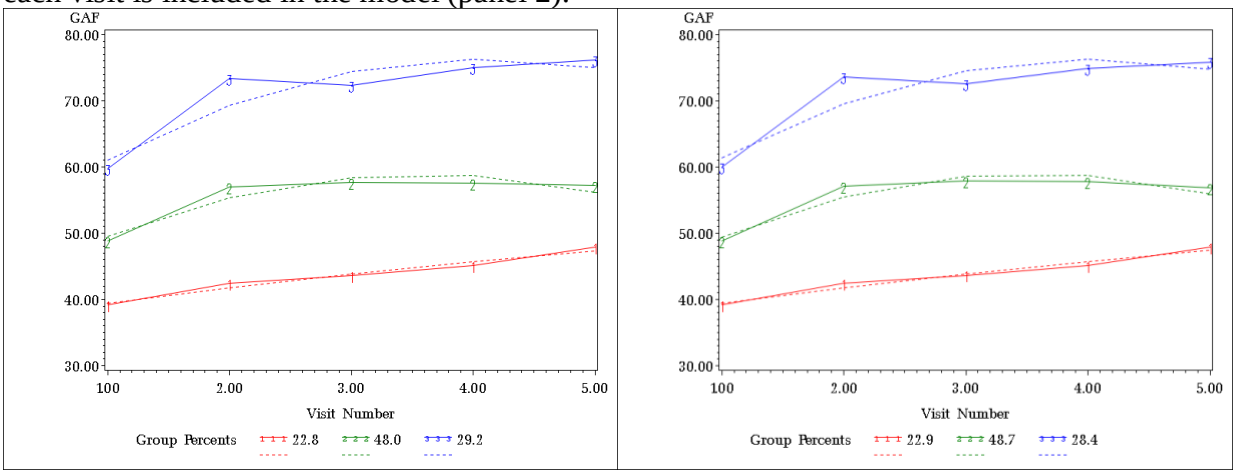


Table S54. Model parameters for the preferred three-group GAF model, including dropout parameters.

Group	Parameter	Estimate	SE	T	P
1	Intercept	37.04	4.57	8.10	<.001
	Linear	2.53	3.62	.70	.485
	Quadratic	-.09	.63	-.15	.885
2	Intercept	40.48	3.08	13.16	<.001
	Linear	10.46	2.51	4.16	<.001
	Quadratic	-1.47	.44	-3.33	<.001
3	Intercept	49.91	3.62	13.80	<.001
	Linear	13.07	2.92	4.48	<.001
	Quadratic	-1.62	.49	-3.30	.001
1	Drop0	-1.53	.28	-5.49	<.001
2	Drop0	-1.56	.19	-8.29	<.001
3	Drop0	-2.41	.34	-7.03	<.001
	Sigma	9.29	.33	28.4	<.001
Group Membership					
1	%	22.89	6.50	3.53	<.001
2	%	48.67	6.49	7.50	<.001
3	%	28.44	5.17	5.50	<.001

BIC (N=496): -2155.63 *BIC (N=133):* -2145.75 *L:* -2109.08

Table S55. Diagnostics of group-based model adequacy for the three-group GAF model with dropout included compared to the base three-group GAF model.

	Base model		*Model with dropout*	
Class	Probability of group membership	Average posterior probability	Probability of group membership	Average posterior probability
1	.228	.819	.229	.833
2	.480	.873	.487	.864
3	.291	.897	.284	.927

Sum of Positive Symptom Severity

Group-based trajectory models of the sum of positive symptom severity were supported up to two trajectory groups (Table S56, Figure S37). The three-group model was rejected due to a negligible increase in BIC value relative to the two-group model (Table S56). However, a four-group model was derived to assess the distribution of data to facilitate comparison with the final four-group model selected in NAPLS2. Upon visual inspection, the Group 1 intercept for NAPLS1 trajectories was higher than the Group 1 intercept observed for NAPLS2, and other intercepts were similar across groups (Figure S38). Steeper rates of improvement were observed for all four trajectory groups in the NAPLS1 sample relative to the NAPLS2 sample, with particularly rapid improvement between the baseline and 6-month visits. It is possible that these higher levels of baseline positive symptoms and more rapid improvement observed in NAPLS1 relative to NAPLS2 could influence pattern replicability and should be kept in mind when comparing and interpreting multi-group models derived in each sample.

In the NAPLS1 sample, the selected two-group model suggests that common trajectories for the sum of positive symptoms include low and improving symptom severity across time that almost reaches zero severity (Group 1) and a higher severity that also improves with time (Group 2) (Figure S39, Table S57). Further assessment of model fit parameters suggest that the selected two-group model adequately fits the underlying data (Table S58). Estimated spaghetti plots of each trajectory group (Figure S40) suggests that the two groups are relatively well separated from one another and that variance is approximately homogeneous within and across groups. A floor effect was observed for Group 1, which could theoretically influence the assumptions of homogeneity. However, all outcomes were modeled using the censored normal (also known as tobit) option in the group-based trajectory modeling code, which accounts for clustering of observation at the minimum or maximum values of a scale and prevents over-extraction of groups or misestimation of trajectory parameters due to associated non-normality (15). Across both groups, rapid improvement was observed in the sum of positive symptom severity between baseline and the 6-month visit. We hope to further explore this phenomenon in future studies with more frequent assessments during this period.

Including dropout and baseline GAF parameters in the final model of positive symptoms did not lead to substantial change in final model estimates or adequacy parameters. Though dropout parameters were significant, estimated trajectories appear very similar across models that exclude and include dropout parameters (Figure S41), model parameters remained consistent (Table S59), and group sizes changed by 1.1% at most (Table S60). Baseline GAF was not

significant as a risk factor. Thus, the effects of attrition are not expected to have an appreciable effect on positive symptom trajectory modeling in multi-trajectory models.

Table S56. Diagnostics of group-based model adequacy for sum of positive symptom severity model comparison.

Number of groups	BIC (*N*=520)	BIC (*N*=133)	Log likelihood	Average posterior probability	Odds of correct classification	Percent assigned to group
1	-1416.82	-1414.09	-1404.31	..	..	100
2	-1363.71	-1358.29	-1338.69	.95 .93	18 15	52 48
3	-1363.50	-1355.32	-1325.97	.92 .89 .85	15 11 36	44 42 14
4	-1373.93	-1362.62	-1323.50	.87 .69 .90 .85	10 56 12 39	40 4 43 13

Figure S37. Model comparison plots of estimated linear trajectories (dashed) and mean sum of positive symptom severity values at each visit (solid) for derived trajectory groups in each model.

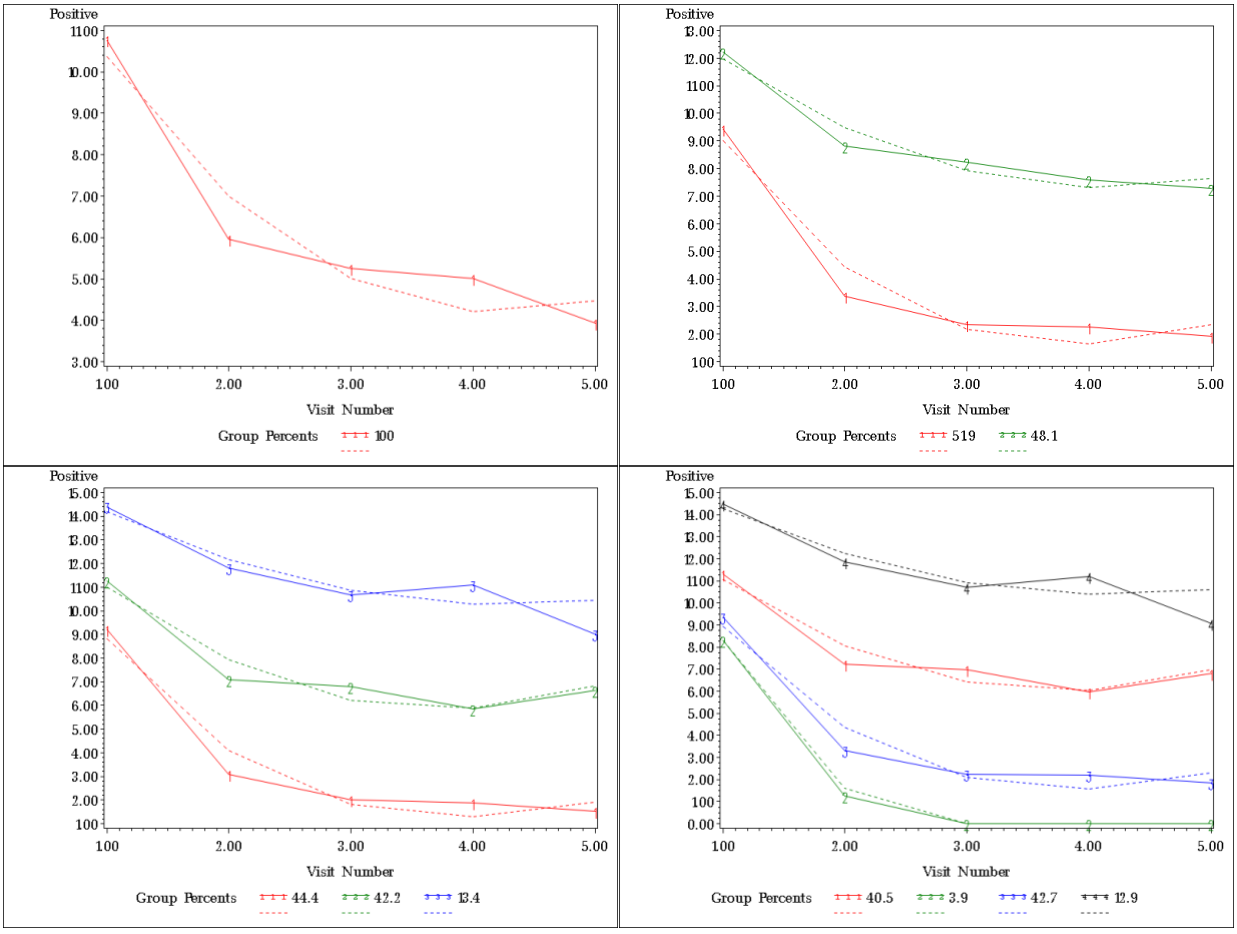


Figure S38. Model comparison plots of the selected four-group model for sum of positive symptom severity in NAPLS2 (left) and the corresponding four-group model in NAPLS1 (right). Note that the y-axis on the NAPLS2 graph ends at 16, whereas the y-axis on the NAPLS1 graph ends at 15.

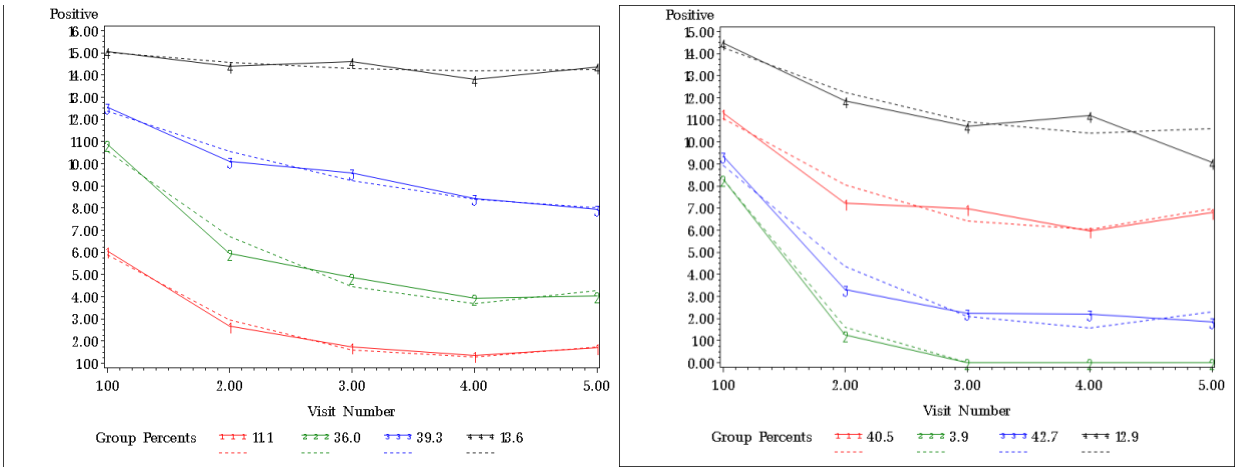


Figure S39. Plot of estimated linear trajectories (dashed) and mean positive symptom severity sum values at each visit (solid) for the preferred two-group sum of positive symptom severity model.

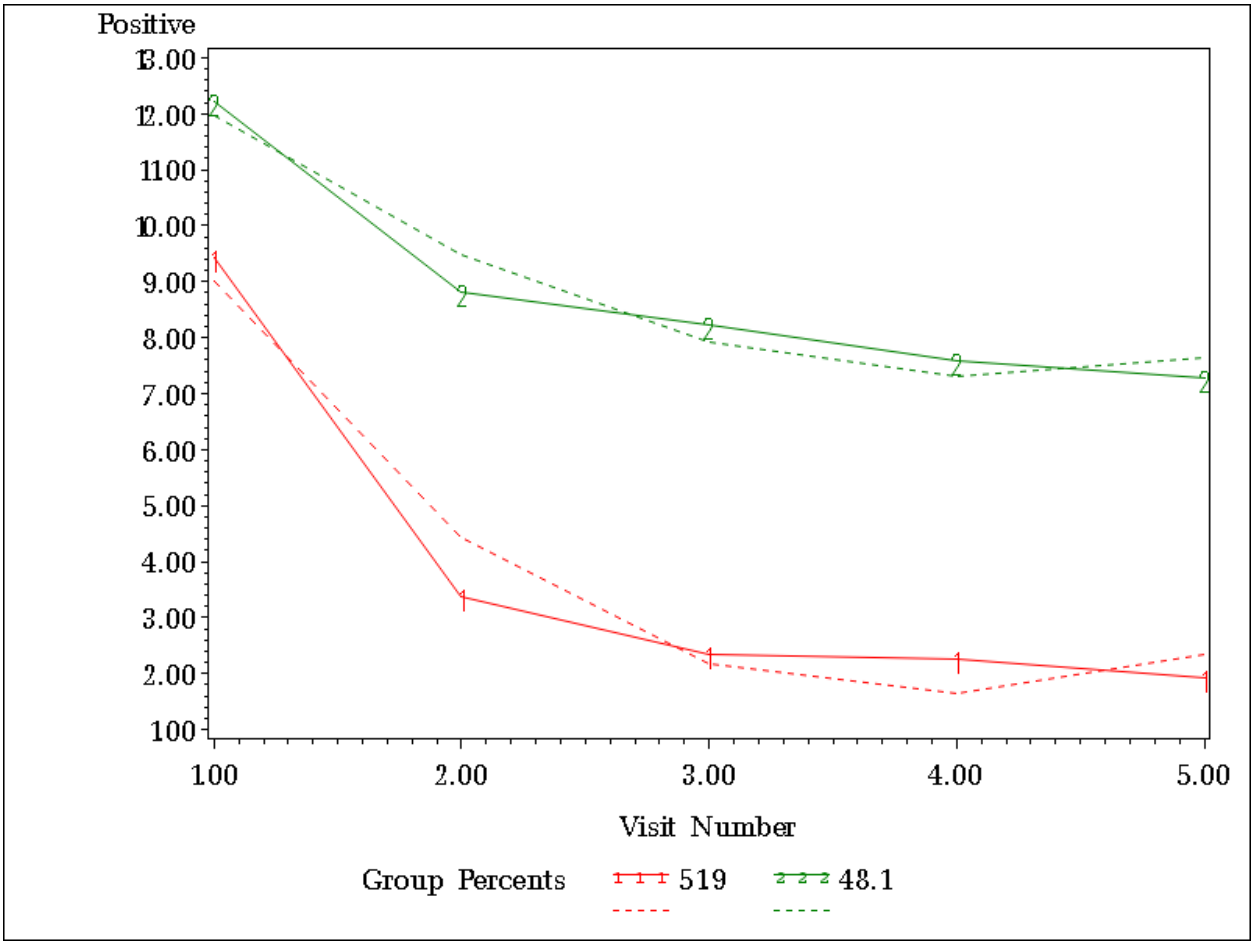


Table S57. Model parameters for the preferred two-group sum of positive symptom severity model.

Group	Parameter	Estimate	SE	T	P
1	Intercept	15.68	.94	16.706	<.001
	Linear	-7.63	.76	-9.987	<.001
	Quadratic	.97	.13	7.508	<.001
2	Intercept	15.42	.98	15.762	<.001
	Linear	-3.92	.82	-4.806	<.001
	Quadratic	.47	.14	3.329	<.001
	Sigma	3.23	.11	28.586	<.001
Group Membership					
1	(%)	51.92	5.39	9.633	<.001
2	(%)	48.08	5.39	8.922	<.001

Table S58. Diagnostics of group-based model adequacy for the preferred two-group sum of positive symptom severity model.

Class	Probability of group membership	95% CI	Proportion assigned to group	Average posterior probability	Odds of correct classification
1	.519	[.413, .625]	.511	.950	17.6
2	.481	[.375, .587]	.489	.932	14.8

Figure S40. Spaghetti plots of individual longitudinal data and estimated trajectories for the preferred two-group sum of positive symptom severity model.

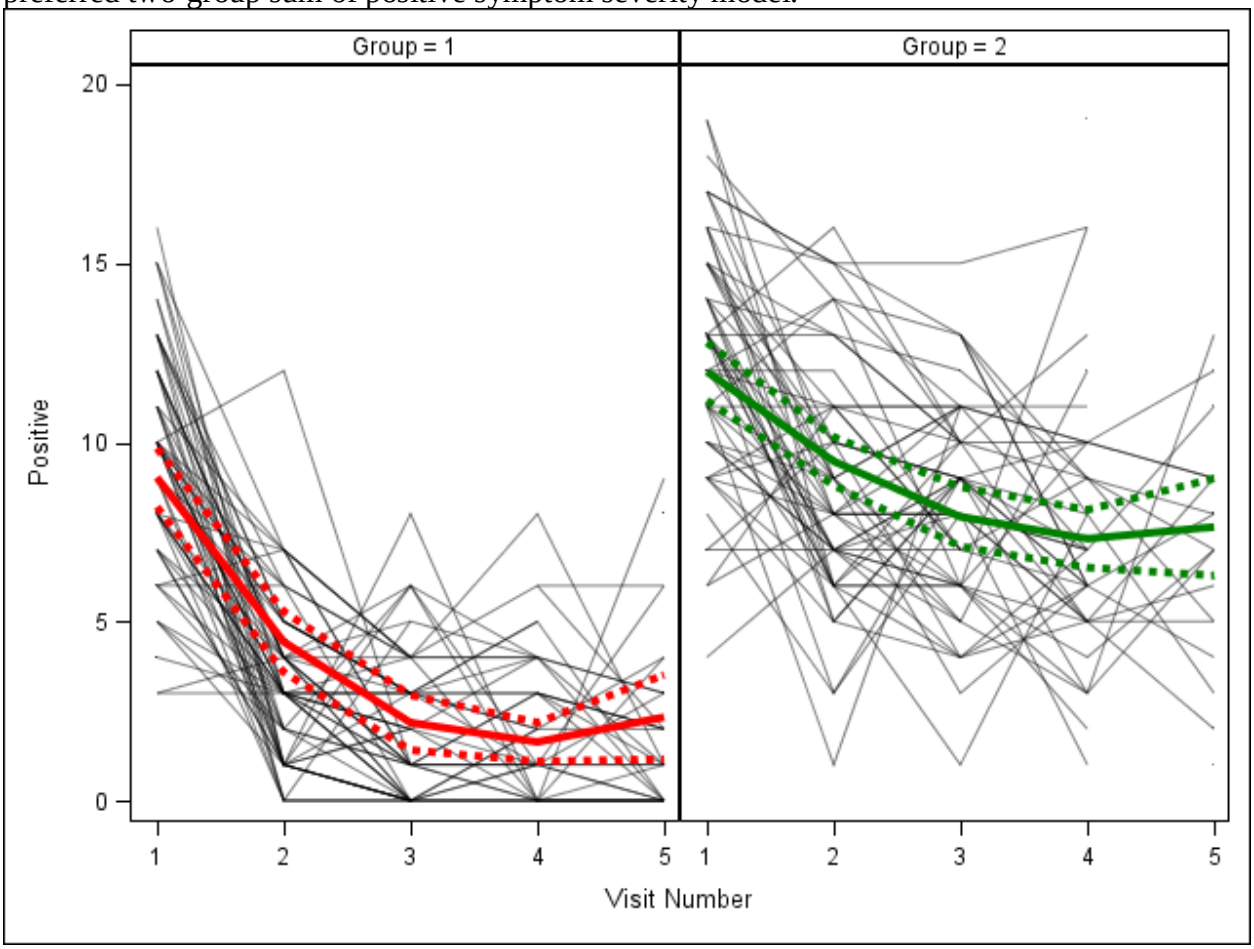


Figure S41. Comparison of the positive symptom final model (panel 1) with the positive symptom model when dropout and baseline GAF are included in the model (panel 2).

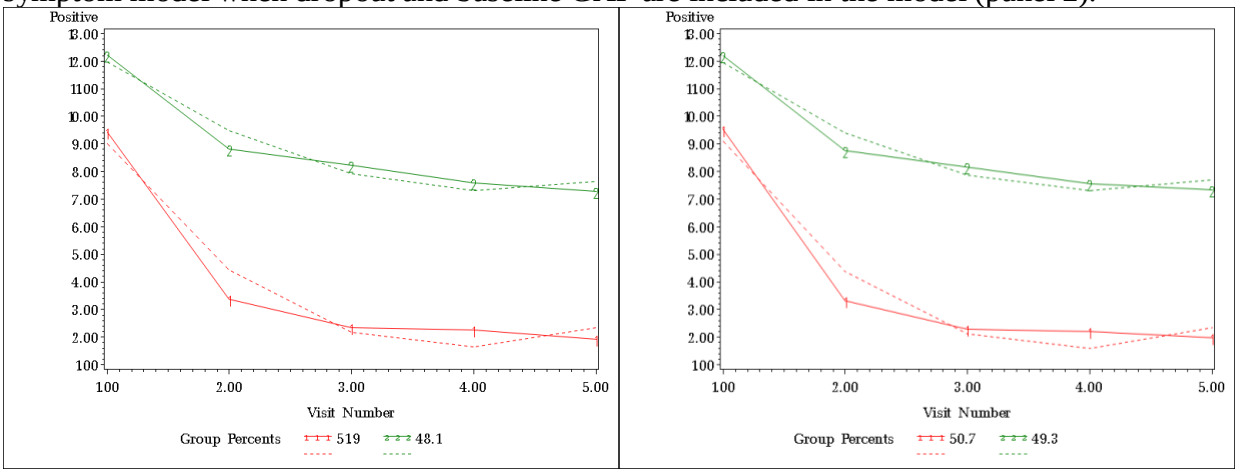


Table S59. Model parameters for the preferred two-group positive symptom model, including dropout and baseline GAF parameters.

Group	Parameter	Estimate	SE	T	P
1	Intercept	15.93	.95	16.75	<.001
	Linear	-7.85	.77	-10.18	<.001
	Quadratic	1.00	.13	7.71	<.001
2	Intercept	15.46	.97	15.92	<.001
	Linear	-4.01	.81	-4.95	<.001
	Quadratic	.49	.14	3.47	<.001
1	Drop0	-2.15	.22	-9.83	<.001
2	Drop0	-1.63	.18	-9.12	<.001
	Sigma	3.23	.11	28.32	<.001
Group Membership					
1	Constant	.	.	.	.
2	Constant	.97	.87	1.12	.264
3	Baseline GAF	-.02	.02	-1.19	.236

BIC (N=517): -1559.20 *BIC (N=132):* -1551.70 *L:* -1524.84

Table S60. Diagnostics of group-based model adequacy for the two-group positive symptom model with dropout and baseline GAF included compared to the base two-group positive symptom model.

	Base model		*Model with dropout and baseline GAF*	
Class	Probability of group membership	Average posterior probability	Probability of group membership	Average posterior probability
1	.519	.950	.508	.946
2	.481	.932	.492	.945

Sum of Negative Symptom Severity

Group-based trajectory models of the sum of negative symptom severity were supported up to three trajectory groups (Table S61, Figure S42). The four-group model was rejected due to a decrease in BIC (Table S61); however, the four-group model was examined in relation to the final four-group model selected in NAPLS2. Upon visual inspection, the four derived trajectories in NAPLS1 appear broadly similar in intercepts and slopes to the final four-group model selected

in NAPLS2 (Figure S43), aside from a more quadratic shape for Groups 3 and 4 in NAPLS1 relative to NAPLS2.

The selected three-group model suggests that common trajectories for the sum of negative symptoms include: very low symptom severity that decreases slightly across time that almost reaches zero severity (Group 1), moderate severity that improves particularly rapidly within the first six months (Group 2), and higher severity that fluctuates (Group 3) (Figure S44, Table S62). Further assessment of model fit parameters suggest that the selected three-group model adequately fits the underlying data (Table S63). Estimated spaghetti plots of each trajectory group (Figure S45) suggests that the three groups are relatively well separated from one another and that variance is approximately homogeneous within and across groups. Similar to the sum of positive symptom severity model, a floor effect is observed for Group 1 and is accounted for by the use of the censored normal option within models. Again, rapid improvement in the sum of negative symptom severity between baseline and the 6-month visit was present in Groups 1 and 2, which we hope to further explore in future studies.

Including dropout and baseline GAF parameters in the final model of negative symptoms did not lead to substantial change in trajectory shapes, but did influence group size estimates. Though dropout parameters were significant, estimated trajectories appear very similar across models that exclude and include dropout parameters (Figure S46) and model parameters remained consistent (Table S64). Group sizes changed by ~4% in Groups 1 and 2 (Table S65). Baseline GAF was significant as a risk factor. Thus, the effects of attrition are not expected to have an appreciable effect on negative symptom trajectory shapes, but may lead to overestimation of the size of Group 1 and underestimation of Group 2 size in multi-trajectory models.

Table S61. Diagnostics of group-based model adequacy for sum of negative symptom severity model comparison.

Number of groups	BIC (*N*=468)	BIC (*N*=133)	Log likelihood	Average posterior probability	Odds of correct classification	Percent assigned to group
1	-1483.18	-1480.66	-1470.88	..	..	100
2	-1399.97	-1394.94	-1375.38	.97 .92	9 35	76 24
3	-1387.37	-1379.82	-1350.48	.92 .86 .96	10 11 178	55 34 11
4	-1390.60	-1380.54	-1341.41	.84 .79 79 .93	11 6 19 118	33 39 17 10

Figure S42. Model comparison plots of estimated linear trajectories (dashed) and mean value of sum of negative symptom severity values at each visit (solid) for derived trajectory groups in each model.

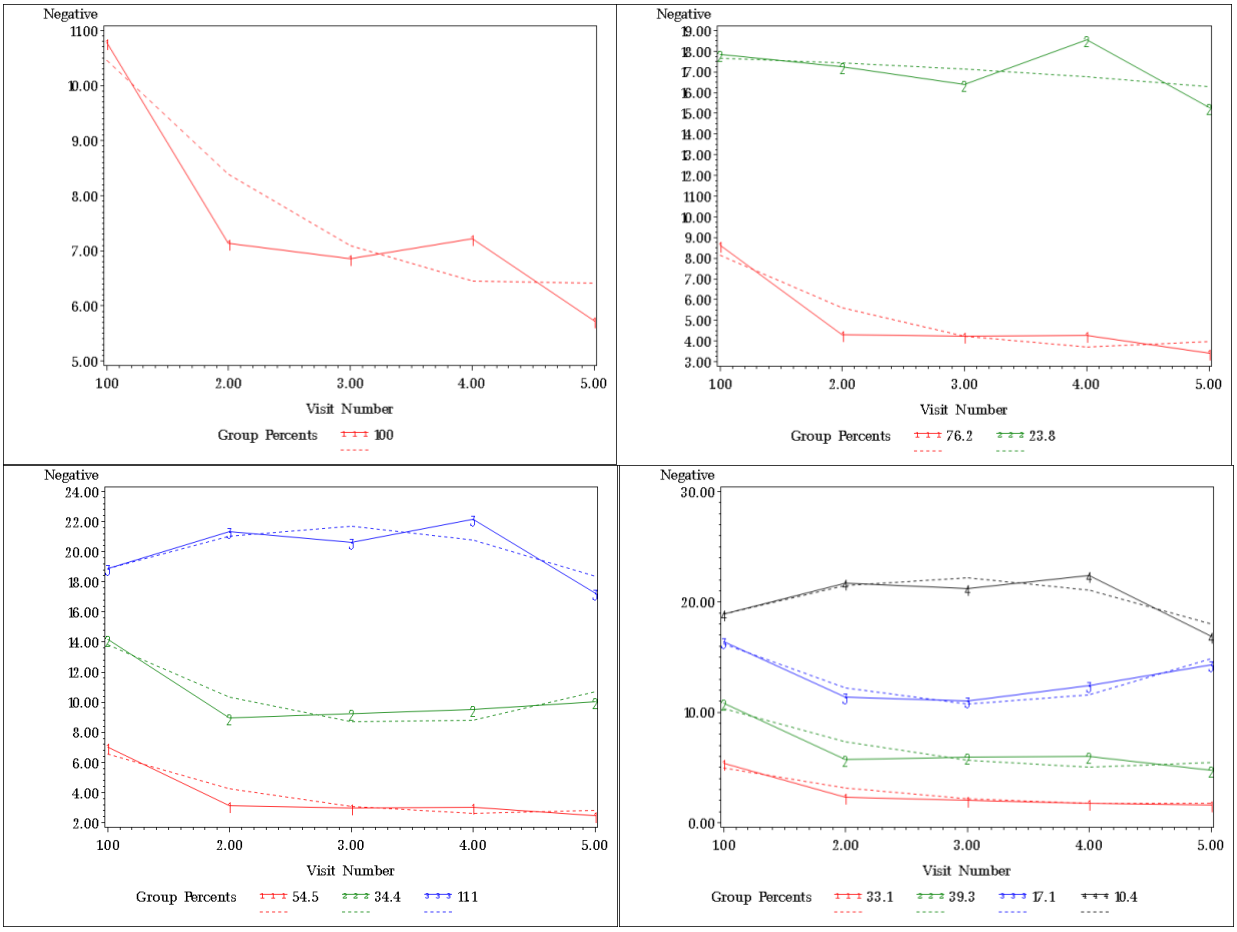


Figure S43. Model comparison plots of the selected four-group model for sum of negative symptom severity in NAPLS2 (left) and the corresponding four-group model in NAPLS1 (right). Note that the y-axis on the NAPLS2 graph ends at 21, whereas the y-axis on the NAPLS1 graph ends at 30.

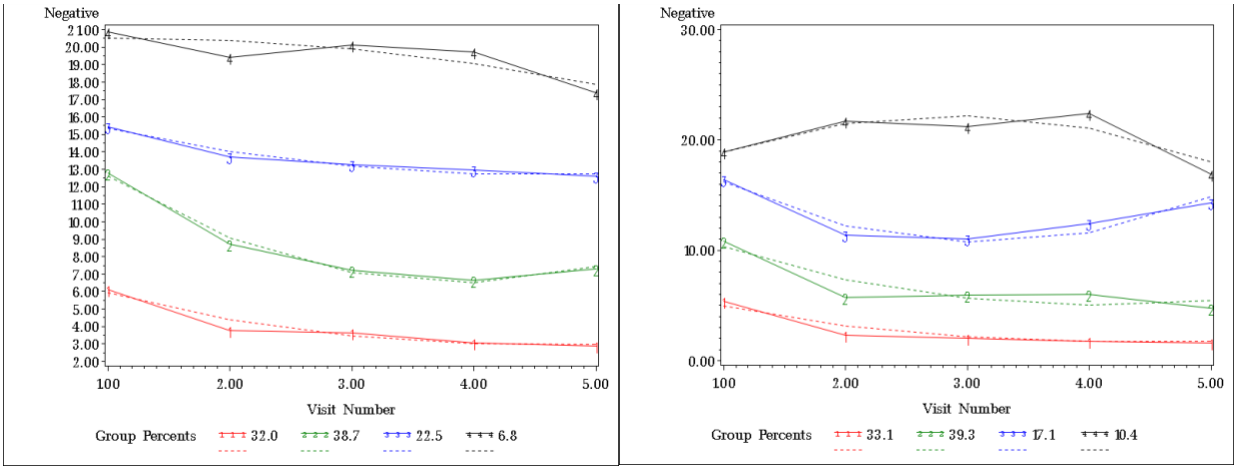


Figure S44. Plot of estimated linear trajectories (dashed) and mean negative symptom severity sum values at each visit (solid) for the preferred three-group sum of negative symptom severity model.

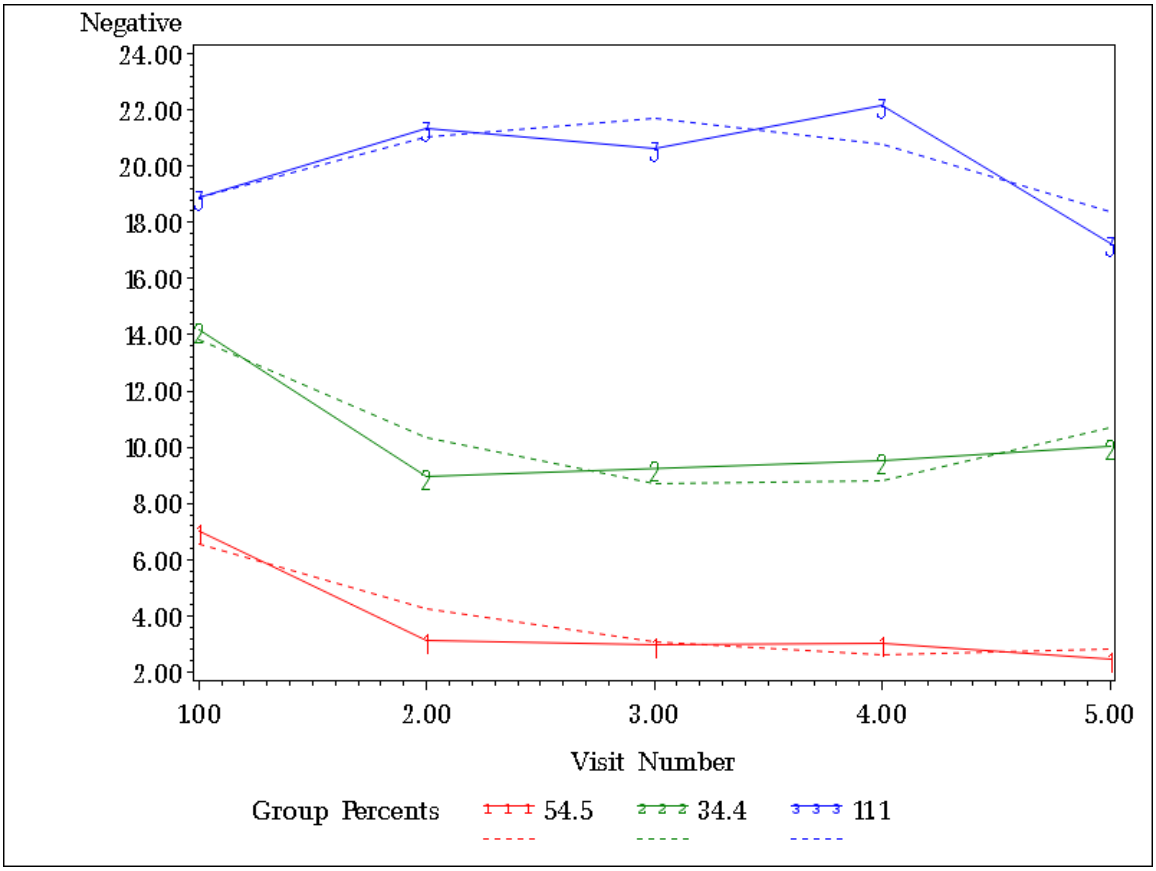


Table S62. Model parameters for the preferred three-group sum of negative symptom severity model.

Group	Parameter	Estimate	SE	T	P
1	Intercept	10.00	1.42	7.049	<.001
	Linear	-4.10	1.09	-3.756	<.001
	Quadratic	.49	.19	2.644	.009
2	Intercept	19.09	2.13	8.957	<.001
	Linear	-6.17	1.60	-3.849	<.001
	Quadratic	.90	.28	3.204	.001
3	Intercept	15.20	3.30	4.605	<.001
	Linear	4.44	2.91	1.522	.129

	Quadratic	-.76	.52	-1.451	.148
	Sigma	4.51	.18	25.734	<.001
Group Membership					
1	(%)	54.51	6.85	7.955	<.001
2	(%)	34.35	6.50	5.288	<.001
3	(%)	11.14	3.18	3.507	<.001

Table S63. Diagnostics of group-based model adequacy for the preferred three-group sum of negative symptom severity model.

Class	Probability of group membership	95% CI	Proportion assigned to group	Average posterior probability	Odds of correct classification
1	.545	[.410, .680]	.403	.922	9.9
2	.344	[.217, .471]	.433	.857	11.4
3	.111	[.048, .174]	.164	.957	178.2

Figure S45. Spaghetti plots of individual longitudinal data and estimated trajectories for the preferred three-group sum of negative symptom severity model.

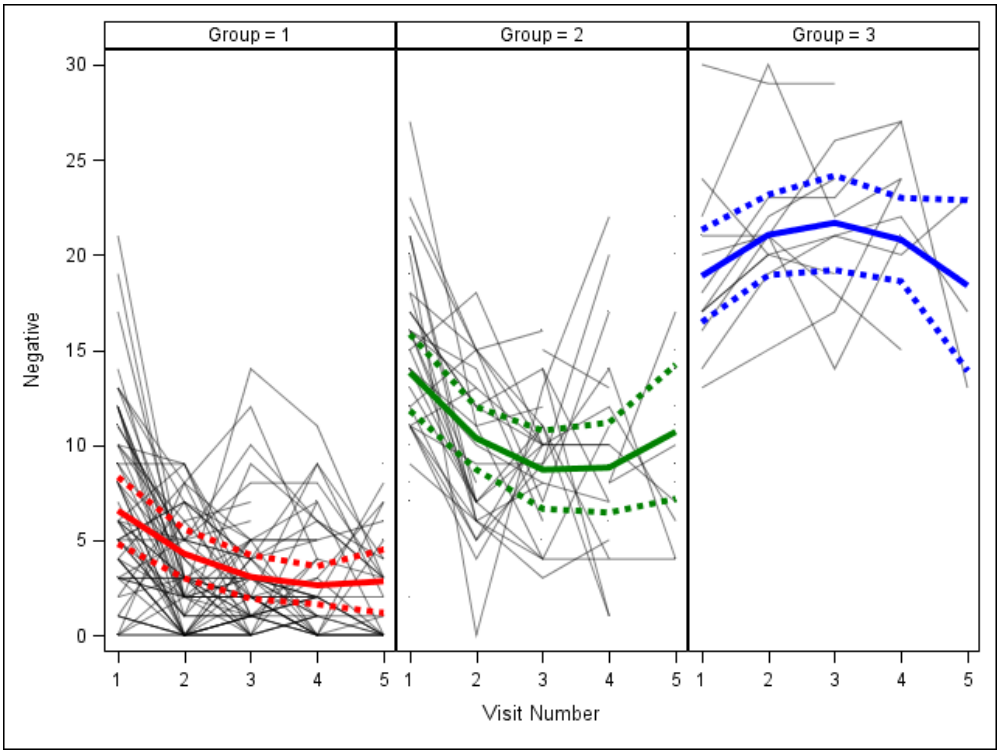


Figure S46. Comparison of negative symptom final model (panel 1) with final negative symptom model when dropout and baseline GAF are included in the model (panel 2).

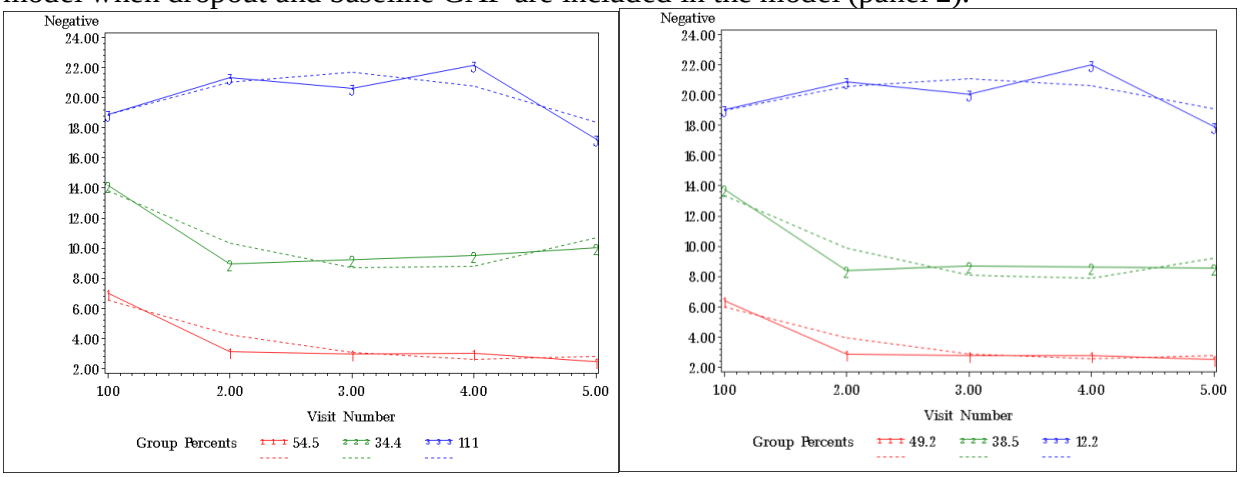


Table S64. Model parameters for the preferred three-group negative symptom model, including dropout and baseline GAF parameters.

Group	Parameter	Estimate	SE	T	P
1	Intercept	9.10	1.47	6.19	<.001
	Linear	-3.81	1.12	-3.41	<.001
	Quadratic	.47	.19	2.48	.014
2	Intercept	18.46	1.86	9.93	<.001
	Linear	-5.92	1.43	-4.14	<.001
	Quadratic	.81	.25	3.20	.002
3	Intercept	16.42	3.77	4.35	<.001
	Linear	3.10	3.62	.86	.392
	Quadratic	-.51	.64	-.80	.427
1	Drop0	-1.90	.20	-9.38	<.001
2	Drop0	-1.51	.21	-7.12	<.001
3	Drop0	-1.23	.35	-3.50	<.001
	Sigma	4.50	.17	25.90	<.001
Group Membership					
1	Constant	.	.	.	.
2	Constant	11.10	2.91	3.82	<.001
	Baseline GAF	-.22	.06	-3.98	<.001
3	Constant	12.8	3.23	3.96	<.001

Baseline GAF	-.29	.07	-4.50	<.001

BIC (N=467): -1575.10 *BIC (N=132):* -1564.36 *L:* -1522.85

Table S65. Diagnostics of group-based model adequacy for the three-group negative symptom model with dropout and baseline GAF included compared to the base three-group negative symptom model.

	Base model		*Model with dropout and baseline GAF*	
Class	Probability of group membership	Average posterior probability	Probability of group membership	Average posterior probability
1	.545	.922	.492	.961
2	.344	.857	.385	.884
3	.111	.957	.122	.905

Sum of Disorganized Symptom Severity

Group-based trajectory models of the sum of disorganized symptom severity were supported up to two trajectory groups (Table S66, Figure S47). The three-group model was rejected due to a lack of substantial improvement in BIC score (Table S66). However, the four-group model was examined in relation to the final four-group model selected in NAPLS2. Upon visual inspection, the four derived trajectories in NAPLS1 appear similar in intercepts and slopes to the final four-group model selected in NAPLS2 for Groups 3 and 4 (Figure S48). In contrast to NAPLS2, in which Group 2 appears to exhibit consistently moderate disorganized symptoms, Group 2 in NAPLS1 appears to describe a small number of individuals who have a high disorganized symptom severity sum score at baseline and decrease rapidly by the six-month visit. Group 1 also differs across samples, consisting of individuals who do not experience any disorganized symptoms in NAPLS1 versus individuals with low levels at baseline that decrease across the study in NAPLS2.

The selected two-group model suggests that common trajectories for the sum of disorganized symptoms include very low symptom severity across time that improves to almost zero severity in the early portion of the study (Group 1) and high severity that remains consistent across visits (Group 2) (Figure S49, Table S67). Further assessment of model fit parameters suggest that the selected two-group model adequately fits the underlying data (Table S68). Estimated spaghetti plots of each trajectory group (Figure S50) suggest that the two groups are relatively well separated from one another and that variance is approximately homogeneous within and across groups. Again, the floor effect observed for Groups 1 is accounted for by use of the censored normal option.

Including dropout and baseline GAF parameters in the final model of disorganized symptoms did not lead to substantial change in trajectory shapes, but did influence group size estimates. Though dropout parameters were significant, estimated trajectories appear very similar across models that exclude and include dropout parameters (Figure S51) and model parameters remained consistent (Table S69). Group sizes changed by ~5% in both Groups 1 and 2 (Table S70). Baseline GAF was significant as a risk factor. Thus, the effects of attrition are not expected

to have an appreciable effect on negative symptom trajectory shapes, but may lead to overestimation of the size of Group 1 and underestimation of Group 2 size in multi-trajectory models.

Table S66. Diagnostics of group-based model adequacy for sum of disorganized symptom severity model comparison.

Number of groups	BIC (*N*=451)	BIC (*N*=133)	Log likelihood	Average posterior probability	Odds of correct classification	Percent assigned to group
1	-1121.96	-1119.52	-1109.74	..	..	100
2	-1079.83	-1074.94	-1055.38	.95 .92	6 39	78 22
3	-1082.09	-1074.77	-1045.42	.90 .77 .89	7 8 59	57 31 12
4	-1090.54	-1080.77	-1041.65	.87 .92 81 .89	5 504 11 56	58 2 28 12

Figure S47. Model comparison plots of estimated linear trajectories (dashed) and mean sum of disorganized symptom severity values at each visit (solid) for derived trajectory groups in each model.

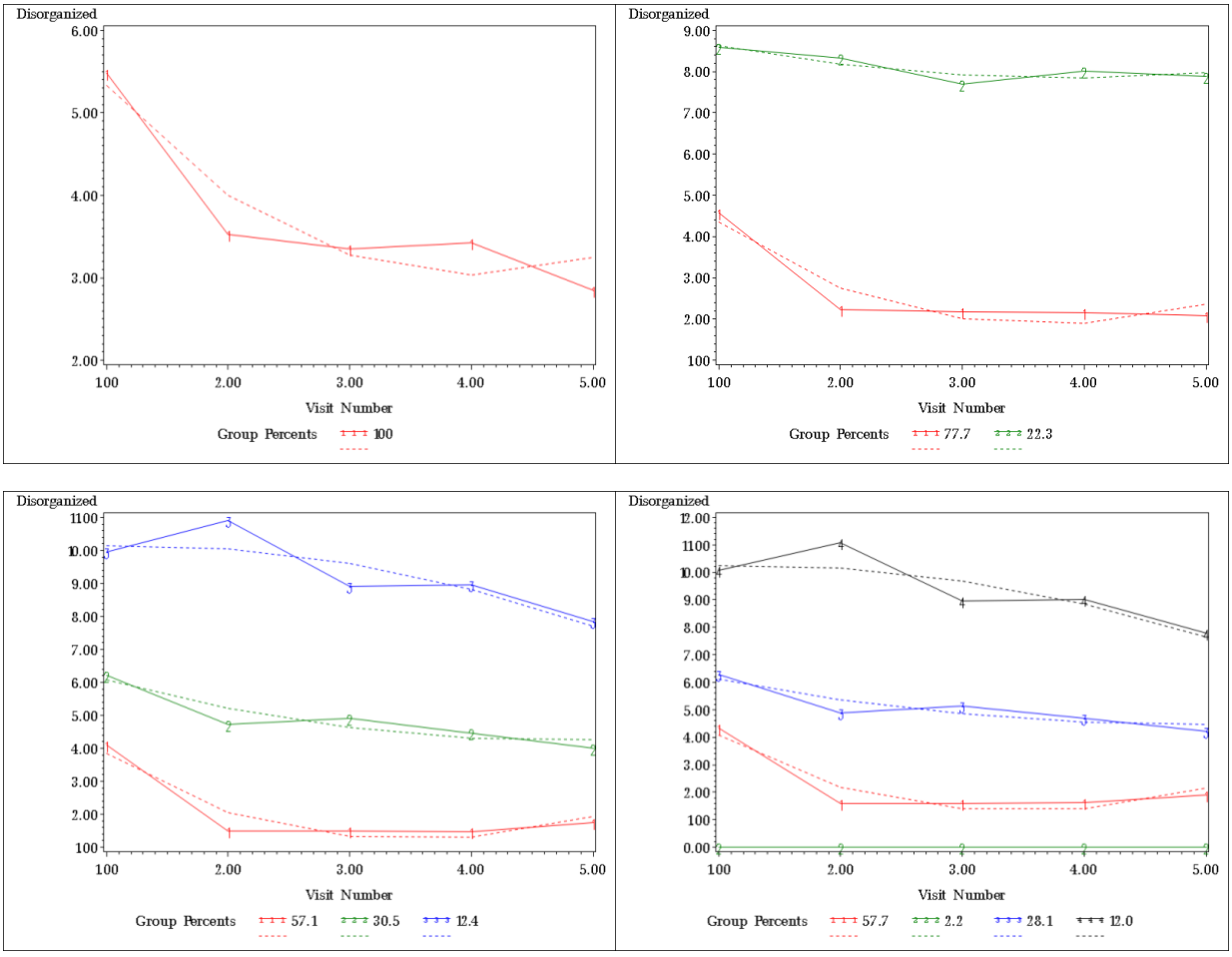


Figure S48. Model comparison plots of the selected four-group model for sum of disorganized symptom severity in NAPLS2 (left) and the corresponding four-group model in NAPLS1 (right).

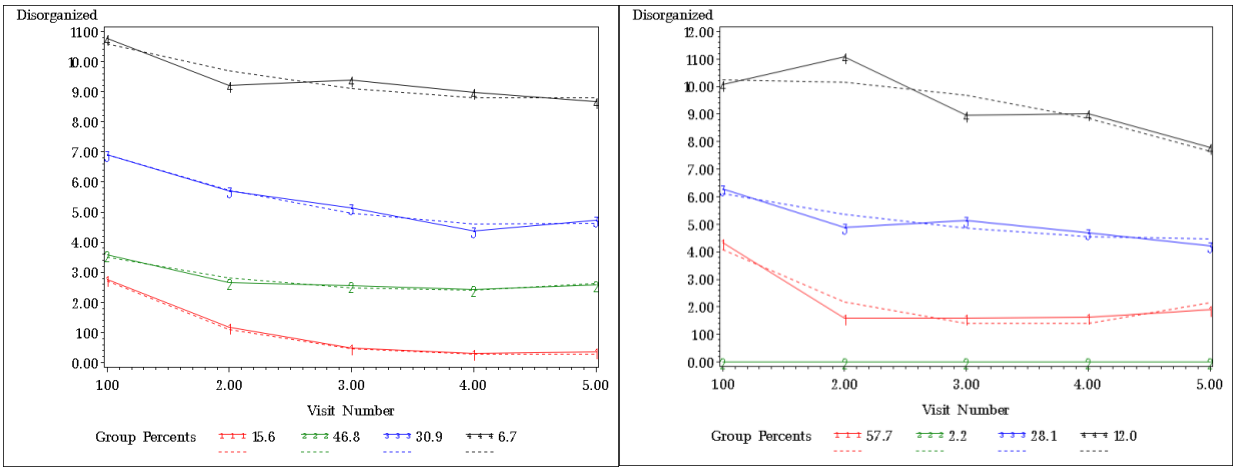


Figure S49. Plot of estimated linear trajectories (dashed) and mean disorganized symptom severity sum values at each visit (solid) for the preferred two-group model.

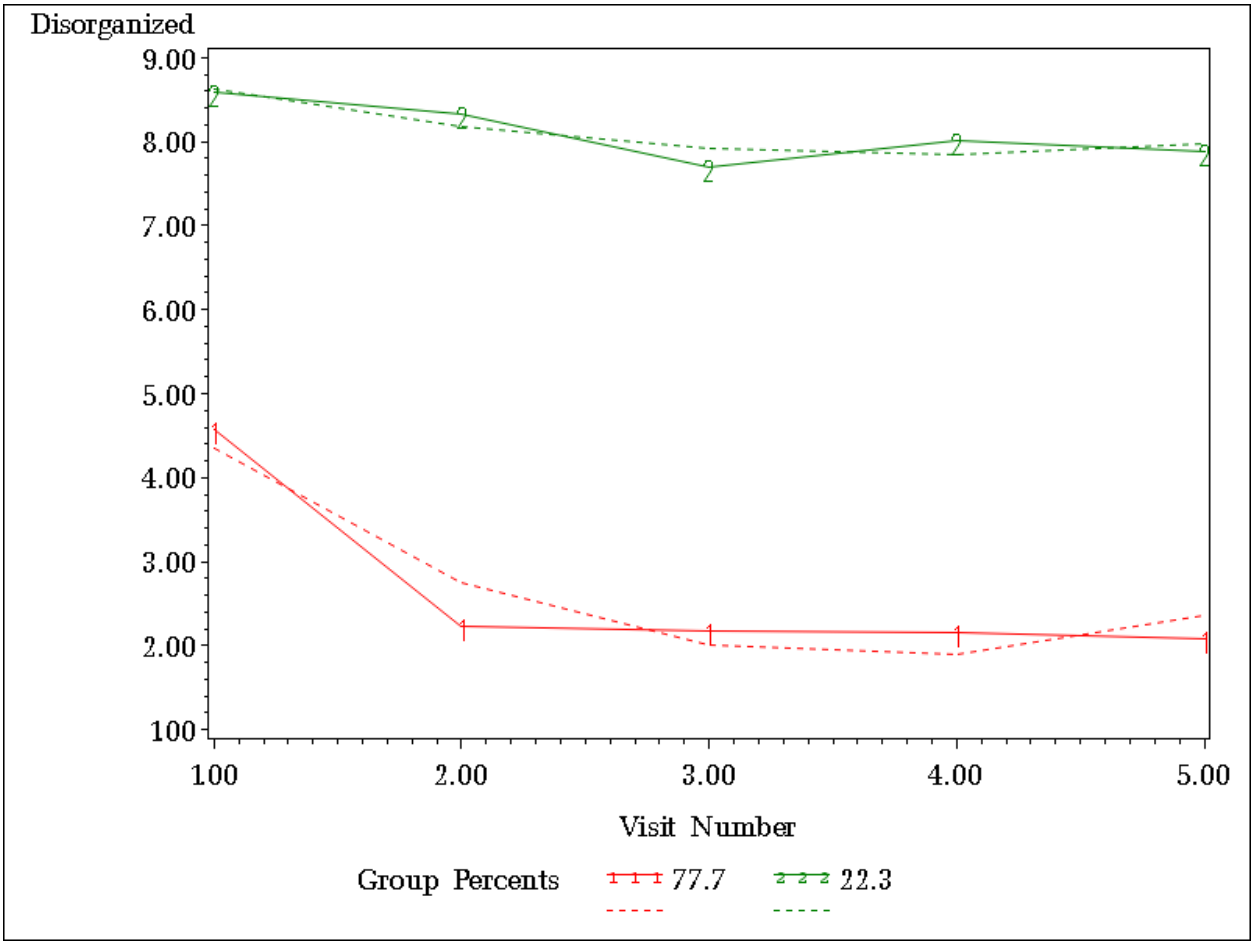


Table S67. Model parameters for the preferred two-group sum of disorganized symptom severity model.

Group	Parameter	Estimate	SE	T	P
1	Intercept	6.94	.72	9.742	<.001
	Linear	-3.10	.59	-5.214	<.001
	Quadratic	.42	.10	4.088	<.001
2	Intercept	9.27	1.46	6.332	<.001
	Linear	-.74	1.24	-.597	.551
	Quadratic	.10	.23	.425	.671
	Sigma	2.93	.12	24.681	<.001
Group Membership					
1	(%)	.777	4.79	16.219	<.001
2	(%)	.223	4.79	4.651	<.001

Table S68. Diagnostics of group-based model adequacy for the preferred two-group sum of disorganized symptom severity model.

Class	Probability of group membership	95% CI	Proportion assigned to group	Average posterior probability	Odds of correct classification
1	.777	[.683, .871]	.797	.954	6.0
2	.223	[.129, .317]	.203	.918	39.0

Figure S50. Spaghetti plots of individual longitudinal data and estimated trajectories for the preferred two-group sum of disorganized symptom severity model

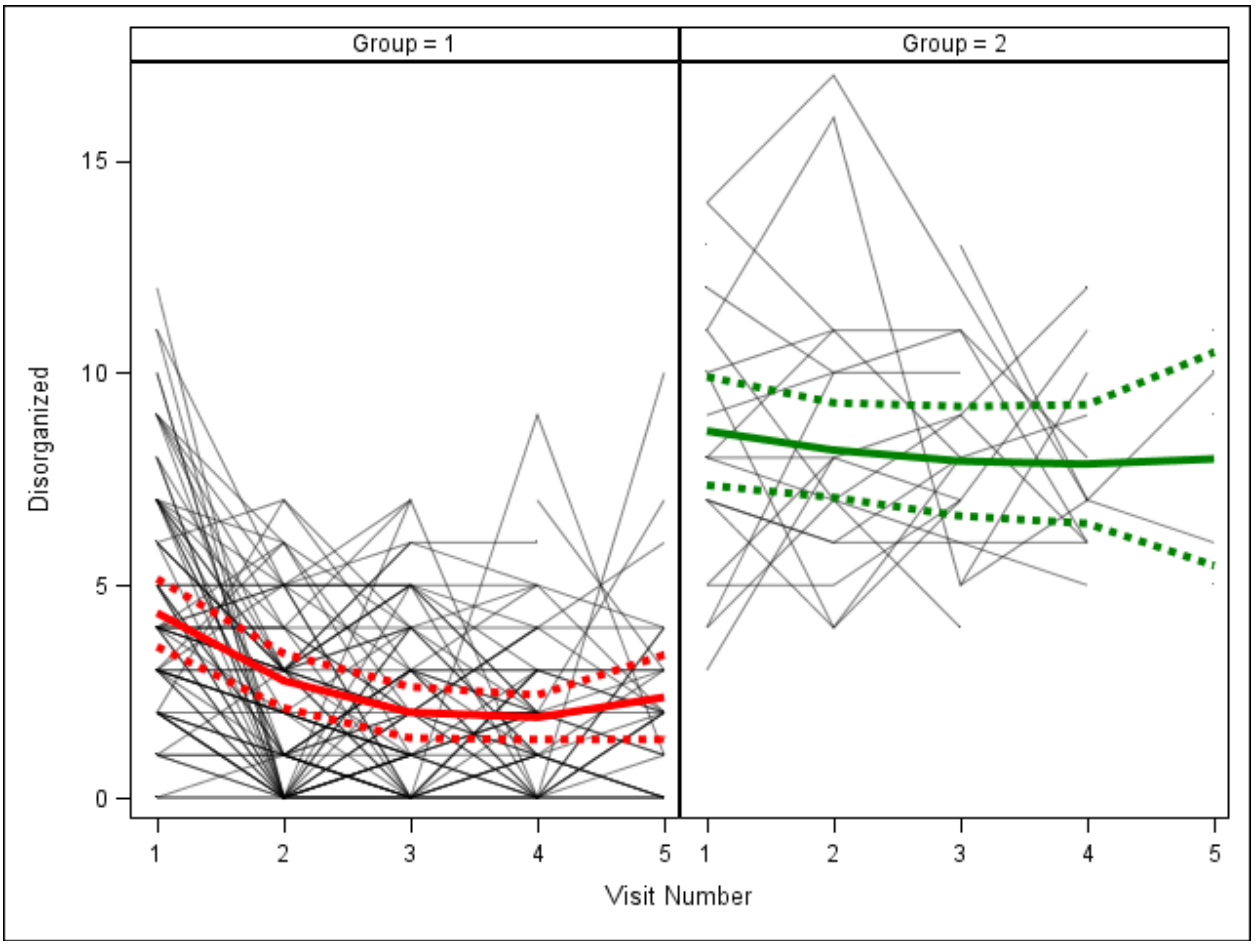


Figure S51. Comparison of the disorganized symptom final model (panel 1) with the disorganized symptom model when dropout and baseline GAF are included in the model (panel 2).

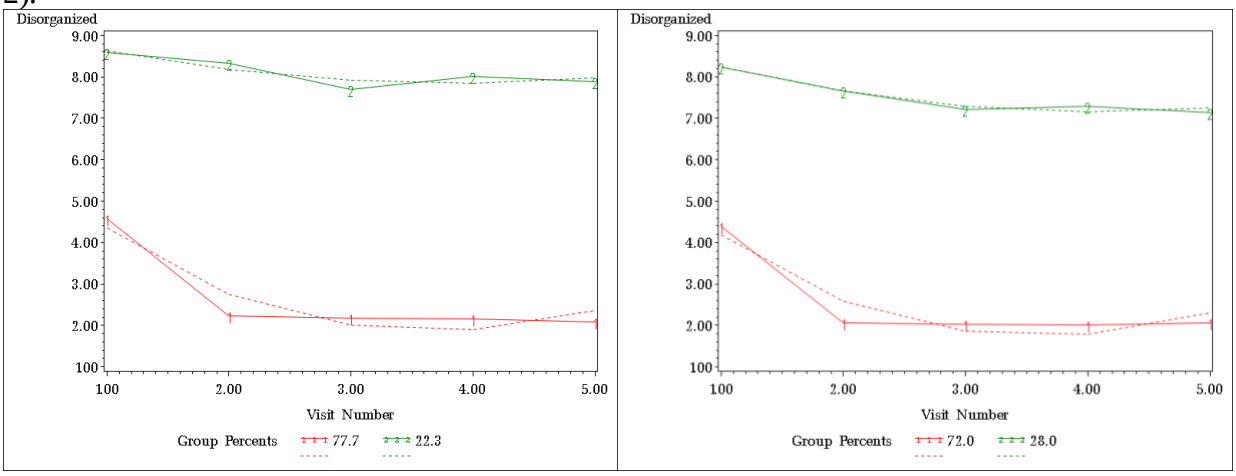


Table S69. Model parameters for the preferred two-group disorganized symptom model, including dropout and baseline GAF parameters.

Group	Parameter	Estimate	SE	T	P
1	Intercept	6.84	.74	9.18	<.001
	Linear	-3.21	.62	-5.20	<.001
	Quadratic	.44	.11	4.15	<.001
2	Intercept	9.04	1.32	6.86	<.001
	Linear	-.92	1.13	-.81	.417
	Quadratic	.11	.21	.54	.590
1	Drop0	-1.65	.16	-10.52	<.001
2	Drop0	-1.19	.23	-5.18	<.001
	Sigma	2.94	.12	24.54	<.001
Group Membership					
1	Constant	.	.	.	.
2	Constant	4.05	1.52	2.67	.008
3	Baseline GAF	-.11	.03	-3.38	<.001

BIC (N=5450): -1291.22 *BIC (N=132):* -1284.48 *L:* -1257.62

Table S70. Diagnostics of group-based model adequacy for the two-group disorganized symptom model with dropout and baseline GAF included compared to the base two-group disorganized symptom model.

	Base model		*Model with dropout and baseline GAF*	
Class	Probability of group membership	Average posterior probability	Probability of group membership	Average posterior probability
1	.777	6.0	.720	.954
2	.223	39.0	.280	.906

Sum of General Symptom Severity

Group-based trajectory models of the sum of general symptom severity were supported up to two trajectory groups (Table S71, Figure S52). The three-group model was rejected due to a negligible improvement in BIC (Table S71). However, the three-group model was examined in relation to the final three-group model selected in NAPLS2. Group 2 in NAPLS1 exhibited a higher intercept and more rapid decrease in symptoms between the baseline and 6-month visits

compared to NAPLS2 Group 2, and NAPLS1 Group 3 exhibited slightly higher rates of improvement across the study than NAPLS2 Group 3 (Figure S53).

The selected two-group model suggests that common trajectories for the sum of general symptoms include a low symptom severity that improves most rapidly during the first portion of the study and almost reaches zero severity (Group 1) and a higher severity that improves relatively steadily across time (Group 2) (Figure S54, Table S72). Further assessment of model fit parameters suggest that the selected two-group model adequately fits the underlying data (Table S73). Estimated spaghetti plots of each trajectory group (Figure S55) suggests that the two groups are relatively well separated from one another and that variance is approximately homogeneous within and across groups. As in other models, the floor effect observed for Group 1 is accounted for by the use of the censored normal option within models.

Including dropout and baseline GAF parameters in the final model of general symptoms did not lead to substantial change in trajectory shapes, but did influence group size estimates. Though dropout parameters were significant, estimated trajectories appear very similar across models that exclude and include dropout parameters (Figure S56) and model parameters remained consistent (Table S74). Group sizes changed by ~3% in both Groups 1 and 2 (Table S75). Baseline GAF was significant as a risk factor. Thus, the effects of attrition are not expected to have an appreciable effect on general symptom trajectory shapes, but may lead to overestimation of the size of Group 1 and underestimation of Group 2 size in multi-trajectory models.

Table S71. Diagnostics of group-based model adequacy for sum of general symptom severity model comparison.

Number of groups	BIC (*N*=458)	BIC (*N*=133)	Log likelihood	Average posterior probability	Odds of correct classification	Percent assigned to group
1	-1228.59	-1226.12	-1216.34	..	..	100
2	-1189.27	-1184.32	-1164.76	.93 .90	15 8	47 53
3	-1187.08	-1179.66	-1150.31	.81 .91 .79	6 17 17	43 38 19

Figure S52. Model comparison plots of estimated linear trajectories (dashed) and mean score for sum of general symptom severity at each visit (solid) for derived trajectory groups in each model.

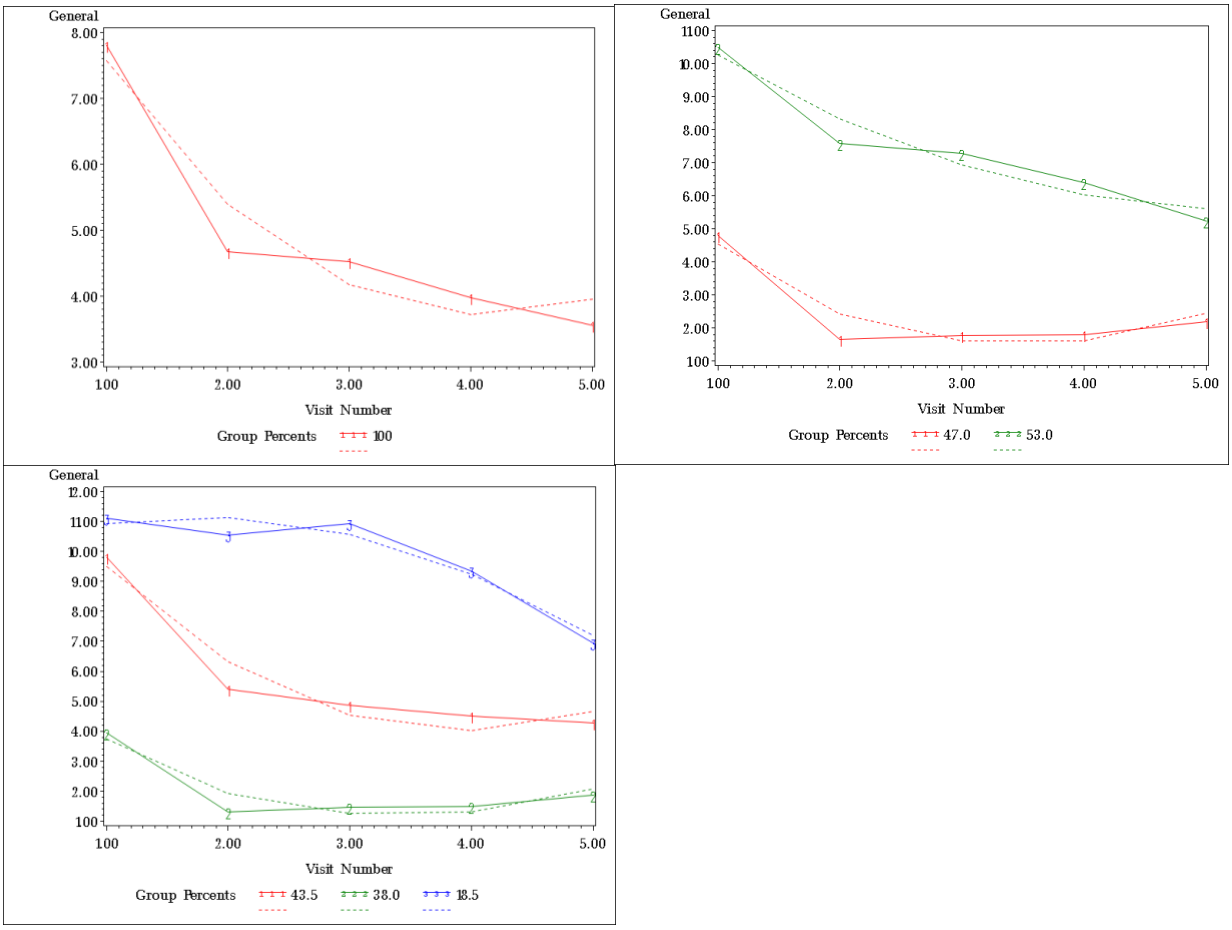


Figure S53. Model comparison plots of the selected three-group model for sum of general symptom severity in NAPLS2 (left) and the corresponding three-group model in NAPLS1 (right).

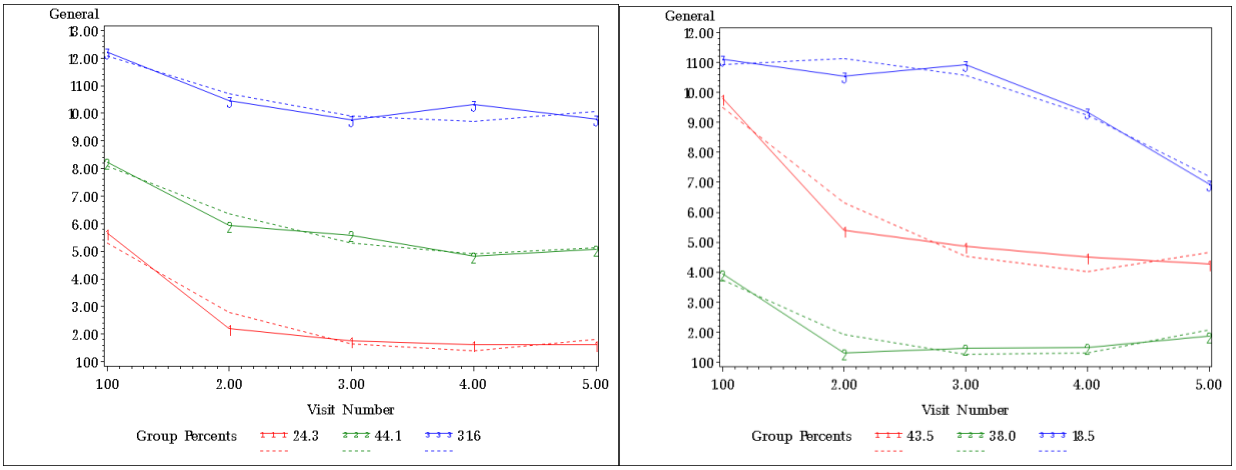


Figure S54. Plot of estimated linear trajectories (dashed) and mean general symptom severity sum values at each visit (solid) for the preferred two-group model.

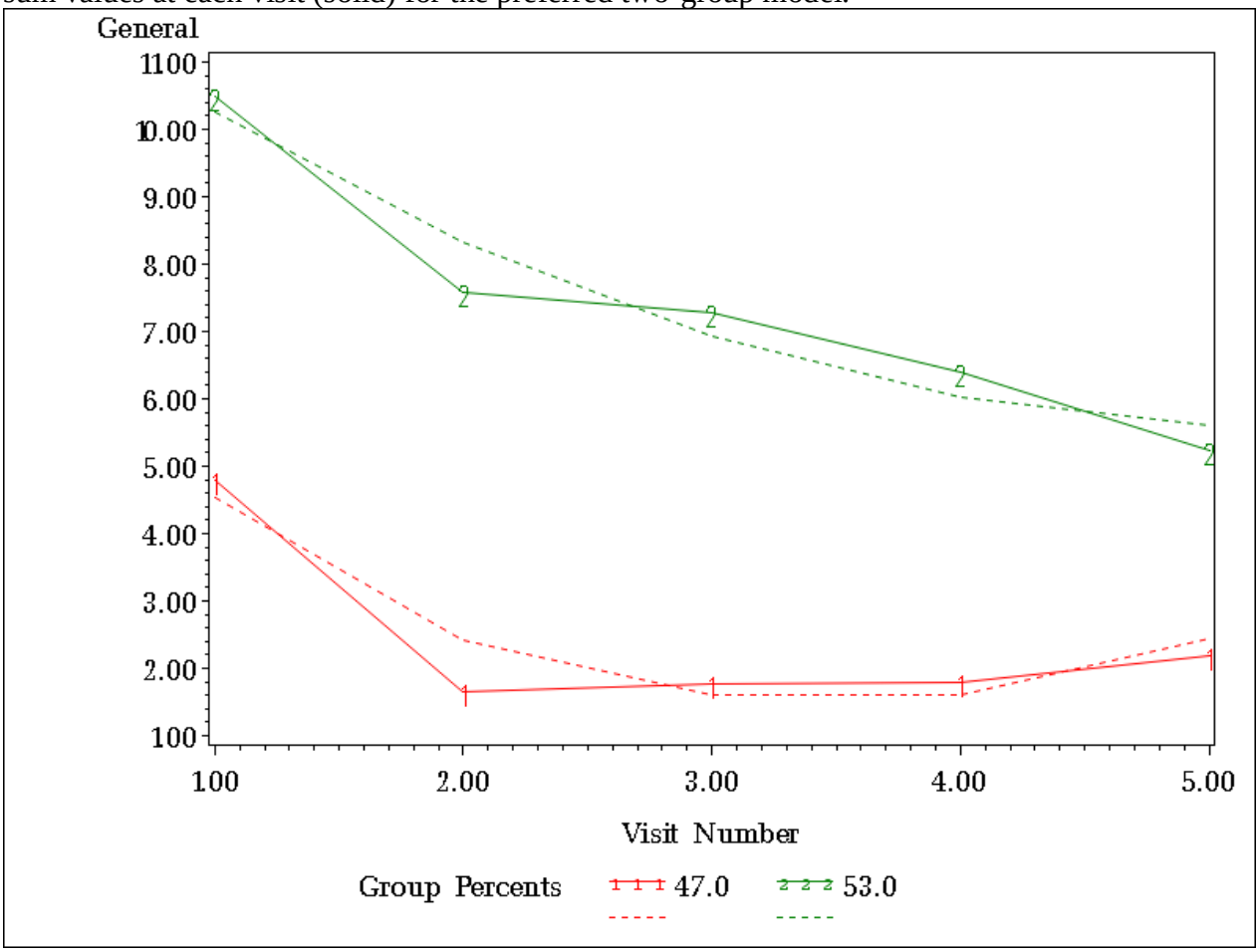


Table S72. Model parameters for the preferred two-group sum of general symptom severity model.

Group	Parameter	Estimate	SE	T	P
1	Intercept	8.41	1.25	6.712	<.001
	Linear	-4.77	.97	-4.905	<.001
	Quadratic	.68	.17	4.067	<.001
2	Intercept	12.67	1.10	11.536	<.001
	Linear	-2.67	.95	-2.825	.005
	Quadratic	.25	.17	1.480	.140
	Sigma	3.67	.15	24.831	<.001
Group Membership					
1	(%)	47.03	6.29	7.477	<.001
2	(%)	52.97	6.29	8.422	<.001

Table S73. Diagnostics of group-based model adequacy for the preferred two-group sum of general symptom severity model.

Class	Probability of group membership	95% CI	Proportion assigned to group	Average posterior probability	Odds of correct classification
1	.470	[.347, .593]	.444	.929	14.8
2	.530	[.407, .653]	.556	.895	7.6

Figure S55. Spaghetti plots of individual longitudinal data and estimated trajectories for the preferred two-group sum of general symptom severity model.

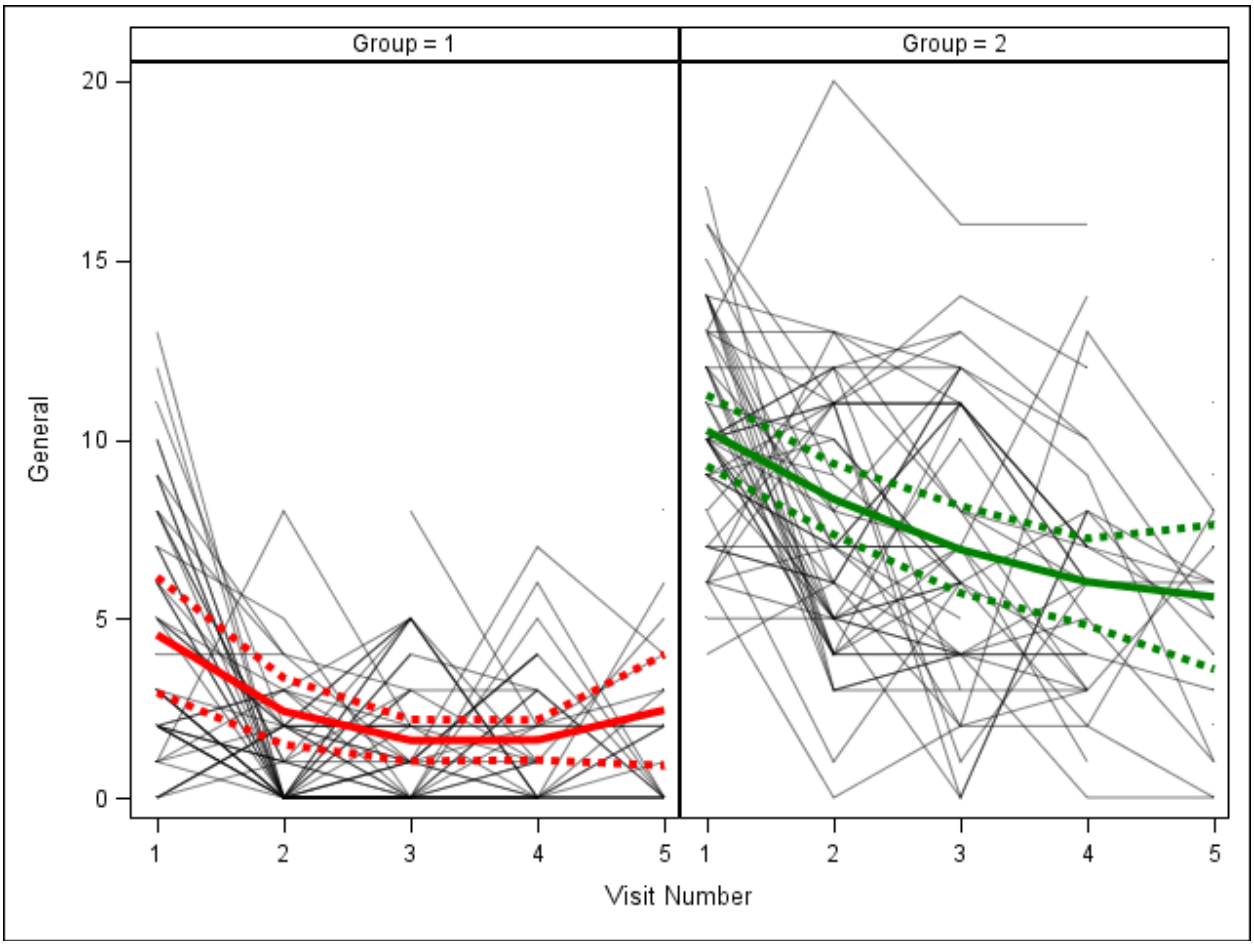


Figure S56. Comparison of general symptom final model (panel 1) with final general symptom model when dropout and baseline GAF are included in the model (panel 2).

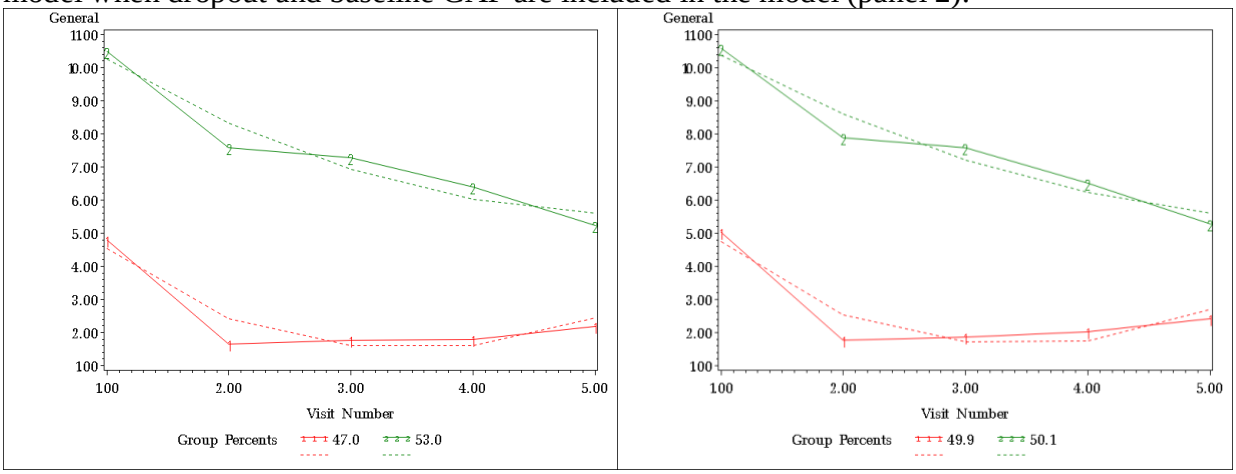


Table S74. Model parameters for the preferred two-group general symptom model, including dropout and baseline GAF parameters.

Group	Parameter	Estimate	SE	T	P
1	Intercept	8.71	1.23	7.10	<.001
	Linear	-4.86	.93	-5.21	<.001
	Quadratic	.70	.16	4.39	<.001
2	Intercept	12.52	1.16	10.82	<.001
	Linear	-2.34	1.03	-2.28	.023
	Quadratic	.19	.18	1.04	.301
1	Drop0	-1.78	.20	-9.02	<.001
2	Drop0	-1.34	.18	-7.54	<.001
	Sigma	3.67	.15	24.65	<.001
Group Membership					
1	Constant	.	.	.	.
2	Constant	4.12	1.17	3.54	<.001
	Baseline GAF	-.08	.02	-3.48	<.001

BIC (N=457): -1400.47 *BIC (N=132):* -1393.64 *L:* -1336.78

Table S75. Diagnostics of group-based model adequacy for the two-group general symptom model with dropout and baseline GAF included compared to the base two-group general symptom model.

	Base model		*Model with dropout and baseline GAF*	
Class	Probability of group membership	Average posterior probability	Probability of group membership	Average posterior probability
1	.470	.929	.499	.938
2	.530	.895	.501	.903

Summary

Overall, preferred univariate models of GAF and the four sum of symptom severity measures in NAPLS1 suggest that, typically, two trajectory groups are statistically supported for

each outcome. This reduction in number of trajectory groups identified in NAPLS1 compared to NAPLS2 likely reflects differences in sample size, given that NAPLS1 (N = 133) is less than half the size of NAPLS2 (N = 422) and sample size can affect the number of trajectory groups derived in samples below 300-500 individuals (Sampson et al., 2004). As the derived trajectory groups are used to conceptualize different aspects of an underlying continuous distribution of outcomes rather than truly distinct subgroups (Nagin, 2005), the observed decrease in derived trajectories observed in NAPLS1 is not surprising. When the same number of groups were derived in NAPLS1 as in NAPLS2, similar ranges of values were observed across samples. This consistency suggests that the underlying structure of the data may be similar between the two datasets, aside from more rapid improvements in symptoms between the baseline and 6-month visits in NAPLS2 relative to NAPLS1 across symptoms (and particularly for positive symptom severity). Based on these univariate analyses, we expected that derivation of a three-group multi-trajectory model in NAPLS1 would yield generally similar patterns to those observed in NAPLS2 but may not be as statistically reliable and/or exhibit less intermediate trajectories (i.e., different combinations of "low" and "high" trajectories) due to reductions in sample size relative to NAPLS2. Additionally, slopes of NAPLS1 trajectories may be somewhat steeper than slopes observed in NAPLS2, particularly for positive symptoms, due to the differences observed between the samples in univariate analyses.

Models that included dropout parameters and accounted for higher rates of missing data among individuals with lower baseline GAF scores in univariate models of GAF and symptom severity sums did not result in appreciable change in trajectory estimates, but did identify inflation of probabilities of group membership of 3-5% for negative, disorganized, and general symptom groups. Though it is unfortunate that dropout and baseline GAF cannot be accounted for in the final multivariate model due to challenges with model conversion, these results suggest that the trajectory shapes identified in the multivariate model are not substantially influenced by missing data and can be interpreted with reasonable confidence. We note in the main manuscript that group memberships may be inflated by 3-5% for Group 1 and deflated for Group 2 by a similar amount, and adjust our interpretation of group sizes accordingly.

References

1. Addington J, Cadenhead KS, Cornblatt BA, Mathalon DH, Mcglashan TH, Perkins DO, et al. North American Prodrome Longitudinal Study (NAPLS 2): Overview and recruitment. Schizophr Res. 2012;142(1–3):77–82.
2. Addington J, Liu L, Buchy L, Cadenhead KS, Cannon TD, Cornblatt BA, et al. North American Prodrome Longitudinal Study (NAPLS 2): The prodromal symptoms. J Nerv Ment Dis. 2016;203(5):328–35.
3. Addington J, Cadenhead KS, Cannon TD, Cornblatt B, McGlashan TH, Perkins DO, et al. North American Prodrome Longitudinal Study: A collaborative multisite approach to prodromal schizophrenia research. Schizophr Bull. 2007;33(3):665–72.
4. Woods SW, Addington J, Cadenhead S, Cannon TD, Cornblatt BA, Heinssen R, et al. Validity of the Prodromal Risk Syndrome for first psychosis: Findings From the North American Prodrome Longitudinal Study. Schizophr Bull. 2009;35(5):894–908.
5. Brandizzi M, Valmaggia L, Byrne M, Jones C, Iwegbu N, Badger S, et al. Predictors of functional outcome in individuals at high clinical risk for psychosis at six years follow-up. J Psychiatr Res. 2015;65:115–23.
6. Lee TY, Kim SN, Correll CU, Byun MS, Kim E, Jang JH, et al. Symptomatic and functional remission of subjects at clinical high risk for psychosis: A 2-year naturalistic observational study. Schizophr Res. 2014;156(2–3):266–71.
7. Verma S, Subramaniam M, Abdin E, Poon L, Chong S. Symptomatic and functional remission in patients with first-episode psychosis. Acta Psychiatr Scand. 2012;126:282–9.
8. Cadenhead KS, Minichino A, Kelsven S, Addington J, Bearden C, Cannon TD, et al. Metabolic abnormalities and low dietary Omega 3 are associated with symptom severity and worse functioning prior to the onset of psychosis: Findings from the North American Prodrome Longitudinal Studies Consortium. Schizophr Res. 2019;204:96–103.
9. Woods SW, Walsh BC, Hawkins KA, Miller TJ, Saksa JR, Souza DCD, et al. Glycine treatment of the risk syndrome for psychosis: Report of two pilot studies. Eur Neuropsychopharmacol. 2013;23(8):931–40.
10. Woods STW, Tully EM, Walsh AC, Hawkins KA, Callahan JL, Cohen SJ, et al. Aripiprazole in the treatment of the psychosis prodrome: An open-label pilot study. Br J Psychiatry. 2007;191(suppl. 51):s96–101.
11. Woods SW, Breier A, Zipursky RB, Perkins DO, Addington J, Miller TJ, et al. Randomized trial of olanzapine versus placebo in the symptomatic acute treatment of the schizophrenic prodrome. Biol Psychiatry. 2003;54:453–64.
12. Webb JR, Addington J, Perkins DO, Bearden CE, Cadenhead KS, Cannon TD, et al. Specificity of incident diagnostic outcomes in patients at clinical high risk for psychosis. Schizophr Bull. 2015;41(5):1066–75.
13. van de Schoot R, Sijbrandij M, Winter SD, Depaoli S, Vermunt JK. The GRoLTS-checklist: Guidelines for reporting on latent trajectory studies. Struct Equ Model. 2017;24(3):451–67.
14. Haviland AM, Jones BL, Nagin DS. Group-based trajectory modeling extended to account for nonrandom participant attrition. Sociol Methods Res. 2011;40(2):367–90.
15. Nagin D. Group-Based Modeling of Development. Cambridge: Harvard University Press; 2005.
16. Nagin DS, Jones BL, Lima Passos V, Tremblay RE. Group-based multi-trajectory

modeling. Stat Methods Med Res. 2016;
17. Wickham H. ggplot2: Elegant graphics for data analysis. New York: Springer-Verlag; 2016.
18. Baptiste A. gridExtra: Miscellaneous functions for „Grid“ graphics. R package. 2015.
19. Muthén BO, Muthén LK. Integrating person centered and variable centered analyses: Growth mixture modeling with latent trajectory classes. Kd 24, Alcoholism: Clinical and Experimental Research. 2000. lk 882–91.
20. Velthorst E, Fett AKJ, Reichenberg A, Perlman G, Van Os J, Bromet EJ, et al. The 20-year longitudinal trajectories of social functioning in individuals with psychotic disorders. Am J Psychiatry. 2017;174(11):1075–85.
21. Celeux G, Soromenho G. An entropy criterion for assessing the number of clusters in a mixture model. J Classif. 1996;13:195–212.
22. Jones BL, Nagin DS. Advances in group-based trajectory modeling and an SAS procedure for estimating them. Sociol Methods Res. 2007;35(4):542–71.
23. Hipp JR, Bauer DJ. Local solutions in the estimation of growth mixture models. Psychol Methods. 2006;11(1):36–53.
24. Gueorguieva R, Chekroud AM, Krystal JH. Trajectories of relapse in randomised, placebo-controlled trials of treatment discontinuation in major depressive disorder: An individual patient-level data meta-analysis. The Lancet Psychiatry. 2017;4(3):230–7.