HIR
Healthcare Informatics Research

Figure S1. Four-topic latent Dirichlet allocation (LDA) symptom distribution. Column colors represent individual subtypes. Symptoms were included here if they were one of the two most common symptoms for a subtype.
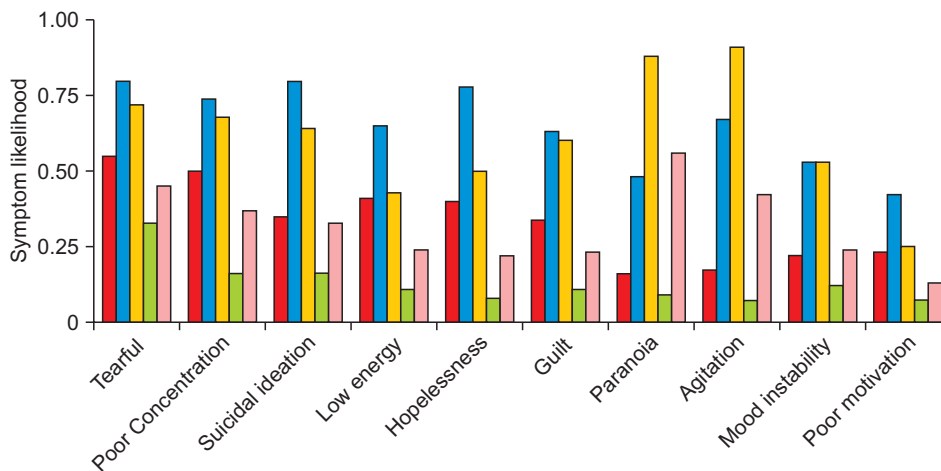

Figure S2. Five-class latent class analysis (LCA) symptom likelihoods. Column colors represent individual subtypes. The top ten most common symptoms in the dataset were included here.

Table S1. Symptoms used to create subtypes

Anergia	Elation	Low mood	Social withdrawal
Anhedonia	Flight of ideas	Mood instability	Stupor
Apathy	Formal thought disorder	Mutism	Suicidal ideation
Aggression	Gradinosity	Negative symptoms	Tangential speech
Agitation	Guilt	Paranoia	Tearfulness
Arousal	Hallucinations	Passivity	Thought block
Blunted affect	Helplessness	Persecutory ideation	Though broadcast
Circumstantial speech	Hopelessness	Poor appetite	Thought insertion
Concrete thinking	Hostility	Poor concentration	Thought withdrawal
Delusions	Insomnia	Poor insight	Treatment resistant
Derailment of speech	Irritability	Poor motivation	Waxy flexibility
Disturbed sleep	Loss of coherence	Poverty of speech	Weight loss
Echolalia	Low energy	Poverty of thought	Worthlessness

We excluded "low mood" and "disturbed sleep" in our work. In our initial analyses, we found that they were not useful for defining subtypes because every subtype featured them. This was not surprising given that they were the most prevalent symptom by large margins.