Supplementary Material

Section I: Description of the Bayesian models for the analysis of ECLIPSE and SPIROMICS cohorts

Section II: Results based on alternative definition of the frequent-exacerbator group

Section III: OpenBUGS code for the final models

Section I: Bayesian models for ECLIPSE and SPIROMICS

Both ECLIPSE(1) and SPIROMICS(2) categorized each patient-year as being associated with 0, 1, or 2+ (2 or more) AECOPDs. This results in 27 different combination of AECOPD patterns over three years. Both studies provided data on the number of individuals who fell into each of the 27 categories.

Let $Y_{i,j}$ be the observed number of AECOPDs for the i^{th} patient in the j^{th} year.

The "unstable underlying rate" model was specified as

$$Y_{i,j} \sim Poisson(\lambda_{i,j})$$

$$\lambda_{i,j} \sim Gamma(\alpha, \beta)$$

In this model, the underlying AECOPD rate, specified by $\lambda_{i,j}$, takes different values for each patient in each year (therefore there is no stable underlying rate).

The "stable underlying rate" model was specified as

$$Y_{i,j} \sim Poisson(\lambda_i)$$

$$\lambda_i \sim Gamma(\alpha, \beta)$$

In this model, the underlying AECOPD rate, specified by λ_i, takes a fixed value for each patient, but remains the same across follow-up years for each patient (therefore there is a stable underlying rate).

The OpenBUGS(3) code for the models is provided below. We note that because $Y_{i,j}$ is truncated (follow-up years with 2 or more AECOPDs are coded as 2), the internally programmed Poisson distribution in OpenBUGS could not be used. Instead, we used the "zero trick" approach to directly implement the probability mass function of the truncated Poisson distribution(4, p.353).

We also examined the choice of the Normal distribution in place of the Gamma to model the underlying AECOPD rates:

$\ln(\lambda_{i,j}) \sim Normal(\mu, v)$ for the 'unstable underlying rate';

$\ln(\lambda_i) \sim Normal(\mu, v)$ for the 'stable underlying rate'.

.

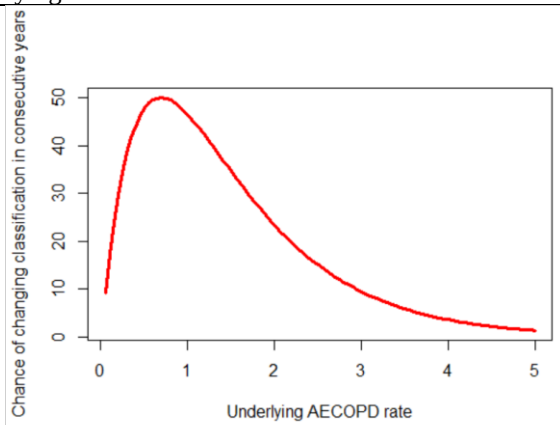
Among these variants, we chose two final models, one for the 'unstable underlying rate', and one for the 'stable underlying rate', based on the likelihood. We note that the use of penalized goodness-of-fit metrics is not required in this context because all four models have the same number of fixed parameters (p=2). ***Table S1*** provides the results. For both cohorts, and for both the 'unstable underlying rate' and 'stable underlying rate' models, the versions based on the gamma distribution better fitted the data.

Table S1: Goodness-of-fit measures (log-likelihood) of the models (the higher the better)		
Model	**ECLIPSE**	**SPIROMICS**
Unstable underlying rate		
Gamma	**-5346**	**-2532**
Normal	-5346	-3992
Stable underlying rate		
Gamma	**-4856**	**-2385**
Normal	-4896	-4337
All results are based on 1,000 Markov Chain Monte Carlo simulations after 1,000 burn-ins (for adaptation)		

Section II: Results based on alternative definition of the frequent-exacerbator group

1. American Thoracic Society (ATS) definition of frequent exacerbator (≥1 moderate/severe AECOPD)

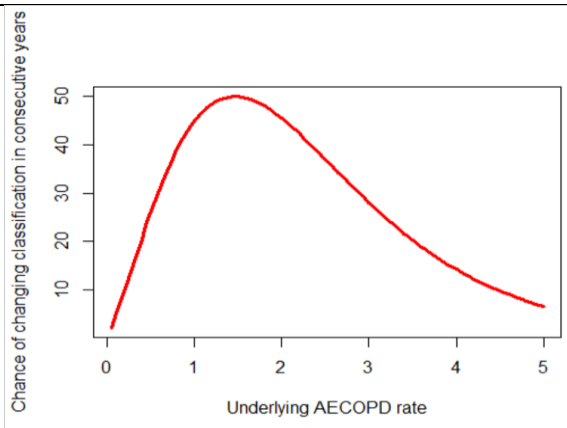
Figure S1. Probability of being classified into different exacerbator phenotypes over two consecutive years given a stable underlying AECOPD rate.



2. The Global Initiative for Chronic Obstructive Lung Disease (GOLD) (≥2 moderate AECOPD or ≥1 severe AECOPD)

Because the GOLD definition of frequent exacerbator uses separate thresholds for moderate and severe AECOPDs, this analysis required one additional parameter: the rate ratio of severe to moderate + severe AECOPDs. For consistency with the overall content, we used data from the ECLIPSE and SPIROMICS studies to estimate this ratio. In ECLIPSE, 18.2% of AECOPDs in the first year were severe; this value was 24.3% for SPIROMICS. This results in a sample-size-weighted average rate ratio of 20.3%.

Figure S2. Probability of being classified into different exacerbator phenotypes over two consecutive years given a stable underlying AECOPD rate.



Section III: OpenBUGS code for the final model

```
#ECLIPSE (data are coded as cumulative sum)
list(n_total=1.67900E+03, cumul_margin=c(0.00000E+00, 3.87000E+02, 4.89000E+02, 5.24000E+02, 6.26000E+02, 6.78000E+02, 7.13000E+02, 7.48000E+02,
7.83000E+02, 8.01000E+02, 8.86000E+02, 9.38000E+02, 9.56000E+02, 1.00700E+03, 1.04200E+03, 1.07700E+03, 1.11200E+03, 1.14700E+03, 1.19800E+03,
1.23300E+03, 1.25100E+03, 1.26900E+03, 1.30400E+03, 1.33900E+03, 1.39000E+03, 1.40800E+03, 1.47600E+03, 1.67900E+03))


#SPIROMICS (data are coded as cumulative sum)
list(n_total=1.10500E+03, cumul_margin=c(0.00000E+00, 5.64000E+02, 6.37000E+02, 6.57000E+02, 7.43000E+02, 7.67000E+02, 7.77000E+02, 8.12000E+02,
8.34000E+02, 8.55000E+02, 9.09000E+02, 9.21000E+02, 9.30000E+02, 9.51000E+02, 9.59000E+02, 9.62000E+02, 9.75000E+02, 9.88000E+02, 1.00000E+03,
1.01600E+03, 1.02300E+03, 1.03200E+03, 1.04100E+03, 1.04900E+03, 1.05600E+03, 1.06800E+03, 1.08000E+03, 1.10500E+03))


#Initial values (for both null and stable underlying rate models)
list(alpha=1, beta=1)



#unstable underlying rate model
  model()
  {
    for(i in 0:2)
    {
      for(j in 0:2)
      {
        for(k in 0:2)
        {
          for(l in  (cumul_margin[9*i+3*j+k+1]+1):cumul_margin[9*i+3*j+k+2])
          {
            Ns[l,1]<-i
            Ns[l,2]<-j
            Ns[l,3]<-k
          }
        }
      }
    }


    for(i in 1:n_total)
    {
      for(j in 1:3)
      {
        rate[i,j]~dgamma(alpha,beta)
        p[i,j]<-step(1.5-Ns[i,j])*pow(rate[i,j],Ns[i,j])*exp(-rate[i,j])+step(Ns[i,j]-1.5)*(1-exp(-rate[i,j])-exp(-rate[i,j])*rate[i,j])
      }

      p1[i]<-p[i,1]*p[i,2]*p[i,3]

      LL[i]<-log(p1[i])
      dummy[i]<-0
      dummy[i]~dloglik(LL[i])
    }
    alpha~dgamma(0.001,0.001)
    beta~dgamma(0.001,0.001)

    mu<-alpha/beta
    var_w<-alpha/beta/beta
  }



#Stable underlying rate model
model()
  {
    for(i in 0:2)
    {
      for(j in 0:2)
      {
        for(k in 0:2)
        {
          for(l in  (cumul_margin[9*i+3*j+k+1]+1):cumul_margin[9*i+3*j+k+2])
          {
```

```
        Ns[l,1]<-i
        Ns[l,2]<-j
        Ns[l,3]<-k
       }
     }
    }
  }


  for(i in 1:n_total)
  {
    my_rate[i]~dgamma(alpha,beta)
    for(j in 1:3)
    {
      rate[i,j]<-my_rate[i]
      p[i,j]<-step(1.5-Ns[i,j])*pow(rate[i,j],Ns[i,j])*exp(-rate[i,j])+step(Ns[i,j]-1.5)*(1-exp(-rate[i,j])-exp(-rate[i,j])*rate[i,j])
    }

    p1[i]<-p[i,1]*p[i,2]*p[i,3]

    LL[i]<-log(p1[i])
    dummy[i]<-0
    dummy[i]~dloglik(LL[i])
  }
  alpha~dgamma(0.001,0.001)
  beta~dgamma(0.001,0.001)

  mu<-alpha/beta
  var_b<-alpha/beta/beta
}
```

References

1. Hurst JR, Vestbo J, Anzueto A, et al. Susceptibility to exacerbation in chronic obstructive pulmonary disease. *N. Engl. J. Med.* 2010;363(12):1128–1138.

2. Han MK, Quibrera PM, Carretta EE, et al. Frequency of exacerbations in patients with chronic obstructive pulmonary disease: an analysis of the SPIROMICS cohort. *Lancet Respir. Med.* 2017;5(8):619–626.

3. Lunn D, Thomas A, Best N, et al. WinBUGS – A Bayesian modelling framework: Concepts, structure, and extensibility. *Stat. Comput.* 2000;10:325–337.

4. Vidakovic B. Engineering biostatistics: an introduction using MATLAB and WinBUGS. Hoboken, New Jersey: Wiley; 2017 964 p.