

# Learning endometriosis phenotypes from patient-generated data: Supplementary information

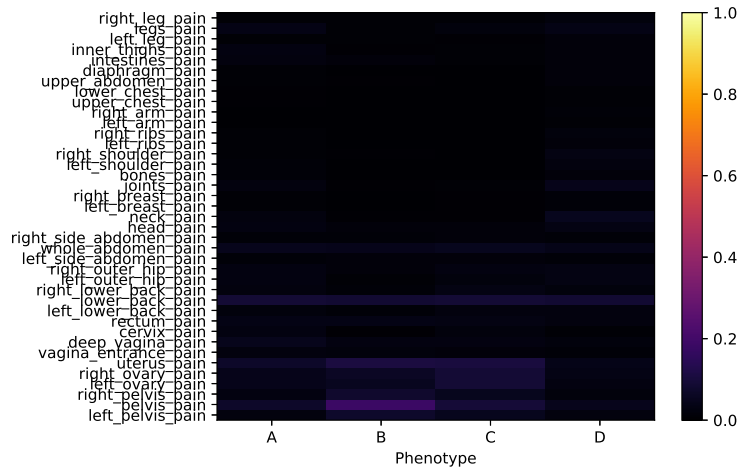
## 1 Supplementary Results

### 1.1 Per-question answer vocabulary

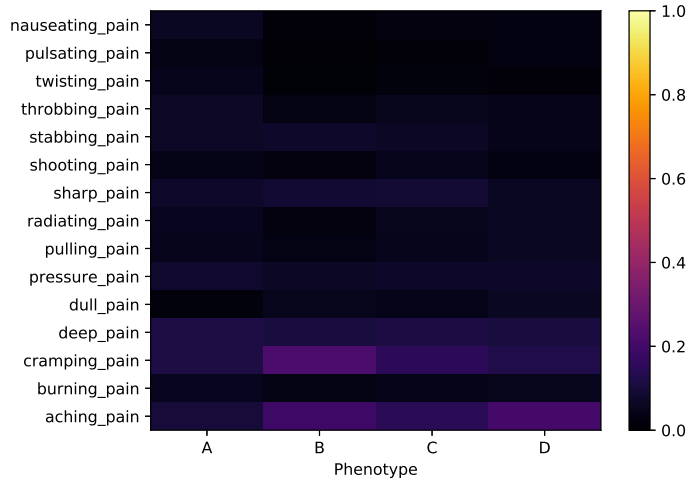
- **Where is the pain?:** bones\_pain, cervix\_pain, deep\_vagina\_pain, diaphragm\_pain, head\_pain, inner\_thighs\_pain, intestines\_pain, joints\_pain, left\_arm\_pain, left\_breast\_pain, left\_leg\_pain, left\_lower\_back\_pain, left\_outer\_hip\_pain, left\_ovary\_pain, left\_pelvis\_pain, left\_ribs\_pain, left\_shoulder\_pain, left\_side\_abdomen\_pain, legs\_pain, lower\_back\_pain, lower\_chest\_pain, neck\_pain, pelvis\_pain, rectum\_pain, right\_arm\_pain, right\_breast\_pain, right\_leg\_pain, right\_lower\_back\_pain, right\_outer\_hip\_pain, right\_ovary\_pain, right\_pelvis\_pain, right\_ribs\_pain, right\_shoulder\_pain, right\_side\_abdomen\_pain, upper\_abdomen\_pain, upper\_chest\_pain, uterus\_pain, vagina\_entrance\_pain, whole\_abdomen\_pain
- **Describe the pain:** aching\_pain, burning\_pain, cramping\_pain, deep\_pain, dull\_pain, nauseating\_pain, pressure\_pain, pulling\_pain, pulsating\_pain, radiating\_pain, sharp\_pain, shooting\_pain, stabbing\_pain, throbbing\_pain, twisting\_pain
- **How severe is the pain?:** mild\_pain, moderate\_pain, severe\_pain
- **What are you experiencing?:** allergies, asthma, blurry\_vision, chest\_pressure, dizziness, eczema, fatigue, fever, headache, hives, hot\_flash, itchy, mentally\_foggy, noise\_sensitivity, numbness, rash, ringing\_in\_ears, sinus\_congestion, sweaty, swelling, touch\_sensitivity
- **How severe is the symptom?:** mild\_symptoms, moderate\_symptoms, severe\_symptoms
- **Describe your period flow:** heavy\_flow, light\_flow, medium\_flow
- **What kind of bleeding:** breakthrough\_bleeding, clots, no\_bleeding, spotting
- **Describe GI/GU system:** blood\_in\_stool, cant\_urinate, constipation, diarrhea, endo\_belly, frequent\_urination, gas, heartburn, mouth\_sores, nausea, painful\_bowel\_movement, painful\_urination, stomach\_upset, uncomfortably\_full, vomiting
- **How severe is it?:** mild\_GI, moderate\_GI, severe\_GI
- **Describe sex:** avoided\_sex, bleeding\_from\_sex, no\_sex, painful\_after\_sex, painful\_during\_sex, sex\_felt\_good
- **Activities difficult to perform:** climb\_stairs, eat, get\_dressed, get\_out\_of\_bed, have\_sex, housework, jump, kneel, lie\_down, lift, no\_trouble, prepare\_food, run, shop, shower, sit\_down, sleep, socialize, stand, stretch, use\_toilet, walk, work
- **How was your day?:** bad\_day, good\_day, great\_day, manageable\_day, unbearable\_day
- **Medications/hormones taken:** adrenergic\_agonists, amphetamine, analgesic, analgesic/narcotic, analgesic/nsaids, analgesic/opioids, anesthetic, anorectic, anti-inflammatory, antacid, antacid/nsaids, antibiotics, anticholinergic, anticoagulant, anticonvulsant, antidepressant, antidiabetic\_medication, antidiarrheal, antiemetic, antihistamine, antihypertensive, antipsychotic, antispasmodic, antispasmodic/sedative, anxiolytic, anxiolytic/anesthetic/muscle\_relaxant, barbituate, barbituate/analgesic, beta\_blocker, bronchodilator, calcium\_channel\_blocker, cough\_medicine, decongestant, diuretic,

dopamine\_agonist, estrogen, estrogen/progestin, gonadotropin-releasing\_hormone\_agonist, gonadotropin-releasing\_hormone\_antagonist, hormone\_based\_chemotherapy, hormone\_replacement\_therapy, human\_chorionic\_gonadotropin, human\_follicle\_stimulating\_hormone, laxative, muscle\_relaxant, narcotic, narcotic/nsaids, neuropathic\_pain\_medication, no\_med\_hormones, noclass, nonbenzodiazepine\_hypnotic, nsaids, opioids, progestin, sedative, statin, steroid, stimulant, thyroid\_hormones, topical\_anti-tumor\_medication, triptan, vasoconstrictor, vitamin\_a\_derivative

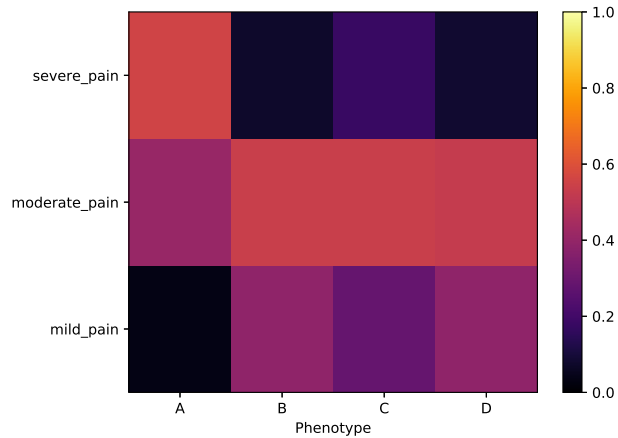
## 1.2 Per-question posteriors



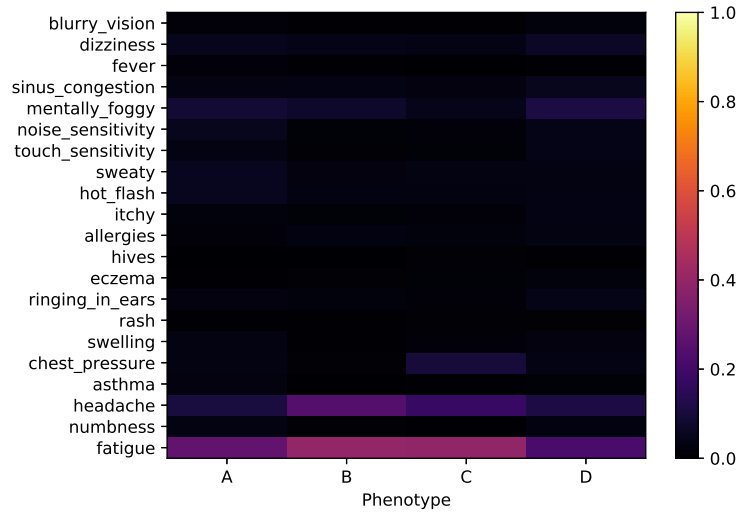
(a) Where is the pain?



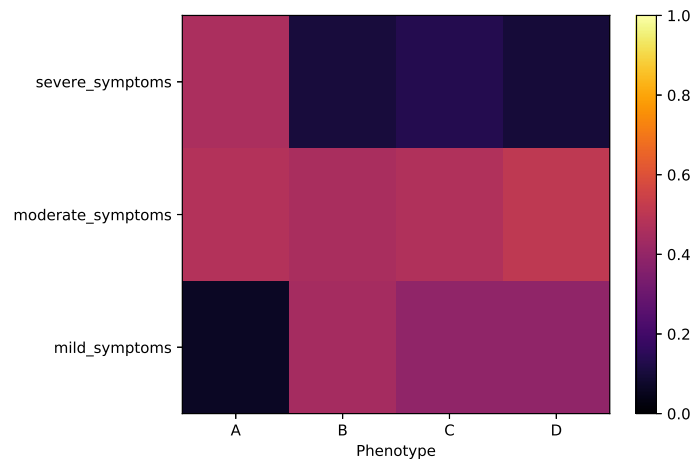
(b) Describe the pain.



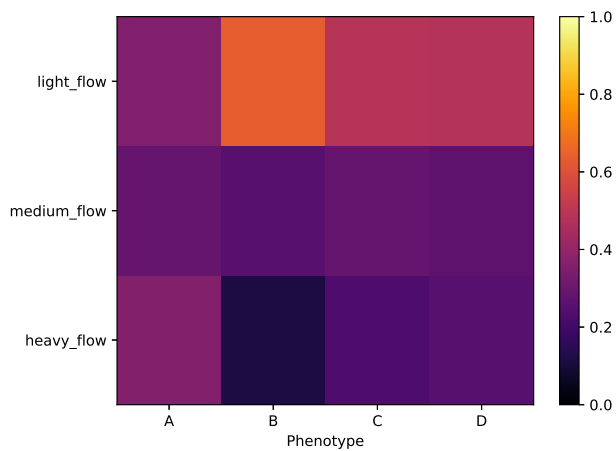
(c) How severe is the pain?



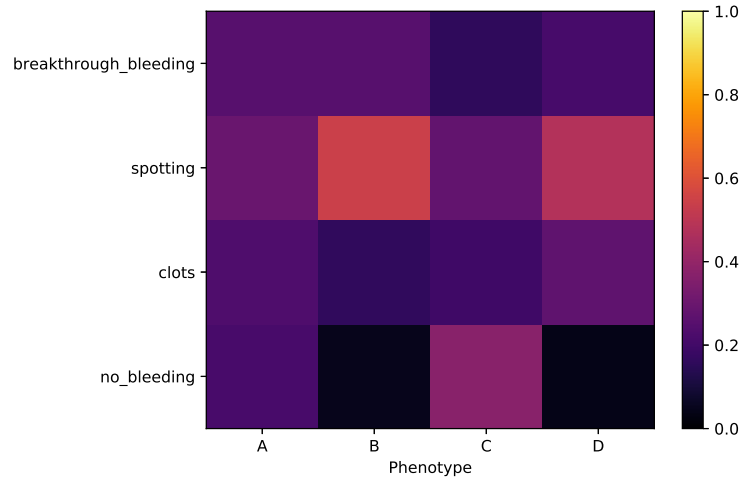
(d) What are you experiencing?



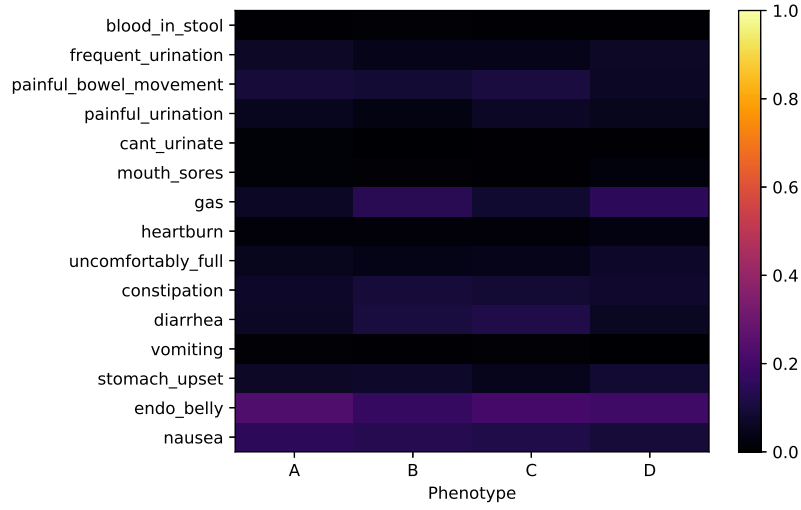
(e) How severe is the symptom?



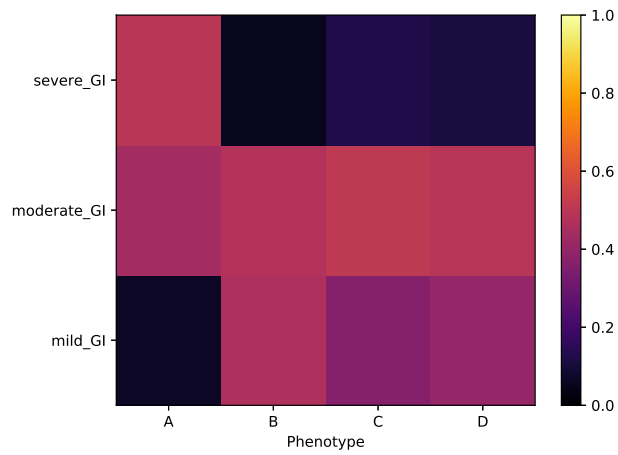
(f) Describe your period flow.



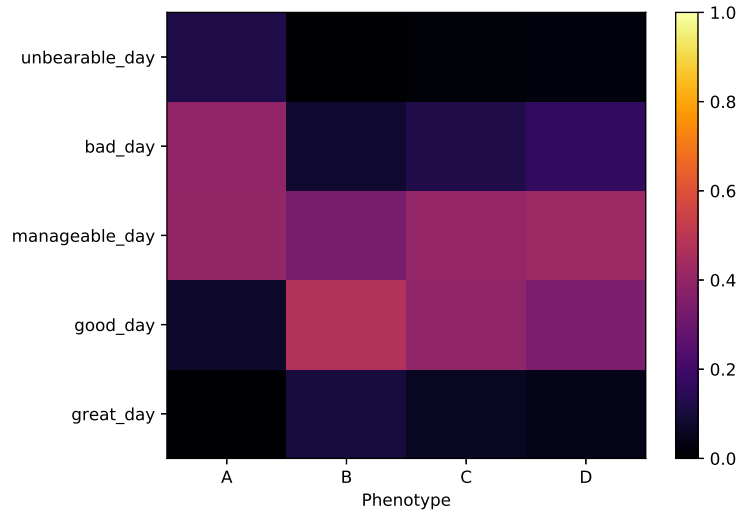
(g) What kind of bleeding.



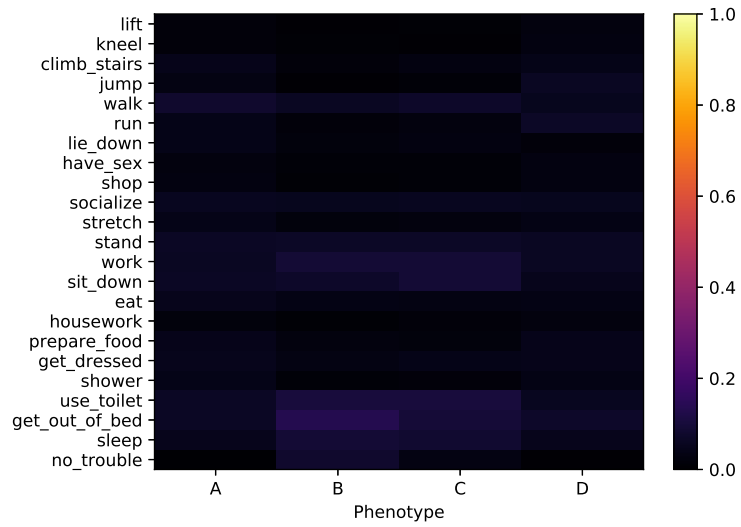
(h) Describe GI/GU system.



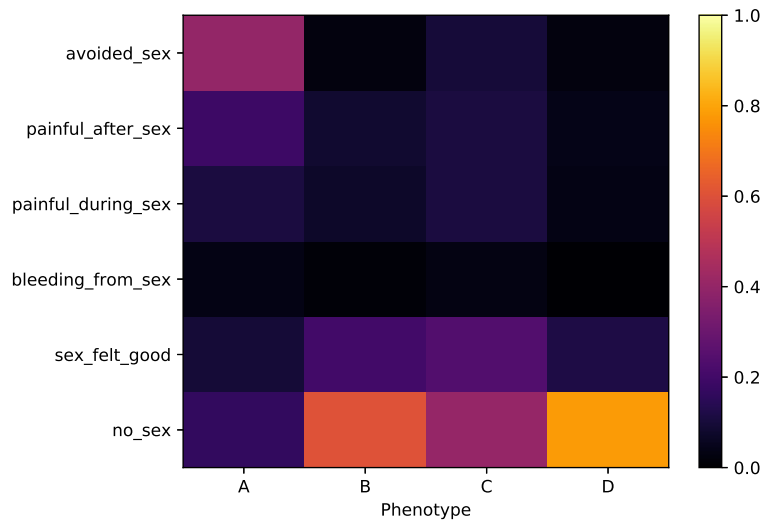
(i) How severe is it?



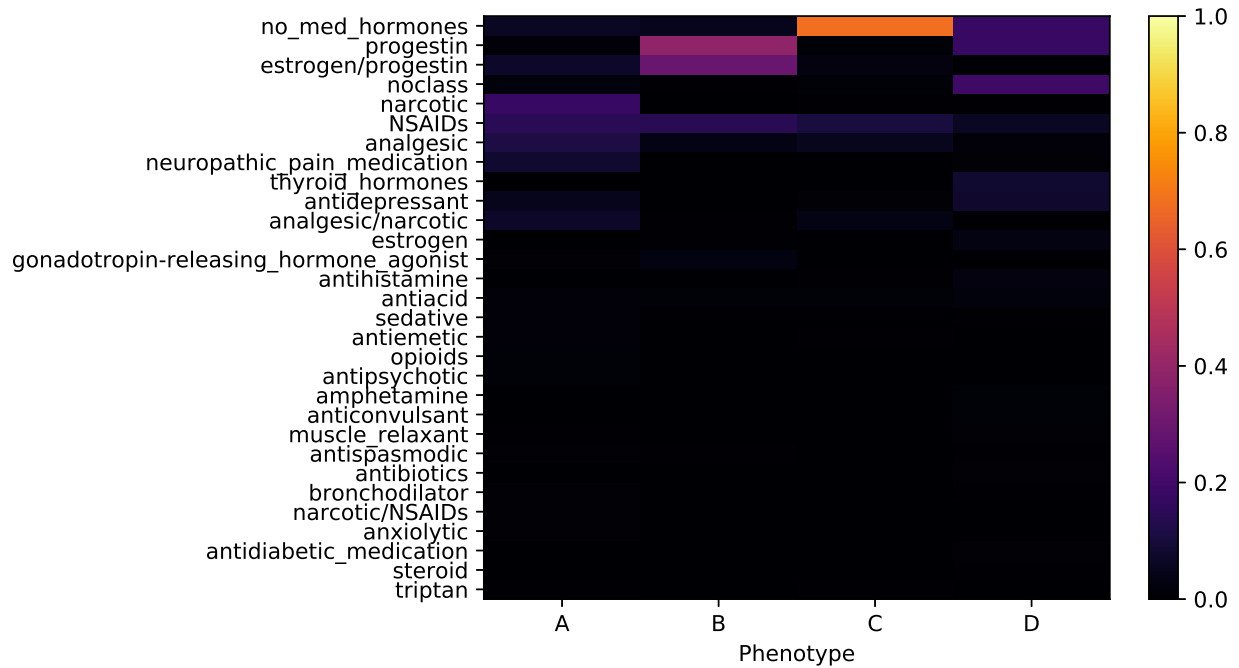
(j) How was your day?



(k) Activities difficult to perform.



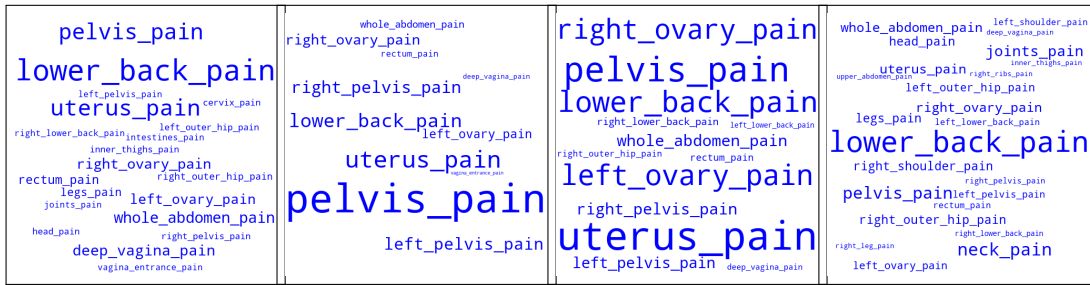
(l) Describe sex.



(m) Medications/hormones taken. For visual clarity, only the top 30 (most likely) vocabulary items of the posterior are displayed.

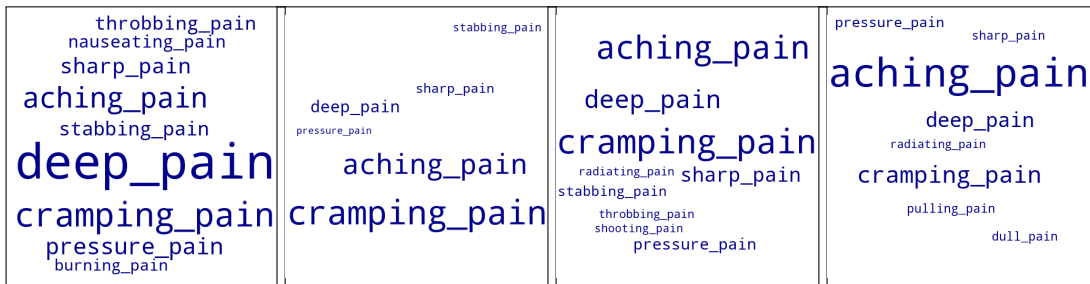
Supplementary Figure 1: Visualization of learned posteriors for endometriosis phenotypes. Each phenotype is defined as a set of per-question probability distributions across the answers to each of the thirteen questions. Each heatmap represents the likelihood of the answers within a question for a given phenotype. For instance, the “*no\_sex*” answer is highly likely to be tracked under phenotype D, and not likely to be tracked under phenotype A —yellow versus purple respectively, in heatmap (l).

### 1.3 Per-question answer-clouds



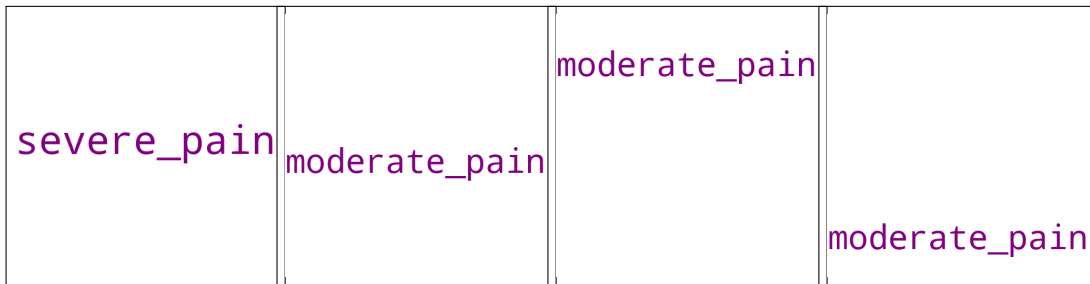
(a) Phenotype A      (b) Phenotype B      (c) Phenotype C      (d) Phenotype D

Supplementary Figure 2: Where is the pain.



(a) Phenotype A      (b) Phenotype B      (c) Phenotype C      (d) Phenotype D

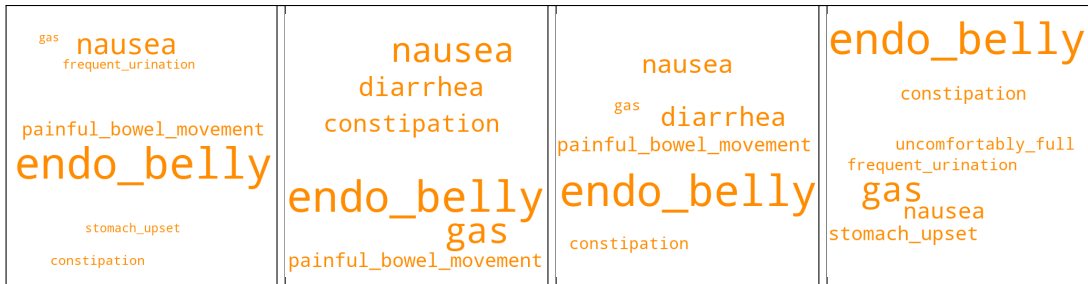
Supplementary Figure 3: Describe the pain.



(a) Phenotype A      (b) Phenotype B      (c) Phenotype C      (d) Phenotype D

Supplementary Figure 4: How severe is the pain.





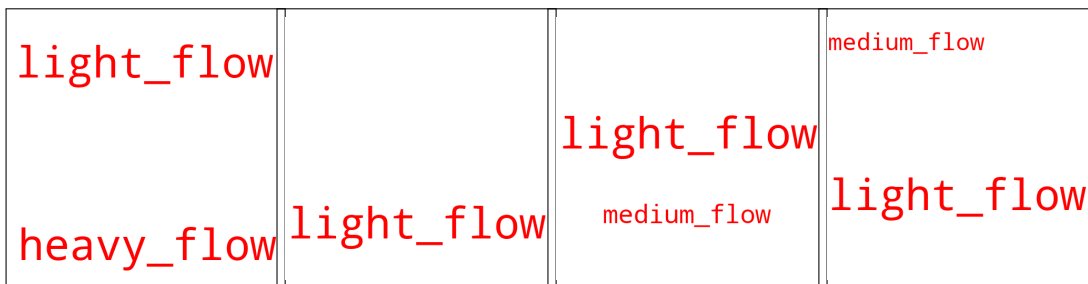
(a) Phenotype A      (b) Phenotype B      (c) Phenotype C      (d) Phenotype D

Supplementary Figure 5: Describe GI/GU system.



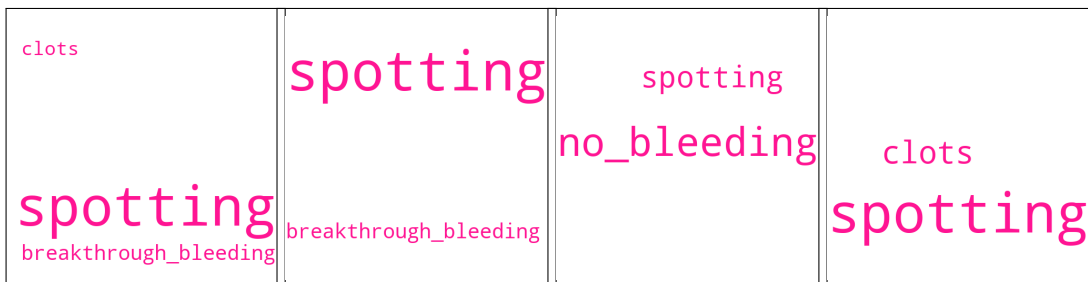
(a) Phenotype A      (b) Phenotype B      (c) Phenotype C      (d) Phenotype D

Supplementary Figure 6: How severe is it.



(a) Phenotype A      (b) Phenotype B      (c) Phenotype C      (d) Phenotype D

Supplementary Figure 7: Describe your period flow.



(a) Phenotype A      (b) Phenotype B      (c) Phenotype C      (d) Phenotype D

Supplementary Figure 8: What kind of bleeding.



(a) Phenotype A      (b) Phenotype B      (c) Phenotype C      (d) Phenotype D

Supplementary Figure 9: Describe sex.



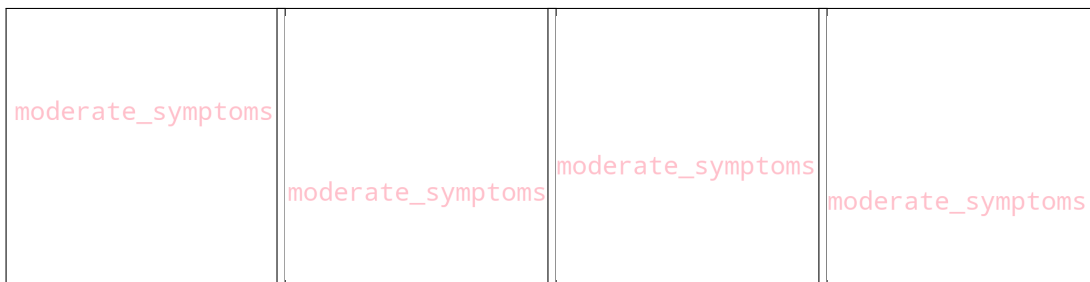
(a) Phenotype A      (b) Phenotype B      (c) Phenotype C      (d) Phenotype D

Supplementary Figure 10: Medications/hormones taken.



(a) Phenotype A      (b) Phenotype B      (c) Phenotype C      (d) Phenotype D

Supplementary Figure 11: What are you experiencing.



(a) Phenotype A      (b) Phenotype B      (c) Phenotype C      (d) Phenotype D

Supplementary Figure 12: How severe is the symptom.



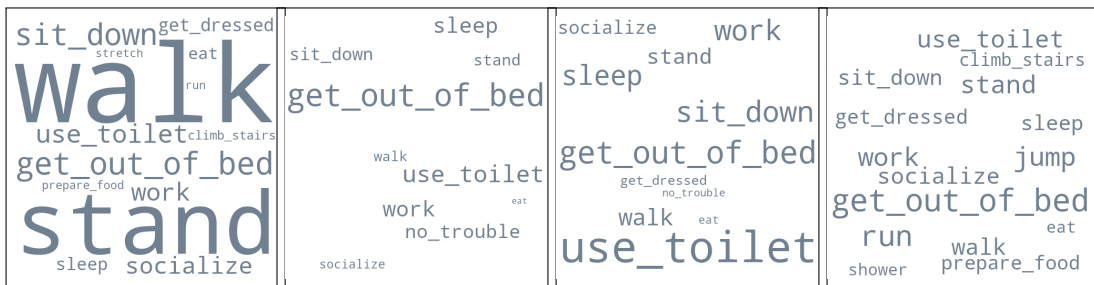
(a) Phenotype A

(b) Phenotype B

(c) Phenotype C

(d) Phenotype D

Supplementary Figure 13: How was your day?



(a) Phenotype A

(b) Phenotype B

(c) Phenotype C

(d) Phenotype D

Supplementary Figure 14: Activities difficult to perform

## 2 Supplementary Methods

### 2.1 K-means baseline clustering

A plausible baseline clustering approach is to group the available data in the Euclidean space formed by the concatenation of all the per-question response frequencies. That is, to count the number of times each user self-tracks a vocabulary item, and to form a count-based (or the corresponding normalized density) based feature vector. For this approach, the collection of responses to different questions  $q = 1, \dots, Q$  with their vocabulary size  $V_q$  is concatenated.

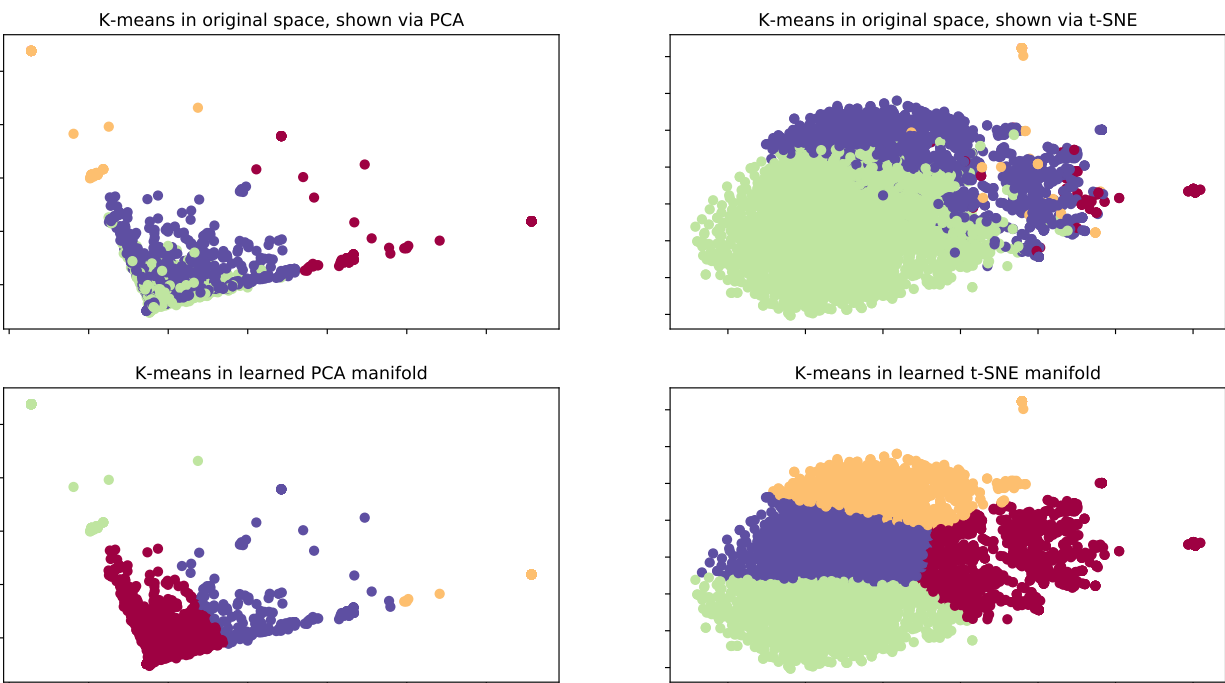
We find that a K-means based clustering in such  $(V_1 + V_2 + \dots + V_Q)$ -dimensional Euclidean space does not attain any insightful endometriosis phenotypes: only two clusters dominate the groupings (see Supplementary Figure 15), and they provide no clinically meaningful associations with the WERF questionnaire responses. Besides, such an approach is hard to interpret clinically, as it is based on distances in a very high dimensional Euclidean space—the input is a  $(V_1 + V_2 + \dots + V_Q)$ -dimensional vector.

An alternative baseline is to first reduce the input feature dimensionality—e.g., learn a map to a lower (2 or 3) dimensional space—and then cluster in this embedded space. We have successfully identified, via K-means, 4 separate cluster in such embedded space (shown in Supplementary Figure 15) both based on linear and non-linear dimensionality reduction techniques. However, none of the learned phenotypes are discriminative on clinically meaningful features. We find that many learned K-means clusters are associated with clinically meaningful outcomes (such as daily living, overall health or number of laparoscopies), and are therefore not discriminative on endometriosis relevant responses.

Besides, the interpretability of the learned phenotypes is hindered by the dimensionality reduction technique used. For example, if using the nonlinear t-Distributed Stochastic Neighbor Embedding (t-SNE) method that attains the most discriminative separation, one is faced with two challenges: (i) that as t-SNE optimizes a non-convex objective function via randomly initialized gradient descent algorithms, different runs result in different solutions; and (ii) that t-SNE does not retain distances from original to embedded spaces, resulting in uninterpretable original high-dimensional Euclidean distances.

On the contrary, we propose an extended mixed-membership model, which is a Bayesian generative model that can accommodate the inherent heterogeneity and uncertainty of the data, to capture the latent structure of collections of groups of self-tracked signs and symptoms.

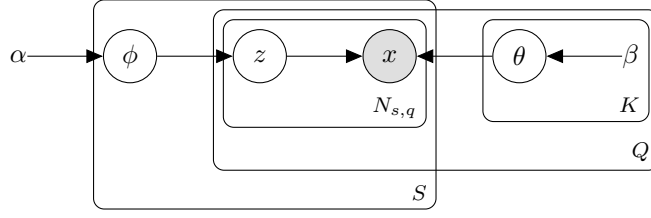
### Clustering alternatives



Supplementary Figure 15: K-means based clustering results: (top left) K-means in original space, shown in 2-d embedding learned by PCA; (top right) K-means in original space, shown in 2-d embedding learned by t-SNE; (bottom left) K-means in embedded 2-dimensional space learned via PCA; (bottom right) K-means in embedded 2-dimensional space learned via t-SNE.

## 2.2 Extended Mixed membership model

The proposed extended mixed-membership model is illustrated in Supplementary Figure 16 and described below.



Supplementary Figure 16: Probabilistic graphical model of the proposed mixed-membership model.

The per-question mixed-membership generative process for subjects  $s = 1, \dots, S$  follows:

1. Draw per-subject phenotypic proportions  $\phi_s \sim \text{Dirichlet}_K(\phi|\alpha)$  with hyperparameter  $\alpha$ .
2. Draw per-phenotype and per-question response proportions  $\theta_{k,q} \sim \text{Dirichlet}_{V_q}(\theta|\beta_{k,q})$ , for all phenotypes  $k = 1, \dots, K$ , and questions  $q = 1, \dots, Q$  with vocabulary size  $V_q$ , and hyperparameters  $\beta_{k,q}$ .
3. Draw per-subject observation phenotype assignments  $z_{s,n} \sim \text{Categorical}_K(z|\phi_s)$ , for  $n = 1, \dots, N_s$ .
4. Draw per-subject questions responses  $x_{s,n}|z_{s,n}, q_{s,n} \sim \text{Categorical}_{V_{q_{s,n}}}(x|\theta_{z_{s,n}, q_{s,n}})$ , for  $n = 1, \dots, N_s$ , where  $q_{s,n}$  indicates the response  $n$  to question  $q$  by subject  $s$ .

After observing a data-set with  $N_{s,q}$  responses per-subject and question, the goal is to infer the phenotypic proportions  $\phi_s$  and the set of  $K$  phenotypes parameterized by  $\theta_{k,q}$  per-question. To that end, and due to the conjugacy assumptions in the generative process, we resort to a collapsed Gibbs sampler that utilizes the following distribution

$$p(z_{s,n^*} = k|x_{s,n^*}, q_{s,n^*}, \alpha_N, \beta_N) \propto p(z_{s,n^*} = k|\alpha_{s,N})p(x_{s,n^*}|q_{s,n^*}, z_{s,n^*}, \beta_{k,q,N}),$$

$$\text{with } \begin{cases} p(z_{s,n^*} = k|\alpha_{s,N}) = \frac{\alpha_{k,s,N}}{\sum_{k=1}^K \alpha_{k,s,N}}, \\ p(x_{s,n^*}|q_{s,n^*}, z_{s,n^*}, \beta_{k,q,N}) = \frac{\beta_{k,q,v_{q,N}}}{\sum_{v_q=1}^{V_q} \beta_{k,q,v_q,N}}, \end{cases} \quad (1)$$

where the parameters for the updated posteriors are

$$p(\phi_s|Z_N, \alpha_0) = \text{Dirichlet}_K(\phi_s|\alpha_{s,N}), \quad \text{with } \alpha_{k,s,N} = \alpha_{k,0} + N_{s,k};$$

$$p(\theta_{k,q}|X_N, Q_N, Z_N, \beta_{k,q,0}) = \text{Dirichlet}_{V_q}(\theta_{k,q}|\beta_{k,q,N}), \quad \text{with } \beta_{k,q,v_q,N} = \beta_{k,q,v_q,0} + N_{k,q,v_q}. \quad (2)$$