

ASSESSING HETEROGENEITY IN MENSTRUAL CYCLES BY MEANS OF A MULTILEVEL LATENT CLASS APPROACH

Francesca Bassi^a, Bruno Scarpa^a

^a*Department of Statistical Sciences, University of Padua, Italy¹*

ABSTRACT

In this paper, we study the problem of heterogeneity in cervical mucus hydration at different times relatively to the mucus peak both between cycles and women, specifying and estimating appropriate multilevel latent class models for longitudinal data. We estimate multilevel and growth latent class models which classify women on the basis of the evolution of cervical mucus characteristics observed over the fertile period of each menstrual cycle taking into account that we observe a different number of cycles per woman, and correlation over time between consecutive observation. The effect of potential covariates on mucus evolution patterns is as well evaluated. Results confirm the existence of heterogeneity in mucus evolution between cycles and women. Moreover, an important significant effect of woman's age is found.

KEYWORDS

cervical mucus, peak day, multilevel models, growth models, heterogeneity

1. Introduction

The observation of the cervical mucus symptom (CMS) is a widely used indicator to identify ovulation and the fertile phase in a menstrual cycle (e.g., Billings et al. 1972). The mucus symptom allows a woman to precisely define the beginning of this phase and informs on the event of ovulation because cervical mucus secretions, stimulated by a rise in oestrogen, are known to increase in volume about 5-6 days prior to ovulation

¹ CONTACT Francesca Bassi Email: francesca.bassi@unipd.it

(Katz, Slade, and Nakajima 1997; Moghissi, Syner, and Evans 1972). Ovulation is the key event in a menstrual cycle that determines the fertile interval during which intercourse can potentially result in pregnancy. Although women typically do not know when they ovulate, the cycle day of ovulation may be detected retrospectively observing mucus characteristics evolution over cycle days (World Health Organisation 2000; Hilgers and Prebil 1979). The mucus peak day is considered to be the last day in the cycle during which at least one characteristic of high fertility in mucus type has been observed or felt, moreover, this day must be preceded by an adequate growth in sensation and appearance of mucus more fertile characteristics which should also show afterwards a clear change to the less fertile (Fehring 2002). Ovulation is expected within two days after the peak. The width of the fertile window around ovulation, that is the number of days during which intercourse has a non-zero probability of resulting in conception is uncertain, ranging in the reference literature from two to ten days. However, some studies shown that the probability that a pregnancy results from a single act of intercourse is small unless intercourse occurs in the interval starting from five days before ovulation and ending on the day of ovulation (see, for example, Wilcox, Dunson, and Baird 2000).

There is a wide interest in predicting the fertile days in a menstrual cycle among couples desiring a pregnancy and among those wishing to avoid conception by periodic abstinence (Scarpa, Dunson and Colombo 2006; Scarpa and Dunson 2007). Cervical mucus detection is potentially an accurate marker of the fertile days (Katz 1991; Dunson, Sinai, and Colombo 2001; Scarpa, Dunson and Colombo 2006). Therefore, it is of substantial interest to assess the magnitude of heterogeneity among women and among cycles from a given woman in the trajectory of the detected mucus secretions at different times during an interval of potential fertility defined relatively to ovulation (Dunson and Colombo 2003). Recently, for example, Pennoni et al. (2017) estimated latent Markov models to determine the fertile window in a woman's cycle, exploiting information on mucus and basal body temperature registered by couples instructed in a number of European centres teaching natural family planning methods (Colombo Masarotto, 2000).

In this paper, we study the problem of heterogeneity in cervical mucus hydration at different times relatively to the mucus peak both between cycles and women, specifying and estimating appropriate multilevel latent class models for longitudinal data. De Bin

and Scarpa (2014) faced a similar problem by using Bayesian nonparametric tools, without easily accounting for the multilevel characteristics of the problem. We dispose of a two-level dataset: observed cervical mucus characteristics in menstrual cycles of women and we address the measurement of heterogeneity in trajectories following two ways. We estimate multilevel latent class models, which classify women on the basis of the evolution of cervical mucus characteristics observed over the fertile period of each menstrual cycle, taking into account that we observe a different number of cycles per woman and that there exists also heterogeneity between cycles. The model includes over time correlation among observations and estimates the effect of potential covariates on mucus evolution patterns. The estimation of a multilevel latent class growth mixture model (Gomes and Dias 2015) classifies women, again on the basis of observed mucus evolution in the fertile phase of their cycles, which is assumed to follow a latent trajectory with random coefficients; this allows to analyse more in depth the effect of the age of the women on mucus heterogeneity.

This paper is based on a previous work on heterogeneity on cervical mucus performed on a much smaller sample of women (Bassi and Scarpa 2015) and it extends the analyses to all women participating in the study described in detail in Section 2. Moreover, in this work, the specification of the multilevel models is improved in various directions, starting from evidences in the data, and the potential effect of covariates on mucus evolution over time is estimated.

Results confirm what emerged in the previous study: heterogeneity in mucus evolution between cycles and women is non-negligible. The refined model specifications reinforces this result. Moreover, an important significant effect of woman's age is found. In particular, as age increases, variation of mucus type over the fertile window diminishes.

2. The study design

The data used in the paper were collected by prof. Bernardo Colombo in a study in collaboration with four Italian Billings centres and are now set at disposal for researchers by the Department of Statistical Sciences of the University of Padua. The four Italian centers that gave their collaboration were: Centro Lombardo Metodo Billings, Milan; Centro Piemontese Metodo Billings, Saluzzo; Associazione Metodo

Billings Emilia Romagna, Parma and Centro Studi e Ricerche per la Regolazione della Fertilità of the Catholic University of Milan. These centres give advice to subjects interested in learning about the fertile phase of the woman and the use of natural family planning methods to avoid and achieve pregnancies.

2.1. Some notes on the Billings method

In the Billings method for natural family planning (Billings and Westmore 1998) it is admitted that the woman describes the mucus symptoms with a personal terminology that the teacher following her is able to understand. Conventional signs are used to indicate the woman's interpretation of her condition, particularly concerning fertile and infertile days: coloured stamps, standard adjectives, symbols for electronic processing. In this exercise, ordered numerical codes reflecting usual phenomenology of the symptom have been adopted. The Billings ovulation method is based on the observation by the woman of the mucus symptoms, that is, the characteristics of her cervical mucus coming out of the vagina, and the assessment of the sensation that the mucus entails at the level of the vulva. The observations, registered on a special chart, allow the determination of the fertile and infertile phase of each cycle. This method divides the menstrual cycle into sections. The days of the period are considered potentially fertile since detection of mucus is impossible. The period may be followed by an infertile pre-ovulatory phase. The appearance of the symptoms marks the beginning of the fertile phase. In general, the characteristics of the signs are not the same for all women. Usually, it is possible to identify the so-called basic infertile pattern (BIP) during the days immediately following menses. The most common pattern is characterized by a dry sensation and the absence of mucus. Typically, a woman can easily identify this pattern from her first cycle of observation (dry BIP). The second one, with an unchanging mucus pattern (u.m. BIP), is distinguished by damp sensation and/or continuous mucus discharge. In the last instance, the characteristics of sensation, appearance and consistency of the discharge remain steadily unchanged day after day. The woman will have learned recognizing such a BIP after a suitable number of cycles (generally three). When cervical mucus with characteristics different from those typical during the BIP is observed, the potentially fertile phase of the cycle is considered to begin.

A crucial indicator is the symptom of mucus peak. A conventional definition of the peak day has been agreed upon among the four centres participating in the study. According to this agreement, the mucus peak day is considered to be “the last day of the cycle during which at least one characteristics of high fertility in mucus type has been observed or felt, considering characteristics of high fertility the wet sensation and/or the observation of slippery, transparent, liquid or watery mucus, or blood trails. Moreover, this day must be preceded by an adequate growth in sensation and appearance of mucus characteristics which should also show afterwards a clear change to the less fertile” (Billings and Westmore 1998). Ovulation is expected within two days following the peak and is then used as a reference for determining the end of the fertile phase. When in a cycle no peak is detected, it is judged that ovulation did not occur, and it is not possible to identify the postovulatory phase.

The four Italian Billings centres that participate in the study sent to the Department of Statistical Sciences of the University of Padua the charts in which women, referring to each centre, registered the characteristics of menstrual cycles (a complete description of the study and the data can be found in Colombo 2007).

The charts include general information on the cycle: date of beginning and length; duration of menses; occurrence of conception; type of BIP; the identification of the day of the detected peak, by the woman, by the teacher and, eventual biological peak; qualification of the cycle, whether there is incomplete information on mucus or missing data, disturbances, stress that do not allow the identification of the peak mucus day or no information on mucus or unprotected intercourse and eventual pregnancy. For each cycle day, the woman was requested to indicate mucus typology, presence of disturbances and of unprotected intercourse acts. The daily observations by the woman are classified according to the numerical codes described in Table 1. The Billings method describes mucus symptoms as ordered: mucus with an assigned higher label (see Table 1) represents more fecund days in the cycle than mucus with lower labels. In the same table, also codes to classify the BIP are reported.

Homogeneity in the information sent by the four centres is assured by the fact that all of them provided instructions on the same method of natural family planning, applying uniform procedures in teaching, practical applications, linguistic descriptions and conventions (Colombo et al. 2006).

Table 1. Classification of cervical mucus sensation and appearance and BIP

CODE	MUCUS SENSATION	MUCUS APPEARENCE
0	No information	No information
1	No sensation or dry sensation	No mucus, no discharge
2	No more dry sensation	No mucus, or insubstantial discharge
3	Damp sensation	Thick, creamy, whitish, yellowish, sticky, stringy mucus
4	Wet, liquid (no slippery) sensation	
5	Wet-lubricated, slippery sensation	Clear, stretchy, liquid, watery mucus, blood trails
BIP		
0	Non identified	
1	Dry (or first type) mucus	
2	Unchanging mucus in sensation or appearance	
3	Unchanging mucus in both sensation and appearance	

Although charts are strictly anonymous, some information was collected on women and partners demographic characteristics and reproductive history. Specifically, woman's and partner's date of birth, date of marriage or beginning of relationship of the couple; and with reference to the woman, number of pregnancies before entering the study; date and type (miscarriage, breastfeeding, childbirth) of last event before entering the study; information on taking hormonal contraception and, eventually, date on which last pill was taken; date on which the woman leaves the study and for which reason (pregnancy, miscarriage not later than 60 days since the beginning of last period, drop-out, end of the study); information about going or not with pregnancy after 60 days after the last period; date of positive pregnancy test; date of childbirth and result of the pregnancy (number of children and gender).

2.3. The data

The four Billings centres provided 2,914 cycles registered by 193 women; this data is recoded in two databases: "ciclo-Billings" containing information on cycles characteristics and "donna-Billings" containing information on women and partners. A unique code assigned to each woman and the cycles she registered allows to integrate the two databases. The first charted cycle began in 1976 and the last one in 1998. Women provided from 1 to 104 cycles, not all of them are consecutive since registration might have been suspended for events as childbirth, breastfeeding, miscarriage or other: 157 women provided only one group of consecutive cycles, 29 women two groups of consecutive cycles, 3 women three groups, 3 women four groups, 1 woman five groups. We selected cycles with identified mucus peak day and with complete information on mucus registered from the fifth day before the peak to day one after the peak. We ended

up with 2,284 cycles provided by 188 women. Table 2 reports descriptive statistics of the main concerned variables in our sample.

Table 2. Descriptive statistics

Variable	Mean	SD
Woman's age	30.37	4.357
Number of previous pregnancies	1.31	1.255
Pre-ovulatory phase length	16.44	4.418
Period length	5.56	1.287
Cycle with conception	0.06	0.242
Mucus at day -5 from peak	3.24	1.359
Mucus at day -4 from peak	3.72	1.234
Mucus at day -3 from peak	4.15	1.100
Mucus at day -2 from peak	4.54	0.851
Mucus at day -1 from peak	4.81	0.594
Mucus at peak day	4.89	0.375
Mucus at day +1 from peak	2.48	0.811

3. Analysis of heterogeneity

The database at our disposal has a multilevel nature: mucus observations (registered by each woman using the codes described in Table 2) in the days composing the fertile window are nested in cycles that are nested in women participating to the study. Starting from the exercise presented in a previous research (Bassi and Scarpa 2015), we extend the analysis to a larger database and we specify and estimate more complex multilevel models that better describe the nature of the collected data. Finally, the potential effect of cycle-level and women-level covariates is explored.

In the previous research, a multilevel level latent class model (Vermunt 2003 and equation 1) identified clusters of cycles (level-1 units) and classes of women (level-2 units) showing the existence of heterogeneity in mucus trajectories over time at both levels.

$$P(\underline{Y}_j = \underline{s}) = \sum_{h=1}^H [P(W_j = h) \prod_{i=1}^{n_j} [\sum_{l=1}^L P(X_{ij} = l | W_j = h) \prod_{k=1}^K P(Y_{ijk} = s_k | X_{ij} = l)]]$$

(1),

where

Y_{ijk} , $i=1, \dots, I$, $j=1, \dots, J$, $k=1, \dots, K$, denote the observation of mucus in day k of the fertile window in cycle i by woman j ;

s_k , $s_k = 1, \dots, S_k$, the particular value of mucus observed in day k ;

X_{ij} , a latent variable with L classes;

199 l , a particular latent class, $l=1, \dots, L$;

200 \underline{Y}_j , the full vector of observations by woman j , containing ordinal variables;

201 \underline{s} , a possible response pattern, i.e., it contains the sequence of mucus symptoms
202 observed in the K days of the fertile window by each woman; its generic element is s_k .

203 W_j denotes the latent variable at the woman level, assuming value h , with $h=1, \dots, H$;

204 n_j the size of group j , i.e., the number of cycles recorded by woman j .

205 With reference to our extended dataset, K is equal to 7, since we follow Barrett and
206 Marshall (1969) in considering mucus observations in the seven days of the fertile
207 window: five before the peak, the peak day and one day after; J is the number of women
208 participating to the study: 193, and I is the total number of cycles about which we
209 dispose of information, 2,914.

210 As earlier discussed, (see Table 1), labels for mucus symptoms are ordered, so that
211 we treat them as an ordinal variable. The multilevel LC model in equation (1) is based
212 on the classical local independence assumption, which implies that observations Y_{ijk} are
213 independent conditional on LC membership. This hypothesis is not realistic for our
214 dataset where mucus observed for day k depends on the observations of the preceding
215 days of the cycle. We introduced over time correlation among mucus observations
216 inserting direct effects (Bassi et al. 2000) into the basic multilevel latent class model:

$$P(\underline{Y}_j = \underline{s}) = \sum_{h=1}^H [P(W_j = h) \prod_{i=1}^{n_j} [\sum_{l=1}^L P(X_{ij} = l | W_j = h) P(Y_{ij-5} = s_{-5} | X_{ij} = l) \prod_{k=-4}^{+1} P(Y_{ijk} = s_k | Y_{ijk-1} = s_{k-1}, X_{ij})]]$$

(2)

219 Equation (2) is obtained allowing that mucus observation for day k of the fertile
220 window depends on observation on the preceding day. These direct effects are imposed
221 to be equal for parsimony reasons but this is also a sensible assumption since there is no
222 need to think that association among couples of mucus observations changes over time.
223 Following the seminal paper by Barrett and Marshall (1969), we assume that the fertile
224 window includes seven days around the ovulation, therefore, we leave k varying
225 between -5 to +1 with reference to the peak day, for which k is equal to 0. The
226 multilevel LC model with direct effects has a better fit to the data than the model
227 assuming local independence, showing that over time correlation among observations
228 has not to be neglected.

The multilevel LC model with direct effects which revealed the best fit to the data² estimates 4 latent classes of women (level-2 units) and 4 classes of level-1 units (clusters of cycles). To evaluate the fit, we looked at the BIC index (Schwartz 1978). Model indicators are the reported codes for mucus, in the potentially fertile window from the fifth day before the mucus peak to the first day after the peak, and the BIP; for the BIP, local independence on the other indicators is assumed. Pre-ovulatory phase length, period length, information if the cycle is with conception, and the number of unprotected intercourse in the fertile window were considered as inactive covariates for clusters; woman's age is treated as an active covariate.

Estimation results with the best-fitting model are shown in Table 3. Four clusters of cycles are identified. The lower panel of the table shows the average level of mucus code on each day of the potentially fertile window in the four clusters. The same results are used in Figure 1 to describe cluster profiles, to aid interpretation. Mucus scores evolve over time quite differently in the four clusters of cycles. There are significant differences in the mucus observed at the various days of the fertile window. In cluster 4 (containing almost 54% of cycles), mucus score evolves quite linearly from the fifth day before the peak to the peak day; in this cluster, cycles show the highest variability between the mucus coded at the beginning and at the end of the fertile window; the BIP is 1, dry mucus. Cycles in cluster 3 (18% of all) show an increase in mucus score until day two before the peak, then the mucus has almost a stable code until the peak day; the BIP is dry mucus. In cluster 2 (15% of cycles), codes evolve again quite linearly over time, for most of the cycles and the BIP is not identified. In cluster 1 (13%), mucus shows a high code registered from the beginning of the fertile window, cycles have the lowest variability among codes; for most of the cycles the BIP is of unchanging type.

²Model parameters are estimated by means of maximum likelihood. The estimation procedure was carried out with different sets of starting values, in order to avoid local maxima. Responses to items were treated as measured on an ordinal scale (Goodman 1979). For the best fitting model, the value of the log-likelihood is equal to -14,105, BIC = 28,868, the parameters are equal to 85. Convergence is reached after 250 iterations of the EM algorithm, followed by 13 iterations of Newton Raphson. All models were estimated with Latent Gold 5.0 (Vermunt and Magidson 2013).

The mucus code registered in the day immediately following the peak is much lower than that observed in the previous days in all groups, this allows to clearly identify the peak, after it occurred; however, significant differences may be observed across clusters. In clusters 3 and 4, the peak may be identified looking at the evolution of the mucus symptom in the days preceding it. In clusters 1 and 2, this is more difficult since variation over time of mucus codes is almost absent in the days immediately preceding the peak.

Table 3. Multilevel LC model – estimation results

	Size	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Size	$P(X_{ij}=l)$	0.1308	0.1479	0.1812	0.5400
	$P(W_j=h)$	Conditional probabilities			
Class 1	0.1341	0.9752	0.0011	0.0231	0.0006
Class 2	0.1353	0.0002	0.9837	0.0001	0.0160
Class 3	0.2209	0.0001	0.0659	0.8057	0.1283
Class 4	0.5095	0.0001	0.0002	0.0003	0.9994
		Mean values			
Mucus at day -5 from peak		3.9190	3.7544	3.7886	2.5830
Mucus at day -4 from peak		4.3108	4.1613	4.3569	3.0662
Mucus at day -3 from peak		4.5076	4.5297	4.8244	3.5699
Mucus at day -2 from peak		4.7274	4.7466	4.9979	4.1763
Mucus at day -1 from peak		4.8617	4.8376	4.9353	4.7185
Mucus at peak day		4.9143	4.9072	4.8747	4.8877
Mucus at day + 1 from peak		2.8348	2.7403	2.6472	2.1881
		Conditional probabilities			
BIP	0	0.0410	0.6904	0.0123	0.0208
	1	0.0070	0.0004	0.9763	0.9792
	2	0.0001	0.3091	0.0114	0.0000
	3	0.9519	0.0001	0.0000	0.0000
		Mean values			
Pre-ovulatory phase length		18.41	15.67	17.11	10.05
Period		5.92	5.77	5.39	5.47
Previous pregnancies		1,03	1,65	1,37	1,15
Cycle with conception		0.06	0.06	0.04	0.07
Unprotected intercourse		0.25	0.35	0.21	0.51
		Effect			
Age		0,1324	0,0006	0,207	-0,1538*

* statistically significant at 5%

The upper part of Table 3 contains the conditional probabilities $P(X_{ij} = l|W_j = h)$ and shows that the four classes of women have quite different distributions of cycles among clusters. In our model, clusters of cycles affect mucus codes observations and classes of women affect clusters of cycles membership: between-women differences in mucus evolution over the fertile window are explained by between-women differences in the likelihood of belonging to the cycle-level clusters. All parameters in the sub-

model that links clusters to classes are significantly different from 0 and all mean values of mucus across clusters and classes are significantly different.

Class 1 is associated mostly with cluster 1, class 2 with cluster 2, class 3 with cluster 3 and class 4 with cluster 4. Looking at these results, we can definitely say that there exists heterogeneity in the evolution over time of the patterns of daily mucus observation among cycles and among women. We tested the assumption of a one-to-one correspondence between classes and clusters by means of a conditional test. We estimated a multilevel LC model with 4 clusters and 4 classes, direct effects and the restriction that all cycles in each cluster belong to only one class. The conditional test rejected this restriction, showing that cycles registered by a woman assigned to a class, may belong to different clusters. The fact that there is not an exact one-to-one correspondence between a class of women and a cluster of cycles is another evidence of heterogeneity. Especially in class 3, a non-negligible percentage of the cycles belong to a different cluster; i.e, taking into account information on mucus and on characteristics of the woman and of the cycle, still variability remains. This result shows that there is a woman effect on the mucus evolution in the clusters of cycles.

The length of the pre-ovulatory phase is significantly different in the four clusters, as well as the length of menses, the number of previous pregnancies and the number of unprotected intercours. There is also a significantly different percentage of cycles that gives rise to pregnancy in each cluster, this percentage is highest in cluster 4. This result may suggest that women with a different mucus pattern over time have different probabilities of conception during a menstrual cycle and/or that the event of pregnancy modifies mucus observation, even though the average number of unprotected intercours is significantly different in the four clusters. This evidence definitely deserves further investigation.

Age at the beginning of the cycle is inserted in the model as an active covariate for clusters. There is a significant effect on cluster 4, showing that the probability of belonging to this group of cycles decreases significantly with the age of the woman. This evidence suggests that mucus variability over the fertile window reduces with age; this result confirms what already stated in previous reference literature (see, for example, Martyn et al. 2014)

Figure 2 reports mucus evolution over time in the cycles of the women belonging to the four classes.

For what concerns the four classes, women have a significantly different average age, oldest women are in class 4 (30.85 years) and youngest in class 3 (29.73). The average number of previous pregnancies is highest in class 2 (1.61), lowest in class 1 (1.03). The fact that in class 4 we find women who are, on average, older than in the other classes and had a low number of previous pregnancies (1,17) is only apparently contradictory; this class is composed of older women who probably entered the study in order to use the Billings method to achieve a pregnancy. In this class, in fact, we find women with the highest proportion of cycles with conception and a high average number of unprotected intercours.

The great majority of cycles of women in class 4 belongs to cluster 4. An apparent contradiction arises between the fact that in class 4 we find the group of oldest women and that the probability of belonging to cluster 4 diminishes with age (there is a negative significant effect of age equal to -0,1538, as reported in Table 3). First of all, it is important to notice that also a non-negligible percentage of cycles of women classified in class 3 – the youngest group, belongs to cluster 4.

Figure 1. Average registered mucus code in the days of the fertile window by cluster of cycles.

about here

Figure 2. Average registered mucus code in the days of the fertile window by class of women.

about here

3.1. *How woman's age affects heterogeneity*

At this point, it is interesting to verify if cycles registered by the same woman change cluster over time and if this change might be affected by the age of the woman. This hypothesis is suggested by the evidence obtained estimating the multilevel LC model that classes of women differ significantly in the average age and that woman's age has a significant effect on mucus variability over the fertile window.

A first order stationary latent class Markov model (equation 3) was specified in order to estimate transition probabilities over clusters of cycles. Cluster belonging at each time point was assessed by the posterior probabilities obtained estimating the multilevel LC model with direct effects and the potential effect of age was estimated using the three-step approach (Vermunt 2010).

$$P(\underline{X}_{ij} = \underline{l}) = \sum_{h_1=1}^H \dots \sum_{h_T=1}^H P(W_{i1} = h_1) \prod_{t=2}^T P(W_{it} = h_t | W_{it-1} = h_{t-1}, Z_{ij} = z) \prod_{t=1}^T P(X_{it} = l_t | W_{it} = h_t) \quad (3),$$

Where \underline{l} is the vector containing l_1, \dots, l_T .

In this case W_{it} is the dynamic latent variable representing cluster belonging and Z_{ij} the age of woman j at the beginning of registration of cycle i .

The three-step approach includes:

- (1) A LC model is estimated for a set of observed variables;
- (2) Subjects are assigned to latent classes based on their posterior class membership probabilities;
- (3) A standard multinomial regression model is estimated using the preceding step assignment as the dependent variable (equation 3).

The result in our application is that there is a very high stability across clusters (99%) over time for cycles. Moreover, woman's age does not have a significant effect on transition probabilities. This result may be explained by the characteristics of our data set (see Table 2): the average age of our sample is around 30, the minimum is 22, the maximum is 40; only for a small percentage of women we observe long sequences of cycles so to be able to detect an increase in age that could lead to a change of cluster towards one containing cycles with lower variability in mucus codes over cycle days.

To better understand the effects of age on mucus evolution pattern over time, a multilevel conditional latent growth mixture model (Vermunt 2007) was estimated. Basically, growth models are regression models for two-level data – time points nested within individuals – in which time enters as a predictor.

Response variables are the six indicators Y_{ijt} of the type of mucus observed and recoded by each woman i for her cycle j , t goes from day five before the peak to the peak day, our defined potentially fertile window. The model assumes that the repeated observations are imperfect measures of an underlying latent trajectory that is linear in our application.

The Multilevel growth mixture model (MGMM) assumes that the population is heterogeneous, and different subpopulations are characterized by different trajectories (Connel and Frey 2006). It can be used to classify individuals based on their growth

trajectories and to classify groups in which individuals are nested based on within-group trajectories (Palardy and Vermunt 2010). The addition of a between-group level of analysis allows to study the association between group characteristics and the growth process when units are nested in groups and to examine if heterogeneity in the between-group growth trajectories is due in part to the presence of unobserved subpopulations at the group level; its general specification is given by the following equations.

A repeated measurement model unit i nested in group j :

$$Y_{ijt} = \alpha_{ij} + a_t \beta_{ij} + \varepsilon_{ijt} \quad (4),$$

where Y_{ijt} indicates the observed variable; α_{ij} and β_{ij} are, respectively, the intercept and the slope; a_t is a variable indicating the passage of time, in this case, $a_t = t - 1$; ε_{ijt} is the random error.

The MGMM is composed by a within-group and a between-group model. The parameters of interest for the within-group component are the intercepts and slopes:

$$\alpha_{ij} = \sum_{l=1}^L \mu_{\alpha l} c_{lij} + \sum_{q=1}^Q \mu_{\alpha q} x_{qij} + \zeta_{\alpha ij},$$

$$\beta_{ij} = \sum_{l=1}^L \mu_{\beta l} c_{lij} + \sum_{q=1}^Q \mu_{\beta q} x_{qij} + \zeta_{\beta ij},$$

where $\mu_{\alpha l}$ and $\mu_{\beta l}$ are, respectively, the mean individual intercept and the slope for latent class l ; c_{lij} is an indicator variable, taking value 1 if individual i is a member of latent class l ; x_{qij} is the value assumed by unit i , belonging to group j , on a covariate q , on which the intercepts and the slopes are regressed to explain variation across individuals; $\mu_{\alpha q}$ and $\mu_{\beta q}$ are coefficients, respectively, for the intercepts and the slopes, $\zeta_{\alpha ij}$ and $\zeta_{\beta ij}$ are the random errors. The model assumes $\zeta_{\alpha ij} \sim N(0, \Psi_\alpha)$, $\zeta_{\beta ij} \sim N(0, \Psi_\beta)$ and $\varepsilon_{it} \sim N(0, \Theta_t)$. $\zeta_{\alpha i}$, $\zeta_{\beta i}$ and ε_{ijt} are mutually independent for every i, j and t .

The between-group model for the intercepts and slopes is given by:

$$\mu_{\alpha j} = \sum_{h=1}^H \gamma_{\alpha h} d_{hj} + \sum_{p=1}^P \gamma_{\alpha p} x_{pj} + \zeta_{\alpha j},$$

$$\mu_{\beta j} = \sum_{h=1}^H \gamma_{\beta h} d_{hj} + \sum_{p=1}^P \gamma_{\beta p} x_{pj} + \zeta_{\beta j},$$

where $\gamma_{\alpha h}$ and $\gamma_{\beta h}$ are, respectively, the mean intercepts and slopes for between-group latent class; d_{hj} is an indicator variable, taking value 1 if group j is a member of latent

class h ; x_{pj} is the value assumed by group j , on a covariate p , on which the intercepts and the slopes are regressed to explain variation across groups; $\gamma_{\alpha p}$ and $\gamma_{\beta p}$ are coefficients, respectively, for the intercepts and the slopes, $\zeta_{\alpha j}$ and $\zeta_{\beta j}$ are the random errors, with assumptions as for the within-group model.

For our application, the specified model is a conditional – with covariates – multilevel growth mixture (MGM) model with between group mixtures and with between group random effects constrained to 0. The variances of the intercept and the slope are set equal across all classes in order to specify a more parsimonious model, not to increase the computational burden and to avoid non-convergence (Palardy and Vermunt 2010). We aim at understanding how the characteristics of the woman, especially age, and of the cycle affect cervical mucus heterogeneity. The model was estimated starting with one latent class and then the number of latent classes was increased till the Bayesian Criterion Index (BIC) began to raise. The best fitting model is that with four latent classes of women³. Table 4 presents estimation results: class sizes, estimated intercept and slope in each class. The quadratic function for the latent trajectory has a worse fit on the data than the linear one.

Table 4. MGM model estimation, between-group component: sizes, intercepts and slopes

	Class 1	Class 2	Class 3	Class 4
Size	0.3544	0.3352	0.1927	0.1177
Intercept	1.8665*	0.9236*	2.7590*	-0.3533*
Slope	0.6721*	0.8518*	0.4791*	1.0631*

* significant at 0.05

Assuming a linear trajectory over time, the intercept and the slope are significantly different across the classes of women, indicating again heterogeneity between women in the evolution of mucus pattern over time. Intercepts range from -0.3533 to 2.8493, slopes from 0.4674 to 1.0805 (Figure 3), clearly the four groups of women observe significantly different patterns of mucus evolution over time. Woman's age affects significantly both the intercept and the slope. This effect is constant across classes and positive on the intercept (0.0394) but negative for the slope (-0.0089), within each class,

³ Model parameters are estimated by means of maximum likelihood. For the best fitting model, the value of the log-likelihood is equal to -14,605, BIC = 29,269, the parameters are equal to 29. Convergence is reached after 30 iterations of Newton Raphson. All models were estimated with Latent Gold 5.0 (Vermunt and Magidson 2013).

older women observed less variation in their mucus in the fertile window. This result is consistent with that found with the estimation of the multilevel latent class model, i.e., that in all groups of women, as age increases, mucus variability significantly diminishes.

The fact that the cycle ends with a pregnancy, the length of the menses and of the pre-ovulatory phase are significant covariates of the intercept and the slope of the latent trajectory of the within-group component of the model (Table 5). The number of unprotected intercours in the fertile window is significantly different across classes. We did not use it as a predictor of the trajectories at cycle level since it is highly correlated with the probability that the cycle ends with a pregnancy.

Table 5. MGM model estimation, within-group component: covariates effects on intercept and slope

	Intercept	Slope
Pregnancy	-0.3314*	0.0657*
Menses length	-0.0877*	0.0179*
Pre-ovulatory phase length	0.0684*	-0.0150*

* significant at 0.05

Figure 3. Estimated trajectory of mucus evolution over the fertile window in the four classes of women
about here

Figure 4 shows clearly the effect of age on the estimated mucus trajectory for women of 20, 30 and 40 years, belonging to class 1.

Figure 4. Estimated trajectory of mucus evolution over the fertile window foe women in class 1 but with different ages

about here

5. Concluding remarks

In this paper, we study the problem of heterogeneity in cervical mucus hydration at different times relatively to the mucus peak. We dispose of a two-level dataset: observed cervical mucus characteristics in menstrual cycles of women and we address the measurement of trajectories heterogeneity following two ways. We explicitly consider the multilevel nature of the data in our models' specification. We estimate a

multilevel latent class model which classifies women on the basis of the evolution of cervical mucus characteristics observed over the fertile period of each menstrual cycle taking into account that we observe a different number of cycles per woman and that there exists also heterogeneity between cycles. In the model, correlation over time among the observations is considered and the potential effect of active and inactive covariates is estimated.

Our research might be improved in various directions by means of further future analysis. We estimated a model where the effects of correlation among mucus observations have been imposed to be equal over time. However, this might be a too strong assumption and we should first consider to test it, possibly to relax it. This operation could result in a larger number of parameters to be estimated with problems of model identification and non-convergence. Thus, this extension should be evaluated carefully.

Our results show that heterogeneity in mucus evolution between cycles and women is non-negligible. Another evidence is that, after taking into account between-cycles variability, still some between-women heterogeneity remains, since a conditional test rejected the hypothesis of a one-to-one correspondence between clusters of cycles and classes of women. We also analysed the effect of some covariates on the mucus symptoms evolution. Unfortunately, by increasing the number of covariates in our model, the number of parameters increases exponentially, making it difficult to obtain reliable estimates. To overcome this limit, either we should obtain a much larger dataset or we should rethink the model with a more parsimonious specification.

The significant effect of the age of the woman on clusters and the fact that classes of women significantly differ in average age suggest that, for a same woman, the mucus evolution pattern over the fertile window, changes with increasing age. This evidence is confirmed by the results of the estimation of a multilevel conditional latent growth mixture model with a linear trajectory.

Another interesting result is that there is a significantly different percentage of pregnancies achieved during the study in groups of cycles and women. The number of pregnancies and of unprotected intercourse has also a statistically significant effect both on the intercept and the slope of the latent trajectory of the within component of the MGM model. This result has to be studied by means of appropriate models as, for example, that proposed by (Schwartz et al. 1980).

References

- Agresti, A., Booth, J.G., Hobert, J.P., and B. Caffo. 2000. Random-effects modeling of categorical response data. *Sociological Methodology* 30 (1): 27-80.
- Bakk, Z., Obersky, D.L., and J.K. Vermunt. 2014. Relating latent class assignments to external variables: standard errors for corrected inference. *Political Analysis* 22 (4): 520-540.
- Barrett J.C., and J. Marshall. 1969. The risk of Conception on Different Days of the Menstrual Cycle, *Population Studies*, 23, 3: 455-461
- Bassi, F., and B. Scarpa. 2015. Latent class modeling of markers of day-specific fertility. *Metron* 73 (2): 263-276.
- Bassi, F., Croon, M., Hagenaars, J.P., and J.K. Vermunt. 2000. Estimating true changes when categorical panel data are affected by uncorrelated and correlated classification errors; an application to unemployment data. *Sociological Methods and Research* 29 (2): 230-268.
- Billings, E.L., Billings, J.J., Brown, J.B., and H.G. Burger. 1972. Symptoms and hormonal changes accompanying ovulation. *Lancet* 1 (7745): 282–284.
- Billings, E.L., and A. Westmore. 1998. *The Billings Method*. Melbourne, Australia: Anne O'Donovan.
- Clogg, C.C., and L.A. Goodman. 1984. Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association* 79 (388): 762-771.
- Colombo : 2007. Characteristics and Limits of the Available Database, Working Paper n.9.2007, Department of Statistical Sciences, University of Padova.
- Colombo, B., and G. Masarotto. 2000. Daily Fecundability: First Results from a New Data Base. *Demographic Research* 3, 5
- Colombo, B., Mion, A., Passarin, K., and B. Scarpa. 2006. Cervical mucus symptom and daily fecundability: first results from a new database. *Statistical Methods in Medical Research* 15 (2): 161-180.
- Connell, A.M., and A.A. Frye. 2006. Growth mixture modelling in developmental psychology: overview and demonstration of heterogeneity in developmental trajectories of adolescent antisocial behaviour. *Infant and Child Development* 15 (6): 609-621.

507 De Bin, R., and B. Scarpa. 2014. Non-parametric Bayesian modeling of cervical mucus
508 symptom, Technical Report 170, Department of Statistics, University of Munich.

509 Dunson, D.B., and B. Colombo. 2003. Bayesian modeling of markers of day-specific
510 fertility. *Journal of the American Statistical Association* 98 (461): 28-37.

511 Dunson, D.B., Sinai, L., and B. Colombo. 2001. The relationship between cervical
512 secretions and the daily probability of pregnancy: effectiveness of the two-day
513 algorithm. *Human Reproduction* 16 (11): 2278-2282.

514 Fehring, R.J. 2002. Accuracy of the peak day of cervical mucus as a biological marker
515 of fertility. *Contraception* 66 (4): 231–235.

516 Goodman, L.A. 1979. Simple models for the analysis of association in cross-
517 classification having ordered categories. *International Journal of Aviation*
518 *Psychology* 74 (367): 537-552.

519 Gomes, A., and J.G. Dias. 2015. Improving the selection of air force pilot candidates
520 using latent trajectories. *Journal of Applied Statistics* 25 (2): 198-121.

521 Hilgers, T.W., and A.M. Prebil. 1979. The ovulation method-vulvar observations as an
522 index of fertility. *Obstetrics and Gynaecology* 53 (1): 12–22.

523 Katz, D.F. 1991. Human cervical-mucus-research update. *American Journal of*
524 *Obstetrics and Gynaecology* 165 (6): 1984–1986.

525 Katz, D.F., Slade, D.A., and S.T. Nakajima. 1997. Analysis of pre-ovulatory changes in
526 cervical mucus hydration and sperm penetrability. *Advances in Contraception* 13
527 (2-3): 143–151.

528 Martyn, F., McAuliffe, M.F., and M. Wingfield. 2004, The role of cervix in fertility: is
529 it time for a reappraisal?. *Human Reproduction* 29 (10): 2092-2098.

530 Moghissi, K.S., Syner, F.N., and T.N. Evans. 1972. A composite picture of the
531 menstrual cycle. *American Journal of Obstetrics and Gynaecology* 114 (4): 405–18.

532 Palardy, G.J., and J.K. Vermunt. 2010. Multilevel growth mixture models for
533 classifying groups. *Journal of Educational and Behavioural Statistics* 35 (5): 532-
534 656.

535 Pennoni, F., M. Barbato, and S. Del Zoppo. 2017. Latent Markov model with covariates
536 to study unobserved heterogeneity among fertility patterns of couples. *Frontiers in*
537 *Public Health* 5: 1-9.

- Scarpa, B., Dunson, D.B., and B. Colombo. 2006. Cervical mucus secretions on the day of intercourse: An accurate marker of highly fertile days. *European Journal of Obstetrics & Gynaecology and Reproductive Behaviour* 125 (1): 72–78.
- Scarpa, B., and D.B. Dunson. 2007. Bayesian selection of optimal rules for timing intercourse to achieve pregnancy. *Statistics in Medicine* 26 (9): 1920-1936.
- Scarpa B., Dunson, D.B., and E. Giacchi. 2007. Bayesian selection of optimal rules for timing intercourse to conceive by using calendar and mucus. *Fertility and Sterility* 88 (4): 915-924
- Schwartz., G. 1978. Estimating the dimension of a model. *The Annals of Statistics* 6 (2): 461-464.
- Schwartz, D., P:D.M. MacDonanld, and V. Henschel. (1980) Fecundability, coital frequency and the viability of ova. *Population Studies* 34(2): 397-400.
- Vermunt, J.K. 2013. Multilevel latent class models. *Sociological Methodology* 33 (1): 213-239.
- Vermunt, J.K., and J. Magidson. 2013. *LG-Syntax User's Guide: Manual for Latent GODL 5.0 Syntax Module*. Belmont, MA: Statistical Innovations Inc.
- Vermunt, J.K. 2007. Growth models for categorical response variables: standard, latent-class, and hybrid approaches. In *Longitudinal Models in the Behavioural and Related Sciences*, K. van Montfort, H. Oud, and A. Satorra, 139-158. Mahwah, NJ: Erlbaum
- Vermunt, J.K. 2010. Mixture models for multilevel data sets. In *Handbook of Advanced Multilevel Analysis*, ed. J. Hox, and J.K. Roberts, 58-91. New York: Routledge.
- Vermunt, J.K. 2010. Latent class modelling with covariates: two improved three-step approaches. *Political Analysis* 18 (4): 450-469.
- Wilcox, A.J., Dunson, D.B., and Baird, D.D. 2000. The timing of the fertile window in the menstrual cycle: day-specific estimates from a prospective study. *British Medical Journal* 321 (7271): 517-521.
- World Health Organization. 2000. A prospective multicentre trial of the ovulation method of natural family planning. I. The teaching phase. *Fertility and Sterility* 36 (1): 152–158.

570

571

572

573