



Department of Statistical Sciences
University of Padova
Italy

UNIVERSITÀ
DEGLI STUDI
DI PADOVA
DIPARTIMENTO
DI SCIENZE
STATISTICHE

Characteristics and Limits of the Available Data Bases

Bernardo Colombo
Department of Statistical Sciences
University of Padova
Italy

June 4, 2007

Characteristics and Limits of the Available Data Bases

Bernardo Colombo

Department of Statistical Sciences
University of Padova
Italy

The aim of this paper is to describe some characteristics and limits of the data bases presented in Table 1, concentrating on the first three. For LONDON, a full description is made in Miolo et al (1993) and for VICENZA in two “Tesi di laurea” (in Italian), by De Nadai A (Acad. Y. 1990-91) and Schiavon A (Acad. Y. 1990-91), which can be consulted in the Library of the Faculty of Statistical Sciences of the University of Padova. See also Jennings V and Sinai I (2001).

Days of coitus –scaled with reference to a marker common to all cycles- responsible for a pregnancy were first presented (see Table 2) by Prof. Marshall in an intervention at the Belgrade Conference of 1964 (Marshall J, 1967)

Following a suggestion by Peter Armitage, J Marshall and C Barrett published the first estimates of daily fecundability using the quantal regression technique (Barrett J and Marshall J, 1969). The base BARRETT– M. in Table 1 contains a little larger sample, from the same origin, which was used with a slightly modified model in a paper of French colleagues (Schwartz D et al., 1980).

In Table 3 is illustrated a sample of data from BARRETT- M. In it, each line refers to the experience of one cycle. In all instances, the last four figures identificate a woman: same figures mean that the cycles belong to the same woman. The 1s in the centre of the page are registrations of days with intercourse, scaled to the shift in the basal body temperature (BBT: see also, f.i., the chart in Note 2 of Colombo and Masarotto, 2000). ID number 6101, f.i., refers to a woman (couple) strictly relying on the stronger suggestion of behaviour for avoiding a pregnancy. Woman (couple) 4950 takes advantage for intercourse in the first cycles only of the infertile phases of each cycle; then she (they) changes her (change their) mind and in the entire (intermenstrual period of a) cycle are registered intercourse episodes. Then come two subjects who behave as the first one. BILLINGS and FERTILI similarly identify days with intercourse and possible intervening pregnancies, the first base using as a marker the cervical mucus symptom (CMS) and FERTILI both BBT and CMS. LONDON and VICENZA lack information on intercourse episodes and pregnancies.

Observations on the good quality of the information provided by the first three bases can be found in the original papers. Here attention is mainly concentrated on some limits which should be considered before making use of them for research purposes.

Figure 1 shows the pattern and level of daily fecundability for each of various subgroups of the whole base FERTILI (Colombo B and Masarotto G, 2000). Two points might be underlined: 1) the daily estimates obtained by application of the Schwartz et al. model fall all within the same window around the reference marker () and show the same shape; 2) there are in general moderate differences in levels between the various groups with one exception, that is for the experience of Auckland, which shows definitely higher levels over the whole period. The origin of such discrepancy is mentioned in Colombo B and Masarotto G (2000). A reliable estimate of the level of daily fecundability can be obtained aggregating the experience of only European centres (see Table 10 of Colombo- Masarotto, 2000).

Ongoing control of the data received in Padova from centres showed at about half-way that several cycles from Düsseldorf reported no days with intercourse after the end of the fertile phase. The

resulting difference in number of days with intercourse between Düsseldorf and the other European centres are shown in Table 4. It appears to be important particularly at younger ages in non-conception cycles. It follows that a reliable evaluation of behaviour in different situations should exclude both centres of Düsseldorf and Auckland, in the last case due to guidelines given in the protocol in view of the specific target of that research (see Colombo- Masarotto, 2000, Section 2.1).

While in FERTILI (as in LONDON) the determination in the shift of BBT has been done uniformly in Padova through a visual evaluation - in a team work- of incoming charts, the identification of the peak mucus day was done autonomously in each of the eight involved centres under the responsibility of the local Principal Investigator. There was a strict guide-line for coding both sensation and appearance of cervical mucus through a numerical classification (Colombo – Masarotto, 2000, Table 1), but subjective judgements had an impact on its application. Evidence of that is found in Table 5. Here Code 3 identifies the instance of *damp feeling and mucus thick, creamy, whitish, yellowish, not stretchy/elastic, sticky*, while Code 4 makes reference to *wet, slippery, smooth feeling, and mucus transparent, like raw egg white, stretchy/elastic, liquid, watery, reddish/with some blood*. One observes a pretty high heterogeneity between centres with respect to the average number of days in a cycle for which it was felt appropriate to characterize the observed features with a specific code. One sees, for instance, in Table 5 that there is an inversion between Milan and Verona in the frequency of use of codes 3 and 4. It is very likely that such differences do not depend on physiological bases. Rather, one might find in them a hint that for certain specific research it is advisable to rely on the information collected centre by centre rather than on aggregates of centres.

BILLINGS offers on fecundability the same information provided by FERTILI, but making use, as a marker, only of the cervical mucus symptom (Colombo et al., 2006).

In Fig. 2, we see that the width of the fertile window and the pattern of the estimates repeat what was found in FERTILI. Also here, there appears to be a marked difference in the general level between one centre (Parma) and the other three (see Fig. 2). An explanation of the origin of this discrepancy can be traced in Section 3.2 and in the Discussion of Colombo et al. (2006).

A big advantage of BILLINGS vs. FERTILI lies in the more complete and homogeneous information about typologies of the cervical mucus symptom. Better homogeneity of classification is here assured by the fact that all four collaborating centres provided instructions on the use of the same method of natural family planning, “applying uniform procedures in teaching, practical applications, linguistic descriptions and conventions” (Colombo et al., 2006, Introduction). In spite of that, also in BILLINGS one finds a certain amount of inter-observer variation in interpreting and coding of mucus typologies (see Bassi F, Mion A and Colombo B, 2003, Table 3).

The difference in the percentage of cycles with identification of mucus reference day between Rome and the other three sites (see Table 6) is amenable to the choice made in that centre- say until about half of the period of data collection- to prefer instances of cycles with a clear and meaningful development of the CMS typology. Also in this instance –for some delicate issues– one might prefer an analysis centre by centre (see, for instance, Dunson DB and Colombo B, 2003) instead of aggregating all data together.

The standardized (with respect to the total number within the window) distribution of days with reported secretions of Fig. 3 puts in evidence another kind of differences between centres in BILLINGS: a variation which might have no relevance in evaluations of fecundability. Comparing the observations of Rome vs. Saluzzo, one might suspect that the discrepancies at both sides of the window could depend on local evaluations of experiences made and related choices in teaching.

Fig. 4 adds, for BILLINGS, one more check of its quality besides those mentioned in Colombo B et al. (2006). One finds in it that in the WHO exercise (Sinai I et al., 1999) most cycles provide

information on secretions for only few days. This concentration appears a little diluted in the old (1978-89) experience of VICENZA and totally disappears in BILLINGS, due to the care in recording the observations over the entire cycle.

Finally, all the bases contain mixtures of observations made on the subjects during their participation to the exercise. In each instance there might be more than one entry during the whole period. And each entry might be divided into distinct groups of consecutive cycles. An analytical information about its structure is provided for each base.

References

- Barrett JC and Marshall J (1969), *The Risk of Conception on Different Days of the Menstrual Cycle*, Population Studies, 23,3: 455-461.
- Bassi F, Mion A, Colombo B (2003), *Interobserver variation in interpreting cervical mucus as an indicator of the fertile phase in a menstrual cycle*, Genus, 53,3-4: 91-102.
- Colombo B, Masarotto G (2000), Daily fecundability: first results from a new data base, Demographic Research 2000, 3/5 (Available at www.demographic-research.org/volumes/vol3/5).
- Colombo B, Mion A, Passarin K, Scarpa B (2006), Cervical mucus symptom and daily fecundability: first results from a new data base, Statistical Methods in Medical Research, 15,2: 161-180.
- Jennings V and Sinai I (2001), Further Analysis of the Theoretical Effectiveness of the Two Day Method of Family Planning, Contraception, 64: 149-153.
- Marshall (1967), Analyse statistique du moment de la conception en relation avec l'élévation de la température sur 5013 cycles, Congrès Mondial de la Population, Belgrade 1965, Vol. II, Nations Unies, 1967, New York.
- Passarin K (1998), Vecchie e nuove stime di fecondabilità e loro applicazioni, Tesi di laurea, Faculty of Statistical Sciences, University of Padova, Acad. Year 1997-98.
- Schwartz D, MacDonald PDM, Heuchel V (1980), Fecundability, Coital Frequency and the Viability of Ova, Population Studies, 34,2: 397-400.
- Sinai I, Jennings V, Arevalo M (1959), The Two Day Algorithm: a New Algorithm to Identify the Fertile Time of the Menstrual Cycle, Contraception, 60: 65-70.

The "Tesi di laurea" can be consulted in the Library of the Faculty of Statistical Sciences of the University of Padova.

Acknowledgements

Table 6 and Figure 2 are reprinted by permission of Sage Publications Ltd from Colombo B et al., 'Cervical mucus symptom and daily fecundability: first results from a new database' in Statistical Methods in Medical Research 15(2), Copyright (©Sage Publications, 2006).

Table 1: *The set of data bases.*

		WOMEN	CYCLES	PREGNAN	MARKERS
*FERTILI	EUR.	782	6,724	487	BBT AND/OR MUCUS
	ALL	881	7,017	575	
*BILLINGS		193	2,754	177	MUCUS
*BARRETT-M.		241	2,192	103	BBT
LONDON		1,798	36,641	-	BBT
VICENZA		282	2,702	-	MUCUS

*With information on daily intercourse.

Table 2: *Day of coitus in association with temperature rise in cycles in which conception occurred.*

	Number of days before temperature rise where coitus occurred										
	10	9	8	7	6	5	4	3	2	1	TOT.
Couples who had no coitus after rise of temperature	1	-	-	-	1	3	1	-	3	3	12
Couples who had coitus after rise of temperature, but not during first 3 temperatures at the higher level	1	3	3	2	1	2	4	1	-	1	18
TOTAL	2	3	3	2	2	5	5	1	3	4	30

Source: Marshall J (1967).

Table 3: A sample of data in Barrett-Marshall

412				11	111				000105514106412	346101
413					1111	1			0002055111164130	346101
414						11			000305511126412	346101
415				11111	11				0004055050165120	346101
416					111111111				0005055020265130	346104
417					1111	11			0006055040365120	346101
418					11111	111			000705531036512	346101
419					11111	111			0008055250465127	346101
420						1111			0009055220565120	346101
421					111111111				0010055160665127	346101
422					11	111			0014055301065127	346101
423				1	1	111	1		0011055100865120	346101
424					11111	111			001205505096512	346101
425					1111	11			0013055031065227	346101
426					1111	1			0015055261165127	346101
427					11	11			0016055221265120	346101
428					11111	11			0017055190166127	346101
429				1	1	1		11	0001056120863120	404950
430				11	1	1		11	0002056070963127	404950
431				1	1			1	0003056041063127	404950
432				1111	1			11	000405631106312	404950
433					111	1		1	0005056281163120	404950
434				1	1	1		1	0006056241263127	404950
435				1	1	1		1	0007056200164320	404950
436						11	1	1	0008056170264127	404950
437				1	11	111	1	11	0009056170364129	404950
438					1	1	1	11	0010056150464127	404950
439				1	11	11	1	11	0011056130564131	404950
440					111	1	1	1	0012056130664127	404950
441				11	1		11	11	0013056120764130	404950
442							111	1	0014056040165120	404950
443					1	11	1	1	0015056010265120	404950
444					1	11	1	11	0016056270265127	404950
445					1	1	1	1	0017056250365120	404950
446					1	1	1	1	0018056200465120	404950
447					1	1	1	1	0019056160565127	404950
448							1111	111	0001057101065127	256423
449							11111	1111	0002057041165127	256423
450							1	1111	0001058171265130	256423
451							11	1	0002058170166130	256423
452							11	111	0003058160266130	256423
453				1			1111	1	0004058260366130	256423
454							1	1	0005058280466130	256423
455							111	1	0006058280566141	256423

Table 4: *Average number per cycle of days with intercourse at March 20, 1996 (Total number of cycles in brackets: FERTILI).*

NON CONCEPTION CYCLES WITH BBT RISE DETECTED					
	Total	<25	25-29	30-34	≥35 years
Düsseldorf	3.59	3.35	3.58	3.81	3.60
	(488)	(79)	(155)	(320)	(170)
Other European Centres	4.94	6.37	5.28	4.76	4.25
	(2657)	(213)	(1115)	(637)	(456)
CONCEPTION CYCLES WITH BBT RISE DETECTED (Acts of intercourse from day 1 through day 18 after BBT rise)					
	Total	<25	25-29	30-34	≥35 years
Düsseldorf	7.79	18.00	9.31	5.29	5.92
	(34)	(2)	(13)	(7)	(12)
Other European Centres	7.04	19.13	7.12	6.98	5.82
	(195)	(18)	(86)	(75)	(16)

Table 5: *Average number of days in a cycle until the peak mucus reference day in which, in FERTILI, were registered Codes 3 and 4.*

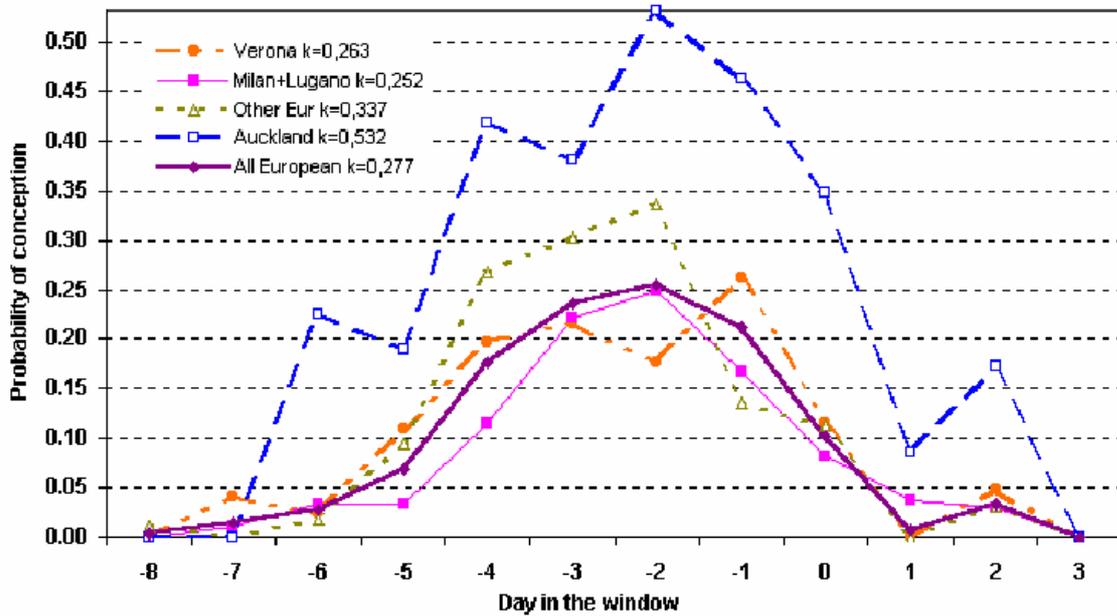
CENTRES	CODE 3		CODE 4	
	Aver.	s.d.	Aver.	s.d.
INER (VR)	3.05	2.63	2.88	1.89
CAMEN (MI)	1.81	1.99	4.76	2.69
LUGANO	2.11	1.64	2.95	1.27
CLER	2.23	2.24	5.04	3.62
DUESSEL	3.34	2.61	4.08	2.36
CMAC	5.00	3.07	3.76	2.58
CAF	2.56	2.52	3.88	1.83
TOTAL	2.45	2.44	4.23	2.68
AUCKLAND	2.18	2.65	4.19	2.93
GRAND TOTAL	2.43	2.45	4.23	2.69

Table 6: *Identification frequency of the peak mucus day in cycles of the BILLINGS centres.*

CENTRES	NO. OF CYCLES	No. of cycles with identification of mucus reference day (% of cycles)
Milan	909	739 (81.3)
Parma	1,060	859 (81.0)
Saluzzo	267	222 (83.1)
Rome	519	463 (89.2)
All	2,755	2,283 (82.9)

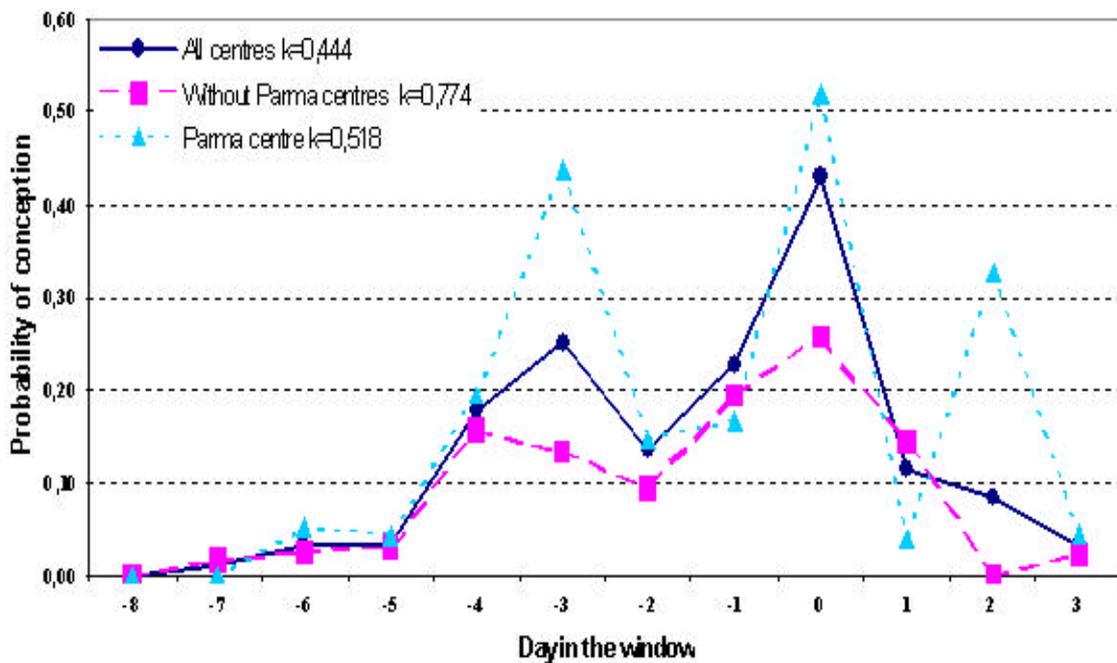
Source: Colombo B et al. (2006).

Fig. 1: Daily fecundability around the BBT reference day (= day 0).



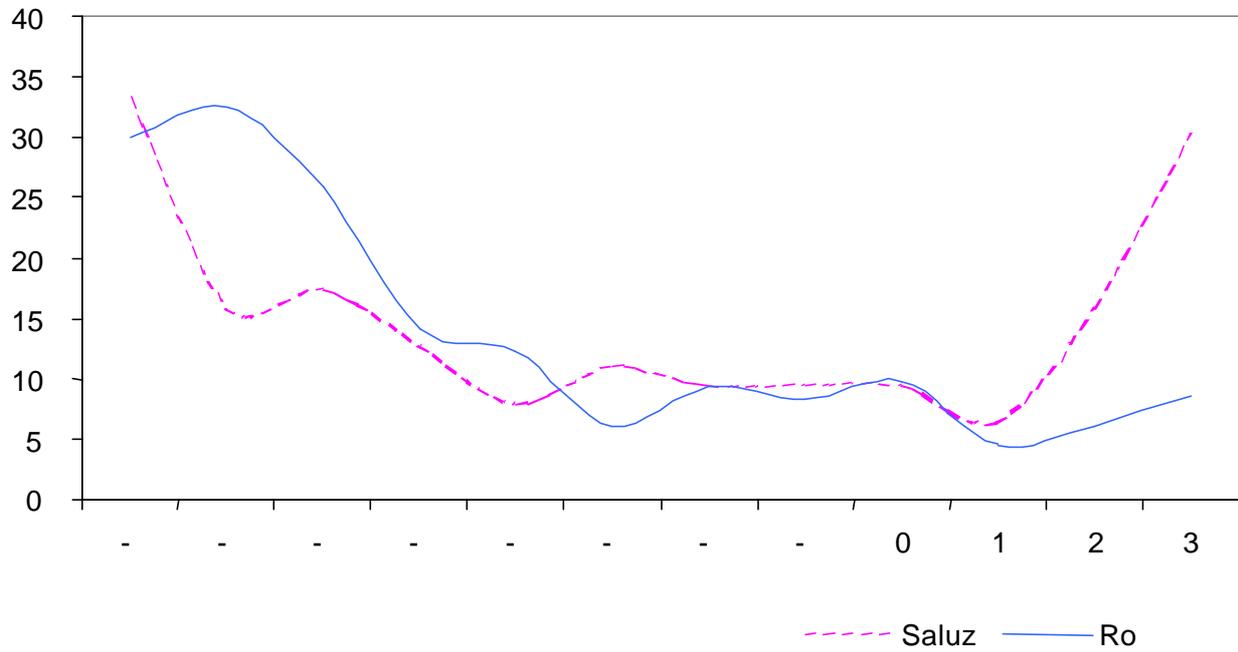
Source: Colombo B and Masarotto G (2000)

Fig. 2: Daily fecundability around the peak reference day (= day 0).



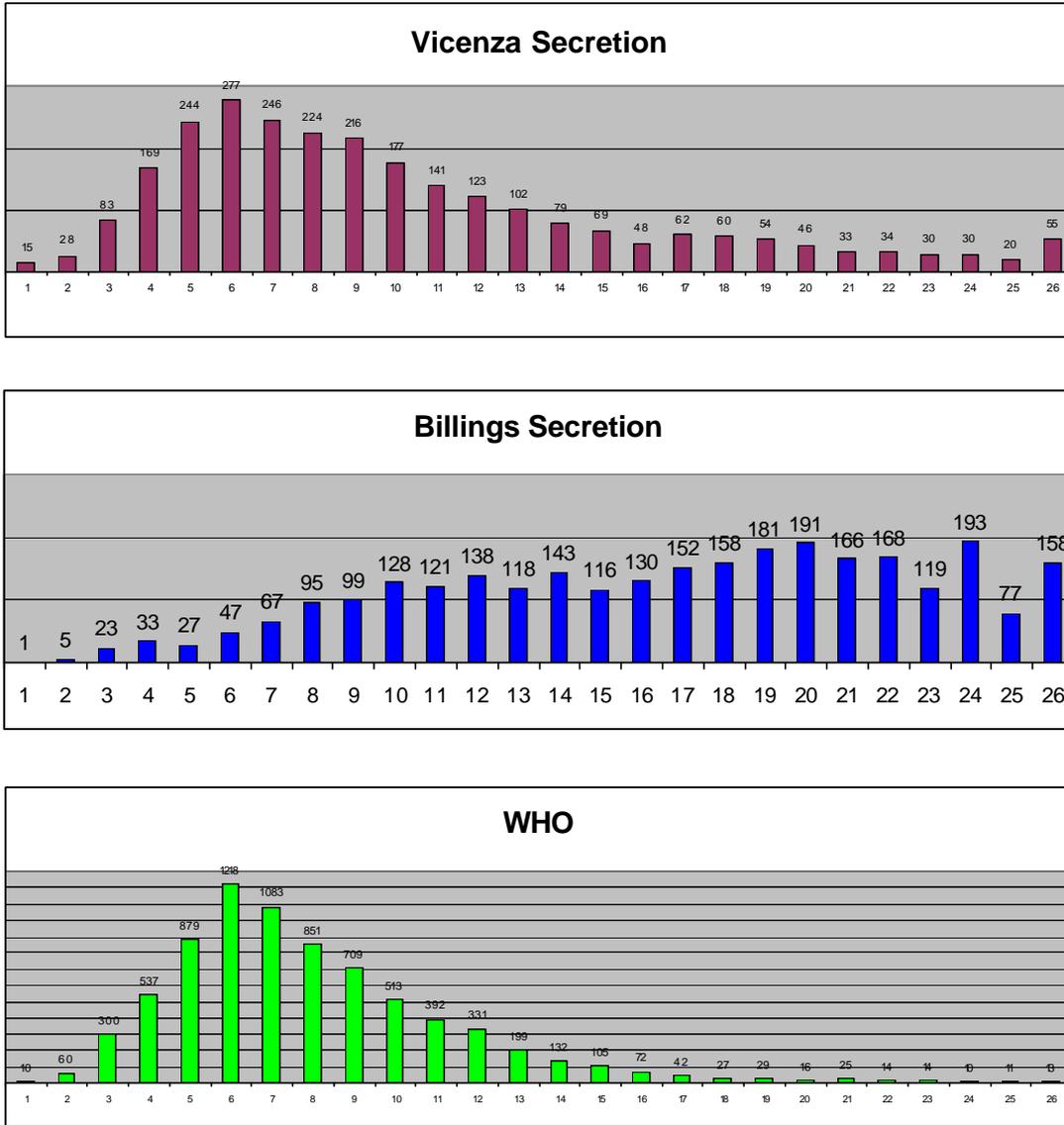
Source: Colombo B et al. (2006)

Fig. 3: *Standardised distribution of days with intercourse in the fertile window (-8,3) around the peak mucus day in the centres of Rome and Saluzzo in BILLINGS.*



Source: Passarin K (1998).

Fig. 4: Number of days in which a “secretion” is registered in three bases.



Source: Sinai I et al. (1999).

Working Paper Series
Department of Statistical Sciences, University of Padova

You may order copies of the working papers from by emailing to wp@stat.unipd.it
Most of the working papers can also be found at the following url: <http://wp.stat.unipd.it>