

Analysis of Contraceptive Failure Data in Intrauterine Device Studies

Modern Competing Risks Approach

S.R. Karia,* Juhani Toivonen,† and Elja Arjas*

The life table method used heretofore in case of intrauterine device (IUD) failure data requires grouping of data into intervals. If the termination times are recorded exactly along with the reason of termination, grouping of data results in some loss of information. Modern competing risks techniques are suggested here for the exact IUD failure data. The uses of cumulative incidence functions, which are essentially the quantity given by Potter's net rate, and cause-specific hazard rates are stressed. Also, this paper focuses on the flaws of life table estimates of net and gross rates, which have been widely used during the past three decades in the analysis of contraceptive failures. The methods suggested in this paper can be used in any other situation where the failure times and reasons of failure are recorded. CONTRACEPTION 1998;58:361-374 © 1998 Elsevier Science Inc. All rights reserved.

KEY WORDS: intrauterine device, life table technique, Potter's net and gross rates, cumulative incidence function, cause-specific termination rate

Introduction

The oldest method used as a summary measure of contraceptive failures, the Pearl index, is based directly on the total number of women-months of exposure.¹ This index can be viewed as the maximum likelihood estimator of the termination rate when constancy of the rate is assumed over time. The main criticism against this index was the apparent invalidity of the constancy assumption. The Pearl index was subsequently modified by assuming the constancy of the rate over monthly intervals, thus making the rate a step function. This method was later replaced by life table methods.²⁻⁵

Multiple decrement life table techniques were recommended by Potter in order to account for the

possibility of several termination reasons, leading to the estimation of so-called net rates.⁶ In contrast, Potter recommended single decrement life tables for the estimation of the termination rate due to a particular reason, such as pregnancy, in a situation in which the other competing causes were assumed "to have been removed" from the population. Such synthetic rates were termed gross rates.⁷ Later on, Tietze and Lewit contributed to this technique.^{2,5,8,9}

Potter, more than three decades ago, viewed the problem of intrauterine contraceptive devices (IUD) as a problem in competing risks.⁷ However, in spite of having numerous statistical techniques available in a competing risks set-up, no systematic attempt has been made in this direction in the context of IUD. A paper by Farley gives the impression of pointing out the faulty methods used and the confusion created because of differences in the nomenclature.¹⁰ But he repeats the same trend, by stressing the use of the daily life table method. We refer to Farley and Trussell and references therein for a detailed account.^{10,11}

The main goal of this work is to place the problem of IUD failure in the modern competing risks framework and suggest appropriate techniques for analysis of contraceptive failure data when the failure times are recorded exactly. Use of life table methods in such a situation will merely result in a loss of information. No attempt has been made to propose new methods as the principal aim is to clarify the confusions inherited throughout. These methods are applied on a particular data set describing IUD failures.

Subjects

We have partially reanalysed the data from a randomized multicenter trial on two IUD, a copper-releasing IUD (Nova T®, Leiras Oy, Turku, Finland) and a levonorgestrel-releasing IUD (LNG-IUD; Levonova®/Mirena®, Leiras Oy, Turku, Finland). The 5-year results of this study, using the life table analysis suggested by Tietze and Lewit, have been published.^{5,12} Our data set consists of 2337 women, 769 in the Nova T group and 1568 in the LNG-IUD group. We examine three coun-

*Rolf Nevanlinna Institute. †Department of Medical Chemistry, University of Helsinki, Helsinki, Finland

Name and address for correspondence: S.R. Karia, Rolf Nevanlinna Institute, P.O. Box 4, FIN 00014 University of Helsinki, Helsinki, Finland; Tel.: +358-9-191-22760; Fax: +358-9-191-22779

Submitted for publication September 9, 1997

Revised November 13, 1998

Accepted for publication November 13, 1998

tries, Finland, Sweden, and Hungary, and five termination reasons, (1) pregnancy, (2) expulsion, (3) amenorrhea, (4) bleeding and pain, and (5) hormonal reasons, for illustration purposes.

It is important to note that the whole analysis can be carried out including other recorded termination reasons without any further complications. Also comparisons of the two devices within the same region and the same device between different regions can be brought about by comparing cumulative termination rates or cumulative termination probabilities (see below).

Statistical Analysis

The data available on each subject enrolled in the study are the termination time and the termination reason: 1, 2, 3, 4, or 5 if uncensored, and the censoring time otherwise (generally the end of the study period). Below, we consider the estimation of the cumulative termination rates due to these five causes based on the data containing individual termination times and their reasons. Let $T_{(1)} < T_{(2)} < \dots < T_{(k)}$ be the ordered distinct uncensored times. Let n_i , d_i , and d_{ij} be the number of women at risk just prior to $T_{(i)}$, the number of terminations at $T_{(i)}$, and the number of terminations of type j at $T_{(i)}$, respectively. Note that $d_i = \sum_{j=1}^5 d_{ij}$. The following definitions are standard in the statistical literature on survival data analysis.

The Kaplan-Meier estimate¹³ of the overall probability of termination of the use of IUD (regardless of its cause) by time t is

$$\hat{F}(t) = 1 - \prod_{\{i|T_{(i)} \leq t\}} \frac{(n_i - d_i)}{n_i} \quad (1)$$

The estimate of probability of continuous use up to time t is then $\hat{S}(t) = 1 - \hat{F}(t)$.

The corresponding estimate of the probability of termination by time t , due to a specific termination reason but in the presence of all the competing causes, is

$$\hat{F}_j(t) = \sum_{\{i|T_{(i)} \leq t\}} (d_{ij}/n_i) \hat{S}(T_{(i)} -). \quad (2)$$

The Nelson-Aalen estimator¹⁴ of cumulative cause-specific termination rate is

$$\hat{H}_j(t) = \sum_{\{i|T_{(i)} \leq t\}} d_{ij}/n_i. \quad (3)$$

Note that the momentary hazard rate concept underlying this cumulative measure refers to individuals who are still using an IUD, whereas the corresponding density concept underlying equation 2 does not involve such a conditioning event. In a sense, equation 2 describes the collective study population, and equation 3 a woman with the IUD still in situ. It

is apparent that equations 2 and 3 are very close to each other as long as only a small proportion of the study population have either terminated the use of the IUD or have been censored, in which case the estimates $\hat{S}(T_{(i)} -)$ are close to 1. Later, however, when only a small proportion of the original population are at risk, equation 3 grows much more rapidly. Note also that $\sum_{j=1}^5 \hat{F}_j(t) = \hat{F}(t)$. We refer to Kalbfleisch and Prentice for more details.¹⁵

For preliminary treatment of the data, and to demonstrate the main characteristics of the different causes of termination, we suggest the following graphical procedures that are available in the literature:

- 1) A plot of $\hat{F}_j(t)$ versus t , for $j = 1, \dots, 5$. This helps in studying the behavior of each termination type with time. The comparison of the termination types can be carried out by $p - p$ plots. A $p - p$ plot is a plot of $\hat{F}_{j_1}(t)$ versus $\hat{F}_{j_2}(t)$ for various t and combinations of (j_1, j_2) .¹⁶
- 2) A cumulative hazard plot: a plot of $\hat{H}_j(t)$ versus t , for $j = 1, \dots, 5$. This gives an insight into the rate of occurrence of termination due to different reasons j for durations that have already lasted up to time t .

Results

Figures 1 to 12 enable us to study in detail the behavior of various termination reasons over the follow-up period. By merely considering the rates at some specific times one fails to see the behavior at intermediate stages. The graphical display of probabilities/rates is a natural way of studying different termination reasons jointly and carrying out comparisons between them.

Figures 1 to 3 give the plot of the estimated cumulative termination probabilities for LNG-IUD. The pregnancy rate was 0 in Sweden and in Hungary, but this was not so in Finland (though it was very low). During the initial stages after an IUD had been inserted, expulsion was the dominating termination cause in Finland. It is interesting to see that the termination rate due to amenorrhea and that due to hormonal reasons in Hungary were quite different from what they were in Finland or in Sweden. The termination rate due to amenorrhea was comparatively higher in Hungary. The changes in the pattern of bleeding, especially occurrences of oligomenorrhea during LNG-IUD use, needed special counselling in order to avoid the unnecessary removal of an IUD. The difference between the two Nordic countries on the one hand and Hungary on the other hand in the termination rate for amenorrhea indicates differences in counselling and in how often amenorrhea was

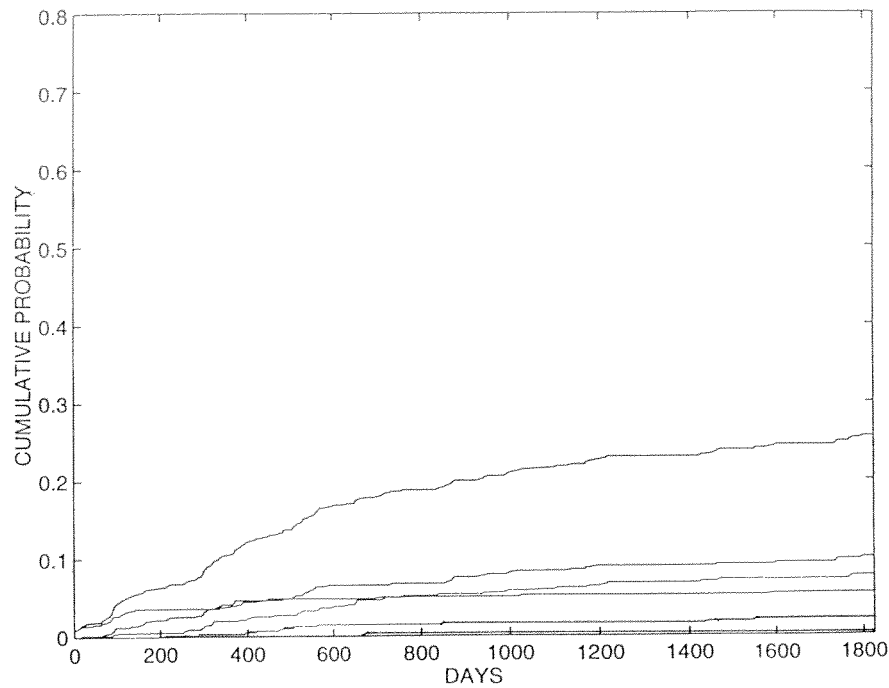


Figure 1. Estimated cumulative probabilities of termination for the LNG-IUD in Finland. The curves correspond to the overall contraceptive method-related termination probability, termination probabilities due to bleeding and pain, hormonal reasons, expulsion, amenorrhea, and pregnancy, respectively, from top to bottom.

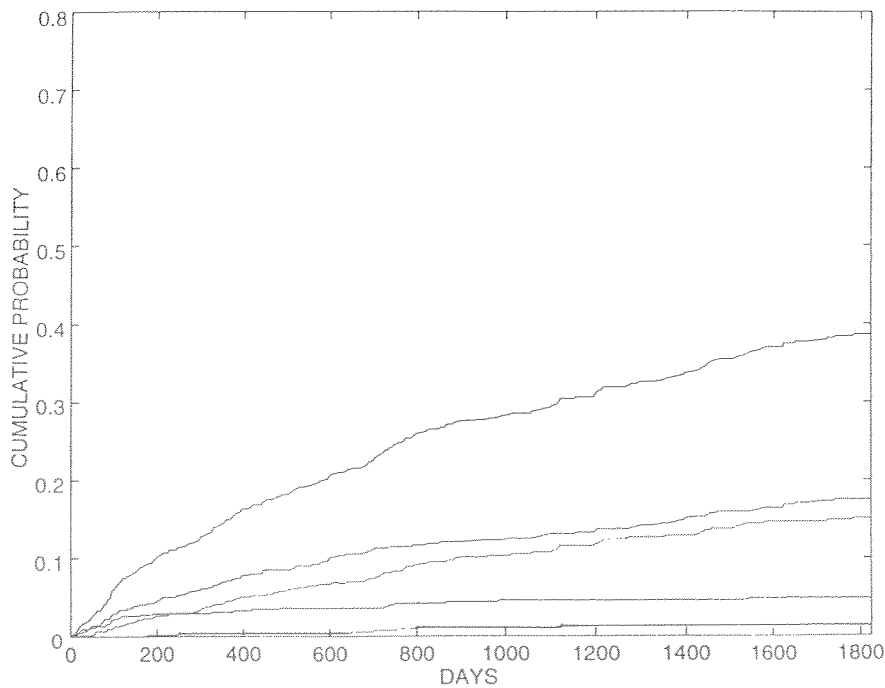


Figure 2. Estimated cumulative probabilities of termination for the LNG-IUD in Sweden. The curves correspond to the overall contraceptive method-related termination probability, termination probabilities due to bleeding and pain, hormonal reasons, expulsion, amenorrhea, and pregnancy (zero), respectively, from top to bottom.

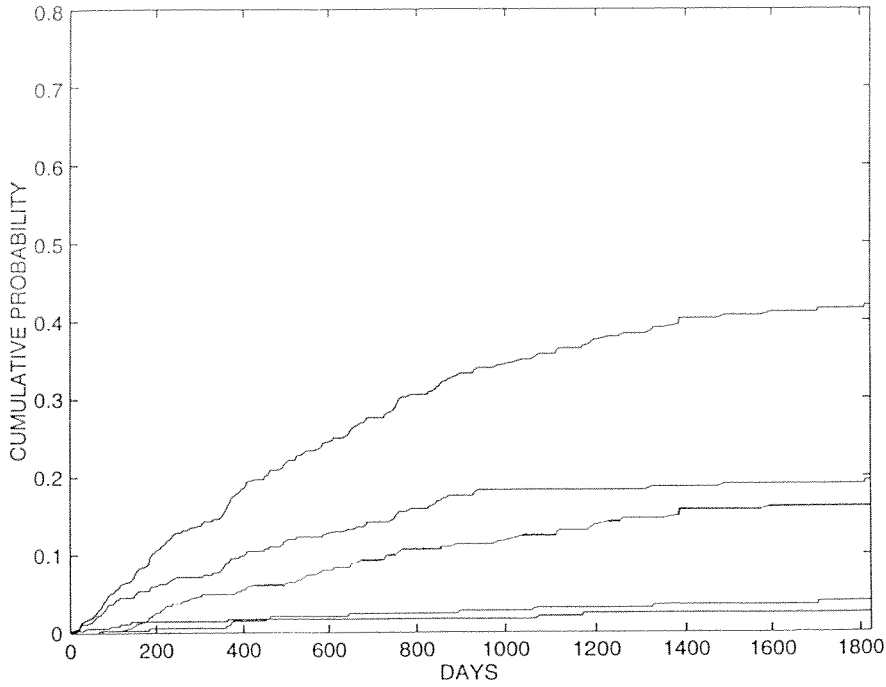


Figure 3. Estimated cumulative probabilities of termination for the LNG-IUD in Hungary. The curves correspond to the overall contraceptive method-related termination probability, termination probabilities due to bleeding and pain, amenorrhea, hormonal reasons, expulsion, and pregnancy (zero), respectively, from top to bottom.

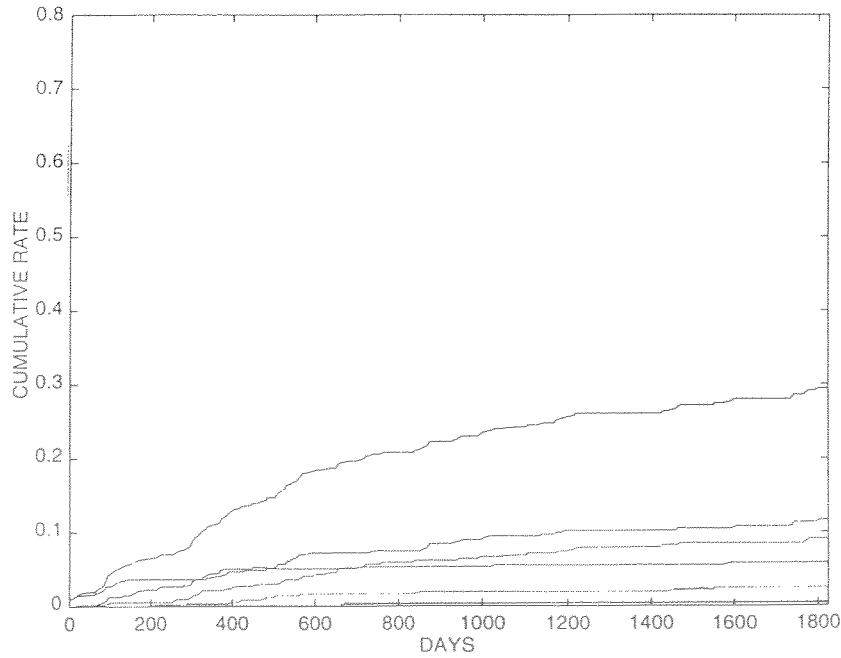


Figure 4. Estimated cumulative rates of termination for the LNG-IUD in Finland. The curves correspond to the overall contraceptive method-related termination rate, termination rates due to bleeding and pain, hormonal reasons, expulsion, amenorrhea, and pregnancy, respectively, from top to bottom.

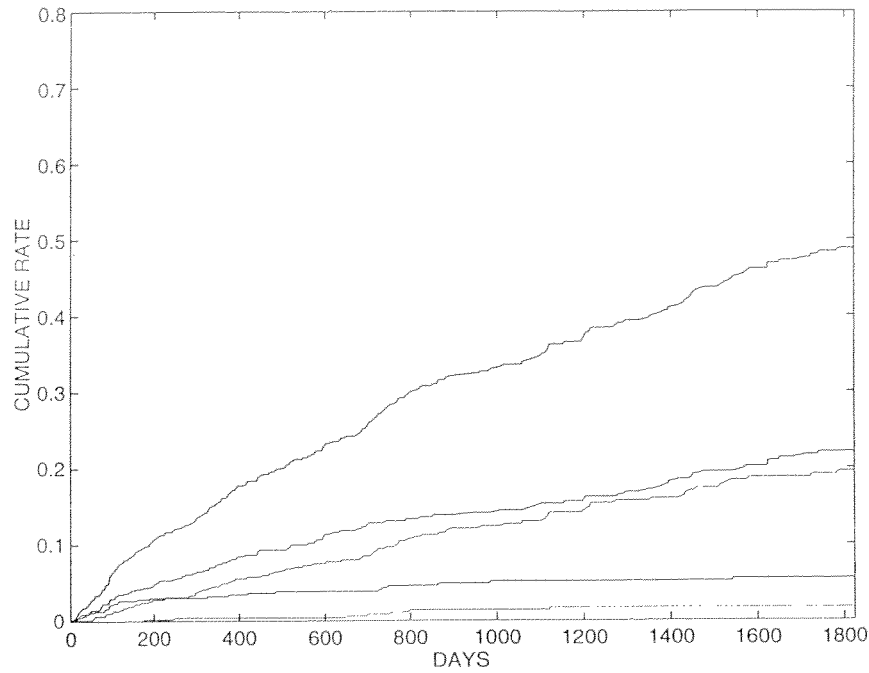


Figure 5. Estimated cumulative rates of termination for the LNG-IUD in Sweden. The curves correspond to the overall contraceptive method-related termination rate, termination rates due to bleeding and pain, hormonal reasons, expulsion, amenorrhea, and pregnancy (zero), respectively, from top to bottom.

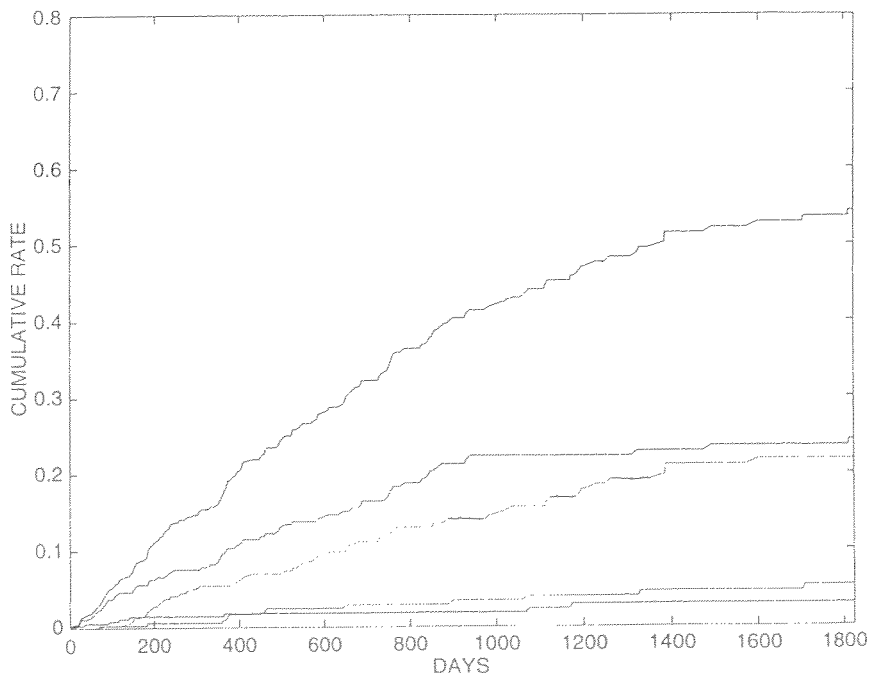


Figure 6. Estimated cumulative rates of termination for the LNG-IUD in Hungary. The curves correspond to the overall contraceptive method-related termination rate, termination rates due to bleeding and pain, amenorrhea, hormonal reasons, expulsion, and pregnancy (zero), respectively, from top to bottom.

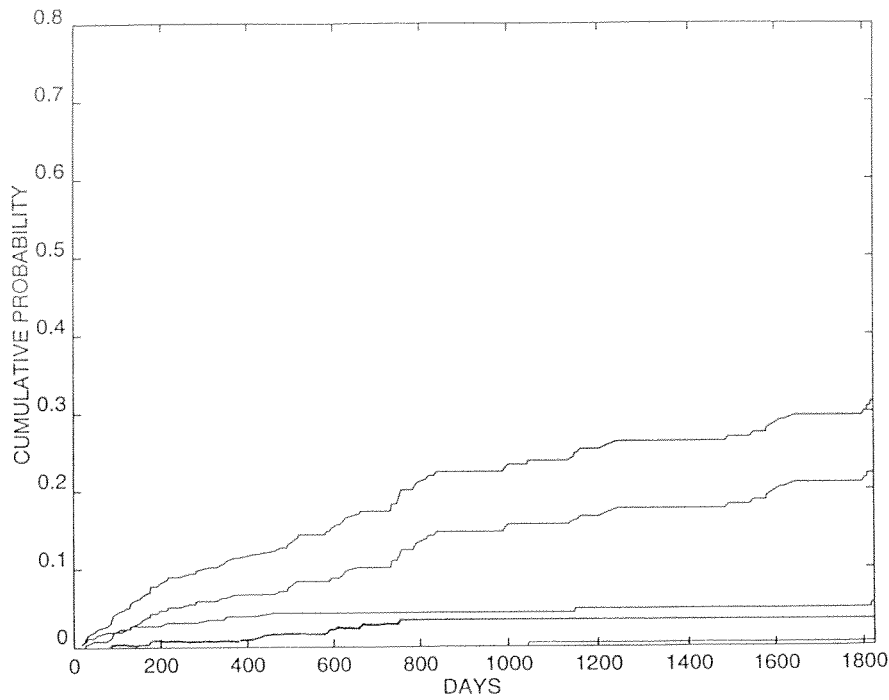


Figure 7. Estimated cumulative probabilities of termination for the Nova T in Finland. The curves correspond to the overall contraceptive method-related termination probability, termination probabilities due to bleeding and pain, expulsion, pregnancy, hormonal reasons, and amenorrhea (zero), respectively, from top to bottom.

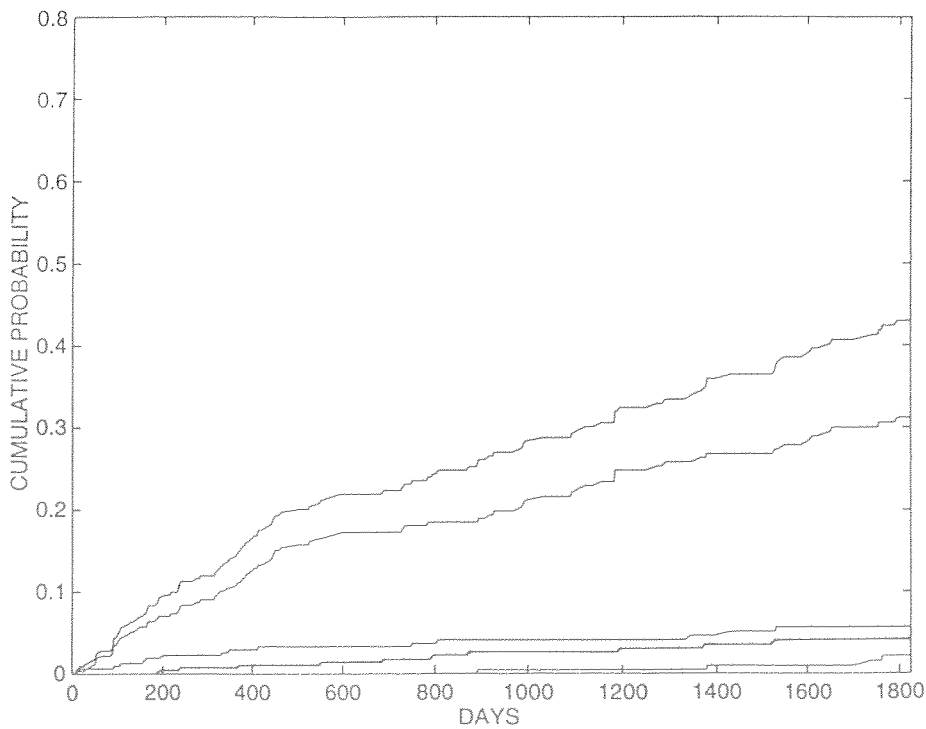


Figure 8. Estimated cumulative probabilities of termination for the Nova T in Sweden. The curves correspond to the overall contraceptive method-related termination probability, termination probabilities due to bleeding and pain, expulsion, pregnancy, hormonal reasons, and amenorrhea (zero), respectively, from top to bottom.

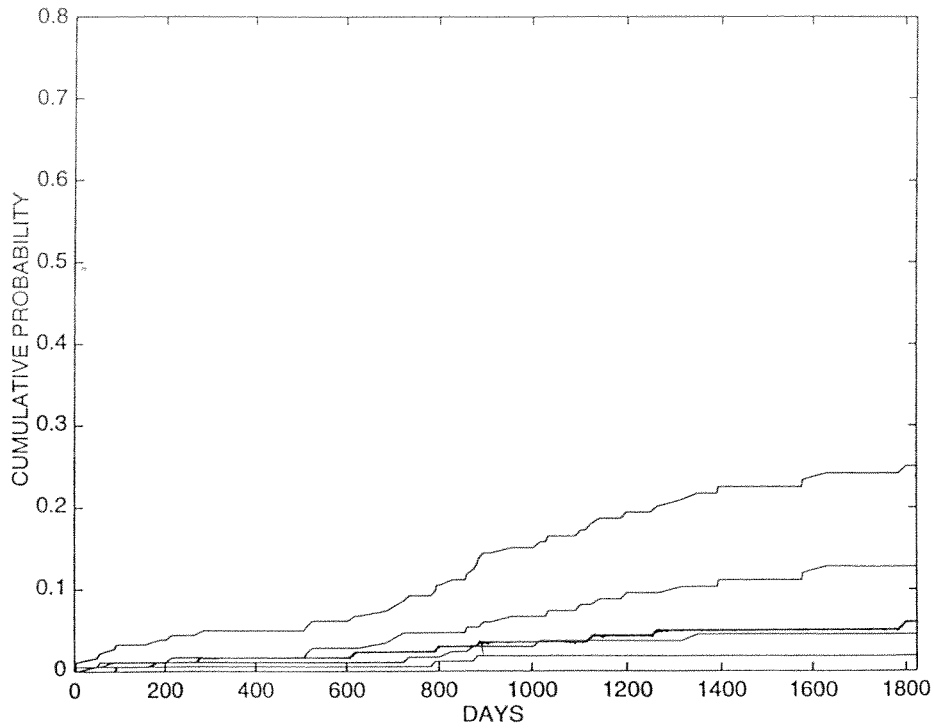


Figure 9. Estimated cumulative probabilities of termination for the Nova T in Hungary. The curves correspond to the overall contraceptive method-related termination probability, termination probabilities due to bleeding and pain, pregnancy, expulsion, hormonal reasons, and amenorrhea (zero), respectively, from top to bottom.

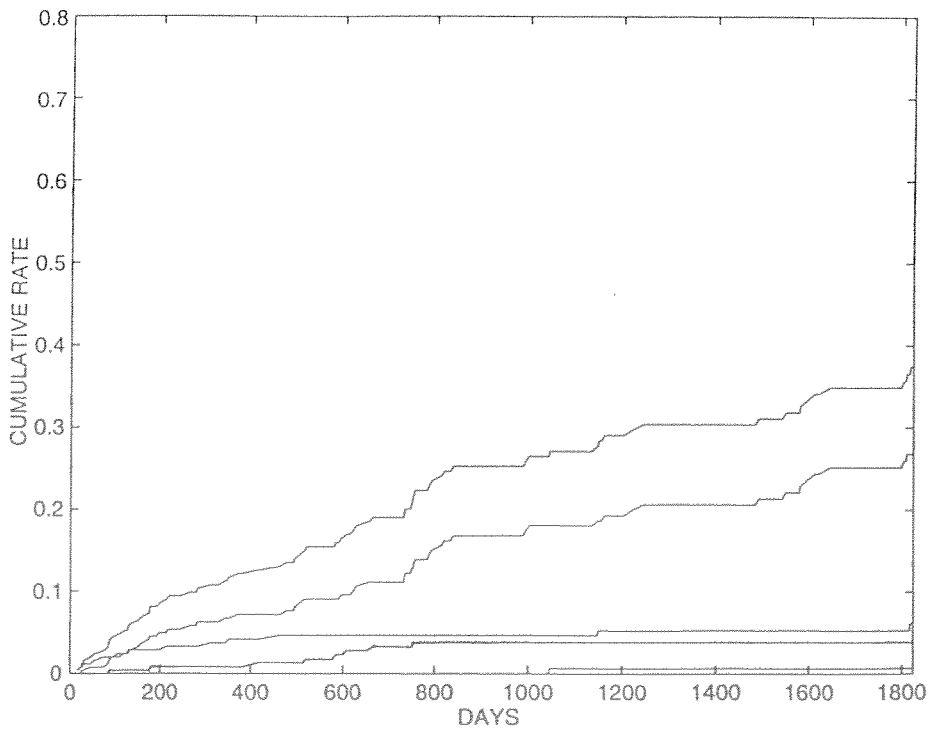


Figure 10. Estimated cumulative rates of termination for the Nova T in Finland. The curves correspond to the overall contraceptive method-related termination rate, termination rates due to bleeding and pain, expulsion, pregnancy, hormonal reasons, and amenorrhea (zero), respectively, from top to bottom.

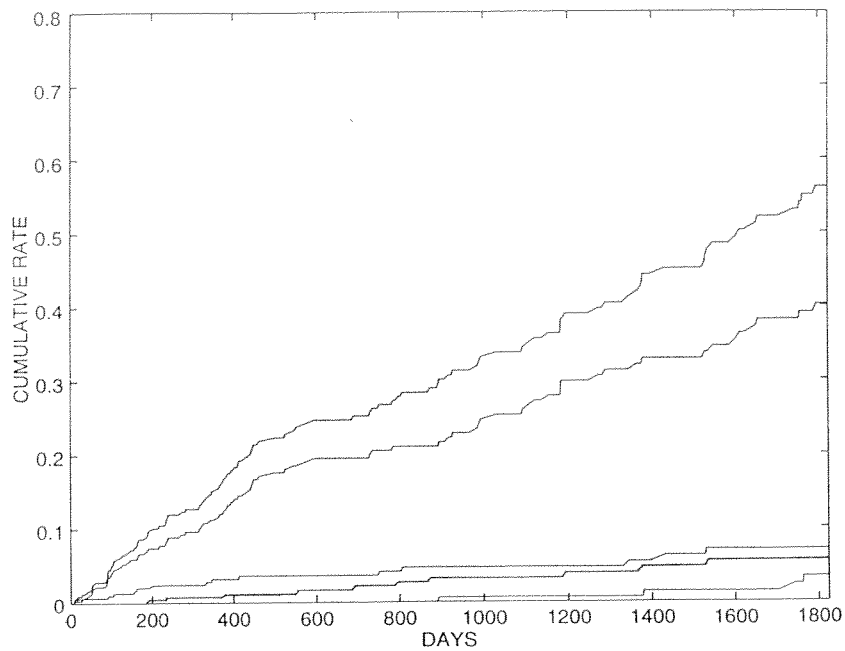


Figure 11. Estimated cumulative rates of termination for the Nova T in Sweden. The curves correspond to the overall contraceptive method-related termination rate, termination rates due to bleeding and pain, expulsion, pregnancy, hormonal reasons, and amenorrhea (zero), respectively, from top to bottom.

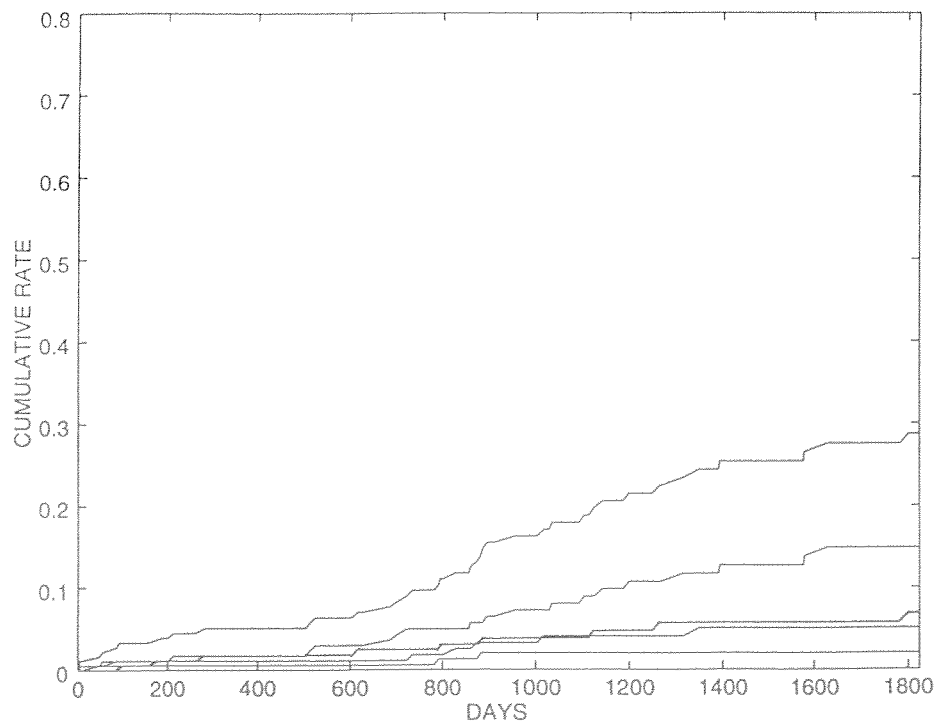


Figure 12. Estimated cumulative rates of termination for the Nova T in Hungary. The curves correspond to the overall contraceptive method-related termination rate, termination rates due to bleeding and pain, pregnancy, expulsion, hormonal reasons, and amenorrhea (zero), respectively, from top to bottom.

Table 1. Estimated five-year cumulative probabilities and rates (Finland)

	Potter net rate	Potter gross rate	$\hat{F}_j(t)$ (probability)	$\hat{H}_j(t)$ (rate)
* LNG-IUD				
Total termination	0.2555	0.2555	0.2554	0.2945
Pregnancy	0.0037	0.0043	0.0037	0.0043
Expulsion	0.0543	0.0568	0.0542	0.0584
Amenorrhea	0.0209	0.0244	0.0209	0.0247
Bleeding/pain	0.1002	0.1101	0.1001	0.1163
Hormonal	0.0764	0.0869	0.0764	0.0908
Nova-T IUD				
Total termination	0.3134	0.3134	0.3129	0.3741
Pregnancy	0.0336	0.0380	0.0336	0.0386
Expulsion	0.0544	0.0596	0.0542	0.0612
Amenorrhea	0.0000	0.0000	0.0000	0.0000
Bleeding/pain	0.2206	0.2362	0.2202	0.2680
Hormonal	0.0048	0.0063	0.0048	0.0063

regarded as being disturbing by the user or her doctor. The same was true for terminations due to other bleeding disturbances or hormonal side-effects. When the termination rates due to reasons 2, 3, and 5 were viewed together, their relative positions in Finland and Sweden were quite different from what they are in Hungary. While cause 3 was common in Hungary and rare in the Nordic countries, the opposite was true about cause 5.

Figures 4 to 6 show the cumulative termination rates. Recall that these rates were obtained by conditioning on the event that the device was in situ. Almost all the observations made above, based on Figures 1 to 3, carry on to these figures. The concavity of the curves corresponds to the decrease of the instantaneous rates of termination with time. The rates were much higher during the first days after insertion and then slowly stabilized. This may be viewed as a consequence of a selection mechanism operating on the study population. Typically, women who were more prone to any cause of termination were removed from the study. Once again, this phenomenon emphasized the impossibility of studying a single termination reason "in isolation."

Figures 7 to 12 show the corresponding results for the Nova T. The observations made above for the LNG-IUD regarding the termination rates due to amenorrhea and hormonal reasons no longer seemed valid. In fact, termination due to bleeding and pain dominated the other termination reasons. The termination rate due to amenorrhea was 0 in all three countries. There was a noticeable difference in the termination rates due to bleeding and pain between Hungary and the Nordic countries. The overall termination rates were also remarkably different.

Comparison of the two devices can be carried out by comparing the rates. The termination rate due to

pregnancy was higher for the Nova T than the LNG-IUD in all three countries. On the other hand, the terminations due to hormonal reasons were more common for the LNG-IUD than for the Nova T in all three countries. This was to be expected since the LNG-IUD is a hormonal contraceptive method. It is important to notice, though, that during the use of the Nova T there are also discontinuations due to hormonal side effects.

Table 1 gives the 5-year cumulative probability and rate estimates, based on the Finnish data, using life table methods and the methods suggested in the earlier section. The Potter rates were evaluated by grouping the data into (approximately) monthly intervals: 0-30 (days), 30-60, 60-90, ..., 1800-1825, and >1825 days. For each interval x , the total number of terminations (d_x), terminations due to pregnancy (p_x), expulsion (e_x), amenorrhea (a_x), bleeding and pain (b_x), hormonal (h_x), and all the remaining reasons plus withdrawal from the study (w_x) were obtained. Also n_x , the number of individuals at risk, was obtained. Using these data, Potter's net and gross rates were calculated.

Notice that if the failure times are recorded exactly the Potter net rates and the probabilities $\hat{F}_j(t)$ will be numerically the same at the grid points, if the time grid is fine. This is because then there are no tied observations, and at each failure time only one of the quantities p_x , e_x , a_x , b_x , h_x , or w_x will be 1, the others being 0. (In our example, even though the failure times in the data are reported in days, some observations are tied.)

It is important not to confuse probabilities and rates. Note that if the time intervals x that are used in constructing the life table are short enough to make the relative frequencies d_x/n_x small, Potter's net rates and the termination-specific probabilities $\hat{F}_j(t)$ have

values that are numerically close to each other. However, as was mentioned before, they will be quite different from the corresponding termination-specific rates (increments of $\hat{H}_i(t)$) if only a small proportion of the original study population are at risk. It is important to note that the Potter gross rates, being probabilities, cannot be compared to the cumulative termination-specific rates $\hat{H}_i(t)$.

Discussion

The inherent inadequacy of the competing risks data to support any contemplation concerning the dependence between different competing risks, and indeed the very nonexistence of latent lifetimes in any real physical sense, calls for the analysis of competing risks data on the basis of the cumulative incidence functions or the cause-specific hazard rates.

From Potter's writing, one can only conclude that gross rates, reflecting a situation in which the other competing causes would not be present, could be estimated from real population life table data without making any further hypotheses, by merely adjusting the sizes of the risk sets used in estimating the monthly net rates. This, quite obviously, cannot be true; no manipulation of the data, be that of a monthly adjustment of the risk set sizes or anything else, can undo the fact that individuals originally in the study population, and who have experienced one of the competing causes of termination, have actually been selected away from the risk set. Thus, unless one wants to create a hypothetical world in which women using an IUD can become pregnant, but in which removal and expulsion of the IUD are impossible, nothing can be said about a situation in which pregnancy risk is considered "in isolation." Even then, the connection between the real world situation and that prevailing in the hypothetical world must be based on an assumption, which simply equates a rate in the hypothetical world with something that can be measured, or at least estimated, from real data. This assumption, which unfortunately is never mentioned explicitly in contraceptive failure literature, cannot be tested. A similar critical view is reflected in Kalbfleisch and Prentice.¹⁵

As noted above, if the failure times are recorded exactly and life table estimates are replaced by corresponding continuous time estimates, then the role of monthly adjustments on the risk sets disappears completely. As a consequence, the expressions corresponding to cumulative Potter's net rate and \hat{F}_i coincide. If the failure/withdrawal times are approximated using a time grid, the net rates and the probabilities \hat{F}_i are numerically close to each other if the grid is so fine that only a few times (a small

proportion) are in the same interval. This explains the closeness of the numerical results obtained in our paper. For exactly recorded times there is no need to use the life table techniques in which the risk set is adjusted using a "midterm" approximation.

The connection between the Potter net rates and the Kaplan-Meier estimate is stated in formulae 1 and 2. Regarding the Potter gross rates, however, their computation for a specific termination reason from exact data would correspond to a Kaplan-Meier estimate, where all competing causes are treated as censoring. In general, for such an estimate to make sense one has to require that censoring is "independent." But, in the present context, there is absolutely no reason to assume that the other termination reasons would be independent from the considered specific reason. In reality, termination reasons are likely to depend on each other in complex ways, which, however, unfortunately cannot be estimated or tested on the basis of data. Very likely, for example, improved counselling in Hungary would have led to fewer terminations in the use of the IUD because of amenorrhea. Consider, therefore, a situation in which counselling were really better and, as a consequence, some women, who otherwise would have been worried by their amenorrhea to the extent that they would have terminated their use of IUD, would actually continue the trial. "Independence" would now mean that we assume that this change has no consequences in the estimation of the rates/probabilities of the other (competing) termination reasons. There is no real way to justify such an assumption.

Finally, these conceptual confusions are only made worse by the fact that the statistical terminology, viz., rate and probability, used in the context of contraceptive failures differs in several ways from what is commonly used in modelling survival or duration data. It is even internally inconsistent, with some authors using the same terms as others, but with a different meaning. For this we refer to Trussell.¹¹

The central idea of this paper is that the modern techniques of competing risks can be used to answer the questions that arise naturally in contraceptive studies, and hence the earlier techniques should be replaced by ones that are now standard in many other areas of applied research.

Acknowledgments

The authors are thankful to Leiras Oy, Turku, Finland, for providing the data and also to Juhani Tuominen, Pekka Lähteenmäki, and Marja Oinonen for their cooperation and helpful suggestions.

References

1. Pearl R. Contraception and fertility in 2000 women. *Hum Biol* 1932;4:363-407.
2. Tietze C. Differential fecundity and effectiveness of contraception. *Eugen Rev* 1959;50:231-7.
3. Potter RG. Additional measures of use-effectiveness of contraception. *Milbank Mem Fund Q* 1963;41:400-18.
4. Potter RG. Applications of life table techniques to measurement of contraceptive effectiveness. *Demography* 1966;3:297-304.
5. Tietze C, Lewit S. Recommended procedures for the statistical evaluation of intrauterine contraception. *Stud Fam Plann* 1973;4:35-42.
6. Potter RG. The multiple decrement life table as an approach to the measurement of use-effectiveness and demographic effectiveness of contraception. Contributed papers. Sydney Conference International Union of the Scientific Study of Population 1967;869-83.
7. Potter RG. Use-effectiveness of intrauterine contraception as a problem of competing risks. In Freedman R, Takeshita JY, eds. *Family Planning in Taiwan*. Princeton, NJ: Princeton University Press, 1969;458-84.
8. Tietze C. Intrauterine contraception: recommended procedures for data analysis. *Stud Fam Plann* 1967;18(suppl):1-6.
9. Tietze C, Lewit S. Evaluation of intrauterine device: ninth progress report of the Cooperative Statistical Program. *Stud Fam Plann* 1970;55:1-40.
10. Farley TMM. Life table methods for contraceptive research. *Stat Med* 1986;5:475-89.
11. Trussell J. Methodological pitfalls in the analysis of contraceptive failure. *Stat Med* 1991;10:201-20.
12. Andersson K, Odland V, Rybo G. Levonorgestrel-releasing and copper-releasing (Nova T) IUDs during five years of use: a randomised comparative trial. *Contraception* 1994;49:56-72.
13. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;53:457-81.
14. Nelson W. Theory and applications of hazard plotting for censored failure data. *Technometrics* 1972;14:945-65.
15. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. New York: John Wiley and Sons, 1980.
16. Deshpande JV, Karia SR. Modelling competing risks data via the cumulative incidence functions. *IAPQR Transactions* (Journal of the Indian Association for Productivity, Quality & Reliability) 1995;20:79-91.
17. Andersen PK, Borgan Ø, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. New York: Springer-Verlag, 1993.

Appendix

Computational Details of the Procedures Suggested by Equations 1-3

As mentioned earlier, the data were written as the termination time (length of exposure, in days) and the termination reason/status (Table 2). For the sake of illustration in this paper, the termination reasons were categorized into (1) pregnancy, (2) expulsion, (3)

Table 2. Termination times and reasons

Subject	Termination time (in days)	Termination reason
1	T_1	5
2	T_2	0
3	T_3	4
4	T_4	3
⋮	⋮	⋮
n	T_n	1

amenorrhea, (4) bleeding and pain, and (5) hormonal reasons. The subjects who terminated the use of IUD due to any of these five reasons within the study period were assigned one of the numbers 1, 2, 3, 4, or 5 corresponding to the termination reason, along with the time of termination. Those who continued the use of IUD after 5 years were assigned 0 as the termination reason and 1825 days as the censoring time. Note that more than one subject may have the same termination time, with possibly different termination reasons. We selected distinct termination times out of n original times and arranged them in ascending order. Let these selected times be $T_{(1)}, T_{(2)}, \dots, T_{(k)}$. Here $k \leq n$. We have reorganized the information in Table 3 by counting the number of terminations due to various reasons (columns 3-7) and the number of women at risk, i.e., those who are still under study (column 2), at each distinct termination time. Column 8 of Table 3 was then obtained by summing columns 3-7, giving the total terminations observed at each distinct time. The estimated probability of termination of the usage of IUD at each time was computed as column 8/column 2 in Table 3. The estimated probability of termination due to pregnancy over time is given by column 3/column 2, due to expulsion by column 4/column 2, and so on (see columns 9-14 of Table 3). Now, it is simple to estimate the cumulative termination-specific rates by adding the probabilities successively. The probability of continuing the use of IUD is estimated by $1 -$ column 8 and an estimate of the probability of continuing the use up to a certain time t is the product of such probabilities corresponding to those termination times which are at most t . The probability of termination due to a specific reason in the presence of all the competing reasons is estimated by multiplying the appropriate column (column 9-column 13) by column 16 and then by taking the cumulative sum of the resulting column (equation 2).

Repeating this procedure for the data from various centers will enable comparison of the centers with respect to the use of IUD.

Table 3. Computation of estimated termination-specific probabilities and rates

Distinct time (in days) (1)	Number of women at risk (2)	Termination due to					Total (8)	Termination-specific rates					Continuation probability (15)	$S(T_{ij}^-)$ (16)	
		1 (3)	2 (4)	3 (5)	4 (6)	5 (7)		(3) ÷ (2) (9)	(4) ÷ (2) (10)	(5) ÷ (2) (11)	(6) ÷ (2) (12)	(7) ÷ (2) (13)			(8) ÷ (2) (14)
$T_{1(1)}$	n_1	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}	d_1	d_{11}/n_1	d_{12}/n_1	d_{13}/n_1	d_{14}/n_1	d_{15}/n_1	d_1/n_1	$1 - (d_1/n_1)$	1
$T_{1(2)}$	n_2	d_{21}	d_{22}	d_{23}	d_{24}	d_{25}	d_2	d_{21}/n_2	d_{22}/n_2	d_{23}/n_2	d_{24}/n_2	d_{25}/n_2	d_2/n_2	$1 - (d_2/n_2)$	$1 - (d_1/n_1)$
$T_{1(3)}$	n_3	d_{31}	d_{32}	d_{33}	d_{34}	d_{35}	d_3	d_{31}/n_3	d_{32}/n_3	d_{33}/n_3	d_{34}/n_3	d_{35}/n_3	d_3/n_3	$1 - (d_3/n_3)$	$\Pi_{j=1}^2 [1 - (d_j/n_j)]$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$T_{1(k)}$	n_k	d_{k1}	d_{k2}	d_{k3}	d_{k4}	d_{k5}	d_k	d_{k1}/n_k	d_{k2}/n_k	d_{k3}/n_k	d_{k4}/n_k	d_{k5}/n_k	d_k/n_k	$1 - (d_k/n_k)$	$\Pi_{j=1}^{k-1} [1 - (d_j/n_j)]$

Table 4. LNG-IUD: Estimated 5-year cumulative probabilities and rates, with standard errors and confidence limits

	$\hat{F}_j(t)$ (probability)	Standard error	Confidence limits	$\hat{H}_j(t)$ (rate)	Standard error	Confidence limits
Total termination	0.255	0.022	{0.216,0.303}	0.295	0.026	{0.249,0.349}
Pregnancy	0.004	0.003	{0.001,0.015}	0.004	0.003	{0.001,0.017}
Expulsion	0.054	0.010	{0.038,0.077}	0.058	0.011	{0.041,0.084}
Amenorrhea	0.021	0.006	{0.012,0.038}	0.025	0.008	{0.014,0.045}
Bleeding/pain	0.100	0.013	{0.078,0.129}	0.116	0.016	{0.089,0.153}
Hormonal	0.076	0.012	{0.057,0.103}	0.091	0.015	{0.067,0.124}

Computation of the Standard Errors of the Estimators Given in Formulae 1-3 and Their Confidence Limits

The details regarding computation of the standard errors and confidence limits are available in the literature.¹⁷ The standard errors of the estimators given in formulae 1-3 can be easily computed on the basis of Tables 2 and 3.

The standard error of the Kaplan-Meier estimator given in equation 1 is the square root of

$$\hat{\sigma}^2(t) = \hat{S}^2(t) \sum_{\{i:T_{(i)} \leq t\}} d_i / (n_i(n_i - d_i)). \quad (4)$$

The standard error at an observed time $T_{(i)}$ is then

$$\hat{\sigma}^2(T_{(i)}) = \hat{S}^2(T_{(i)}) \sum_{l=1}^i d_{il} / (n_l(n_l - d_l)).$$

Here, $\hat{S}(T_{(i)})$ is obtained by taking the product of entries in column 15 from row 1 to row i . The second term is obtained by summing terms of the form column 14/(column 2 - column 8) over rows 1 to i . Then the standard error is simply the square root of the product of these two quantities.

The standard error of the estimator (equation 2) is the square root of

$$\hat{\sigma}_j^2(t) = \sum_{\{i:T_{(i)} \leq t\}} (\hat{F}_j(t) - \hat{F}_j(T_{(i)}))^2 d_{ij} / (n_i(n_i - d_i)) + \sum_{\{i:T_{(i)} \leq t\}} \hat{S}(T_{(i)}) (\hat{S}(T_{(i)}) - 2\hat{F}_j(t) + 2\hat{F}_j(T_{(i)})) d_{ij} / n_i^2. \quad (5)$$

This gives

$$\hat{\sigma}_j^2(T_{(i)}) = \sum_{l=1}^i (\hat{F}_j(T_{(i)}) - \hat{F}_j(T_{(l)}))^2 d_{il} / (n_l(n_l - d_l)) + \sum_{l=1}^i \hat{S}(T_{(i)}) (\hat{S}(T_{(i)}) - 2\hat{F}_j(T_{(i)}) + 2\hat{F}_j(T_{(l)})) d_{il} / n_l^2.$$

The computational procedure for $\hat{F}_j(t)$ is explained above. The other quantities can be obtained by dividing the appropriate columns of Table 3.

The standard error of the estimator (equation 3) is the square root of

$$\hat{V}_j^2(t) = \sum_{\{i:T_{(i)} \leq t\}} d_{ij} / n_i^2. \quad (6)$$

The standard error at time $T_{(i)}$ is obtained by summing terms obtained by dividing the term in the appropriate column between 9 and 13 by the corresponding term in column 2, over rows 1 to i .

The approximate 95% confidence limits are then given by computing

$$\begin{aligned} &\hat{F}(t) \exp \{ \pm 1.96 \hat{\sigma}(t) / \hat{F}(t) \}, \\ &\hat{F}_j(t) \exp \{ \pm 1.96 \hat{\sigma}_j(t) / \hat{F}_j(t) \}, \\ &\hat{H}_j(t) \exp \{ \pm 1.96 \hat{V}_j(t) / \hat{H}_j(t) \}, \end{aligned}$$

where $\hat{\sigma}(t)$, $\hat{\sigma}_j(t)$, and $\hat{V}_j(t)$ are square roots of $\hat{\sigma}^2(t)$, $\hat{\sigma}_j^2(t)$, and $\hat{V}_j^2(t)$, respectively.

Tables 4 and 5 give estimates, standard errors, and confidence limits for the 5-year cumulative probabilities and rates for the LNG-IUD and Nova-T IUD for Finland, as obtained by using these methods.

Table 5. Nova-T IUD: Estimated 5-year cumulative probabilities and rates, with standard errors and confidence limits

	$\hat{F}_j(t)$ (probability)	Standard error	Confidence limits	$\hat{H}_j(t)$ (rate)	Standard error	Confidence limits
Total termination	0.313	0.038	{0.247,0.397}	0.374	0.046	{0.295,0.475}
Pregnancy	0.034	0.012	{0.017,0.066}	0.039	0.014	{0.019,0.078}
Expulsion	0.054	0.015	{0.032,0.093}	0.061	0.018	{0.035,0.108}
Amenorrhea	0	0	0	0	0	0
Bleeding/pain	0.220	0.001	{0.171,0.283}	0.268	0.039	{0.201,0.357}
Hormonal	0.005	0.005	{0.001,0.034}	0.006	0.006	{0.001,0.045}

Different types of contraception will generally lead to different classifications of termination reasons, but the basic principles of the analysis remain the same. In particular, one can include as many termination reasons as one wants to study, without complicating the procedure. Inclusion of more termination reasons

will simply lead to increasing the number of columns in Table 3. A test procedure for the joint comparison of several termination reasons, for example, of a particular device in two different countries, and based on the termination-specific rates will be considered in a future work.