Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2020;6(2):290-304

Original Article

Estimation of the Distribution of Duration of Breastfeeding from Cross-Sectional Data: Some Methodological Issue.

Anup Kumar[1], Sachin Kumar[1], Jai Kishun[1]*, Uttam Singh[1], Pushpraj[1,2]

[1]Department of Biostatistics & Health Informatics, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow - 226014, India.
[2]Department of Statistics, *Babasaheb Bhimrao Ambedkar University*, Lucknow-226014, India.

ARTICLE INFO

Received 14.11.2020
Revised 23.11.2020
Accepted 01.12.2020
Published 10.12.2020

Key words:
Breastfeeding;
Epidemiological;
Life table;
Kaplan-Meier estimator;
Sampling frame.

ABSTRACT

Background: Duration of breastfeeding is an important health indicator of mother and child. There are various indirect epidemiological methods available to estimate the duration of breastfeeding from cross sectional data.

Objective: To estimate the distribution of duration of breastfeeding at national level cross sectional data and compare various available technique. The impact of the sampling frame (ascertain of the individual understudy) is also evaluated.

Method: National Family Health Survey (NFHS-IV) data is used. Duration of breastfeeding of only those children who were born before 60 months from survey date were included in the study. The technique of Current Status Data, Life Table Analysis, and Kaplan Meier (KM) estimator is applied to assess the distribution of duration of breastfeeding.

Result: The mean estimate is 32.84, 33.14 and 33.64 months by Kaplan Maier Estimator, Current Status Data and Life Table Analysis respectively. The Current Status and Life Table method are better than Kaplan Meier Estimator as it is doesn't based on recall data and heaping present in the data.

Conclusion: One must be very cautions while estimating the various epidemiological parameters from cross section data set. The assumptions of the methodology as per data available should be evaluate. If such data is not available, the available methodology may be modified. Regression analysis based on Current Status data technique may be used to assess the impact of various clinical and epidemiological factors (such as nutrition of mother, health status of mother etc.) on duration of breastfeeding.

* Corresponding Author's Email: jaikishan.stat@gmail.com

Introduction

Distribution of duration of breastfeeding is an important indicator of health of mother and child. Breastfeeding is one of the most effective ways to ensure child health and survival(1,2) in first year breastfeeding, and up to one-third of the second year of life of the child. Breastfed children have better intelligence tests, are less overweight or obesity and less prone to diabetes later in life. The mother who breastfeed also have a reduced risk of breast and ovarian cancers. Mothers worldwide are recommended to exclusively breastfeed infants for the child's first six months to achieve optimal growth, development, and health of child.

Generally, for estimating the distribution of duration of breastfeeding, a cohort of birth say of size N is followed, till all have completed breastfeeding. The data thus obtain may provide us the distribution of duration of breastfeeding of the cohort. Practically such data is unavailable and also difficult to obtain. On the other hand, cross-sectional data on the duration of breastfeeding is available in different national level health survey. Various literature explained different techniques for estimation of duration of breastfeeding(3–5). Generally, Life Table, Kaplan Meier, and Current Status Estimator are very commonly used methods. In cross-section data, the duration of breastfeeding is not appropriately reported due to recall lapse of women. Hence confirmed age heaping at 6,12,36 months etc are common(6–10) . Generally studies rely on reported duration/recollection of duration, in first approach, duration of breastfeeding is considered to refer to the age of the children at the time of complete termination, regardless of the time when consumption of other foods began(6; 7). Other approach sees the use of current status or interval-censored data, whereby just the current breastfeeding status along with the age of the child at the time of survey interview is considered in building a picture of termination age (8; 9). The major advantages of retrospectively reported breastfeeding data are the ease of data collection, researchers often opting for cross-sectional approaches in order to save time and cost, as well as the ability to capture relatively larger sample sizes (14). Drawback of recalled data is associated with age heaping, with participants tending to round up or down (15) the exact age of the child when the breastfeeding termination took place. This limits the ability to draw valid inferences (16). With current status data, the likelihood of age heaping is comparatively lower, except for the case of heaping in the reported age of children. This remains a problem due to misreporting of age. In general, more errors occur the greater the time-lag between an event and its recall. Some previous studies, concerning the distribution of breastfeeding termination times, have observed that the current-status measures lead to unbiased estimates of the survival function for a sample of births that occur during a fixed period (14). While, the present approach promises more reliable measures, several studies on breastfeeding have been conducted using this approach(13,16–18) due to computational complexity (8).

Apart from the types of estimation techniques, one another important factor affecting the distribution of duration of breastfeeding is the sampling frame. Sampling frame is defined as the method by which an individual is ascertained or identified as a member sample population and the time's reference for the duration variable to be measured(14, 15). The distribution of duration of breastfeeding also depends much upon the sampling frame as shown in table 6. As the sampling frame changes the distribution also change. For estimating the duration of breastfeeding there are two sampling frames and consequently, the distribution will also change (21). For example, the duration of breastfeeding would be different if considering child as unit or mother as a unit. On the basis of these two sampling frame the distribution will be change.

Further, the sampling frame is also decided on the feasibility and considering the non-sampling error. In all these cases, the distribution is likely will vary in such a situation. Hence, certainly, a careful evaluation of the sampling frame is needed for analyzing the observed data and drawing inferences about population characteristics

The three techniques (Current Status, Kaplan Meier and Life Table Techniques) for estimating the distribution of duration of breastfeeding from cross-sectional data is compared. The effect of the sampling frame is also evaluated. Consequently, the appropriate technique and feasible sampling frame in cross-sectional data is examined.

Methodology

Data

Birth record data of National Family Health Survey (NFHS-IV), collected during year 2015-16 is used. It is cross sectional data. From this dataset, variables such as child is alive, index of birth history, date of the interview, date of birth of the child, currently breastfeeding (‘yes’, ’no’), months of breastfeeding are used. Only those children included in study, whose age was in-between 0-60 months at the time of survey. Among this group of children or duration of breastfeeding many children have completed the breastfeeding and many are still continuing breastfeeding. A total 176335 number of children found whose age were less than 60 months. Figure 1 shows the extraction of dataset from Birth record data of National Family Health Survey (NFHS-IV) data.

Figure 1 Between these duration,children are taken for the analysis

Following three techniques explained below were used to estimate the duration of breastfeeding on available data (NFHS-IV).

1. Current status technique
2. Kaplan Meier Estimator
3. Life table technique

Current status technique

In observation, (t) is the duration of breastfeeding which is restricted to knowledge of whether or not (t) exceeds the date of survey. This structure is known as current status data and sometimes referred to as referred case I interval censored data. In cross sectional study, the age of the child along with his current status of breastfeeding is noted. For example, there will n_0 of number of child of 0-1 month, n1 number of child of 1-2 month, n_t

is the number of child of age (t,t+1) months (t=1,2,….T) and among that n_t , y_t number of children are still breastfeeding.
So, S_t proportion of child still breastfeeding

$$S_t = \frac{y_t}{n_t} = 1 - F(t) \quad (1)$$

denote the proportion of children has breastfeeding more than t. The value of S_t can be obtained for different (t) from the data and the duration of breastfeeding can be obtained by spline smoothing, the plot between t and S_t. S_t is obtained by the equation (1). S_t is smoothed by spline, and after smoothing the cumulative distribution and distribution is obtained (see table 2).

Kaplan Meier estimator

The retrospectively reported durations of breastfeeding for weaned children, along with censored durations of breastfeeding for children still being breast-fed at the time of the survey. Analysis of this type of data is done by Kaplan Maier estimator.
Let $t > 0$ be the duration of breastfeeding if a child stops breastfeeding then it is events. of interest takes place. As indicated above, the goal is to estimate the survival function S_t underlying.

S_t= *Prob* (t > T),where t = 0,1,...is the time .(2)

The estimator of the survival function S_t (the probability that life is longer than T) is given by:

$$S_t = S_t = \prod_{i:ti \le T} \left(1 - \frac{di}{ni}\right), \quad (3)$$

With t_i a duration of breastfeeding when at least one event happened, d_i the number of events (e.g., number of the child who weaned) that happened at time $t_i$and n_i the individuals known to have survived (have not yet had an event or been censored) up to time $t_i$as shown in table 3.
For applying the Kaplan Meier estimator, in National Family Health Survey (NFHS-IV) the months of duration of breastfeeding is given and the event has been occurred or not is given by "whether baby is current breastfeeding or not" is given by these two variable the Kaplan Meier estimator is estimated.

Life Table Technique

The simplest analysis of the data of duration of breastfeeding irrespective of how they are ascertained in subject to serious limitation when the observation is truncated at some point before the age or due to survey date. To minimize the bias resulting in incomplete observation of duration of breastfeeding. A brief description of the method and notation are given in below table1:

Table 1: Summary of Notations and their Descriptions used in Life Table Technique

Notation	Description
$\boldsymbol{N_o}$	The number of an eligible child in the study
$\boldsymbol{t}$	Completed number of months since birth.
$\boldsymbol{N_t}$	Number of child having breastfeeding more than t^{th} month:
$\boldsymbol{B_t}$	Number of child complete the breastfeeding between t & $t+n$ months:
$\boldsymbol{V_{t;i}}$	The time of breastfeeding of i^{th} child within the interval$(t \& t + n)$.
$\boldsymbol{W_t}$	No. of child withdrawn breastfeeding within the interval between t & $t + n$.
$\boldsymbol{w_{t;j}}$	The time of withdrawn of j^{th} child with in the interval $(t;\ t + n)$.
$\boldsymbol{q_t}$	Conditional probability of breastfeeding between t & $t + n$ months of breastfeeding.

Notation	Description
$\boldsymbol{p_t}$	$1 - q_t$.
$\boldsymbol{a_t}$	average time of breastfeeding in the interval $(t;\ t + n)$.
$\boldsymbol{M_t(dt)}$	Conditional probability of breastfeeding between t & $t + n$.
$\boldsymbol{n_t}$	Length of interval
$\boldsymbol{E_t}$	Expected duration of breastfeeding after t months
$\boldsymbol{r_t}$	Number of child weaned at the t months

Table 2 Distribution function of Current status data for duration of breastfeeding of ≤ 60 months

Time	No. of child whose breastfeeding duration is greater than that interval.	Total	S_t		*F(t)* after smoothing	*f(t)* after smoothing
			Unsmoothed	Spline smoothed		
0-1	1674	1702	0.984	0.987	0.013	0.004
1-2	3642	3688	0.988	0.983	0.017	0.002
2-3	3867	3944	0.980	0.981	0.019	0.004
3-4	4064	4141	0.981	0.977	0.023	0.005
4-5	4159	4270	0.974	0.972	0.028	0.002
5-6	4226	4348	0.972	0.970	0.030	0.006
6-7	4088	4221	0.968	0.965	0.035	0.004
7-8	4104	4265	0.962	0.960	0.040	0.009
8-9	4128	4308	0.958	0.951	0.049	0.005
9-10	3956	4181	0.946	0.946	0.054	0.002
10-11	3725	3931	0.948	0.944	0.056	0.017
11-12	3496	3724	0.939	0.928	0.072	0.015
12-13	3630	3951	0.919	0.913	0.087	0.019
13-14	3720	4100	0.907	0.895	0.105	0.016
14-15	3412	3865	0.883	0.878	0.122	0.007
15-16	3365	3841	0.876	0.871	0.129	0.013
16-17	3391	3923	0.864	0.858	0.142	0.019
17-18	3258	3836	0.849	0.839	0.161	0.014
18-19	3320	4005	0.829	0.825	0.175	0.023
19-20	3065	3746	0.818	0.802	0.198	0.019
20-21	2958	3770	0.785	0.783	0.217	0.010
21-22	2784	3556	0.783	0.773	0.227	0.024
22-23	2520	3328	0.757	0.749	0.251	0.029
23-24	2238	3040	0.736	0.720	0.280	0.055
24-25	2212	3214	0.688	0.665	0.335	0.028
25-26	2056	3230	0.637	0.636	0.364	0.011
26-27	2003	3143	0.637	0.626	0.374	0.036
27-28	1877	3129	0.600	0.590	0.410	0.009
28-29	1732	2988	0.580	0.581	0.419	0.007
29-30	1677	2881	0.582	0.574	0.426	0.039
30-31	1634	2912	0.561	0.534	0.466	0.025

Time	No. of child whose breastfeeding duration is greater than that interval.	Total	S_t		*F(t)* after smoothing	*f(t)* after smoothing
			Unsmoothed	Spline smoothed		
31-32	1479	2841	0.521	0.509	0.491	0.006
32-33	1347	2660	0.506	0.504	0.496	0.010
33-34	1324	2642	0.501	0.494	0.506	0.001
34-35	1228	2499	0.491	0.495	0.507	0.039
35-36	1220	2482	0.492	0.454	0.546	0.063
36-37	1085	2559	0.424	0.392	0.608	0.024
37-38	923	2453	0.376	0.368	0.632	0.028
38-39	916	2544	0.360	0.339	0.661	0.013
39-40	780	2397	0.325	0.327	0.673	0.003
40-41	771	2335	0.330	0.324	0.676	0.011
41-42	744	2355	0.316	0.313	0.687	0.011
42-43	709	2289	0.310	0.301	0.699	0.008
43-44	660	2244	0.294	0.293	0.707	0.011
44-45	650	2224	0.292	0.282	0.718	0.027
45-46	573	2141	0.268	0.255	0.745	0.005
46-47	522	2093	0.249	0.258	0.750	0.006
47-48	531	2016	0.263	0.244	0.756	0.043
48-49	435	2012	0.216	0.201	0.799	0.018
49-50	364	1925	0.189	0.183	0.817	0.006
50-51	324	1815	0.179	0.178	0.822	0.013
51-52	340	1952	0.174	0.165	0.835	0.006
52-53	289	1868	0.155	0.159	0.841	0.006
53-54	298	1820	0.164	0.153	0.847	0.003
54-55	255	1892	0.135	0.132	0.850	0.017
55-56	256	1928	0.133	0.133	0.867	0.001
56-57	253	1902	0.133	0.135	0.868	0.001
57-58	244	1786	0.137	0.131	0.869	0.003
58-59	210	1772	0.119	0.118	0.872	0.004
59-60	211	1708	0.124	0.124	0.876	0.124
>60	0		0	0	1	0

$$P_t = exp^{-\int_t^{t+n} m_t dt} \qquad \text{T=0,1,2, n=0,1,2, …} \qquad (4)$$

$$a_t = \left\{\frac{1}{nm_t} - \frac{exp^{-nm_t}}{1-exp^{-nm_t}}\right\} \qquad \text{T=0,1,2, …; n=0,1,2, …} \qquad (5)$$

The maximum likelihood estimation of m_t is obtained as, once the estimates of m_t are obtained, then for different values of t, the estimated values of q_t and a_t can be obtained easily using equation respectively and consequently other columns of the life table can also be obtained. The other important column $exp^{\,t}$ is computed using the procedure described in (22).

$$m_t = \frac{B_t}{nN_t + a_t B_T + a_t W_t} \qquad t = 0,1,2, \ldots \quad (6)$$

$n = 0,1,2, \ldots$

$$L_i = n_i(l_i - d_i) + a_i n_i d_i \quad (7)$$

$$r_t = l_t * Q_t \quad (8)$$

$$E_i = \frac{T_i}{l_i} \quad (9)$$

In fact, $exp^{\,t}$ represents the expected duration of breastfeeding after t months. It is pertinent to mention that m_t is almost the same λ i.e. λ is assumed to be constant over time while m_t may vary for different values of t as shown in table 4.

For the life table analysis, the duration of breastfeeding and the specific question is asked whether a child was "still breastfeeding" at the

Table 3 Distribution function of Kaplan Meier estimator for the duration of breastfeeding of ≤ 60 months

Time	**Number of child who start breastfeeding during these interval**	**no. of child who stop breastfeeding**	**no. of child who are censored**	**S_t)**	**S_t after smoothing**	**$F(t)$ after smoothing**	**$f(t)$ after smoothing**
0-1	176335	573	1674	0.997	0.991	0.009	0.021
1-2	174088	2363	3642	0.983	0.970	0.030	0.016
2-3	168083	3506	3867	0.963	0.954	0.046	0.008
3-4	160710	2073	4064	0.950	0.946	0.054	0.004
4-5	154573	869	4159	0.945	0.943	0.057	0.020
5-6	149545	981	4226	0.939	0.923	0.077	0.016
6-7	144338	4021	4088	0.913	0.907	0.093	0.007
7-8	136229	1023	4104	0.906	0.901	0.099	0.010
8-9	131102	1351	4128	0.896	0.891	0.109	0.009
9-10	125623	1347	3956	0.887	0.882	0.118	0.010
10-11	120320	885	3725	0.880	0.882	0.128	0.030
11-12	115710	504	3496	0.876	0.842	0.158	0.037
12-13	111710	8213	3630	0.812	0.805	0.195	0.000
13-14	99867	578	3720	0.807	0.805	0.195	0.014
14-15	95569	895	3412	0.800	0.791	0.209	0.015
15-16	91262	1994	3365	0.782	0.776	0.224	0.003
16-17	85903	1122	3391	0.772	0.773	0.227	0.029
17-18	81390	536	3258	0.767	0.744	0.256	0.030
18-19	77596	4918	3320	0.718	0.714	0.286	0.002
19-20	69358	369	3065	0.715	0.712	0.288	0.005
20-21	65924	679	2958	0.707	0.706	0.294	0.006
21-22	62287	194	2784	0.705	0.701	0.299	0.013
22-23	59309	331	2520	0.701	0.709	0.312	0.054
23-24	56458	378	2238	0.696	0.634	0.366	0.106
24-25	53842	12027	2212	0.541	0.528	0.472	0.003

Time	**Number of child who start breastfeeding during these interval**	**no. of child who stop breastfeeding**	**no. of child who are censored**	**S_t)**	**S_t after smoothing**	**$F(t)$ after smoothing**	**$f(t)$ after smoothing**
25-26	39603	569	2056	0.533	0.529	0.475	0.004
26-27	36978	644	2003	0.524	0.521	0.479	0.007
27-28	34331	328	1877	0.519	0.515	0.485	0.006
28-29	32126	643	1732	0.508	0.509	0.491	0.020
29-30	29751	181	1677	0.505	0.489	0.511	0.019
30-31	27893	1788	1634	0.473	0.471	0.529	0.000
31-32	24471	49	1479	0.472	0.470	0.530	0.002
32-33	22943	154	1347	0.469	0.468	0.532	0.004
33-34	21442	53	1324	0.468	0.464	0.536	0.014
34-35	20065	121	1228	0.465	0.470	0.550	0.052
35-36	18716	94	1220	0.462	0.398	0.602	0.051
36-37	17402	4019	1085	0.356	0.348	0.652	0.006
37-38	12298	52	923	0.354	0.354	0.658	0.008
38-39	11323	108	916	0.351	0.350	0.666	0.007
39-40	10299	34	780	0.350	0.327	0.673	0.001
40-41	9485	131	771	0.345	0.344	0.674	0.001
41-42	8583	17	744	0.344	0.341	0.675	0.001
42-43	7822	125	709	0.339	0.338	0.676	0.004
43-44	6988	11	660	0.338	0.338	0.680	0.002
44-45	6317	9	650	0.338	0.337	0.682	0.002
45-46	5658	40	573	0.335	0.334	0.684	0.002
46-47	5045	16	522	0.334	0.336	0.686	0.001
47-48	4507	6	531	0.334	0.318	0.687	0.001
48-49	3970	397	435	0.300	0.298	0.688	0.012
49-50	3138	5	364	0.300	0.300	0.700	0.002
50-51	2769	16	324	0.298	0.298	0.702	0.000
51-52	2429	3	340	0.298	0.297	0.703	0.001
52-53	2086	5	289	0.297	0.297	0.703	0.001
53-54	1792	1	298	0.297	0.296	0.704	0.001
54-55	1493	6	255	0.296	0.295	0.705	0.001
55-56	1232	3	256	0.295	0.294	0.706	0.001
56-57	973	5	253	0.293	0.293	0.707	0.001
57-58	715	0	244	0.293	0.292	0.708	0.002
58-59	471	1	210	0.291	0.291	0.709	0.000
59-60	260	0	211	0.290	0.290	0.710	0.290
>60	49	49	0	0	0	1	0

Table 4 Life table calculation for duration of breastfeeding for age at ≤ 60 months

Interval	N_t	W_t	B_t	m_t	P_t	Q_t	a_t
0-1	176335	1674	573	0.003	0.997	0.003	0.500
1-2	174088	3642	2363	0.013	0.984	0.016	0.499
2-3	168083	3867	3506	0.020	0.964	0.036	0.498
3-4	160710	4064	2073	0.013	0.952	0.048	0.499
4-5	154573	4159	869	0.006	0.946	0.054	0.500
5-6	149545	4226	981	0.006	0.940	0.060	0.499
6-7	144338	4088	4021	0.027	0.915	0.085	0.498
7-8	136229	4104	1023	0.007	0.908	0.092	0.499
8-9	131102	4128	1351	0.010	0.899	0.101	0.499
9-10	125623	3956	1347	0.011	0.890	0.110	0.499
10-11	120320	3725	885	0.007	0.883	0.117	0.499
11-12	115710	3496	504	0.004	0.880	0.120	0.500
12-13	111710	3630	8213	0.070	0.820	0.180	0.494
13-14	99867	3720	578	0.006	0.816	0.184	0.500
14-15	95569	3412	895	0.009	0.808	0.192	0.499
15-16	91262	3365	1994	0.021	0.791	0.209	0.498
16-17	85903	3391	1122	0.013	0.781	0.219	0.499
17-18	81390	3258	536	0.006	0.776	0.224	0.499
18-19	77596	3320	4918	0.060	0.731	0.269	0.495
19-20	69358	3065	369	0.005	0.727	0.273	0.500
20-21	65924	2958	679	0.010	0.720	0.280	0.499
21-22	62287	2784	194	0.003	0.718	0.282	0.500
22-23	59309	2520	331	0.005	0.714	0.286	0.500
23-24	56458	2238	378	0.007	0.709	0.291	0.499
24-25	53842	2212	12027	0.197	0.582	0.418	0.484
25-26	39603	2056	569	0.014	0.574	0.426	0.499
26-27	36978	2003	644	0.017	0.565	0.435	0.499
27-28	34331	1877	328	0.009	0.559	0.441	0.499
28-29	32126	1732	643	0.019	0.549	0.451	0.498
29-30	29751	1677	181	0.006	0.545	0.455	0.500
30-31	27893	1634	1788	0.060	0.513	0.487	0.495
31-32	24471	1479	49	0.002	0.512	0.488	0.500
32-33	22943	1347	154	0.006	0.509	0.491	0.499
33-34	21442	1324	53	0.002	0.508	0.492	0.500
34-35	20065	1228	121	0.006	0.505	0.495	0.500
35-36	18716	1220	94	0.005	0.503	0.497	0.500
36-37	17402	1085	4019	0.201	0.411	0.589	0.483
37-38	12298	923	52	0.004	0.409	0.591	0.500
38-39	11323	916	108	0.009	0.405	0.595	0.499
39-40	10299	780	34	0.003	0.404	0.596	0.500

Interval	N_t	W_t	B_t	m_t	P_t	Q_t	a_t
40-41	9485	771	131	0.013	0.399	0.601	0.499
41-42	8583	744	17	0.002	0.398	0.602	0.500
42-45	7822	2019	145	0.006	0.350	0.65	0.499
45-48	5658	1626	62	0.003	0.321	0.679	0.499
49-51	3970	1123	418	0.033	0.284	0.716	0.492
51-54	2429	927	9	0.001	0.254	0.746	0.500
54-57	1493	764	14	0.003	0.194	0.806	0.499
57-60	715	665	1	0.000	0.135	0.865	0.500
>60	49	0	49	0.286	0	1	

time of the survey, is used for the interval of one month for the life table analysis and in the last, take interval of three months because in interval of one month the number of child who still breastfeeding is very small due to which estimated survival is not estimated.

A spline(13,16,23) $S(X)$ is a smooth piecewise defined function whose "pieces" are low-degree polynomials defined on separate intervals of the range of x. The pieces are joined together in a suitably smooth fashion at joint points called knots. It is represented by a limited number of parameters and are smoothens the function that are extremely flexible in shape. It bridges the gap between parametric and nonparametric methods in statistics.

A large number literature presents an algorithm for calculating splines of various degrees. Cubic splines (splines of degree 3)(24,25) are often used in practice, since they are reasonably flexible in shape and reliable algorithms are available for their calculation. A simple method for calculating cubic splines, which involves rescaling the time axis to the unit interval, is given by (18) and is used in this paper. "The S_t parameters, which represent the cumulative survival function up to various time points, were considered as a function of the age of the child at the time of the survey and were represented by a cubic spline". Three knots at .25, .50, and .75 with 3 degrees were found to be sufficient for the model fitting.

Result and Discussion

The figure (2) shows the overall pattern of duration of breastfeeding by the three methods. The survival curve represents the probability of mothers who continued to breastfeed at any given time. From figure (2) it is observed that the survival curve of Kaplan Meier Estimator is fluctuated with time and these are mainly occurring in multiple of six months although the survival curve by other two methods does not fluctuate timely.

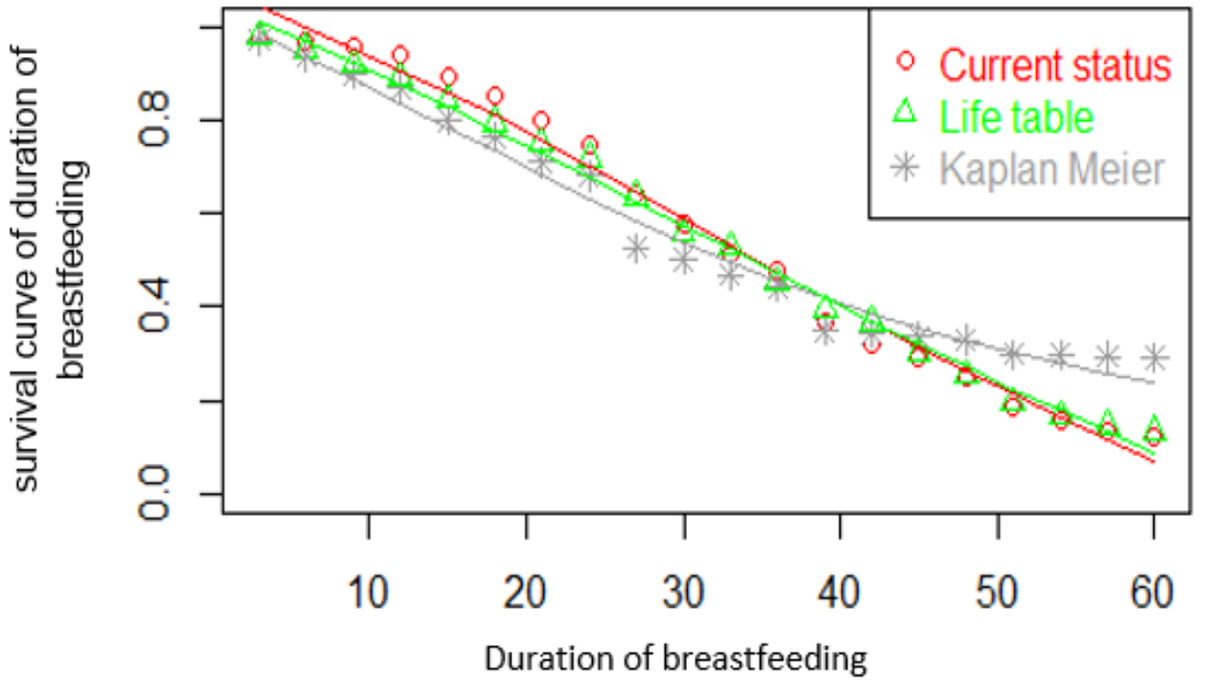

Figure 2 The survival curve of duration of breastfeeding by all method with interval of 3 months with spline smoothen

As S_t is obtained by all the three methods as shown in table (2,3,4) after that it is smoothen by spline for 60 months of duration of breastfeeding. Then the cumulative distribution $F(t)$ and distribution function $f(t)$is obtained for all the method.

For 60 months of duration of breastfeeding mean estimate by Kaplan Maier Estimator is 32.84, by Current Status Data is 33.14 and by Life Table Analysis is 33.64 then for mean duration of breastfeeding for 60 months because after 60 months there are only 13.5(%) of proportion of child still doing breastfeeding, so we assume that, S_t is zero after the 60 months of duration of breastfeeding. Quartiles are obtained by the graph. Similarly, for 36 months of duration of breastfeeding is obtained.

As shown in table 5 for 60 months the mean duration of breastfeeding is approximately equal but the median and quartiles are different by all the methods. This may occurred due to possible reasons, Kaplan Maier Estimator the information about duration of breastfeeding is depend on recall bases, the survival time is estimated at the point at which event occur, and censored event are uniformly distributed over the time however in current status data the information about duration of breastfeeding is find out by subtracting 60 months from date of interview (in CMC) and the survival time is estimated at the point at which event occur, but the roll of censored event is in interval of duration of breastfeeding is not uniformly distributed over time although in life table the information about duration of breastfeeding is also taken on recall bases but the roll of censored event is in interval of duration of breastfeeding.

The way of taking sampling frame is also very important aspect of determining the distribution of estimation of duration of

Table 5 Mean and Quartiles obtained for the duration of breastfeeding by all three methods

	duration of breastfeeding for 36 months			duration of breastfeeding for 60 months		
	Current Status Data	Life Table Analysis	Kaplan Meier Estimator	Current Status Data	Life Table Analysis	Kaplan Meier Estimator
Mean	27.08	28.56	28.33	33.14	33.64	32.84
Q_1	24.25	22.35	23.65	22.65	23.62	24.58
Q_2/Median	30.98	32.53	NA	31.64	34.26	30.00
Q_3	NA	NA	NA	46.54	46.47	59.00

Table 6-Estimation on different sampling frame

Birth's record file(Mother)								
	duration of breastfeeding for ≤36 months				duration of breastfeeding for ≤60 months			
	Mean	Q_1	Q_2/Median	Q_3	Mean	Q_1	Q_2/Median	Q_3
Current Status Data	27.08	24.25	30.98	NA	33.14	22.65	31.64	46.54
Life Table Analysis	28.56	22.35	32.53	NA	33.64	23.62	34.26	46.47
Kaplan Meier Estimator	28.33	23.65	NA	NA	32.84	24.58	30	59
Kids record file(Child)								
Current Status Data	29.58	23.56	30.58	NA	33.24	21.65	32.21	58
Life Table Analysis	29.54	21.24	29.56	36	33.58	20.59	33.56	60
Kaplan Meier Estimator	27.51	18	NA	NA	34	18	30	60

breastfeeding. If we are taking child information data from National Family Health Survey (NFHS-IV) then we collect the information all the child which was taken birth in last the five years, then determining the estimation of distribution of duration of breastfeeding. But if we take mother information data from National Family Health Survey (NFHS-IV) then we collect the information of all the women who are currently breastfeeding in the last five years, then determining the estimation of the distribution of duration of breastfeeding. Because an inverse relationship exists between birth rates and the time between the ways you are taking the data its affects the distribution.

The distribution in the form of 1- $F(t)$ (Survival function) is obtained for the duration of breastfeeding by current status, life table, and Kaplan Meier techniques. Further, the value of mean and quartiles is also obtained. The distribution is quite close in each other in life table and current status data, whereas in Kaplan Meier estimator the distribution is different (obtain by Kolmogorov smirnov test). It was found that difference in current status and life table K-S test is 0.016, for current status and Kaplan Meier estimator K-S test is 0.027 and for life table method and Kaplan Meier estimator K-S test is 0.025. This may be because there is age heaping due to recall lapse. Life table techniques method adjusts the effect of recall bias up to some extent. But the distribution obtained based on current status review very less effort to collect the data. The chance of recall bias is almost zero. Hence the distribution obtained from this technique is most feasible and appropriate in the contrast of cross sectional data. One must take care of the sampling frame while estimating such distribution as shown in table 6. Ideally, it would be appropriate to involve only those children whose age is at least 60 months. But such data will suffer from recall bias. So current status data of children whose age is 0 to 60 months should be used to evaluate duration of breastfeeding.

Conclusion

All the above mentioned three techniques are of a non-parametric approach. Although a parametric approach may also be used to evaluate the distribution under some suitable assumptions.

The current status and life table method are better than Kaplan Meier estimator as it is doesn't based on recall data and heaping present in the data. One must be very cautions while estimating the various epidemiological parameters from available data set. The assumptions of the methodology as per data available should be evaluate. If such data is not available, the available methodology may be modified. Regression analysis based on current status data technique should be used to assume the impact of various clinical and epidemiological factors (such as nutrition of mother, health status of mother etc.) on duration of breastfeeding.

Conflict of Interest

The author declares that there is not conflict of interest.

Reference

1. Recommendations O, Response WHO. Breastfeeding. 2020;1–14. Available from: https://www.who.int/news-room/facts-in-pictures/detail/breastfeeding
2. Breastfeeding [Internet]. [cited 2020 Nov 7]. Available from: https://www.who.int/news-room/facts-in-pictures/detail/breastfeeding
3. Ma S. Additive risk model with case-cohort sampled current status data. Stat Pap. 2007;48(4):595–608.
4. Jewell NP, Van Der Laan M, Lei X. Bivariate current status data with univariate monitoring times. Biometrika. 2005;92(4):847–62.
5. Smith DP. Regression analysis of " current status " life tables on duration of breastfeeding in Sri Lanka. Soc Biol. 1985;32(April 2015):90–101.
6. Agampodi SB, Agampodi TC, Kankanamge U, Piyaseeli D. Breastfeeding practices in a public health field practice area in Sri Lanka : a survival analysis. Int Breastfeed. 2007;2(1):13.
7. Augusto J, Carrazedo DA, Antonio F, Colugnati B. Exclusive breastfeeding duration and determinants among Brazilian children under two years of age Duração e determinantes do aleitamento materno exclusivo entre crianças brasileiras menores de dois anos. Rev Nutr. 2013;26(3):259–69.
8. Kasahun AW, Wako WG, Gebere MW, Neima GH. Predictors of exclusive breastfeeding duration among 6 – 12 month aged children in gurage zone , South Ethiopia : a survival analysis. Int Breastfeed J. 2017;12(20):1–9.
9. Rahman S, Rahman A. Prevalence of undernutrition in Bangladeshi children. J Biosoc Sci. 2020;52(4):596–609.
10. Robert E, Coppieters Y, Swennen B, Dramaix M. Breastfeeding Duration : A Survival Analysis — Data from a Regional Immunization Survey. Biomed Res Int. 2014;2014.
11. Abada TSJ, Trovato F, Lalu N. Determinants of breastfeeding in the Philippines : a survival analysis. Soc Sci Med. 2001;52(1):71–81.
12. Giashuddin MS, Kabir M. Duration of breast-feeding in Bangladesh. Indian J Med Res. 2004;119(June):267–72.
13. Taylor P, Grummer-strawn LM, Grummer-strawn LM. Regression Analysis of Current-Status Data : An Application to Breast-Feeding Regression Analysis of Current-Status Data : An Application to Breast-Feeding. J Am Stat Assoc. 1993;88(42):758–65.
14. Chakrabarty TK. Deconvolving Kernel Density Estimation of Right Censored Duration Data with Recall Errors. Am J Appl Math Stat. 2016;(January 2014).
15. Suchard MA. Sex, lies and self-reported counts: Bayesian mixture models for heaping in longitudinal count data via birth–death processes. Ann Appl Stat. 2015;9(2):572–96.
16. Diamond ID, Mcdonald JW. Proportional hazards models for current status data : application to the study of differentials in age at weaning in pakistan. Demography. 1986;23(4).
17. Vanderhoeft C. Accelerated failure time models: an application to cur&ent status breast-feeding data from pakistan 1. Genus. 1982;38(1):135–57.
18. Trussell J, Grummer-strawn L. Trends and Differentials in Breastfeeding Behaviour: Evidence from the WFS and DHS*. Popul Stud (NY). 1992;46(2):285–307.
19. Sheps MC. Distribution of Birth Intervals to the Sampling. Theor Popul Biol. 1972;1(1972):1–26.
20. Sheps MC, Menken JA, Ridley JC, Lingner JW. Truncation Effect in Closed

and Open Birth Interval Data. J Am Stat Assoc. 1972;65(330):678–93.
21. Wolfers D. Determinants of birth intervals and their means. Popul Stud (NY). 1968;22(2):253–62.
22. Chiang long chin. Constructing Current Life Tables. J Am Stat Assoc. 1972;67(339):538–41.
23. Wegman EJ, Wright IW. Splines in statistics. J Am Stat Assoc. 1983;78(382):351–65.
24. Rao CR. Handbook of Statistics [Internet]. Handbook o. 29, editor. elsevier; 2009. Available from: www.elsevier.com/books/sample-surveys-design-methods-and-applications/pfeffermann/978-0-444-53124-7
25. Klein JP. Handbook of Survival Analysis. Handbook of Survival Analysis. 2016.