

Suggested Modification of Previous Published Log- Incidence Mathematical Model of Breast Cancer Incidence

Jingyi Xie^{1*}

¹*Pre-Biology and Computer Science Major, University of California, Santa Barbara, Santa Barbara, California, CA, United States, 93105*

**Corresponding author. Email: jingyixie@ucsb.edu*

ABSTRACT

Pike et al. developed an outstanding mathematical model to relate the risk factors related to reproduction to the incidence of breast cancer in 1983. In 1994, Bernard Rosner and Graham A. Colditz modified Pike's model to improve some of the deficiencies such as a lack of terms for spacing of births and an inability to accommodate births after 40 years and came up with a log-incidence mathematical model. The study modifies the 1994 model to account for the effect of women's first period age on breast cancer risk, because an early period indicates a greater lifetime exposure to estrogen, and a high estrogen level is an important factor in breast cancer development. A simple modification was conducted by simply multiplying the overall equation by the increased risk correlated to the age of the first period. Due to the lack of support from the real-life dataset, the validity of such modification could not be determined; however, if it successfully applies to real-world data, a much better screening strategy could be applied to women to greatly reduce the mortality of breast cancer.

Keywords: *Breast cancer; risk factors; log-incidence mathematical model; reproductive risk; periodic risk*

1. INTRODUCTION

Breast cancer is one of the most common women's cancers in the United States, only after skin cancer. Typically, it is caused by a malignant tumor that has developed in the breast, originating in breast cancer, as a result of a failure in the cell division control center, resulting in abnormally massive division and multiplication. Numerous research related to breast cancer have been done in order to better screen and cure such diseases. Breast cancer could be detected in both men and women, but in women, the risk is significantly greater: about 1 in 8 United States women will develop invading breast cancer. but only 1 in 833 in men, which is significantly less than other cancers[1]. It is the most commonly diagnosed cancer among U.S. women: about 30% of newly diagnosed cancer in women in 2021 will be breast cancer [1]. In 2021, it is estimated that 281,550 new cases of invading breast cancer are expected to be diagnosed in U.S. Women, along with 49,290 new cases of non-invasive breast cancer, but only 2,650 new cases of invasive breast cancer are expected to be diagnosed in men [1]. For U.S. women, breast cancer death rates are

higher than those of any other cancer besides lung cancer. About 43,600 women in the U.S. are expected to die in 2021 from breast cancer [1]. As of 2021, according to the WHO, breast cancer accounts for 12% of all new annual cancer cases worldwide, becoming the most common cancer globally [1]. Nowadays, as treatments advance and screening advance, the death rate decreases by 1% each year, and the breast cancer incidence rates decrease in the year 2000, as more accurate and advanced screening methods are developed [1]. There are few risk factors related to the development of breast cancer. In this article, the age and reproductive status of women are focused on. If the relationship between age and reproductive status with breast cancer risk could be better determined and understood, then when screening, the high-risk groups could be taken more seriously and the group itself could be more cautious.

2. THE LOG-INCIDENCE MATHEMATICAL MODEL OF BREAST CANCER INCIDENCE

In 1996, an advanced log-incidence mathematical

model of breast cancer incidence based on Pike’s model in 1983 was introduced by Bernard Rosner and Graham A. Colditz. Compared with the original model, this newly developed model included terms for time intervals between pregnancies and the ability to accommodate births after age 40 years. In the new model, instead of considering factors associated with reduced risk of breast cancer as factors lower the rate of “breast tissue aging”, they transform it on its the molecular level to accumulate genetic damages in the pathway to breast cancer. They used an extension of the Pike et al. model to fit independent intending dataset between 1976 to 1990, and added a term to summarize the time intervals between births, and realized this term was significantly and effectively related to reduced breast cancer risk: the shorter time intervals between each birth, the lower the risk. This model allows for the transient increase in risk with first pregnancy to increase with age at first pregnancy and applies the model to a cohort of 89,132 middle-aged women in the Nurses’ Health Study [2]. In particular, the multiple-birth log-incidence model can be expressed as follows:

$$\ln(\text{incidence}) = \alpha + \beta_0 T_0 + \beta_1 (T^* - T_0) + \beta_2 (T - T_m)m + \beta_3 (T_1 - T_0)B_1 + \beta_4 B + \beta_5 B(T - T_m)m,$$

where T^* =minimum (age, age at menopause); B =birth index = $\sum_{i=1}^B (T^* - T_i)B_i$ = total years from each birth to a minimum (age, age at menopause) summed over all births in parous women and $B=0$ for nulliparous women. β_0 =increase in log incidence/year from birth to menarche; β_1 =increase in log incidence/year after menarche before menopause in nulliparous women or before first birth in parous women; β_2 =increase in log incidence/year after menopause in nulliparous women; $\beta_3(T_1 - T_0)$ =immediate increase in log incidence after the first birth that is assumed to increase linearly with $T_1 - T_0$; $\beta_1 + i\beta_4$ =increase in log incidence/year after menarche, before menopause for women with i births, $i=1$ to 8 ; $\beta_2 + B\beta_5$ =increase in log incidence/year after menopause in women with birth index= B . The rationale for this model is that the number of potentially malignant cells increases multiplicatively with time, but that different events in the reproductive history of a woman affect the rate of increase. This model was fit using iteratively reweighted least squares using PROC NLIN of SAS, and the parameters were easily interpreted in a relative risk (RR) context. The equations are shown below.

- $\exp(\beta_0 - \beta_1)$ = RR of breast cancer for a 1-year increase in age at menarche among nulliparous women
- $\exp(\beta_0 - \beta_1 - \beta_3)$ = RR of breast cancer for a 1-year increase in age at menarche among parous women
- $\exp(\beta_3 - \beta_4)$ = RR of breast cancer for a 1-year increase in age at first birth among parous, premenopausal women
- $\exp[\beta_3 - \beta_4 - \beta_5(t - t_m)]$ = RR of breast cancer for a 1-year increase in age at first birth among parous, postmenopausal women
- $\exp(-\beta_4)$ = RR of breast cancer for a 1-year increase in age at i th birth ($i \geq 2$) among multiparous, premenopausal women
- $\exp[-\beta_4 - \beta_5(t - t_m)]$ = RR of breast cancer for a 1-year increase in age at i th birth ($i \geq 2$) among multiparous, postmenopausal women
- $\exp(\beta_1 - \beta_2)$ = RR of breast cancer for a 1-year increase in age at menopause for nulliparous, postmenopausal women
- $\exp\{\beta_1 - \beta_2 + s[\beta_4 + \beta_5(t - t_m - 1 - \bar{h})]\}$ = RR of breast cancer for a 1-year increase in age at menopause for parous postmenopausal women, where $\bar{h} = b/s$
- $\exp[\beta_3(t_1 - t_0) + \beta_4 b]$ = RR of breast cancer for parous versus nulliparous premenopausal woman
- $\exp[\beta_3(t_1 - t_0) + b[\beta_4 + \beta_5(t - t_m)]]$ = RR of breast cancer for a parous versus nulliparous postmenopausal woman

Figure 1 : equations used in the model with terms explained.

Overall, the observed and expected relative risks by age at first birth and parity group versus nulliparous women (table 2) shows that women who had their first child at age 35 or younger tend to have a lower risk of breast cancer, and the risk would be increasing for about 10 years after, then it would drop below the risk of nulliparous women. Age matters because during pregnancy, breast cells grow rapidly; therefore, if genetically the breast cells were sickle, then this increased sickle breast cells could lead to the development of breast cancer. The likelihood of this increases with age, but the model shows that the more births a woman has had, the lower her risk of breast cancer tends to be. For nulliparous women, they have a slightly subtle higher risk. Also, after adding the term of births, it was much more clear that women who have children consecutively tend to have more benefits than women whose childbirths are spread out, partially due to the stable estrogen level. The result of the log-incidence model was pretty ideal. It provided a better fit to the data than the Pike model, with parameter estimates interpretable in terms of RRs, and it can be fitted using standard commercially available statistical software back then. However, there still existed some serious disadvantages: difficulty in translating model parameters into RR, and an incomplete fit to data that slightly

overestimated incidence for pre-menopausal women with an early age at first birth and that underestimated

incidence for post-menopausal women with a latter age at first birth [3-7].

Age, y	Age at 1st birth, y	Parity							
		1		2		3		4+	
		Observed RR (No. of cases)	Expected RR	Observed RR (No. of cases)	Expected RR	Observed RR (No. of cases)	Expected RR	Observed RR (No. of cases)	Expected RR
30-44	20-24	0.94 (7)	1.08	0.75 (55)	1.03	0.88 (71)	1.00	0.71 (44)	0.95
	25-29	1.14 (20)	1.13	1.01 (88)	1.11	0.87 (45)	1.09	0.84 (21)	1.05
	30-39	1.07 (11)	1.28	1.11 (17)	1.26	0.88 (4)	1.23	0.96 (1)	1.17
45-54	20-24	1.48 (16)	1.03	0.93 (90)	0.96	0.88 (122)	0.90	0.75 (151)	0.80
	25-29	0.92 (23)	1.12	0.87 (111)	1.05	1.02 (129)	0.99	0.67 (90)	0.90
	30-34	1.83 (30)	1.24	1.06 (38)	1.19	1.38 (30)	1.14	0.64 (6)	1.05
	35-39	1.21 (10)	1.37	1.40 (11)	1.33	2.09 (4)	1.30	2.11 (1)	1.20
55-64	20-24	0.78 (9)	0.97	0.89 (51)	0.90	0.85 (63)	0.81	0.72 (90)	0.68
	25-29	0.84 (20)	1.08	0.99 (105)	0.99	0.97 (111)	0.91	0.87 (125)	0.78
	30-34	1.44 (24)	1.21	1.43 (53)	1.13	1.21 (30)	1.05	1.08 (21)	0.94
	35-39	1.66 (16)	1.33	1.32 (13)	1.27	1.57 (7)	1.22	1.39 (2)	1.11

*No. of cases for nulliparous women was 27, 64, and 66 in age groups 30-44, 45-54, and 55-64 years.

Figure 2: observed and expected relative risks by age at first birth and parity group versus nulliparous women

3. MODEL MODIFICATION

Even though the model has improved a lot based on the previous one, as mentioned by the authors, there were still many things that were inaccurate. One thing the model did not take into consideration was the age at the first period. Giving births lowers the risk of estrogen receptor-positive type breast cancers, in which cancer cells grow as a response to the presence of the hormone estrogen [6-7], and the earlier a woman starts having menstruations, the longer her breast tissue is exposed to estrogens released during the periods; therefore, greater exposure to estrogen throughout her lifetime [8]. Starting menstruation at a young age is linked to an increase in breast cancer risk. Specifically, women whose menstruation cycles begins before the age of 11 have about 15-20 percent higher breast cancer risk compared to those who begin their menstrual cycles at age 15 or older, and for every year younger a woman was when she began her period, her breast cancer risk increased by 5 percent since more estrogen would be produced throughout their lifetime [9-10]. Also, an early first period at 11 or younger leads to an 80 percent higher chance of premature menopause, and at menopause, estrogen levels fall to very low levels [10-11]. In conclusion, starting early increases breast cancer risk, late pregnancy or nulliparous increases breast cancer risk, and post-menopause increases breast cancer risk, and all of these factors have something to do with a sudden increase in the amount of estrogen. Since estrogen level is an important factor in determining breast cancer risk, the age at first period should also be taken into consideration into such mathematical models in order to obtain a more accurate result. Since for every year younger, a woman was when she began her period, her breast cancer risk increased by 5% as mentioned above, when applying the equation to real world women, the age of their first period needed to be known. If they have their

period after the age of 15, then nothing needs to be added to the equation, since according to most researchers, having a period after the age of 15 has a negligible effect on breast cancer incidence [9]. For every year younger than the age of 15, an extra 0.05 is multiplied to the overall equation since a 5% increase in chance:

$$\ln(\text{incidence}) = (\alpha + \beta_0 T_0 + \beta_1 (T^* - T_0) + \beta_2 (T - T_m)m + \beta_3 (T_1 - T_0)B_1 + \beta_4 B + \beta_5 B (T - T_m)m)(1 + 0.5(x)), x=15\text{-age of the first period.}$$

This modified model, however, could not be applied to the original dataset, since the age of those women of their first period was not obtained. However, in the near future, as the medical records of females within the United States get more detailed and precise, this modified model could be applied to real life scenarios.

4. DISCUSSION

Since currently, there is no such dataset that accounts for breast cancer patients' first-period age, this equation could not be applied to real-world scenarios, but only to virtual statistics. In the meanwhile, in the real world, many other factors, especially unhealthy life habits such as being overweight: higher levels of all hormones, smoking: higher levels of all hormones, or drinking: higher levels of all hormones [8]. These are too complex and unstable to be included in a simple mathematical model like this one since the exact amounts of extra hormones largely depend on daily behavior. However, it is highly likely that partial differential equations could be used in such research under further study. If later, a better dataset of female breast cancer patients is collected, and this equation could be applied, further conclusions could be drawn. If this modified equation gives an accurate result that matches the actual dataset, then the screening policy could be largely adjusted based on the result of this equation. A modified model/equation is in need right

now. As for now, the most common ways to diagnose breast cancer are mammography, magnetic resonance imaging, ultrasound, computerized tomography, positron emission tomography and biopsy. However, these techniques have a lot of limitations: being expensive, time-consuming, and not suitable for women of all ages; therefore, if we could assign the appropriate method of detecting breast cancer incidence to those high-risk groups (girls who have their period at an early age, for example) as part of their health insurance, the detection of breast cancer at an early stage would be much easier, and the death rates would greatly decrease long-term at the lowest cost [12]. If this equation is not accurate in real life, then the relationship between the age of the first period, first pregnancy, and age of menopause needed to be further studied in order to find the balance between these risk factors. It is necessary to take this serious.

5. CONCLUSION

In conclusion, without the collection of a dataset that accounts for breast cancer patients' first period age, this mathematical model is not applicable in real life; however, despite the fact that it seems useless under the current circumstance, as the awareness of the importance of first period age rises, and the record of first period age becomes common and required, this model would become really useful in preventing or diagnosing gynecological diseases and/or diseases that are common in females. It would be likely that in the future, after such awareness spreads, the chance of women dying from breast cancer would significantly decrease. All the technologies that are currently used in diagnosing breast cancer are prohibitive to lower-class people. If a more confidential method of determining a woman's risk of developing breast cancer existed, it would be more likely that such a program could be included in the health examination for high-risk groups covered by insurance.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to Prof.T.J, lifetime professor at University of Michigan, Ann Arbor, for providing me with an opportunity to learn mathematical modeling in biology and biomedical sciences and inspiring me to choose the most appropriate topic I worked on.

REFERENCES

- [1]Breastcancer.org. (2021) U.S. breast cancer statistics. https://www.breastcancer.org/symptoms/understand_bc/statistics.
- [2]Rosner B, Colditz GA. (1996) Nurses' Health Study: Log-Incidence Mathematical Model of Breast Cancer Incidence. *Journal of the National Cancer Institute*, Volume 88, Issue 6:359–364.
- [3]Rosner B, Colditz GA, Willett WC. (1994) Reproductive risk factors in a prospective study of breast cancer: the Nurses' Health Study. *Am J Epidemiol*. 139: 819-835.
- [4]Lambe M, Hsieh C, Trichopoulos D, Ekblom A, Pavia M, Adami HO. (1994) Transient increase in the risk of breast cancer after giving birth. *N Engl J Med*. 331(1):5-9.
- [5]Willett WC, Tamimi R, Hankinson SE, Hazra A, Eliassen AH, Colditz GA. (2014) Chapter 18: Nongenetic Factors in the Causation of Breast Cancer, in Harris JR, Lippman ME, Morrow M, Osborne CK. *Diseases of the Breast*, 5th edition, Lippincott Williams & Wilkins.
- [6]Nichols HB, Schoemaker MJ, Cai J, et al. (2019) Breast cancer risk after recent childbirth: a pooled analysis of 15 prospective studies. *Ann Intern Med*. 170(1):22-30.
- [7]Fortner RT, Sisti J, Chai B, et al. (2019) Parity, breastfeeding, and breast cancer risk by hormone receptor status and molecular phenotype: results from the Nurses' Health Studies. *Breast Cancer Res*, 21(1):40.
- [8]Key TJ, Appleby PN, Reeves GK, et al. (2011) For the Endogenous Hormones and Breast Cancer Collaborative Group. Circulating sex hormones and breast cancer risk factors in postmenopausal women: reanalysis of 13 studies. *Br J Cancer*, 105(5):709-22.
- [9]Ritte R, Tikk K, Lukanova A, et al. (2013) Reproductive factors and risk of hormone receptor positive and negative breast cancer: a cohort study. *BMC Cancer*, 13:584.
- [10]Li K, Anderson G, Viallon V, et al. (2008) Risk prediction for estrogen receptor-specific breast cancers in two large prospective cohorts. *Breast Cancer Res*, 20(1):147.
- [11]Boddy J. (2017) An early first menstrual period may lead to premature menopause. *NPR*. 10(1):1.
- [12]Santiago-Montero R, Sossa H, Gutiérrez-Hernández DA, Zamudio V, Hernández-Bautista I, Valadez-Godínez S. (2020) Novel Mathematical Model of Breast Cancer Diagnostics Using an Associative Pattern Classification. *Diagnostics*. 10(3):136.