การเปรียบเทียบตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกสำหรับการรอดชีพของผู้ป่วยมะเร็งเต้านมที่มีการแพร่กระจายไปยังต่อมน้ำเหลือง 1-3 ต่อม

Comparison of Cox Regression and Parametric Models for Survival of Breast Cancer Patients with 1-3 Positive Lymph Nodes

ณัฐวิชญ์ ทองเพชร[1*], พิษณุ เจียวคุณ[2], วลัยทิพย์ บุญญาติศัย[2] และ อิ่มใจ ชิตาพนารักษ์[3]

Nuttawich Thongphet[1*], Phisanu Chiawkhun[2], Walaithip Bunyatisai[2] and Imjai Chitapanarux[3]

[1]*บัณฑิตวิทยาลัย มหาวิทยาลัยเชียงใหม่*

[2]*ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่*

[3]*รังสีวิทยาและมะเร็งวิทยา คณะแพทยศาสตร์ มหาวิทยาลัยเชียงใหม่*

[1]*Graduate School, Chiang Mai University*

[2]*Department of Statistics, Faculty of Science, Chiang Mai University*

[3]*Division of Therapeutic Radiology and Oncology, Faculty of Medicine, Chiang Mai University*

Received : 13 October 2018

Revised : 27 November 2018

Accepted : 16 January 2019

บทคัดย่อ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพระหว่างตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูลและการแจกแจงล็อกลอจิสติก โดยประยุกต์ใช้กับข้อมูลผู้ป่วยมะเร็งเต้านมที่มีการแพร่กระจายไปยังต่อมน้ำเหลือง 1-3 ต่อม จำนวนทั้งหมด 90 ราย ที่เข้ารับการรักษา ณ คณะแพทยศาสตร์ มหาวิทยาลัยเชียงใหม่ ตั้งแต่ปี พ.ศ. 2544 ถึงปี พ.ศ. 2550 ตัวแปรอิสระที่ใช้ในการศึกษา ได้แก่ ระดับของเนื้องอก ขนาดของเนื้องอก จำนวนต่อมที่ได้รับการประเมิน การหมดประจำเดือน เอสโตรเจนรีเซพเตอร์ โปรเจสเตอโรนรีเซพเตอร์ การได้รับรังสีรักษา การได้รับยาเคมีบำบัด สูตรยาในการให้เคมีบำบัด และการรักษาด้วยฮอร์โมน เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพระหว่างตัวแบบคือ เกณฑ์สารสนเทศของอะกะอิเกะ (เอไอซี) ผลการศึกษาเมื่อใช้การวิเคราะห์ตัวแปรเดียวด้วยตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริก พบว่า ขนาดของเนื้องอก การได้รับรังสีรักษา และการรักษาด้วยฮอร์โมนมีอิทธิพลต่อระยะเวลาการรอดชีพอย่างมีนัยสำคัญทางสถิติ เมื่อใช้การวิเคราะห์หลายตัวแปร พบว่า ขนาดของเนื้องอกมีอิทธิพลต่อระยะเวลาการรอดชีพอย่างมีนัยสำคัญทางสถิติสำหรับตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูลและการแจกแจงล็อกลอจิสติก การรักษาด้วยฮอร์โมนมีอิทธิพลต่อระยะเวลาการรอดชีพอย่างมีนัยสำคัญทางสถิติสำหรับตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูล เมื่อพิจารณาจากค่าเอไอซี พบว่า ตัวแบบการถดถอยค็อกซ์มีความเหมาะสมกับชุดข้อมูลมากที่สุดเนื่องจากมีค่าเอไอซีต่ำที่สุด

คำสำคัญ : ตัวแบบการถดถอยค็อกซ์, ตัวแบบพาราเมตริก, การแจกแจงไวบูล, การแจกแจงล็อกลอจิสติก

*Corresponding author. E-mail : nuttawich.th@gmail.com

Abstract

The purpose of this study was to compare the performance of Cox regression and two parametric models under weibull and log-logistic distributions by applying with 90 breast cancer patients with 1-3 positive lymph nodes treated at the Faculty of Medicine, Chiang Mai University from 2001 to 2007. The independent variables to be studied were as follows: tumor grade, tumor size, number of examined nodes, menopause, estrogen receptor, progesterone receptor, radiotherapy, chemotherapy, regimen and endocrine therapy. The Akaike information criterion (AIC) was used for comparing model efficiency. The study results of univariate analysis with Cox regression and parametric models indicated tumor size, radiotherapy, and endocrine therapy were statistically significant effect on survival time. For multivariate analysis, it showed that tumor size was the statistically significant factor on survival time for both parametric models, under weibull and log-logistic distributions, while the endocrine therapy was the statistically significant effect on survival time for Cox regression and the parametric model under weibull distribution. Based on AIC, the Cox regression model was the best appropriate model with the smallest value of AIC.

Keywords : Cox regression model, Parametric model, Weibull distribution, Log-logistic distribution

บทนำ

การวิเคราะห์การรอดชีพเป็นการติดตามการเกิดเหตุการณ์ในช่วงระยะเวลาใดเวลาหนึ่งเพื่อสังเกตว่าเหตุการณ์ที่สนใจจะเกิดขึ้นหรือไม่ ซึ่งต้องมีการกำหนดวันเริ่มต้นและวันสิ้นสุดของการศึกษา การวิเคราะห์การรอดชีพด้วยตัวแบบความเสี่ยง (Hazard model) เป็นวิธีการศึกษาความสัมพันธ์ระหว่างฟังก์ชันความเสี่ยง (Hazard function) กับตัวทำนายเพื่อหาปัจจัยที่ส่งผลต่อฟังก์ชันความเสี่ยง การวิเคราะห์ด้วยตัวแบบความเสี่ยงนั้นจะใช้หลักการของการวิเคราะห์การถดถอย (Singer & Willett, 1991) โดยตัวแบบความเสี่ยงที่นิยมใช้ คือ ตัวแบบการถดถอยค็อกซ์ (Cox regression model) และตัวแบบพาราเมตริก (Parametric model)

ตัวแบบการถดถอยค็อกซ์ เป็นตัวแบบกึ่งพาราเมตริก (Semi-parametric model) ใช้เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระกับระยะเวลาการรอดชีพ โดยการประมาณความเสี่ยงต่อการเกิดเหตุการณ์ที่สนใจ ณ เวลา t และสร้างตัวแบบความสัมพันธ์ระหว่างระยะเวลาการรอดชีพกับตัวแปรอิสระ ตัวแบบการถดถอยค็อกซ์มีข้อตกลงเบื้องต้น คือ ฟังก์ชันความเสี่ยงเป็นสัดส่วนคงที่ตลอดเวลาการศึกษาและตัวแปรอิสระแต่ละตัวไม่ขึ้นกับเวลา ซึ่งในการวิเคราะห์นั้นจะไม่มีการระบุลักษณะของฟังก์ชันความเสี่ยงพื้นฐาน (Baseline hazard function)

ตัวแบบพาราเมตริกเป็นตัวแบบที่มีตัวแปรสุ่มเกี่ยวข้องกับระยะเวลาซึ่งอยู่ภายใต้การแจกแจงต่างๆ เช่น การแจกแจงแบบเลขชี้กำลัง การแจกแจงไวบูล และการแจกแจงแกมมา เป็นต้น ตัวแบบพาราเมตริกมีข้อดีคือ มีรูปแบบของการแจกแจง มีฟังก์ชันการรอดชีพ (Survival function) และฟังก์ชันความเสี่ยง อย่างครบถ้วน แต่ตัวแบบพาราเมตริกนั้นเป็นตัวแบบที่ต้องมีการระบุลักษณะของฟังก์ชันความเสี่ยงพื้นฐานตามเวลาที่เปลี่ยนไป หากระบุฟังก์ชันความเสี่ยงพื้นฐานของเวลาไม่เหมาะสมจะทำให้ได้ผลการวิเคราะห์ข้อมูลที่ผิดพลาด จากการศึกษางานวิจัยของ Zhu *et al.* (2011) ซึ่งได้เปรียบเทียบประสิทธิภาพระหว่างตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูล โดยใช้

ข้อมูลผู้ป่วยมะเร็งกระเพาะอาหารจำนวน 1,715 ราย ตั้งแต่ปี ค.ศ. 1995 ถึงปี ค.ศ. 2006 และใช้เกณฑ์สารสนเทศของอะกะอิเกะ (Akaike's information criterion) หรือเอไอซี (AIC) ในการเปรียบเทียบประสิทธิภาพระหว่างตัวแบบ พบว่าตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูลมีค่า AIC ต่ำกว่าตัวแบบการถดถอยค็อกซ์ จากการศึกษางานวิจัยของ Zulkifli *et al.* (2013) ซึ่งทำการเปรียบเทียบประสิทธิภาพระหว่างตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูล เพื่อหาตัวแบบที่มีความเหมาะสมที่สุด ซึ่งจะนำไปใช้ค้นหาตัวแปรอิสระที่อาจจะเป็นดัชนีในการพยากรณ์โรค (Prognostic factor) โดยศึกษาจากผู้ป่วยมะเร็งเต้านม จำนวน 277 ราย ตั้งแต่ปี ค.ศ. 2000 ถึงปี ค.ศ. 2008 ผลการวิเคราะห์ข้อมูล พบว่า ตัวแบบการถดถอยค็อกซ์มีค่า AIC ต่ำที่สุด และจากการศึกษางานวิจัยของ Habibi *et al.* (2018) ซึ่งได้เปรียบเทียบประสิทธิภาพระหว่างตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกภายใต้การแจกแจงต่างๆ ได้แก่ การแจกแจงแบบเลขชี้กำลัง การแจกแจงไวบูล การแจกแจงกอมเพิรตซ์ การแจกแจงล็อกนอร์มอล การแจกแจงล็อกลอจิสติก และการแจกแจงแกมมานัยทั่วไป โดยใช้ข้อมูลผู้ป่วยมะเร็งกระเพาะอาหารจำนวน 104 ราย ตั้งแต่ปี ค.ศ. 2000 ถึงปี ค.ศ. 2010 ผลการเปรียบเทียบประสิทธิภาพระหว่างตัวแบบ พบว่า ตัวแบบพาราเมตริกมีประสิทธิภาพดีกว่าตัวแบบการถดถอยค็อกซ์ โดยตัวแบบพาราเมตริกภายใต้การแจกแจงล็อกนอร์มอลและการแจกแจงล็อกลอจิสติกมีค่า AIC ต่ำที่สุด เนื่องจากตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกต่างก็เป็นตัวแบบที่มีประสิทธิภาพในการวิเคราะห์ข้อมูลการรอดชีพ ดังนั้นในการศึกษาครั้งนี้ ผู้วิจัยจึงสนใจที่จะเปรียบเทียบประสิทธิภาพระหว่างตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูลและการแจกแจงล็อกลอจิสติก ด้วยเกณฑ์ AIC โดยประยุกต์ใช้กับข้อมูลผู้ป่วยมะเร็งเต้านมซึ่งมีการแพร่กระจายไปยังต่อมน้ำเหลือง 1-3 ต่อม

วิธีดำเนินการวิจัย

1. ข้อมูลที่ใช้ในการศึกษา

ข้อมูลที่ใช้ในการศึกษาครั้งนี้เป็นข้อมูลผู้ป่วยมะเร็งเต้านมที่มีการแพร่กระจายไปยังต่อมน้ำเหลือง 1-3 ต่อมที่เข้ารับการรักษา ณ คณะแพทยศาสตร์ มหาวิทยาลัยเชียงใหม่ ตั้งแต่ปี พ.ศ. 2544 ถึง ปี พ.ศ. 2550 จำนวนทั้งหมด 90 ราย ซึ่งเป็นข้อมูลทุติยภูมิที่ได้รับการอนุมัติจริยธรรมงานวิจัยจากคณะกรรมการจริยธรรมการวิจัย (Research Ethics Committee) คณะแพทยศาสตร์ มหาวิทยาลัยเชียงใหม่ โดยมีตัวแปรอิสระที่ศึกษา 10 ตัวแปร ได้แก่

1.1 ระดับของเนื้องอก (Tumor grade) แบ่งเป็น

1.1.1 Well differentiated (WD) หมายถึง เซลล์มะเร็งมองเหมือนเซลล์ปกติ ส่วนใหญ่จะเจริญเติบโตช้า และ Moderately differentiated (MD) หมายถึง เซลล์มะเร็งจะไม่เหมือนเซลล์ปกติโดยเซลล์จะติดกันและเจริญเติบโตเร็วกว่าเซลล์ปกติ

1.1.2 Poorly differentiated (PD) หมายถึง เซลล์มะเร็งมีรูปร่างไม่เป็นระเบียบและเซลล์ติดกันมีการเจริญเติบโตเร็วมาก

1.2 ขนาดของเนื้องอก (Tumor size) แบ่งเป็น

1.2.1 น้อยกว่าเท่ากับ 3 ซม.

1.2.2 มากกว่า 3 ซม.

1.3 จำนวนต่อมที่ได้รับการประเมิน (Number of examined nodes) แบ่งเป็น

1.3.1 น้อยกว่าเท่ากับ 15 ต่อม

1.3.2 มากกว่า 15 ต่อม

1.4 การหมดประจำเดือน (Menopause) แบ่งเป็น

1.4.1 ระยะก่อนหมดประจำเดือน (Premenopause)

1.4.2 ระยะหลังหมดประจำเดือน (Postmenopause)

1.5 เอสโตรเจนรีเซพเตอร์ (Estrogen receptor) แบ่งเป็น

1.5.1 ไม่มีตัวรับฮอร์โมนเอสโตรเจน (ER-negative)

1.5.2 มีตัวรับฮอร์โมนเอสโตรเจน (ER-positive)

1.6 โปรเจสเตอโรนรีเซพเตอร์ (Progesterone receptor) แบ่งเป็น

1.6.1 ไม่มีตัวรับฮอร์โมนโปรเจสเตอโรน (PgR-negative)

1.6.2 มีตัวรับฮอร์โมนโปรเจสเตอโรน (PgR-positive)

1.7 การได้รับรังสีรักษา (Radiotherapy) แบ่งเป็น

1.7.1 ไม่ได้รับรังสีรักษา (Radio-no)

1.7.2 ได้รับรังสีรักษา (Radio-yes)

1.8 การได้รับยาเคมีบำบัด (Chemotherapy) แบ่งเป็น

1.8.1 ไม่ได้รับยาเคมีบำบัด (Chemo-no)

1.8.2 ได้รับยาเคมีบำบัด (Chemo-yes)

1.9 สูตรยาในการให้เคมีบำบัด (Regimen) แบ่งเป็น

1.9.1 Anthracycline (A) หรือ Anthracycline & Taxane (AT)

1.9.2 Cyclophosphamide, Methotrexate, 5-FU (CMF)

1.10 การรักษาด้วยฮอร์โมน (Endocrine therapy) แบ่งเป็น

1.10.1 ไม่ได้การรักษาด้วยฮอร์โมน (Endocrine-no)

1.10.2 ได้รับการรักษาด้วยฮอร์โมน (Endocrine-yes)

และมีตัวแปรตาม คือ ฟังก์ชันความเสี่ยงต่อการเสียชีวิตของผู้ป่วยมะเร็งเต้านมที่มีการแพร่กระจายไปยังต่อมน้ำเหลือง 1-3 ต่อม ณ ช่วงเวลาหนึ่งๆ

2. ฟังก์ชันที่เกี่ยวข้องกับการวิเคราะห์การรอดชีพ

2.1 ฟังก์ชันการรอดชีพ คือ ความน่าจะเป็นของการรอดชีพ ณ เวลา t มีฟังก์ชันเป็นดังนี้

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t) \tag{1}$$

โดยที่ $S(t)$ คือ ฟังก์ชันการรอดชีพ มีค่าตั้งแต่ 0 ถึง 1, $F(t)$ คือ ฟังก์ชันการแจกแจงสะสม มีค่าตั้งแต่ 0 ถึง 1 และ t คือ ระยะเวลาการรอดชีพ

2.2 ฟังก์ชันความเสี่ยง คือ อัตราการเกิดเหตุการณ์ที่สนใจ ณ เวลา t มีฟังก์ชัน เป็นดังนี้

$$h(t) = \frac{f(t)}{S(t)} = \lim_{\Delta t \to 0} \frac{P(t < T \le t + \Delta t | T > t|)}{\Delta t} \tag{2}$$

โดยที่ $h(t)$ คือ ฟังก์ชันความเสี่ยง, t คือ จุดเริ่มต้นของเวลาที่สนใจในการศึกษา, T คือ จุดสิ้นสุดของเวลาที่ทำการศึกษา และ Δt คือ การไม่เกิดเหตุการณ์ในช่วงเวลา $t - T$

3. ตัวแบบการถดถอยค็อกซ์

Cox (1972) ได้นำเสนอตัวแบบกึ่งพาราเมตริกสำหรับการประมาณความเสี่ยงต่อการเกิดเหตุการณ์ที่สนใจ ณ เวลา t และสร้างตัวแบบความสัมพันธ์ระหว่างระยะเวลาการรอดชีพกับตัวแปรอิสระ โดยตัวแปรอิสระเป็นได้ทั้งตัวแปรเชิงปริมาณหรือตัวแปรเชิงกลุ่มที่ไม่ขึ้นกับเวลา ถ้ามีตัวแปรอิสระ p ตัว คือ $X_1, X_2, ..., X_p$ สามารถเขียนตัวแบบได้ดังนี้

$$h(t) = h_0(t) \cdot \exp\left(\beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p\right) \tag{3}$$

โดยที่ $h(t)$ คือ อัตราการเกิดเหตุการณ์ที่สนใจ ณ เวลา t, $h_0(t)$ คือ ฟังก์ชันความเสี่ยงพื้นฐาน เมื่อให้ค่า X ทุกตัวเป็นศูนย์ และ $\beta_1, \beta_2, ..., \beta_p$ คือ สัมประสิทธิ์การถดถอย (Regression coefficient) ของตัวแบบ

4. ตัวแบบพาราเมตริก

4.1 ภายใต้การแจกแจงไวบูล

การแจกแจงไวบูล (Weibull distribution) เป็นการแจกแจงของตัวแปรสุ่มแบบต่อเนื่องที่เสนอโดยนักคณิตศาสตร์ชาวสวีเดนชื่อ Waloddi Weibull ในปี ค.ศ. 1951 การแจกแจงไวบูลเป็นการแจกแจงหนึ่งที่ใช้ศึกษาเกี่ยวกับข้อมูลชั่วชีวิต (Lifetime data) ซึ่งจะใช้อธิบายเหตุการณ์ต่างๆ ในกรณีที่อัตราความเสี่ยงมีค่าเพิ่มขึ้นหรือลดลง การแจกแจงไวบูลถูกใช้อย่างแพร่หลายในหลายสาขาวิชา โดยเฉพาะทางวิศวกรรมศาสตร์และทางการแพทย์ เช่น ใช้อธิบายเกี่ยวกับการทดสอบอายุการใช้งานของอุปกรณ์ การศึกษาระยะเวลาการเกิดของเนื้องอกในมนุษย์ ระยะเวลาการเติบโตของเซลล์มะเร็ง ระยะเวลาการแพร่ระบาดของโรค และระยะเวลาที่รอดชีวิตของผู้ป่วยจากการได้รับการรักษาทางการแพทย์ เป็นต้น

ถ้าตัวแปรสุ่ม T มีการแจกแจงไวบูล ที่มี 2 พารามิเตอร์ คือ α และ β จะมีฟังก์ชันความหนาแน่นความน่าจะเป็น ฟังก์ชันการรอดชีพ และฟังก์ชันความเสี่ยง ดังนี้

$$f(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta - 1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] \qquad ;t \ge 0, \alpha > 0, \beta > 0 \tag{4}$$

$$S(t) = \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] \qquad ;t \geq 0, \alpha > 0, \beta > 0 \tag{5}$$

$$h(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} \qquad ;t \geq 0, \alpha > 0, \beta > 0 \tag{6}$$

โดยที่ α คือ พารามิเตอร์บ่งขนาด (Scale parameter), β คือ พารามิเตอร์บ่งรูปร่าง (Shape parameter) และ t คือ ตัวแปรสุ่มแทนระยะเวลาการรอดชีพ

4.2 ภายใต้การแจกแจงล็อกลอจิสติก

การแจกแจงล็อกลอจิสติก (Log-logistic distribution) เป็นการแจกแจงความน่าจะเป็นแบบต่อเนื่องสำหรับตัวแปรสุ่มแบบไม่เป็นลบ และเป็นการแจกแจงหนึ่งที่ใช้ในการวิเคราะห์การรอดชีพสำหรับกรณีที่เหตุการณ์ที่สนใจมีอัตราการเกิดขึ้นสูงในช่วงเริ่มต้นและลดลงในช่วงเวลาต่อมา เช่น อัตราการเสียชีวิตจากมะเร็งก่อนที่จะได้รับการวินิจฉัยหรือได้รับการรักษา เป็นต้น

ถ้าตัวแปรสุ่ม T มีการแจกแจงล็อกลอจิสติก ที่มี 2 พารามิเตอร์ คือ α และ β จะมีฟังก์ชันความหนาแน่นความน่าจะเป็น ฟังก์ชันการรอดชีพ และฟังก์ชันความเสี่ยง ดังนี้

$$f(t) = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{\left(1+(t/\alpha)^{\beta}\right)^{2}} \qquad ;t \geq 0, \beta > 0, \alpha > 0 \tag{7}$$

$$S(t) = \left[1+(t/\alpha)^{\beta}\right]^{-1} \qquad ;t \geq 0, \beta > 0, \alpha > 0 \tag{8}$$

$$h(t) = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{1+(t/\alpha)^{\beta}} \qquad ;t \geq 0, \beta > 0, \alpha > 0 \tag{9}$$

โดยที่ α คือ พารามิเตอร์บ่งขนาด, β คือ พารามิเตอร์บ่งรูปร่าง และ t คือ ตัวแปรสุ่มแทนระยะเวลาการรอดชีพ

5. เกณฑ์การเปรียบเทียบประสิทธิภาพระหว่างตัวแบบ

Akaike (1974) ได้เสนอเกณฑ์การคัดเลือกตัวแบบซึ่งสร้างจากการประมาณความแปรปรวนของสารสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback Leibler Information) ระหว่างตัวแบบที่แท้จริงกับตัวแบบที่พิจารณา ตัวแบบที่เหมาะสมที่สุดจะเป็นตัวแบบที่มีค่า AIC ต่ำที่สุด ซึ่ง AIC มีสูตรในการคำนวณดังนี้

$$AIC = -2 \log \left(L\left(\hat{\theta} \right) \right) + 2k \quad (10)$$

โดยที่ $L\left(\hat{\theta}\right)$ คือ ฟังก์ชันภาวะน่าจะเป็นสูงสุดในการประมาณค่าพารามิเตอร์ของตัวแบบที่พิจารณา และ k คือ จำนวนพารามิเตอร์ที่ต้องประมาณค่าของตัวแบบที่พิจารณา

6. วิธีดำเนินการศึกษา

ในการศึกษาครั้งนี้ ตัวแปรอิสระที่ศึกษาทั้งหมดจะเป็นตัวแปรเชิงกลุ่มซึ่งจะนำมาวิเคราะห์ด้วยการวิเคราะห์ตัวแปรเดียว (Univariate analysis) และการวิเคราะห์หลายตัวแปร (Multivariate analysis) ด้วยตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูลและการแจกแจงล็อกลอจิสติก โปรแกรมที่ใช้ในการวิเคราะห์ข้อมูลได้แก่ โปรแกรม R เวอร์ชัน 3.3.3 ซึ่งมีวิธีการดำเนินการศึกษาดังต่อไปนี้

1. วิเคราะห์ข้อมูลทั่วไปของผู้ป่วย จำนวนทั้งหมด 90 ราย โดยใช้สถิติพรรณนา

2. ตรวจสอบข้อตกลงเบื้องต้นของตัวแบบการถดถอยค็อกซ์ ซึ่งคือ ฟังก์ชันความเสี่ยงเป็นสัดส่วนคงที่ตลอดเวลาการศึกษาและตัวแปรอิสระแต่ละตัวไม่ขึ้นกับเวลา ด้วยการทดสอบความสัมพันธ์ระหว่างค่า scale Schoenfeld residuals กับระยะเวลาการรอดชีพ โดยใช้สถิติทดสอบไคกำลังสอง (Chi-square)

3. วิเคราะห์ตัวแปรอิสระต่างๆ ของผู้ป่วยมะเร็งเต้านมด้วยการวิเคราะห์ตัวแปรเดียวโดยใช้ตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกที่ระดับนัยสำคัญ 0.20 เพื่อคัดเลือกตัวแปรอิสระที่จะนำไปวิเคราะห์หลายตัวแปร

4. นำตัวแปรอิสระที่มีค่า p-value < 0.20 ในข้อ 3. มาทำการวิเคราะห์หลายตัวแปรโดยใช้ตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริก ที่ระดับนัยสำคัญ 0.05 เพื่อค้นหาตัวแปรอิสระที่มีอิทธิพลต่อระยะเวลาการรอดชีพ

5. เปรียบเทียบประสิทธิภาพระหว่างตัวแบบโดยพิจารณาจากค่า AIC

ผลการวิจัย

จากข้อมูลทั่วไปของผู้ป่วยมะเร็งเต้านม จำนวน 90 ราย พบว่า ผู้ป่วยส่วนใหญ่เป็นมะเร็งเต้านมที่กำเนิดจากเซลล์ของท่อน้ำนม ร้อยละ 95.6 มีเนื้องอกอยู่ในระดับ WD&MD ร้อยละ 65.5 มีขนาดของเนื้องอกน้อยกว่าเท่ากับ 3 ซม. ร้อยละ 64.4 มีจำนวนต่อมที่ได้รับการประเมินน้อยกว่าเท่ากับ 15 ต่อม ร้อยละ 64.4 และอยู่ในระยะหลังหมดประจำเดือน ร้อยละ 54.4 การตรวจสอบดูว่าเซลล์มะเร็งของผู้ป่วยมีตัวรับฮอร์โมนหรือไม่ โดยการตรวจจากชิ้นเนื้อมะเร็ง พบว่า ผู้ป่วยส่วนใหญ่มีตัวรับฮอร์โมนเอสโตรเจน ร้อยละ 47.8 และมีตัวรับฮอร์โมนโปรเจสเตอโรน ร้อยละ 45.5 สำหรับการรักษามะเร็งเต้านม พบว่า ผู้ป่วยได้รับรังสีรักษา ร้อยละ 44.4 ได้รับยาเคมีบำบัด ร้อยละ 82.2 โดยสูตรยาในการให้เคมีบำบัดที่พบมากที่สุดคือ CMF ร้อยละ 64.4 และได้รับการรักษาด้วยฮอร์โมนร้อยละ 71.1 สำหรับสถานะของผู้ป่วย พบว่า ผู้ป่วยโดยส่วนใหญ่ไม่มีอาการของโรค ร้อยละ 70.0 มีอาการของโรค ร้อยละ 22.2 และเสียชีวิต ร้อยละ 7.8 ดังตารางที่ 1

ตารางที่ 1 ข้อมูลทั่วไปของผู้ป่วย

ตัวแปร	จำนวน	ร้อยละ	ตัวแปร	จำนวน	ร้อยละ
1. ชนิดของมะเร็งเต้านม			7. โปรเจสเตอโรนรีเซพเตอร์		
ท่อน้ำนม	86	95.6	ไม่มีตัวรับฮอร์โมน	32	35.6
ต่อมน้ำนม	1	1.1	มีตัวรับฮอร์โมน	41	45.5
อื่นๆ	2	2.2	ไม่ระบุ	17	18.9
ไม่ระบุ	1	1.1	8. การได้รับรังสีรักษา		
2. ระดับของเนื้องอก			ไม่ได้รับ	50	55.6
WD&MD	59	65.5	ได้รับ	40	44.4
PD	14	15.6	9. การได้รับยาเคมีบำบัด		
ไม่ระบุ	17	18.9	ไม่ได้รับ	16	17.8
3. ขนาดของเนื้องอก			ได้รับ	74	82.2
น้อยกว่าเท่ากับ 3 ซม.	58	64.4	10. สูตรยาในการให้เคมีบำบัด		
มากกว่า 3 ซม.	19	21.1	A หรือ A&T	31	34.5
ไม่ระบุ	13	14.5	CMF	58	64.4
4. จำนวนต่อมที่ได้รับการประเมิน			ไม่ระบุ	1	1.1
น้อยกว่าเท่ากับ 15 ต่อม	58	64.4	11. การรักษาด้วยฮอร์โมน		
มากกว่า 15 ต่อม	32	35.6	ไม่ได้รับ	26	28.9
5. การหมดประจำเดือน			ได้รับ	64	71.1
ระยะก่อนหมดประจำเดือน	35	38.9	12. สถานะของผู้ป่วย		
ระยะหลังหมดประจำเดือน	49	54.4	มีอาการของโรค	20	22.2
ไม่แน่นอน	6	6.7	ไม่มีอาการของโรค	63	70.0
6. เอสโตรเจนรีเซพเตอร์			เสียชีวิต	7	7.8
ไม่มีตัวรับฮอร์โมน	30	33.3			
มีตัวรับฮอร์โมน	43	47.8			
ไม่ระบุ	17	18.9			

ในการวิเคราะห์ด้วยตัวแบบการถดถอยค็อกซ์นั้นจะต้องทำการตรวจสอบข้อตกลงเบื้องต้น คือ ฟังก์ชันความเสี่ยงเป็นสัดส่วนคงที่ตลอดเวลาการศึกษาและตัวแปรอิสระแต่ละตัวไม่ขึ้นกับเวลา ด้วยการทดสอบความสัมพันธ์ระหว่างค่า scale Schoenfeld residuals กับระยะเวลาการรอดชีพ โดยใช้สถิติทดสอบไคกำลังสอง ซึ่งพบว่า ตัวแปรอิสระทั้งหมดมีค่า p-value > 0.05 แสดงว่า ฟังก์ชันความเสี่ยงเป็นสัดส่วนคงที่ตลอดเวลาการศึกษาและตัวแปรอิสระแต่ละตัวไม่ขึ้นกับเวลา ดังนั้น ตัวแปรอิสระทั้งหมดจึงสามารถนำไปวิเคราะห์ด้วยตัวแบบการถดถอยค็อกซ์ได้

ตารางที่ 2 การวิเคราะห์ตัวแปรเดียวด้วยตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริก

ตัวแปร	ค็อกซ์		ไวบูล		ล็อกลอจิสติก	
	$\hat{\beta}$	p-value	$\hat{\beta}$	p-value	$\hat{\beta}$	p-value
ระดับของเนื้องอก						
WD&MD						
PD	0.29	0.8174	-0.14	0.8100	-0.18	0.7600
ขนาดของเนื้องอก						
น้อยกว่าเท่ากับ 3 ซม.						
มากกว่า 3 ซม.	1.77	0.0541*	-1.02	0.0500*	-1.04	0.0500*
จำนวนต่อมที่ได้รับการประเมิน						
น้อยกว่าเท่ากับ 15 ต่อม						
มากกว่า 15 ต่อม	-0.17	0.8375	0.17	0.7800	0.15	0.8100
การหมดประจำเดือน						
ระยะก่อนหมดประจำเดือน						
ระยะหลังหมดประจำเดือน	-0.42	0.6049	0.15	0.7900	0.19	0.7500
เอสโตรเจนรีเซพเตอร์						
ไม่มีตัวรับฮอร์โมน						
มีตัวรับฮอร์โมน	-1.37	0.2496	0.55	0.3400	0.60	0.3100
โปรเจสเตอโรนรีเซพเตอร์						
ไม่มีตัวรับฮอร์โมน						
มีตัวรับฮอร์โมน	-0.87	0.4621	0.42	0.4800	0.41	0.5000
การได้รับรังสีรักษา						
ไม่ได้รับ						
ได้รับ	-1.38	0.1378*	1.04	0.1400*	1.08	0.1300*
การได้รับยาเคมีบำบัด						
ไม่ได้รับ						
ได้รับ	-0.66	0.4573	0.44	0.5100	0.48	0.4800
สูตรยาในการให้เคมีบำบัด						
A หรือ A&T						
CMF	0.90	0.3647	-0.74	0.3200	-0.75	0.3200
การรักษาด้วยฮอร์โมน						
ไม่ได้รับ						
ได้รับ	-1.41	0.0678*	1.05	0.0640*	1.13	0.0480*

จากการวิเคราะห์ตัวแปรเดียวโดยใช้ตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูล และการแจกแจงล็อกลอจิสติก พบว่า ขนาดของเนื้องอก การได้รับรังสีรักษา และการรักษาด้วยฮอร์โมนมีอิทธิพลต่อระยะเวลาการรอดชีพอย่างมีนัยสำคัญทางสถิติ ที่ระดับนัยสำคัญ 0.20 ดังตารางที่ 2 โดยจะนำตัวแปรอิสระดังกล่าวไปวิเคราะห์หลายตัวแปร ที่ระดับนัยสำคัญ 0.05

ตารางที่ 3 การวิเคราะห์หลายตัวแปรด้วยตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริก

ตัวแปร	ค็อกซ์	ไวบูล	ล็อกลอจิสติก
	HR (CI : 95%)	HR (CI : 95%)	HR (CI : 95%)
ขนาดของเนื้องอก			
น้อยกว่าเท่ากับ 3 ซม.	1	1	1
มากกว่า 3 ซม.	7.06 (0.98 – 50.8)	0.35 (0.14 – 0.87)*	0.37 (0.15 – 0.90)*
การได้รับรังสีรักษา			
ไม่ได้รับ	1	1	1
ได้รับ	0.39 (0.04 – 3.60)	1.90 (0.68 – 5.33)	1.79 (0.64 – 5.03)
การรักษาด้วยฮอร์โมน			
ไม่ได้รับ	1	1	1
ได้รับ	0.13 (0.02 – 0.96)*	2.57 (1.05 – 6.31)*	2.42 (0.99 – 5.91)
ค่า AIC	31.99	77.35	77.49

จากการวิเคราะห์หลายตัวแปรโดยใช้ตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูล และการแจกแจงล็อกลอจิสติก ที่ระดับนัยสำคัญ 0.05 พบว่า ขนาดของเนื้องอกมีอิทธิพลต่อระยะเวลาการรอดชีพอย่างมีนัยสำคัญทางสถิติสำหรับตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูลและการแจกแจงล็อกลอจิสติก และการรักษาด้วยฮอร์โมนมีอิทธิพลต่อระยะเวลาการรอดชีพอย่างมีนัยสำคัญทางสถิติสำหรับตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูล ผลการเปรียบเทียบประสิทธิภาพระหว่างตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริก พบว่า ตัวแบบการถดถอยค็อกซ์มีความเหมาะสมกับชุดข้อมูลนี้มากที่สุด เนื่องจากเป็นตัวแบบที่มีค่า AIC ต่ำที่สุด และเมื่อพิจารณาเฉพาะตัวแบบพาราเมตริก พบว่า ตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูลและการแจกแจงล็อกลอจิสติกมีค่า AIC ใกล้เคียงกัน ดังตารางที่ 3

วิจารณ์ผลการวิจัย

ในการศึกษาข้อมูลที่เกี่ยวกับระยะเวลาการรอดชีพด้วยตัวแบบความเสี่ยงนั้น นักวิจัยสามารถเลือกใช้ตัวแบบการถดถอยค็อกซ์หรือตัวแบบพาราเมตริกในการวิเคราะห์ข้อมูล โดยตัวแบบการถดถอยค็อกซ์ เป็นตัวแบบที่ไม่ต้องระบุรูปแบบของการแจกแจงของระยะเวลาการรอดชีพ แต่จะต้องตรวจสอบข้อตกลงเบื้องต้น นั่นคือ ฟังก์ชันความเสี่ยงเป็น

สัดส่วนคงที่ตลอดเวลาการศึกษาและตัวแปรอิสระแต่ละตัวไม่ขึ้นกับเวลา เมื่อทำการวิเคราะห์ข้อมูลด้วยตัวแบบการถดถอยค็อกซ์แล้วพบว่า ข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้น จะส่งผลให้ผลการวิเคราะห์ที่ได้ไม่มีความน่าเชื่อถือ การวิเคราะห์ด้วยตัวแบบพาราเมตริกภายใต้การแจกแจงต่างๆ เช่น การแจกแจงเลขชี้กำลัง การแจกแจงไวบูล และการแจกแจงล็อกลอจิสติก เป็นต้น เป็นทางเลือกหนึ่งที่ใช้ในการวิเคราะห์ข้อมูลที่เกี่ยวกับระยะเวลาการรอดชีพ ซึ่งตัวแบบพาราเมตริกสามารถให้ผลสรุปที่มีพื้นฐานมาจากการแจกแจงของระยะเวลาการรอดชีพและไม่จำเป็นต้องตรวจสอบข้อตกลงเบื้องต้นที่เกี่ยวกับสัดส่วนของฟังก์ชันความเสี่ยง

การศึกษาในครั้งนี้ ได้ทำการเปรียบเทียบประสิทธิภาพระหว่างตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูลและการแจกแจงล็อกลอจิสติก ด้วยเกณฑ์ AIC โดยประยุกต์ใช้กับข้อมูลผู้ป่วยมะเร็งเต้านมที่มีการแพร่กระจายไปยังต่อมน้ำเหลือง 1-3 ต่อม ตัวแปรอิสระที่ใช้ในการศึกษาจะเป็นตัวแปรเชิงกลุ่มซึ่งจะนำมาวิเคราะห์ตัวแปรเดียวและหลายตัวแปรที่ระดับนัยสำคัญต่างๆ การวิเคราะห์ตัวแปรเดียวจะวิเคราะห์ที่ระดับนัยสำคัญ 0.20 เพื่อคัดเลือกตัวแปรอิสระที่จะนำไปวิเคราะห์หลายตัวแปร และการวิเคราะห์แบบหลายตัวแปรจะวิเคราะห์ที่ระดับนัยสำคัญ 0.05 เพื่อค้นหาตัวแปรอิสระที่มีอิทธิพลต่อระยะเวลาการรอดชีพของผู้ป่วยมะเร็งเต้านม และในการเปรียบเทียบประสิทธิภาพระหว่างตัวแบบต่างๆ เพื่อหาตัวแบบที่มีความเหมาะสมที่สุดนั้นจะพิจารณาจากตัวแบบที่มีค่า AIC ต่ำที่สุด จากผลการวิเคราะห์ข้อมูลทั่วไปของผู้ป่วย พบว่า เหตุการณ์ที่สนใจศึกษา ซึ่งคือ การเสียชีวิตของผู้ป่วยมะเร็งเต้านมมีจำนวนน้อย จึงทำให้มีข้อมูลไม่สมบูรณ์ทางด้านขวา (Right censoring) เป็นจำนวนมาก และยังพบว่า ตัวแปรอิสระบางตัวมีข้อมูลสูญหาย (Missing data) ซึ่งอาจทำให้ได้ตัวประมาณที่มีความเอนเอียงและทำให้ตัวแบบมีประสิทธิภาพลดลง (Bunyatisai *et al.*, 2017) จากผลการวิเคราะห์หลายตัวแปร พบว่า ตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกมีค่า $\hat{\beta}$ ไม่เป็นไปในทิศทางเดียวกัน ซึ่งการที่มีค่า $\hat{\beta}$ ที่ไม่เป็นไปในทิศทางเดียวกันนั้นจะส่งผลต่อค่าอัตราส่วนความเสี่ยงและอาจทำให้เกิดข้อผิดพลาดในการสรุปผล

สรุปผลการวิจัย

ผลการวิเคราะห์ตัวแปรเดียวโดยใช้ตัวแบบค็อกซ์และตัวแบบพาราเมตริกมีตัวแปรอิสระที่มีอิทธิพลต่อระยะเวลาการรอดชีพอย่างมีนัยสำคัญทางสถิติ ที่ระดับนัยสำคัญ 0.20 คือ ขนาดของเนื้องอก การได้รับรังสีรักษา และการรักษาด้วยฮอร์โมน ผลการวิเคราะห์หลายตัวแปรโดยใช้ตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริก ที่ระดับนัยสำคัญ 0.05 พบว่า ขนาดของเนื้องอกมีอิทธิพลต่อระยะเวลาการรอดชีพอย่างมีนัยสำคัญทางสถิติสำหรับตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูลและการแจกแจงล็อกลอจิสติก และการรักษาด้วยฮอร์โมนมีอิทธิพลต่อระยะเวลาการรอดชีพอย่างมีนัยสำคัญทางสถิติสำหรับตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูล ในการเปรียบเทียบประสิทธิภาพระหว่างตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูลและการแจกแจงล็อกลอจิสติก พบว่า ตัวแบบการถดถอยค็อกซ์มีความเหมาะสมกับชุดข้อมูลมากที่สุด เนื่องจากเป็นตัวแบบที่มีค่า AIC ต่ำที่สุด ซึ่งสอดคล้องกับงานวิจัยของ Zulkifli *et al.* (2013) ที่ทำการเปรียบเทียบระหว่างตัวแบบการถดถอยค็อกซ์และตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูล เพื่อหาตัวแบบที่มีความเหมาะสมที่สุด ซึ่งจะนำไปใช้ค้นหาตัวแปรอิสระที่อาจจะเป็นดัชนีในการพยากรณ์โรค โดยศึกษาจากข้อมูลผู้ป่วยมะเร็งเต้านม จำนวน 277 ราย ตั้งแต่ปี ค.ศ. 2000 ถึงปี

ค.ศ. 2008 จากผลการวิเคราะห์ข้อมูล พบว่า ตัวแบบการถดถอยค็อกซ์มีค่า AIC ต่ำที่สุด จึงมีความเหมาะสมกับชุดข้อมูลมากกว่าตัวแบบพาราเมตริกภายใต้การแจกแจงไวบูล

เนื่องจากตัวแบบการถดถอยค็อกซ์มีค่า AIC ต่ำที่สุด ดังนั้น ในการศึกษาครั้งนี้จะทำการสรุปผลโดยใช้ตัวแบบการถดถอยค็อกซ์ ซึ่งมีตัวแปรอิสระที่นัยสำคัญ คือ การรักษาด้วยฮอร์โมน โดยมีค่าอัตราส่วนความเสี่ยงน้อยกว่า 1 จึงสามารถสรุปได้ว่า ผู้ป่วยมะเร็งเต้านมที่มีการแพร่กระจายไปยังต่อมน้ำเหลือง 1-3 ต่อม ที่ได้รับการรักษาด้วยฮอร์โมนจะมีความเสี่ยงในการเสียชีวิตลดลง ซึ่งการรักษาด้วยฮอร์โมนนั้นเป็นการรักษามาตรฐานที่ต้องให้ทั้งในผู้ป่วยมะเร็งเต้านมระยะต้นและระยะกระจายที่มีการตอบสนองทางฮอร์โมนเป็นบวก โดยเลือกพิจารณาจากสภาวะประจำเดือนของผู้ป่วยเป็นหลัก นอกจากนี้การรักษาด้วยฮอร์โมนยังสามารถใช้เป็นการรักษาหลักแทนการใช้ยาเคมีบำบัดในผู้ป่วยบางรายที่มีความเสี่ยงต่ำได้อีกด้วย (Dechaphunkul *et al.*, 2011)

กิตติกรรมประกาศ

งานวิจัยฉบับนี้สำเร็จลุล่วงไปด้วยดี เนื่องจากผู้วิจัยได้รับความอนุเคราะห์จากบุคคลหลาย ๆ ท่าน รวมถึงโครงการพัฒนากำลังคนด้านวิทยาศาสตร์ (ทุนเรียนดีวิทยาศาสตร์แห่งประเทศไทย) ที่ให้ทุนสนับสนุนการศึกษาและการทำวิจัย ผู้วิจัยขอขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

เอกสารอ้างอิง

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans Automatic Control*, *19*, 716-723.

Bunyatisai, W., Prasitwattanaseree, S., & Ingsrisawang, L. (2017). Assessing Frailty Survival Models in Describing Variations Caused by Unobserved Covariates. *Chiang Mai J. Sci, 44,* 1191-1200.

Cox, D.R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B (Methodological)*, *34*, 187-220.

Dechaphunkul, A., Chacranon, M., Chaiwiriyawong, S., & Sunpaweravong, P. (2011). Antihormonal Therapy in Breast Cancer. *Songkla Med J*, *29*, 127-142.

Habibi, D., Rafiei, M., Chehrei, A., Shayan, Z., & Tafaqodi, S. (2018). Comparison of Survival Models for Analyzing Prognostic Factors in Gastric Cancer Patients. *Asian Pacific Journal of Cancer Prevention*, *19*, 749-753.

Singer, J.D., & Willett, J.B. (1991). Modeling the Days of Our Lives: Using Survival Analysis When Designing and Analyzing Longitudinal Studies of Duration and the Timing of Events. *Psychological Bulletin, 110,* 268-290.

Zhu, H.P., Xia, X., Yu C.H., Adnan, A., Lui, S.F., & DU, Y.K. (2011). Application of Weibull model for survival of patients with gastric cancer. *BMC Gastroenterology*, *11*, 1-6.

Zulkifli, I.Z., Haron, H., Abd Rahman, H.A., Sarbandi, M.S., & Abdullah, N.N. (2013). Identifying Prognostic Factors of Breast Cancer: Comparison between Cox Proportional Hazard and Weibull Model. In *International Symposium on Mathematical Sciences and Computing Research.* (pp. 142-6).