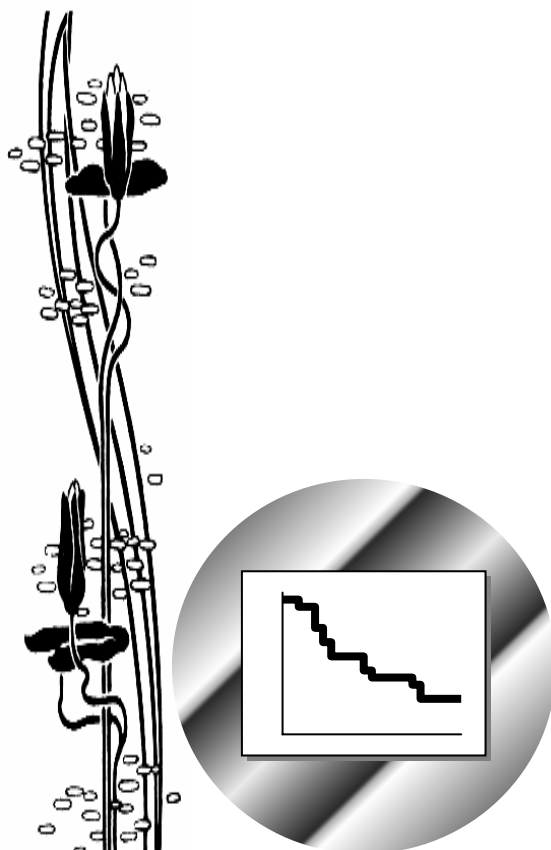


แนวปฏิบัติสำหรับ
การวิเคราะห์ระยะรอดเหตุการณ์
(A PRACTICAL GUIDE FOR SURVIVAL ANALYSIS)



ดร. บัณฑิต ถิ่นคำรพ

แนวปฏิบัติสำหรับ
การวิเคราะห์ระยะปลอดเหตุการณ์
(A PRACTICAL GUIDE FOR SURVIVAL ANALYSIS)

รองศาสตราจารย์ ดร. บัณฑิต ถิ่นคำพร
Ph.D. (Statistics)

ภาควิชาชีวสถิติและประชากรศาสตร์
คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น

ชื่อเรื่อง : แนวปฏิบัติสำหรับการวิเคราะห์ระยะปลอดเหตุการณ์
(A PRACTICAL GUIDE FOR SURVIVAL ANALYSIS)

ผู้แต่ง : บัณฑิต ถิ่นคำรพ

การพิมพ์ : ISBN : 974-654-093-9
พิมพ์ครั้งที่ 1
มกราคม 2544
จำนวน 500 เล่ม

พิมพ์ : โรงพิมพ์มหาวิทยาลัยขอนแก่น
จังหวัดขอนแก่น

คำนำ

แนวปฏิบัติสำหรับการวิเคราะห์ระยะปลอดเหตุการณ์ (A practical guide for survival analysis) นี้ ได้ผนวกไว้ซึ่งความรู้พื้นฐานที่จำเป็นพร้อมกับตัวอย่างที่มีขั้นตอนที่สอดคล้องกับการปฏิบัติในชีวิตจริง เพื่อช่วยให้ผู้อ่านที่พื้นฐานทางสถิติจำกัดได้เข้าใจหลักการ สามารถเกิดแนวคิดรวบยอดเกี่ยวกับวิธีการทางสถิตินี้ และปฏิบัติได้จริง ทั้งนี้เนื้อหาเน้นหนักด้านการประยุกต์ใช้กับสถานการณ์ที่พบบ่อยทางการแพทย์และสาธารณสุข และเลี่ยงการใช้สูตรโดยไม่จำเป็น แน่ใจว่าหนังสือเล่มนี้ไม่สมบูรณ์ในตัวเอง แต่คาดหวังว่าจะเป็นผู้อ่านจะคุ้นเคยและเข้าใจศัพท์และวิธีการเบื้องต้นในการวิเคราะห์ข้อมูลประเภทนี้ ซึ่งจะเป็จุดเริ่มต้นให้ผู้อ่านสามารถศึกษาเพิ่มเติมจากตำราที่ได้อ้างถึงในที่นี้ได้ง่ายยิ่งขึ้น

ผู้เขียนใคร่ขอกราบขอบพระคุณผู้รู้และครูบาอาจารย์ เฉพาะอย่างยิ่ง Professor Annette Dobson ที่ University of Queensland ประเทศออสเตรเลีย ที่ประสิทธิ์ประสาทวิชาความรู้แก่ผู้เขียน รวมทั้งตำราของ Hosmer and Lemeshow (1999) กับของ Kleinbaum (1996) ที่ผู้เขียนศึกษาจนเกิดความเข้าใจและนำไปใช้กับหลายงานวิจัย ขอขอบคุณนักวิจัยหลายท่านที่มีอาจกล่าวนามได้หมดสิ้นที่นำปัญหามาให้ไตร่ตรอง นำข้อมูลมาให้วิเคราะห์จนได้ประสบการณ์และมีส่วนให้เกิดหนังสือเล่มนี้ขึ้นได้ อย่างไรก็ตามความผิดพลาดอาจมีเกิดขึ้นได้นั้นเกิดจากผู้เขียนทั้งสิ้น และใคร่ขอน้อมรับเพื่อแก้ไขด้วยความยินดียิ่ง

บัณฑิต ถิ่นคำพร

มกราคม 2544

สารบัญ

หน้า

บทที่ 1

ความรู้พื้นฐานสำหรับการวิเคราะห์ข้อมูลระยะปลอดเหตุการณ์	1
1.1 การวิเคราะห์ระยะปลอดเหตุการณ์ (Survival analysis) คืออะไร?	2
1.2 คำและความหมายที่ควรทราบ	3
1.2.1 จุดเวลาเริ่มต้น (Begin date).....	3
1.2.2 เหตุการณ์ (Event).....	3
1.2.3 การวัดการเกิดเหตุการณ์.....	3
1.2.4 ประเภทของ Censoring.....	4
1.2.5 ฟังก์ชันที่เกี่ยวข้องกับ Survival analysis.....	5
1.3 เป้าหมาย Survival analysis	9
1.3.1 เพื่อประมาณค่าและพรรณนาลักษณะ $h(t)$ และ / หรือ $S(t)$	9
1.3.2 เพื่อเปรียบเทียบค่า $h(t)$ และ / หรือ $S(t)$ ระหว่างกลุ่มที่ศึกษา.....	9
1.3.3 เพื่อหาความสัมพันธ์ระหว่างปัจจัยต่างๆ กับระยะปลอดเหตุการณ์	9
1.4 สิ่งที่ต้องนำเสนอในรายงานวิจัย	11
1.4.1 สิ่งที่ต้องนำเสนอภายใต้หัวข้อระเบียบวิธีวิจัย	11
1.4.2 ข้อมูลที่ต้องนำเสนอภายใต้หัวข้อผลการศึกษา.....	11

บทที่ 2

การวิเคราะห์ข้อมูลระยะปลอดเหตุการณ์ในกลุ่มเดียว	13
2.1 ลักษณะข้อมูลและศัพท์ที่เกี่ยวข้อง	13
2.2 การวิเคราะห์ข้อมูลที่ไม่เหมาะสมที่พบบ่อย	17
2.3 การวิเคราะห์ที่เหมาะสมที่พบบ่อย.....	20

2.3.1	การวิเคราะห์โดยคำนวณเองด้วยมือ.....	20
2.3.2	การวิเคราะห์โดยใช้คอมพิวเตอร์	24
2.3.3	สรุปผลการพรรณนา Survival experience.....	33

บทที่ 3

การวิเคราะห์ข้อมูลระยะปลอดเหตุการณ์ที่มีมากกว่าหนึ่งกลุ่ม.....	34
3.1 การวิเคราะห์เปรียบเทียบระหว่างกลุ่มที่ไม่เหมาะสม.....	35
3.1.1 วิเคราะห์โดยใช้ระยะปลอดเหตุการณ์เป็นตัวแปรตาม	35
3.1.2 วิเคราะห์โดยใช้การเกิดเหตุการณ์เป็นตัวแปรตาม.....	37
3.2 การวิเคราะห์เปรียบเทียบระหว่างกลุ่มที่เหมาะสม	40
3.2.1 เปรียบเทียบ Survival experience โดยใช้ Log-rank test	40
3.2.2 คำนวณ Overall incidence rate และ Median survival time.....	45
3.2.3 คำนวณ Overall incidence rate พร้อมช่วงเชื่อมั่น.....	46
3.2.4 คำนวณช่วงเชื่อมั่นของ Median survival time.....	47
3.2.5 คำนวณ Survival probability ในแต่ละจุดเวลา พร้อมช่วงเชื่อมั่น.....	47
3.2.6 สร้างกราฟ Kaplan-Meier survival curve	48
3.2.7 Relative risk พร้อมช่วงเชื่อมั่น	49
3.2.8 อัตราส่วนของ Median survival time พร้อมช่วงเชื่อมั่น	50
3.2.9 ผลต่างของ Survival probability ระหว่างกลุ่ม	50
3.2.10 Hazard ratio พร้อมช่วงเชื่อมั่น (จาก Altman, 1991 หน้า 383).....	51
3.2.11 Cox regression analysis.....	52
3.2.12 Parametric model for survival analysis	60

IV

บทที่ 4

แนวทางการวิเคราะห์ขั้นสูง : บทสรุปเนื้อหาที่สำคัญ	64
4.1 ตรวจสอบข้อมูล	64
4.2 เปรียบเทียบลักษณะพื้นฐานของสองกลุ่มที่ศึกษา	64
4.3 ทำการ Set ข้อมูลเพื่อการวิเคราะห์ Survival analysis.....	64
4.4 พรรณนารูปแบบการศึกษา	64
4.5 พรรณนา Overall survival experience	64
4.6 ทำการวิเคราะห์อย่างหยาบ (Bivariate analysis).....	64
4.7 ทำการวิเคราะห์ตัวแปรเชิงพหุ (Multivariable analysis)	66
4.7.1 The initial multivariable model	66
4.7.2 หลักการตัดหรือคงไว้ซึ่งตัวแปร (เป้าหมายเพื่อ Simplify model)	66
4.7.5 การพิจารณา Interaction effect.....	70
4.7.6 Assessment of model adequacy.....	70
4.7.7 Interpretation and presentation of final model.....	72

บทที่ 5

ตัวอย่างรวบรวมแนวทางการวิเคราะห์ข้อมูล.....	74
ตอนที่ 1 : แนวทางการวิเคราะห์ข้อมูล.....	76
5.1 ตรวจสอบข้อมูล	76
5.2 เปรียบเทียบลักษณะพื้นฐานของสองกลุ่มที่ศึกษา	77
# การนำเสนอผลการศึกษาส่วนที่ 1	80
5.3 ทำการ Set ข้อมูลเพื่อการวิเคราะห์ Survival analysis.....	82
5.4 พรรณนารูปแบบการศึกษา	82
5.5 พรรณนา Overall survival experience	83

5.6 ทำการวิเคราะห์อย่างหยาบ (Bivariate analysis).....	86
# การนำเสนอผลการศึกษาค้นคว้าที่ 2.....	107
5.7 สร้างโมเดล (Model fitting)	109
5.7.1 The initial multivariable model	109
5.7.2 เลือกแบบเจาะจง (Purposeful selection)	
โดย Backward elimination	110
5.7.3 Scale selection กรณีตัวแปร Continuous	114
5.7.4 นำทุกตัวแปรที่ถูกตัดออกไปกลับเข้าไปใน Model อีกครั้ง.....	122
5.7.5 การพิจารณา Interaction effect.....	124
5.7.6 Assessment of model adequacy.....	131
# การนำเสนอผลการศึกษาค้นคว้าที่ 3.....	141
5.8 การสร้างกราฟของ Predicted survival curves	
จาก Cox Proportional Hazards model	150
5.9 ส่วนเพิ่มเติมการวิเคราะห์ข้อมูล.....	151
5.9.1 การวิเคราะห์เพื่อ Prediction of survival	151
5.9.2 ตัวอย่างคำสั่งสำหรับการ Fit Model ที่เป็น Parametric Model.....	155
5.9.3 การ Fit Model กรณีมี Time varying covariates	157
แบบฝึกหัด.....	162
1. การศึกษาประสิทธิผลของภูมิคุ้มกันในการรักษามะเร็ง.....	162
2. การทดลองวิธีการรักษามะเร็ง.....	166
เอกสารอ้างอิง.....	168

บัญชีตาราง

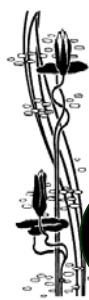
หน้า

ตารางที่ 2.1	การคำนวณ SURVIVAL PROBABILITY โดยวิธี KAPLAN-MEIER	21
ตารางที่ 3.1	การคำนวณ LOG-RANK TEST	42
ตารางที่ 5.1	ลักษณะของตัวอย่างจำแนกตามกลุ่มทดลอง.....	81
ตารางที่ 5.2	ผลของปัจจัยต่าง ๆ ต่อระยะปลอดการติดยาซ้ำ สำหรับตัวแปรแรงนับ	108
ตารางที่ 5.3	ผลของปัจจัยต่าง ๆ ต่อระยะปลอดการติดยาซ้ำ สำหรับตัวแปรต่อเนื่อง	109
ตารางที่ 5.4	ผลของปัจจัยต่าง ๆ ต่อระยะปลอดการติดยาซ้ำ	142
ตารางที่ 5.5	ผลของปัจจัยต่าง ๆ ต่อระยะปลอดการติดยาซ้ำ (ไม่มี FRACTIONAL POLYNOMIAL TERMS)	149

บัญชีภาพ

หน้า

รูปที่ 1.1	Survival curve	5
รูปที่ 1.2	ลักษณะHazard function ที่เป็น Exponential model	7
รูปที่ 1.3	ลักษณะHazard function ที่เป็น Increasing Weibull model	7
รูปที่ 1.4	ลักษณะHazard function ที่เป็น Decreasing Weibull model	8
รูปที่ 1.5	ลักษณะHazard function ที่เป็น Lognormal model.....	8
รูปที่ 1.6	ลักษณะ Survival experience.....	9
รูปที่ 1.7	Survival experience เปรียบเทียบสองกลุ่ม.....	10
รูปที่ 2.1	ลักษณะหน่วยสังเกต 6 ประเภทที่พบบ่อย.....	13
รูปที่ 2.2	สรุประยะปลอดเหตุการณ์	15
รูปที่ 2.3	คำอธิบายลักษณะข้อมูล	16
รูปที่ 2.4	รูปแบบข้อมูลสำหรับนำไปวิเคราะห์ในคอมพิวเตอร์.....	17
รูปที่ 2.5	การวิเคราะห์ข้อมูลที่ไม่เหมาะสมที่พบบ่อย.....	18
รูปที่ 2.6	การวิเคราะห์ข้อมูลที่ไม่เหมาะสมที่พบบ่อย.....	21
รูปที่ 2.7	Kaplan-Meier survival curve	22
รูปที่ 2.8	Median survival time จาก Kaplan-Meier survival curve	24
รูปที่ 2.9	การป้อนข้อมูลเข้าคอมพิวเตอร์ ชั้นที่หนึ่ง.....	25
รูปที่ 2.10	การป้อนข้อมูลเข้าคอมพิวเตอร์ ชั้นที่สอง	25
รูปที่ 2.11	การสั่งเพื่อขอดูรายการข้อมูล	27
รูปที่ 2.12	รายการข้อมูลอันเป็นผลจากการป้อนข้อมูล	28
รูปที่ 2.13	รายการข้อมูลอันเป็นผลจากคำสั่ง stset.....	29
รูปที่ 2.14	สรุปผลการพรรณนา Survival experience.....	33



ความรู้พื้นฐานสำหรับ การวิเคราะห์ข้อมูล

1

ระยะปลอดเหตุการณ์

การวิจัยโดยทั่วไปสนใจว่ามีเหตุการณ์เกิดขึ้นหรือไม่ เช่น การป่วย การหาย หรือการตาย เป็นต้น การวิเคราะห์จึงใช้ค่าร้อยละหรืออัตราของการเกิดเหตุการณ์ดังกล่าว แต่บ่อยครั้งพบว่าระยะเวลาจนกระทั่งเกิดเหตุการณ์ หรือที่เรียกกันว่าระยะปลอดเหตุการณ์มีความสำคัญมากกว่า เช่น แทนที่จะดูว่าผลการรักษามีตายหรือไม่ ก็มาดูว่ายี่ตอายุไปได้ (รอดชีพ) นานเท่าใด เป็นต้น กรณีนี้ข้อมูลมักไม่สมบูรณ์ (มี Censored observation) เนื่องจากหลายสาเหตุเช่นระยะเวลาที่วิจัยมีจำกัด ทำให้บางหน่วยตัวอย่างอาจเกิดเหตุการณ์หลังจากวันที่หยุดเก็บข้อมูล บางหน่วยตัวอย่างอาจขาดหายจากการติดตาม หรือตายจากสาเหตุอื่น เหตุการณ์ที่ศึกษาจึงยังไม่เกิดซึ่งมักพบได้บ่อยในการศึกษาทางการแพทย์และสาธารณสุข

ข้อมูลระยะปลอดเหตุการณ์นี้ แม้เป็นข้อมูลต่อเนื่องโดยธรรมชาติ แต่การวิเคราะห์ด้วยวิธีการทางสถิติธรรมดาเช่น หาค่าเฉลี่ย ทดสอบโดยใช้ t-test หรือวิเคราะห์โดย Correlation หรือ Linear regression นั้นไม่ถูกต้องเหมาะสม เนื่องจากข้อมูลไม่สมบูรณ์ แม้จะเลือกวิเคราะห์โดยใช้ค่าร้อยละหรืออัตราของการเกิดเหตุการณ์ ก็จะต้องประสิทธิภาพเพราะไม่ได้ใช้ข้อมูล

เกี่ยวกับระยะเวลา รูปแบบการวิเคราะห์ที่เหมาะสมเฉพาะกับข้อมูลประเภทนี้เรียกว่า Survival analysis ซึ่งมีวิธีการทางสถิติหลายวิธี เป็นวิธีการวิเคราะห์ที่กำลังทวีความสำคัญขึ้นเป็นลำดับ และมีการใช้ในงานวิจัยมากเป็นอันดับสองรองจาก Logistic regression

บทนี้จะนำความรู้เบื้องต้นเกี่ยวกับ Survival analysis โดยชี้ให้เห็นระหว่างการวิเคราะห์ที่ไม่เหมาะสมแต่พบได้บ่อย และสิ่งที่ถูก พร้อมเหตุผล ดังต่อไปนี้

1.1 การวิเคราะห์ระยะปลอดเหตุการณ์ (Survival analysis) คืออะไร?

ระยะปลอดเหตุการณ์ (Time to event หรือ Survival time) คือ ระยะเวลาระหว่างเมื่อเริ่มต้นจนถึงเมื่อเกิดเหตุการณ์

ค่าที่ใช้บ่อยเกี่ยวกับระยะปลอดเหตุการณ์คือ Survival time ซึ่งหมายถึงตั้งแต่เริ่มต้นการสังเกตจนถึงตาย แต่อาจหมายถึงอย่างอื่นได้เช่น

- Relapse-free period
- Remission duration
- Progression-free survival

สัญลักษณ์ที่ใช้

T = Survival time ($T \geq 0$)

t = Specific value of T

เช่นระยะปลอดเหตุการณ์มากกว่า 10 ปี เขียนเป็น $T > t=10$

1.2 คำและความหมายที่ควรทราบ

1.2.1 จุดเวลาเริ่มต้น (Begin date)

- เป็นจุดเริ่มต้นจริงของกระบวนการที่นำไปสู่การเกิดเหตุการณ์ที่ศึกษา
- ไม่จำเป็นต้องเป็นเวลาตามนาฬิกา เช่น อุณหภูมิ ความยาว น้ำหนัก
- มีความหมายเดียวกันกับทุกคน กล่าวคือทุกคนมีลักษณะคล้ายกันด้าน
พยากรณ์ของการเกิดเหตุการณ์ ณ จุดเริ่มต้น
- ตัวอย่าง เวลาเกิด วันเริ่มต้นใช้ยา วันเริ่มต้นรับปัจจัย

1.2.2 เหตุการณ์ (Event)

- การตาย การติดเชื้อ การเกิดอาการ การกลับซ้ำของอาการ
- การหายจากโรค การปฏิสนธิ การจำหน่ายจากผู้ป่วยใน
- การเลิกบุหรี่ การหย่านม
- ดำน้ำไม่เกินระยะเวลาที่อาจเกิดอันตราย
- วัดความร้อน แต่เทอร์โมมิเตอร์สามารถวัดได้สูงสุดเพียง 100°C

1.2.3 การวัดการเกิดเหตุการณ์

- เมื่อเกิดเหตุการณ์ เรียกว่า Failure
- เหตุการณ์ยังไม่เกิดเมื่อสิ้นสุดระยะเวลาที่ศึกษา เรียกว่า Censored
- Withdraw เช่น เลิกกินยาต่อเนื่อง โดยที่เหตุการณ์ยังไม่เกิด เรียกว่า
Censored
- Lost to follow-up โดยที่เหตุการณ์ยังไม่เกิด เรียกว่า Censored
- ตายเนื่องจากสาเหตุอื่นก่อนเหตุการณ์เกิด หรือ Competing event เช่น
ผู้ป่วยที่ติดตามการป่วยด้วยโรคหัวใจแต่ตาย ก่อน ด้วยอุบัติเหตุเรียกว่า
Censored
- ตัวแปรคือ Event เป็นตัวแปรสุ่ม มีค่า 1 = Failure และ 2 = Censored

พึงตระหนักว่า เพราะ Censoring จึงไม่สามารถใช้วิธีการทางสถิติที่ใช้โดยทั่วไปได้

1.2.4 ประเภทของ Censoring

1.2.4.1 Left censoring

Survival time ไม่สมบูรณ์ทางด้านซ้ายของ Follow-up period กล่าวคือ ไม่สามารถบอกได้ว่าเหตุการณ์เกิดเมื่อไร รู้เพียงว่าเกิดก่อนระยะเวลาที่เริ่มสังเกต เช่น ศึกษาระยะเวลาปลอดโรคเอดส์ในกลุ่มผู้ป่วยโรคเอดส์ เหตุการณ์ที่ศึกษาคือการติดเชื้อเอดส์ วันที่เริ่มสังเกตคือวันที่ได้รับการวินิจฉัยว่าป่วยคือแสดงอาการและผลตรวจเลือดเป็นบวก แต่ไม่ทราบวันที่ได้รับเชื้อ ข้อสังเกตคือ ระยะเวลาที่สังเกตจะยาวนานกว่าระยะปลอดเหตุการณ์ (ไม่กล่าวถึงวิธีการวิเคราะห์ในที่นี้)

1.2.4.2 Right censoring

Survival time ไม่สมบูรณ์ทางด้านขวาของ Follow-up period เป็นประเภทของ Censoring ที่พบได้บ่อยและจะกล่าวถึงตลอดคู่มือเล่มนี้ ข้อสังเกตคือจะตรงข้ามกับ Left-censoring และระยะเวลาที่สังเกตจะสั้นกว่าระยะปลอดเหตุการณ์

1.2.4.3 Interval Censoring

เหตุการณ์ที่ศึกษาสามารถเกิดได้ ณ เวลาใดเวลาหนึ่งของช่วงเวลาที่ระบุ แต่เหตุการณ์นั้นยังไม่เกิด (ถ้าเหตุการณ์เกิด ก็บอกไม่ได้ว่าเมื่อไร รู้แต่ว่าเกิด ณ เวลาใดเวลาหนึ่งของช่วงเวลาที่ระบุ) ไม่กล่าวถึงวิธีการวิเคราะห์ในที่นี้อีกเช่นกัน

การที่ระยะปลอดเหตุการณ์ไม่สมบูรณ์ นอกจาก Censoring แล้ว ยังมีสาเหตุมาจาก Truncation ได้ด้วย ผู้อ่านสามารถศึกษารายละเอียดเพิ่มเติมจาก Hosmer and Lemeshow (1999) หน้า 253 - 257

1.2.5 ฟังก์ชันที่เกี่ยวข้องกับ Survival analysis

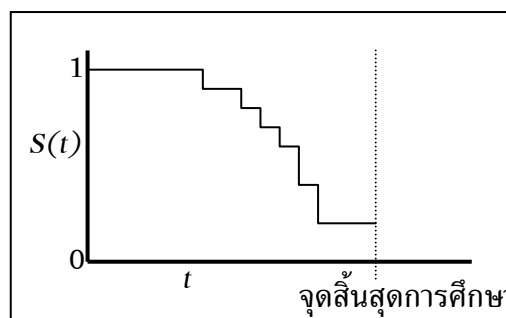
$S(t)$ คือ Survival function

$h(t)$ คือ Hazard function

1.2.5.1 Survival function

- เป็นพื้นฐานของ Survival analysis
- บอกความน่าจะเป็นที่คนคนหนึ่งยังคงปลอดเหตุการณ์มากกว่าเวลา t โดยที่ปลอดเหตุการณ์มาโดยตลอดก่อนหน้านี้ เขียนแทนด้วย $S(t) = P(T > t)$
- กล่าวอีกนัยหนึ่งคือบอกโอกาสที่คนคนหนึ่งจะเกิดเหตุการณ์หลังจากเวลา t
- $S(t)$ เป็น Step functions

รูปที่ 1.1 Survival curve



- ที่ $t = 0$ ค่า $S(t) = S(0) = 1$ เพราะยังไม่มีใครที่เกิดเหตุการณ์ กล่าวอีกนัยหนึ่ง ค่าความน่าจะเป็นของการรอดชีพ หรือของการที่ยังไม่เกิดเหตุการณ์ (probability of survival or probability of event-free) เมื่อเริ่มศึกษานั้น เท่ากับ 1
- ที่ $t = \infty$ ค่า $S(t) = S(\infty) = 0$ เพราะไม่มีใครรอด ทุกคนเกิดเหตุการณ์ กล่าวอีกนัยหนึ่ง ค่าความน่าจะเป็นของการรอดชีพ หรือของการที่ยังไม่เกิดเหตุการณ์เมื่อระยะเวลาที่ศึกษาเป็นอนันต์นั้น เท่ากับ 0

1.2.5.2 Hazard function

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

$h(t)$ ให้ค่าอัตราการเกิดเหตุการณ์ ณ จุดเวลา t โดยที่มีการปลอดเหตุการณ์จนกระทั่งถึงเวลานั้น อัตรานี้เรียกว่า Hazard rate มีค่า 0 ถึง ∞

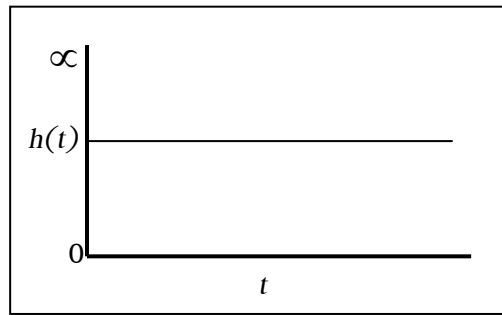
$h(t)$ ให้ค่าเกี่ยวกับการเกิดเหตุการณ์ แต่ $S(t)$ ให้ค่าเกี่ยวกับการปลอดเหตุการณ์

$h(t)$ สามารถนำเสนอเป็นกราฟได้ และมีได้หลายลักษณะ ต่อไปนี้เป็นตัวอย่าง

(1) คงที่ เรียก Exponential model

เช่น $h(t)$ ของผู้มีสุขภาพดี ไม่มีเหตุการณ์ตลอดระยะเวลาที่ศึกษา ถ้าเหตุการณ์ที่ศึกษาคือ การตาย ไม่ว่าเวลาจะเปลี่ยนไปนานเพียงใด ก็สมารถที่จะตายได้เท่าเดิม เหมือนเมื่อแรกเริ่ม

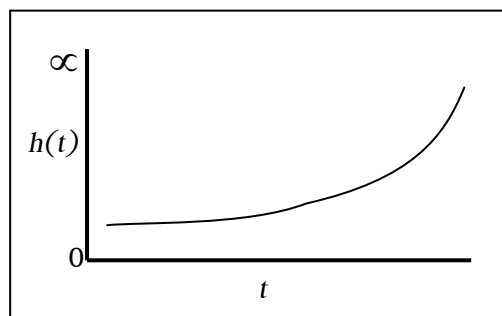
รูปที่ 1.2 ลักษณะ Hazard function ที่เป็น Exponential model



(2) เพิ่มขึ้นตามเวลา เรียก Increasing Weibull model

เช่น $h(t)$ ของผู้ป่วยโรคเอดส์ระยะสุดท้าย ถ้าเหตุการณ์ที่ศึกษาคือ การตาย ยิ่งนานวันยิ่งสามารถที่จะตายได้มากขึ้น

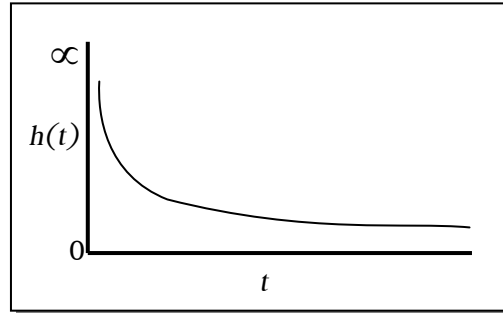
รูปที่ 1.3 ลักษณะ Hazard function ที่เป็น Increasing Weibull model



(3) ลดลงตามเวลา เรียก Decreasing Weibull model

เช่น $h(t)$ ของผู้ป่วยเอดส์ไปโรซีสหลังได้รับการรักษา ถ้าเหตุการณ์ที่ศึกษาคือ การตาย ยิ่งนานวันยิ่งสามารถที่จะตายได้น้อยลง

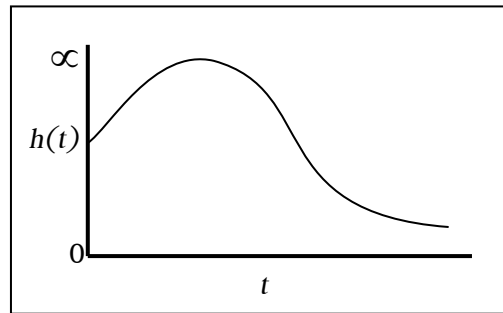
รูปที่ 1.4 ลักษณะ Hazard function ที่เป็น Decreasing Weibull model



(4) ระยะแรกเพิ่มจากนั้นลดลง เรียก Lognormal model

เช่น $h(t)$ ของคนงานก่อสร้าง ถ้าเหตุการณ์ที่ศึกษาคือ การบาดเจ็บ เมื่อเริ่มทำงานใหม่สามารถที่จะบาดเจ็บได้มาก แต่ยิ่งนานวันยิ่งสามารถที่จะบาดเจ็บได้น้อยลงเพราะมีทักษะการทำงานมากขึ้น

รูปที่ 1.5 ลักษณะ Hazard function ที่เป็น Lognormal model



1.2.5.3 ความสัมพันธ์ระหว่าง Survivor function กับ Hazard function

- $h(t)$ เป็น Mathematical model สำหรับ Survival analysis
- $h(t)$ เป็นปฏิภาคผกผันกับ $S(t)$ ตัวหนึ่งเพิ่มอีกตัวจะลด และเมื่อรู้ตัวหนึ่งก็สามารถรู้อีกตัวได้

$$H(t) = -\ln(S_{(t)}) \text{ หรือ}$$

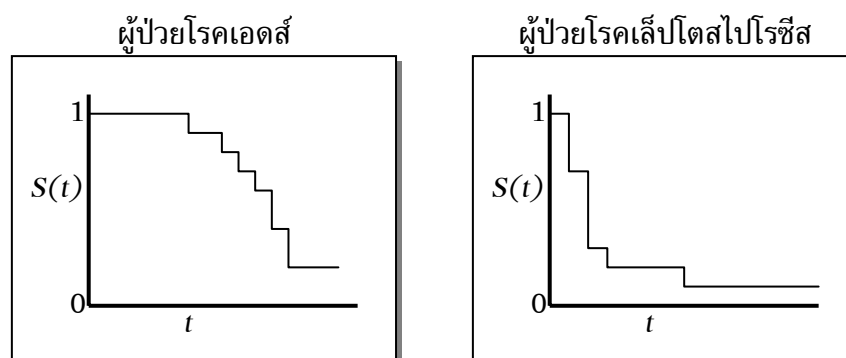
$$(S_{(t)}) = \text{EXP}(-H_{(t)})$$

1.3 เป้าหมาย Survival analysis

- 1.3.1 เพื่อประมาณค่าและพรรณนาลักษณะ $h(t)$ และ / หรือ $S(t)$
- 1.3.2 เพื่อเปรียบเทียบค่า $h(t)$ และ / หรือ $S(t)$ ระหว่างกลุ่มที่ศึกษา
- 1.3.3 เพื่อหาความสัมพันธ์ระหว่างปัจจัยต่างๆ กับระยะปลอดเหตุการณ์

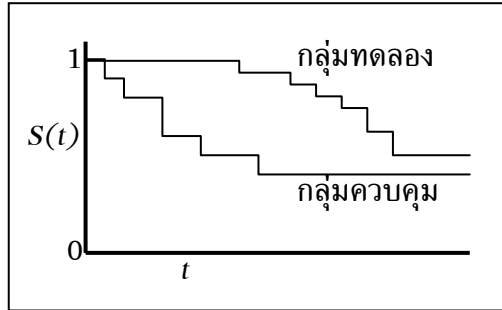
ตัวอย่างเป้าหมายที่ 1 เพื่อประมาณค่าและพรรณนาลักษณะ $h(t)$ และ / หรือ $S(t)$ รูปที่ 1.6 ซึ่งให้เห็นว่าผู้ป่วยโรคเอดส์มีโอกาสรอดชีพหลังติดเชื้อระยะยาวพอประมาณ พอถึงระยะหนึ่งโอกาสดังกล่าวจะลดลงอย่างรวดเร็ว ในขณะที่ผู้ป่วยโรคเล็ปโตสไปโรชีสนั้น โอกาสรอดชีพไม่กี่ป้นหลังติดเชื้อลดลงอย่างรวดเร็ว หลังจากนั้นระยะหนึ่งจะไม่มีเปลี่ยนแปลง

รูปที่ 1.6 ลักษณะ Survival experience



ตัวอย่างเป้าหมายที่ 2 เพื่อเปรียบเทียบค่า $h(t)$ และ / หรือ $S(t)$ ระหว่างกลุ่มที่ศึกษา

รูปที่ 1.7 Survival experience เปรียบเทียบสองกลุ่ม



ตัวอย่างเป้าหมายที่ 3 เพื่อหาความสัมพันธ์ระหว่างปัจจัยต่างๆ กับ ระยะปลอดเหตุการณ์ โดยสถิติที่เรียกว่า Cox proportional hazard model ซึ่งจะกล่าวรายละเอียดตอนท้าย

Cox regression -- Breslow method for ties

No. of subjects =	150	Number of obs =	150
No. of failures =	58		
Time at risk =	4971		
Log likelihood =	-253.02492	LR chi2(3) =	2.34
		Prob > chi2 =	0.5053

	_t	_d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
treat			.6904472	.1877797	-1.362	0.173	.4051638 1.176604
age			1.03496	.0750555	0.474	0.636	.8978302 1.193034
sex			.8033445	.5959047	-0.295	0.768	.1877145 3.438

1.4 สิ่งที่ต้องนำเสนอในรายงานวิจัย

1.4.1 สิ่งที่ต้องนำเสนอภายใต้หัวข้อระเบียบวิธีวิจัย

- 1) นิยามของ
 - จุดเริ่มต้น (Time origin)
 - เหตุการณ์ที่ศึกษา (Event)
 - กรณีที่จัดว่าเป็น Censored
- 2) อธิบายที่มาของขนาดตัวอย่าง
- 3) บอกชื่อวิธีการทางสถิติที่ใช้คำนวณค่า Survival probability
- 4) บอกชื่อวิธีการทางสถิติที่ใช้ทดสอบสมมติฐาน
- 5) ระบุชื่อตัวแปรทุกตัวที่ทำการวิเคราะห์
- 6) ถ้ามีการใช้ Cox regression analysis ให้บอกวิธีการเลือกตัวแปรเข้าใน Model เริ่มต้น กระบวนการในการได้มาซึ่ง Model สุดท้าย และอธิบายวิธีการสำหรับการตรวจสอบข้อกำหนดเบื้องต้น (Assumption) ของ Model ด้วย
- 7) ระบุชื่อโปรแกรมทางสถิติที่ใช้

1.4.2 ข้อมูลที่ต้องนำเสนอภายใต้หัวข้อผลการศึกษา

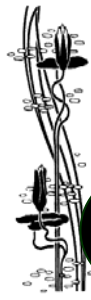
ส่วนที่ 1 ลักษณะทั่วไปทางประชากรของกลุ่มที่ศึกษา

ส่วนที่ 2 สรุปข้อมูลเบื้องต้นของ Survival analysis

- 1) Date of start accrual
- 2) Date of end accrual
- 3) Date of end follow-up และ Median follow-up time
- 4) จำนวนคนที่ lost to follow-up พร้อมอธิบายวิธีการจัดการในขั้นตอนการวิเคราะห์
- 5) จำนวนคนที่ยังคงไม่เกิดเหตุการณ์ (Number at risk) ในแต่ละช่วงเวลา

ส่วนที่ 3 ผลการวิเคราะห์เพื่อตอบคำถามวิจัย

- 1) Overall survival
 - คำนวณระยะปลอดเหตุการณ์ (Median survival time)
 - ค่าร้อยละที่ปลอดเหตุการณ์เป็นระยะเวลาที่ระบุ (Percent surviving)
- 2) Kaplan-Meier survival curve ที่มีการระบุ Number at risk ในแต่ละเวลาที่เลือก หรืออาจเสนอเส้นแสดงค่าช่วงเชื่อมั่นเฉพาะเวลาที่เลือกร่วมด้วยก็ได้
- 3) ถ้าเป็นการศึกษาแบบ Randomized controlled trial ให้รายงาน Overall survival จำแนกเป็นของแต่ละกลุ่มด้วย
- 4) ในการรายงานผลการทดสอบทางสถิติ ให้รายงานค่าสถิติ ค่า degree of freedom (ถ้ามี) และค่า p-value ที่เป็นค่าจริงที่ได้จากการทดสอบ
- 5) ถ้ามีการเปรียบเทียบ Survival ระหว่างตั้งแต่สองกลุ่มขึ้นไป ให้รายงานค่า Overall survival ในแต่ละกลุ่ม ตามด้วย Hazard ratio ช่วงเชื่อมั่นของ Hazard ratio และ p-value
- 6) ถ้ามีการใช้ Cox regression analysis ให้รายงานผล Crude hazard ratio จาก Univariate model ร่วมกับผล Adjusted hazard ratio จาก Multivariable model ตามด้วยช่วงเชื่อมั่นและ p-value



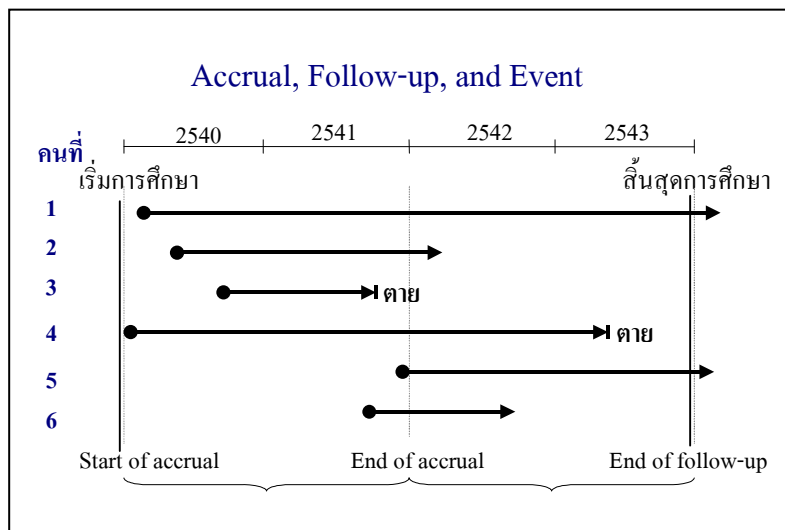
การวิเคราะห์ข้อมูล ระยะปลอดเหตุการณ์ 2 ในกลุ่มเดียว

2.1 ลักษณะข้อมูลและศัพท์ที่เกี่ยวข้อง

ในการศึกษาระยะปลอดเหตุการณ์ มีลักษณะหน่วยสังเกต 6 ประเภทที่พบบ่อยแสดงในรูปที่ 2.1

สมมติการศึกษาหนึ่ง ศึกษาการรอดชีพผู้ป่วยมะเร็งรังไข่หลังผ่าตัด เหตุการณ์ที่ศึกษาคือ “ตาย” เพื่อทราบระยะเวลารอดชีพ ได้ทำการศึกษาโดยมีระยะเวลาทั้งหมด 4 ปี จากต้นปี พ.ศ. 2540 ถึงปลายปี พ.ศ. 2543

รูปที่ 2.1 ลักษณะหน่วยสังเกต 6 ประเภทที่พบบ่อย



จากรูปที่ 2.1 ผู้ป่วยทุกคนที่เข้ารับการผ่าตัดตั้งแต่ต้นปี 2540 (Start of accrual) ถึงปลายปี 2541 (End of accrual) จะได้รับการบรรจุเข้าเป็นหน่วยศึกษา เรียกระยะเวลา 2 ปีนี้ว่า Accrual period ระยะเวลากจากผู้ป่วยรายสุดท้ายที่ได้รับการบรรจุเข้าเป็นหน่วยศึกษาจนถึงสิ้นสุดการศึกษา (End of follow-up) เรียกว่า Follow-up period ในที่นี้คือ 2 ปี นับจากปลายปี 2541 ถึงปลายปี 2543

ระยะเวลา Accrual period และ Follow-up period นี้มีความสำคัญต่อการคำนวณขนาดตัวอย่าง

ผู้ป่วยคนที่ 1 เข้ารับการผ่าตัดต้นปี 2540 จนสิ้นสุดการศึกษายังมีชีวิตอยู่ ข้อมูลระยะเวลารอดชีพรายนี้ถือเป็น Censored observation คือเราทราบว่าเขายังมีชีวิตอยู่เมื่อปลายปี 2543 ซึ่งสิ้นสุดการศึกษา เขาอาจตายไม่กี่ปวันหลังจากนั้น หรืออาจมีชีวิตอยู่ต่อไปอีกหลายปี ไม่มีข้อมูลเหล่านี้

ผู้ป่วยคนที่ 2 เข้ารับการผ่าตัดกลางปี 2540 แต่ตายเมื่อต้นปี 2542 เนื่องจากอุบัติเหตุซึ่งเป็นสาเหตุการตายอื่นที่ไม่ใช่สิ่งที่ศึกษา กรณีนี้ถือเป็น Censored observation เช่นกัน เพราะเขาอาจมีชีวิตต่อไปอีกนานถ้าหากไม่เสียชีวิตจากอุบัติเหตุเสียก่อน ในที่นี้ อุบัติเหตุเรียกว่า Competing risk

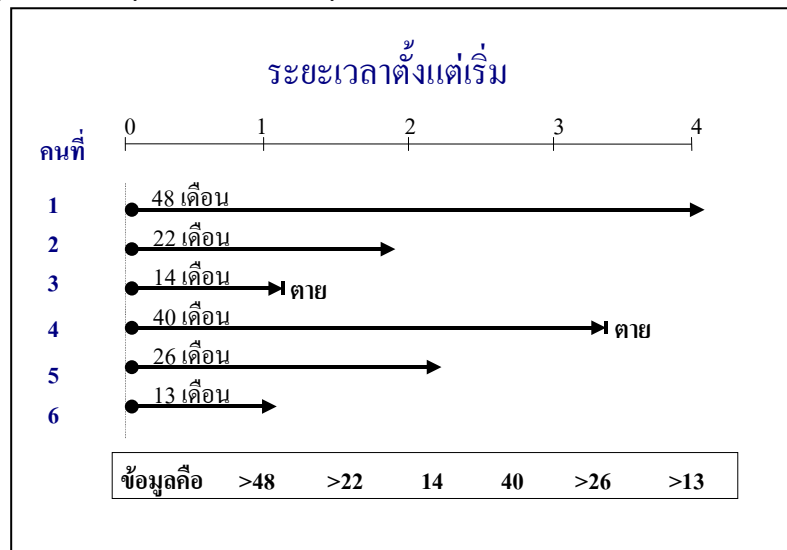
ผู้ป่วยคนที่ 3 เข้ารับการผ่าตัดต้นปี 2540 และตายจากโรคมะเร็งตับเมื่อปลายปี 2541 เช่นเดียวกับรายที่ 4 แต่รายที่ 4 นี้มีระยะเวลารอดชีพยาวกว่ามากเพราะเข้ารับการผ่าตัดก่อนและตายทีหลัง ข้อมูลระยะเวลารอดชีพของทั้งสองรายนี้ถือเป็น Complete observation

ผู้ป่วยรายที่ 5 และ 6 ได้รับการผ่าตัดปลายปี 2541 รายที่ 5 นั้นยังมีชีวิตอยู่เมื่อสิ้นสุดระยะเวลาศึกษา ส่วนรายที่ 6 ทราบว่ายังมีชีวิตอยู่เมื่อ

กลางปี 2542 แต่ไม่สามารถติดต่อกับได้หลังจากนั้น ข้อมูลระยะเวลารอดชีพของทั้งสองรายนี้ถือเป็น Censored observation เหมือนกัน แต่ระยะเวลาต่างกัน

ระยะเวลาที่ต่างกันนั้นเป็นไปได้ทั้งจุดเวลาที่เริ่มและสิ้นสุดการสังเกตต่างกัน เพื่ออำนวยความสะดวกเข้าใจเกี่ยวกับระยะเวลาดังกล่าว จึงให้นับวันเริ่มสังเกตเป็นเวลา 0 แล้วนับต่อไปจนเท่ากับระยะเวลาที่ศึกษาทั้งหมด ดังรูปที่ 2.2

รูปที่ 2.2 สรุประยะปลอดเหตุการณ์



ข้อมูลระยะปลอดเหตุการณ์ที่ได้คือ >48, >22, 14, 40, >26 และ >13 เดือน ที่ต้องมีเครื่องหมายมากกว่า (>) หน้าตัวเลขนั้น เพื่อแสดงให้เห็นว่าเป็นระยะเวลาอย่างน้อยที่ยังมีชีวิตอยู่ ไม่ใช่ระยะเวลารอดชีพ หรือเป็น Censored observation นั้นเอง

การเกิดเหตุการณ์จากตัวอย่างข้างต้นคือ “ตาย” ใน Survival analysis เรียกว่า “Failure” ซึ่งหมายถึงผู้ป่วยรายที่ 3 และ 4 เท่านั้น นอกเหนือจากนี้เรียกว่า “Censored” (รูปที่ 2.3)

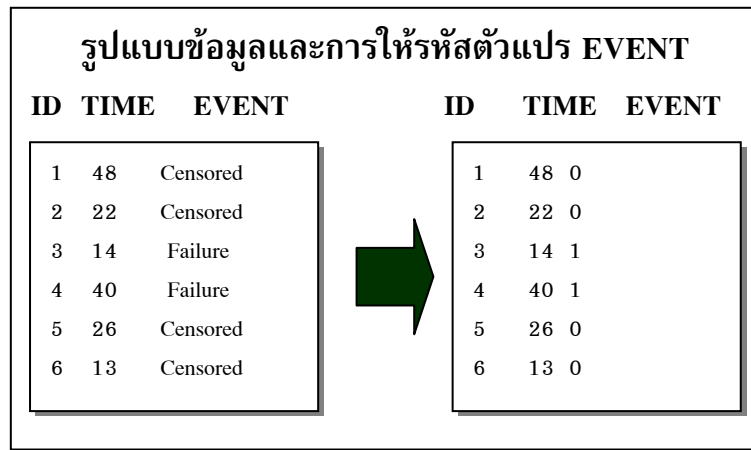
รูปที่ 2.3 คำอธิบายลักษณะข้อมูล

คนที่	ระยะปลอดเหตุการณ์ (เดือน)	ผลเมื่อวันสุดท้าย	เหตุการณ์
1	48	ยังมีชีวิตอยู่เมื่อสิ้นสุดระยะเวลาศึกษา	Censored
2	22	ตายเนื่องจากสาเหตุอื่นที่ไม่ใช่สิ่งที่ศึกษา	Censored
3	14	ตายจากสาเหตุที่ศึกษา	Failure
4	40	ตายจากสาเหตุที่ศึกษา	Failure
5	26	ยังมีชีวิตอยู่เมื่อสิ้นสุดระยะเวลาศึกษา	Censored
6	13	Lost to follow-up	Censored

ข้อมูลที่จะนำไปวิเคราะห์ต่อไปคือ “ระยะปลอดเหตุการณ์ (Event-free period)” รูปแบบข้อมูลสำหรับนำไปวิเคราะห์ในคอมพิวเตอร์ ต้องกำหนดตัวแปร ในที่นี้กำหนดให้ ID เป็นเลขที่ผู้ป่วย (Identification Number) ให้ TIME เป็นระยะปลอดเหตุการณ์ และให้ EVENT เป็นเหตุการณ์ที่ศึกษา โดยค่าเป็น 1 คือตาย (Failure) และ 0 คือ Censored (รูปที่ 2.4)

การกำหนดค่าตัวแปร EVENT ตามนี้เป็นข้อกำหนดที่ต้องทำ ส่วนค่าตัวแปรอื่นสามารถกำหนดได้อย่างอิสระ แม้กระทั่ง TIME ที่สามารถวัดหน่วยเป็นปี เดือน วัน หรือแม้กระทั่งละเอียดกว่านั้นได้

รูปที่ 2.4 รูปแบบข้อมูลสำหรับนำไปวิเคราะห์ในคอมพิวเตอร์



2.2 การวิเคราะห์ข้อมูลที่ไม่เหมาะสมที่พบบ่อย

เมื่อข้อมูลได้รับการเก็บรวบรวม และลงรหัสเรียบร้อยแล้ว ข้อมูลตัวแปร TIME ไม่มีเครื่องหมาย > หน้าตัวเลขเหมือนอย่าง que แสดงในบรรทัดท้ายสุดของรูปที่ 2.2 แล้ว แต่ถูกกำกับด้วยตัวแปร EVENT แทน จึงเป็นเหตุที่มักพบมีการวิเคราะห์ที่ไม่เหมาะสมได้บ่อย

การวิเคราะห์ข้อมูลที่ไม่เหมาะสมที่พบบ่อย มีสรุปในรูปที่ 2.5 บ้าง ก็คำนวณค่าความชุก (Prevalence) ของการตาย บ้างคำนวณอัตราการตาย (Rate) ซึ่งไม่เหมาะสม

จากที่ศึกษาทั้งหมด 6 คน ตาย 2 คน ตามความเข้าใจโดยทั่วไป แล้ว $\text{Prevalence} = 2/6$ แต่ผิดเพราะการตายขึ้นอยู่กับระยะเวลาในการ

สังเกต ควรคำนวณอัตรา แต่เมื่อคำนวณอัตราตายเป็น Rate = 2/6 ต่อ 4 ปี แต่ก็ผิดอีกเพราะระยะสังเกตของทุกคนไม่เท่ากับ 4 ปี

รูปที่ 2.5 การวิเคราะห์ข้อมูลที่ไม่เหมาะสมที่พบบ่อย

วิเคราะห์ข้อมูลอย่างไร?			
ID	TIME	EVENT	
1	48	0	
2	22	0	
3	14	1	
4	40	1	
5	26	0	
6	13	0	

Prevalence = 2/6	✗
Rate = 2/6 per 4 years	✗
Incidence density = 2/163 person-months	✓
Proportion of surviving at month 't'	✓
Median survival time	✓

เพื่อให้ถูกต้อง อัตราควรคำนวณเป็น Incidence Density ซึ่งเท่ากับ 2 หาดด้วย 163 ซึ่งได้มาจากระยะเวลาปลอดเหตุการณ์ของทุกคนรวมกัน (48+22+14+40+26+13) หน่วยจึงเป็น Person-months

แม้ Incidence Density จะไม่ผิด แต่ต้องรู้ความหมาย เช่น จากตัวอย่าง ค่าดังกล่าวนี้เท่ากับ 1.2/100 Person-months เราทราบเพียงว่า ถ้าติดตามคนไป 1000 Person-months จะมีการตายเกิดขึ้น 12 ครั้ง กล่าวอีกนัยหนึ่ง ตามคน 1000 คนไป 1 เดือน หรือตามคน 100 คนไป 10 เดือน จะมีคนตาย 12 คน เราไม่ทราบแม้กระทั่งตายช้าหรือเร็วเท่าไร หรือ

ความเสี่ยงต่อการตายไม่กี่เดือนหลังผ่าตัด กับหลาย ๆ เดือนหลังผ่าตัดเป็นอย่างไร นั่นคือที่มาของ Survival analysis

การวิเคราะห์ Survival analysis ให้องค์ความรู้ที่ค้ำึงถึงโอกาสตาย ณ จุดเวลาใด ๆ ที่สนใจได้ด้วย (Proportion of surviving at month 't') เป็นการให้องค์ความรู้เกี่ยวกับความเสี่ยงต่อการเกิดเหตุการณ์ นอกจากนั้นยังให้ทราบระยะเวลารอดชีพโดยเฉลี่ยได้ด้วย

ระยะเวลารอดชีพโดยเฉลี่ย เป็นอีกค่าหนึ่งที่มีกพบคำนวณผิตบ่อยโปรดพิจารณาการคำนวณต่อไปนี้

ก. หาค่ามัธยฐาน โดย Median = $(22+26)/2 = 24$ คือหาจากค่ากลางของลำดับข้อมูลต่อไปนี้ : 13, 14, 22, 26, 40, 48) จะเห็นว่าคำนวณถูกต้อง แต่จะถูกก็ต่อเมื่อข้อมูลไม่มี Censored เท่านั้น ค่า 24 เดือนในที่นี้ไม่สามารถนำมาสรุปเป็นระยะเวลารอดชีพโดยเฉลี่ยได้

ข. คำนวณจาก $(48+22+14+40+26+13) / 6 = 27.2$ ค่านี้คือค่าเฉลี่ยเลขคณิต (Mean) ซึ่งไม่ถูกต้องอีกเช่นกัน เพราะข้อมูลมี Censored observation

ค. คำนวณจาก $(48+22+26+13) / 4$ คือนำเอาเฉพาะผู้ยังไม่ตาย (EVENT=0) ซึ่งมี 4 คน มาคำนวณหาค่า Mean พบทำกันบ่อย แต่ผิตอย่างมาก เพราะข้อมูลเหล่านี้เป็นระยะเวลาเท่าที่สังเกต ไม่ใช่ระยะเวลาเท่าที่เขามีชีวิตอยู่

ง. คำนวณจาก $(14+40) / 2$ คล้ายข้อ ค. แต่เอาเฉพาะผู้ที่รู้ว่าตายแล้วแน่นอนมาคำนวณ ไม่ถูกต้องอีกเช่นกันเพราะไม่คำนึงถึงกลุ่มที่ยังมีชีวิตอยู่ถึง 4 คน

ค่าที่ได้จากการคำนวณทั้ง 4 วิธีที่กล่าวข้างต้น ล้วนไม่สามารถนำมาสรุปเป็นระยะเวลารอดชีพโดยเฉลี่ยได้ แต่ต้องคำนวณเป็นค่า Median survival time ตามวิธีการ Survival analysis เท่านั้น

2.3 การวิเคราะห์ที่เหมาะสมที่ใช้บ่อย

ต่อไปนี้จะกล่าวถึงวิธีการวิเคราะห์ที่ถูกต้องที่นิยมใช้กันทั่วไป โดยใช้ข้อมูลจากตัวอย่างข้างต้น ในบทนี้เป็นการวิเคราะห์ในตัวอย่างกลุ่มเดียว กรณีมากกว่าสองกลุ่มกล่าวในบทที่ 3

ในการวิเคราะห์ระยะปลอดเหตุการณ์ในตัวอย่างกลุ่มเดียว ที่เป็นที่ยอมรับกันคือการพรรณนา Survival experience โดยวิธีการของ Kaplan-Meier ในที่นี้ขอเสนอให้ดูทั้งการวิเคราะห์โดยคำนวณเองเพื่อเกิดความเข้าใจ และทำด้วยคอมพิวเตอร์เพื่อเปรียบเทียบผล และถือปฏิบัติในชีวิตจริง ดังนี้

2.3.1 การวิเคราะห์โดยคำนวณเองด้วยมือ

ขั้นแรกจัดเรียงลำดับข้อมูลเวลาก่อน จากนั้นคำนวณค่าความน่าจะเป็น หรือโอกาสในการรอดชีพ ณ จุดเวลาที่ข้อมูลมี ค่าเหล่านี้สามารถนำเสนอเป็นกราฟต่อไป เป็นลำดับดังนี้

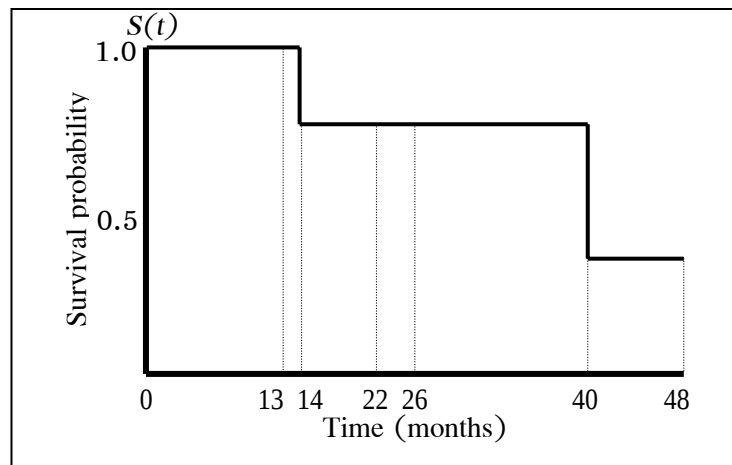
4/5 เท่ากับ 0.8 แต่ด้วยโอกาสรอดชีพดังกล่าวมีเงื่อนไขว่าต้องรอดชีพมาก่อน จึงต้องคูณด้วยค่าความน่าจะเป็นของการรอดชีพเมื่อ 13 เดือน ซึ่งเท่ากับ 1 ได้เท่ากับ 0.8 นี่คือ Conditional probability ทำเช่นนี้เรื่อยไปจนครบ

จากตารางข้างต้น ทำให้เราทราบได้ว่า โอกาสรอดชีพของผู้ป่วยมะเร็งตับเมื่อ 48 เดือนหลังการผ่าตัดเท่ากับ 0.4 กล่าวอีกนัยหนึ่ง มีเพียงร้อยละ 40 ของผู้ป่วยที่ยังคงมีชีวิตอยู่ถึงปีที่ 4 หลังผ่าตัด เราสามารถทราบค่าเหล่านี้ได้ทุกจุดเวลาที่มีข้อมูล เช่น 80% ของผู้ป่วยที่ได้รับการผ่าตัดมีโอกาสมีชีวิตอยู่ได้ถึงเดือนที่ 26 เป็นต้น

2.3.1.3 นำเสนอเป็น Kaplan–Meier survival curve

การสร้าง Kaplan–Meier Survival Curve ทำได้โดยนำค่าความน่าจะเป็นจากตาราง 2.1 นำเสนอเป็นกราฟที่มีลักษณะขั้นบันได (Survival function เป็น Step function) ดังรูปที่ 2.7

รูปที่ 2.7 Kaplan–Meier survival curve



Kaplan–Meier Survival Probability ไม่มี Assumption เกี่ยวกับการแจกแจงข้อมูล จึงเป็น Nonparametric

Assumption for Kaplan–Meier Survival Curve คือสาเหตุของ Censoring ต้องไม่เกี่ยวข้องกับการเกิดเหตุการณ์ ถ้า Censored เพราะผลการรักษาไม่ดีขึ้นจึงเลิกการติดต่อกับผู้ป่วย Kaplan–Meier Survival Curve ที่ได้ถือว่าไม่เหมาะสม

ปัญหาชวนคิดซึ่งเกี่ยวเนื่องกับข้อกำหนดอีกอย่างหนึ่งของ Kaplan–Meier methods คือ “การศึกษาผู้ที่เสียชีวิตภายในปี พ.ศ. 2544 โดยเก็บข้อมูลจากใบมรณบัตรเพื่อหาค่าเฉลี่ยระยะเวลามีชีวิตเปรียบกันระหว่างผู้ที่เคยกับไม่เคยใช้อินเตอร์เน็ต ผลน่าจะออกมาอย่างไร ทำไมเป็นเช่นนั้น”

คำตอบคือ ผลต้องออกมาว่า “ผู้เคยใช้อินเตอร์เน็ตอายุสั้นกว่าผู้ไม่เคยใช้อินเตอร์เน็ต” อย่างแน่นอน เพราะอินเตอร์เน็ตมีเมื่อไม่นานมานี้เอง ผู้อายุมาก ๆ แล้วมักไม่ใช้ แต่หนุ่มสาวใช้อินเตอร์เน็ตมากกว่า ดังนั้นผู้ที่ตายในปี 2544 นั้นผู้เคยใช้อินเตอร์เน็ตจึงมีอายุสั้นกว่า

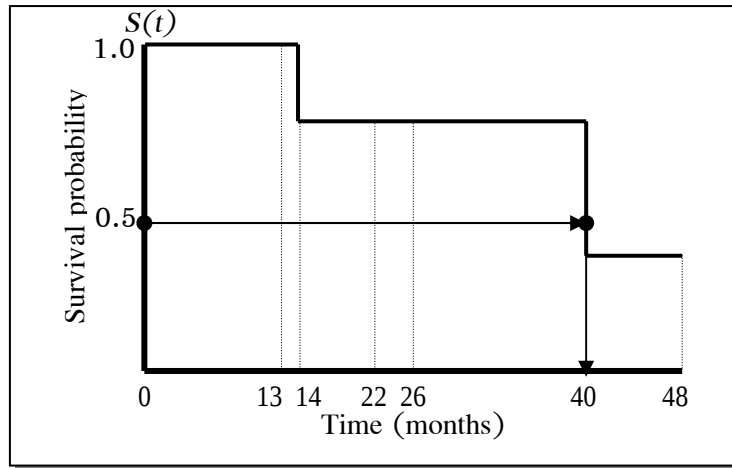
ปัญหานี้ชี้ให้เห็นว่า ถ้า ณ จุดเริ่มต้นการศึกษา ลักษณะผู้ที่ศึกษาแตกต่างกันแล้ว ค่าที่ได้จาก Kaplan–Meier methods ต้องแปลความหมายอย่างระมัดระวัง

2.3.1.4 การคำนวณหาค่าเฉลี่ยระยะปลอดเหตุการณ์

ค่าเฉลี่ย (Average) ของ ระยะปลอดเหตุการณ์ ต้องใช้ค่า Median survival time ซึ่งหาได้จาก Kaplan–Meier survival curve ค่าของจุดเวลาที่

ตรงกับค่า Survival probability 0.5 คือค่าดังกล่าว ดังนั้นค่าเฉลี่ยระยะปลอดเหตุการณ์จากรูป 2.8 คือ 40 เดือน

รูปที่ 2.8 Median survival time จาก Kaplan–Meier survival curve

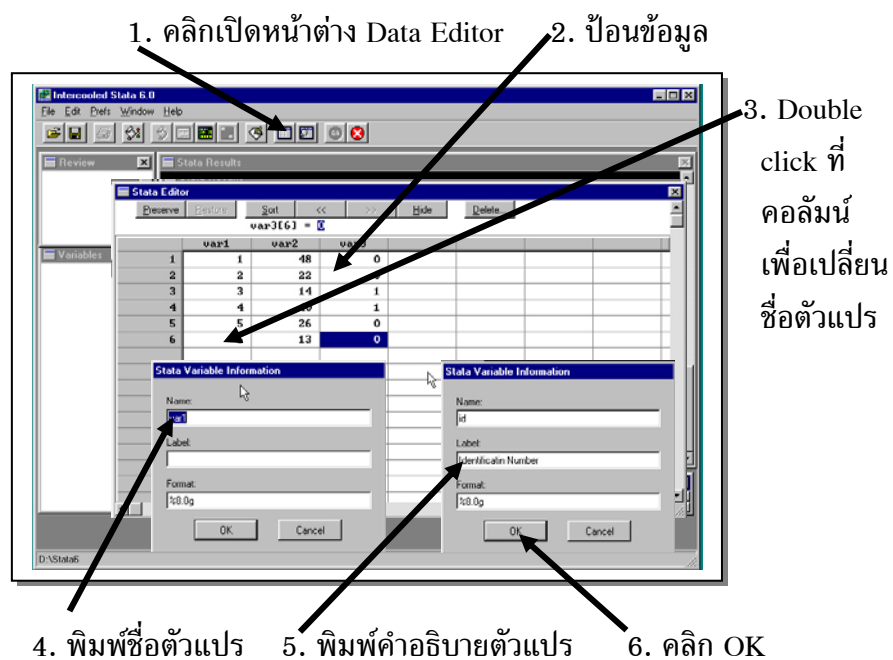


2.3.2 การวิเคราะห์โดยใช้คอมพิวเตอร์

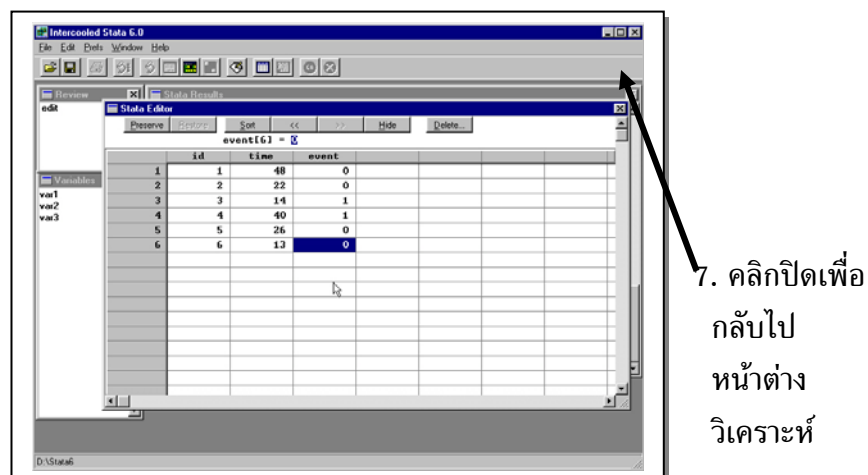
ที่กล่าวข้างต้นแสดงการคำนวณด้วยมือเพื่อให้เกิดความเข้าใจ ที่ทำได้เพราะข้อมูลไม่มากนัก แต่ถ้าข้อมูลมากต้องใช้คอมพิวเตอร์ มีหลายโปรแกรมทางสถิติที่สามารถใช้วิเคราะห์ Survival analysis ได้ แต่ในที่นี้เลือกโปรแกรม STATA 6 เพราะสามารถวิเคราะห์ได้ครอบคลุมมากในขณะที่ยังราคาไม่สูงนัก (ดูรายละเอียดใน <http://www.stata.com>) ในบทนี้ขอแนะนำการใช้เบื้องต้นกับข้อมูลตัวอย่างข้างต้น ดังนี้

ขั้นที่ 1 ป้อนข้อมูลเข้าคอมพิวเตอร์ด้วยโปรแกรม STATA

รูปที่ 2.9 การป้อนข้อมูลเข้าคอมพิวเตอร์ ขั้นที่หนึ่ง



รูปที่ 2.10 การป้อนข้อมูลเข้าคอมพิวเตอร์ ขั้นที่สอง



เมื่อดำเนินการตามข้อ 1 ถึง 7 ตามรูปที่ 2.9 และ 2.10 ที่กล่าวข้างต้น จะได้ข้อมูลที่ต้องการวิเคราะห์อยู่ในหน่วยความจำของ STATA เพื่อให้ข้อมูลสามารถนำมาใช้ได้อีกในภายหลัง ควรบันทึกข้อมูลโดยคำสั่ง "save" หรือคลิกที่แถบเครื่องมือบนจอได้

การป้อนข้อมูล ควรให้เป็นตัวเลข (Numeric) ทั้งหมด เพื่อง่ายต่อการจัดการและการวิเคราะห์ข้อมูลในระดับสูงขึ้นได้

การให้ชื่อตัวแปรควรกระทำหลังจากป้อนข้อมูลเสร็จ โดยเปลี่ยนชื่อจาก "var1" และ "var2" จนถึงตัวสุดท้ายที่ตั้งให้โดยอัตโนมัติจาก STATA ให้เป็นชื่อตัวแปรตามต้องการตามข้อ 3 รูปที่ 9 การตั้งชื่อตัวแปรดังกล่าวควรกำหนดเป็นอักษรตัวเล็กเพื่อง่ายต่อการสังเคราะห์ต่อไป (STATA ให้ความหมายอักษรตัวเล็กหรือใหญ่ต่างกัน)

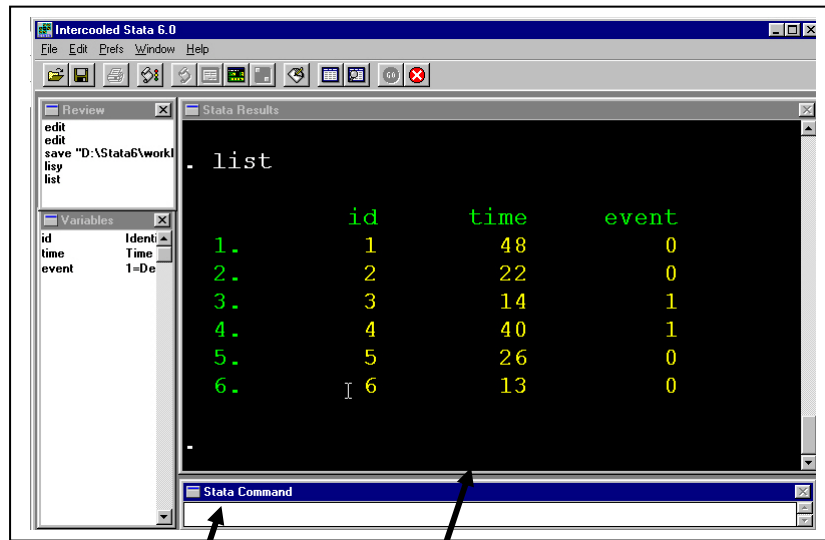
อย่างไรก็ตาม ผู้อ่านสามารถป้อนข้อมูลโดยใช้โปรแกรมอื่น แม้กระทั่งใช้ Microsoft Excel ที่มีในคอมพิวเตอร์ส่วนมาก จากนั้นจึงนำเข้ามาใน STATA โดยเพียงก๊อปปี้และวาง (Copy and Paste)

ขั้นที่ 2. เตรียมข้อมูลเพื่อการวิเคราะห์ Survival analysis

ก่อนเริ่มการวิเคราะห์ Survival analysis จำเป็นต้องเตรียมข้อมูลก่อน ก่อนอื่นควรดูรายการข้อมูล จากนั้นต้องแน่ใจว่าข้อมูลมีอย่างน้อย 2 ตัวแปรคือ ตัวแปรที่เป็นระยะปลอดเหตุการณ์ และตัวแปรที่เป็นเหตุการณ์ ดังรูปที่ 2.11 และแสดงให้เห็นชัดเจนในรูป 2.12

ขั้นที่ 2.1 ดูรายการข้อมูล

รูปที่ 2.11 การสั่งเพื่อขอดูรายการข้อมูล



1. พิมพ์คำสั่ง list แล้ว
กด <ENTER>

2. หน้าต่างแสดงผล

การเรียกดูรายการข้อมูล สามารถใช้คำสั่ง `list` หรืออาจสั่ง `edit` หรือคลิกที่ Data Editor Menu

ผลที่ปรากฏหลังจากสั่งคำสั่ง `list` เป็นดังรูปที่ 2.12

รูปที่ 2.12 รายการข้อมูลอันเป็นผลจากการป้อนข้อมูล

คำสั่งปรากฏหลังจุด (Dot command)



	id	time	event
1.	1	48	0
2.	2	22	0
3.	3	14	1
4.	4	40	1
5.	5	26	0
6.	6	13	0

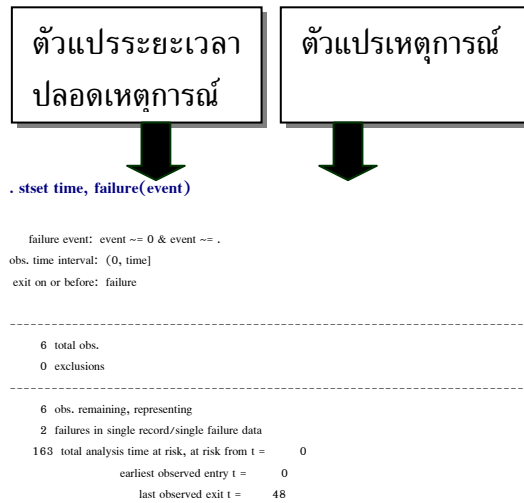
ตลอดหนังสือเล่มนี้ คำสั่ง STATA พร้อมผลที่ได้ แสดงเป็นตัวอักษรที่มีลักษณะต่างไปและมีขนาดเล็กกว่าปกติ โดยคำสั่งจะแสดงหลังจุดเช่นคำสั่ง `.list` ตามรูปที่ 2.12 บางคำสั่งอาจยาวกว่าหนึ่งบรรทัด บางคำสั่งอาจมีคำอธิบายส่วนประกอบของคำสั่งประกอบไปด้วย การฝึกหัดเพื่อทำซ้ำกับตัวอย่างในหนังสือนี้จะช่วยให้เข้าใจยิ่งขึ้น อย่างไรก็ตามผู้อ่านสามารถใช้คำสั่ง `help` ตามด้วยชื่อคำสั่ง เช่น `help stset` เป็นต้น (แนะนำให้สั่ง `help st` ก่อนเพื่อดูรายการของคำสั่งทุกคำสั่งสำหรับ Survival analysis) หรืออาจศึกษารายละเอียดเพิ่มเติมในคู่มือการใช้โปรแกรม StataCorp. (1999) หรือ ในอินเทอร์เน็ตที่ <http://www.stata.com> ได้

อนึ่ง หากผู้อ่านจะทำซ้ำกับคำสั่งที่ใช้ในหนังสือเล่มนี้ ให้ลอกคำสั่งพิมพ์เข้าไปโดยตรงโดยไม่ต้องพิมพ์จุดก่อน อักษรตัวเล็กหรือใหญ่มีความหมายต่างกัน

ขั้นที่ 2.2 ทำการ Set ข้อมูลเพื่อการวิเคราะห์ Survival analysis

คำสั่งสำหรับ Survival analysis ใน STATA ทุกคำสั่งจะขึ้นต้นด้วย `st` ซึ่งมาจากคำว่า Survival time แต่คำสั่งแรกสุดคือ

"stset" เพื่อ Set ข้อมูลเพื่อการวิเคราะห์ Survival analysis ดังนี้



ผลจากคำสั่ง stset จากการคลิกที่หน้าต่าง Data editor เป็นดังนี้

รูปที่ 2.13 รายการข้อมูลอันเป็นผลจากคำสั่ง stset

id	time	event	_st	_d	_t	_t0
1	48	0	1	0	48	0
2	22	0	1	0	22	0
3	14	1	1	1	14	0
4	40	1	1	1	40	0
5	26	0	1	0	26	0
6	13	0	1	0	13	0

ข้อมูลเดิม

ตัวแปรใหม่ที่ถูกสร้างขึ้นโดย
อัตโนมัติ

คำอธิบาย:

_st ค่า 1 = นำเข้าวิเคราะห์ ค่า 0 = ไม่นำเข้าวิเคราะห์

_d คำ 1 = Failure คำ 0 = Censored
 _t0 และ _t ระยะปลอดเหตุการณ์ _t0 คือจุดเวลาเริ่มต้น
 และ _t คือจุดสิ้นสุด

ขั้นที่ 2.3 คำสั่งวิเคราะห์ข้อมูลที่ใช้บ่อย

. stsum

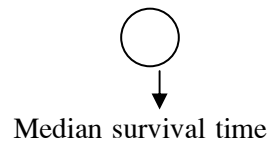
failure _d: event
 analysis time _t: time

	incidence		no. of	----- Survival time -----		
	l time at risk	rate		25%	50%	75%
total	163	.0122699	6	40	40	.

อุบัติการณ์ = $2/163 = 0.012$ ต่อเดือน

หรือ 1.2 ต่อ 100 คน-เดือน

นี่คือ Incidence density



คำสั่ง stsum ใช้สรุป (Summarize) ค่าสถิติที่สำคัญเกี่ยวกับ
 Survival experience

. cii 163 0.0122699 , p

		-- Poisson Exact --			
Variable Exposure	Mean	Std. Err.	[95% Conf. Interval]		
163	.0122699	.0086762	.0014888	.0443053	

คำสั่ง cii เป็นคำสั่งคำนวณช่วงเชื่อมั่น (Confidence interval) ที่มี
 i อีกตัวเพราะเป็น Immediate command สำหรับคำสั่ง ci เนื่องจากในที่นี้
 เราคำนวณจากตัวเลขที่ใส่ลงไปโดยตรง ไม่ใช่ชื่อตัวแปร ตัวเลขที่ใส่เข้าไป

คือจำนวน Person-time ตามด้วย Incidence ที่ได้จากคำสั่ง `stsum` ทำคำสั่งคือ `, p` นั้นเพื่อบอกให้ STATA ทราบว่าให้คำนวณบนการแจกแจงแบบ Poisson

สรุปผลได้ดังนี้ "จากผู้ป่วยที่ศึกษาทั้งหมด 6 ราย รวมเป็นหน่วยศึกษา 163 คน-เดือน (person-months) มีผู้ป่วยตาย 2 ราย คิดเป็นอุบัติการณ์ 1.2 ต่อ 100 คน-เดือน (95%CI: 0.15 ถึง 4.4)"

`. stquant`

Number of deaths = 2
Number of ties = 0

```

+-----+
| 50 percentile      95% Confidence Interval      |
| survival time =   40.000   19.650   60.350   (Large Sample Approx.) |
|                   14.000   48.000   (Brookmeyer-Crowley) |
+-----+

```

คำสั่ง `stquant` เป็นคำสั่งคำนวณช่วงเชื่อมั่นของ Median survival time

สรุปผลได้ดังนี้ "ระยะปลอดเหตุการณ์โดยเฉลี่ย (Median survival time) เท่ากับ 40 เดือน เราเชื่อมั่น 95% ว่าค่าเฉลี่ยดังกล่าวในประชากรอยู่ระหว่าง 19.7 ถึง 60.4 เดือน"

`. sts list, at(12, 24, 36, 48)`

failure _d: event
analysis time _t: time

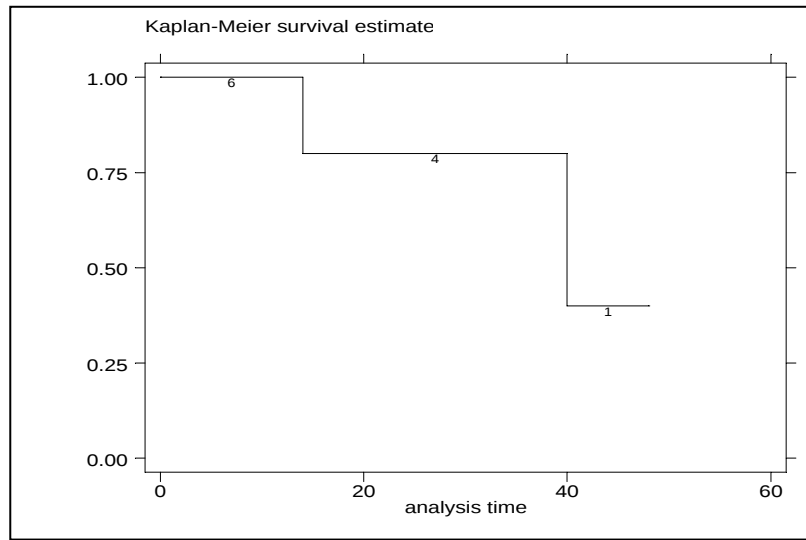
Time	Beg.		Survivor		Std.	
	Total	Fail	Function	Error	[95% Conf. Int.]	
12	0	0	1.0000	.	.	.
24	4	1	0.8000	0.1789	0.2038	0.9692
36	3	0	0.8000	0.1789	0.2038	0.9692
48	1	1	0.4000	0.2966	0.0114	0.8290

Note: Survivor function is calculated over full data and evaluated at indicated times; it is not calculated from aggregates shown at left.

คำสั่ง `sts list` เป็นคำสั่งคำนวณ Survival probability และช่วงเชื่อมั่น โดยสามารถระบุจุดเวลาได้จากการเพิ่มคำสั่ง `, at()` แล้วใส่ตัวเลขในวงเล็บตามตัวอย่างข้างต้น

สรุปผลได้ดังนี้ "โอกาสที่จะอยู่โดยปลอดเหตุการณ์ในระยะ 24 เดือน เท่ากับ 80% (95%CI: 20.4% ถึง 96.9%)"

```
. sts g, atrisk
      failure _d: event
      analysis time _t: time
```



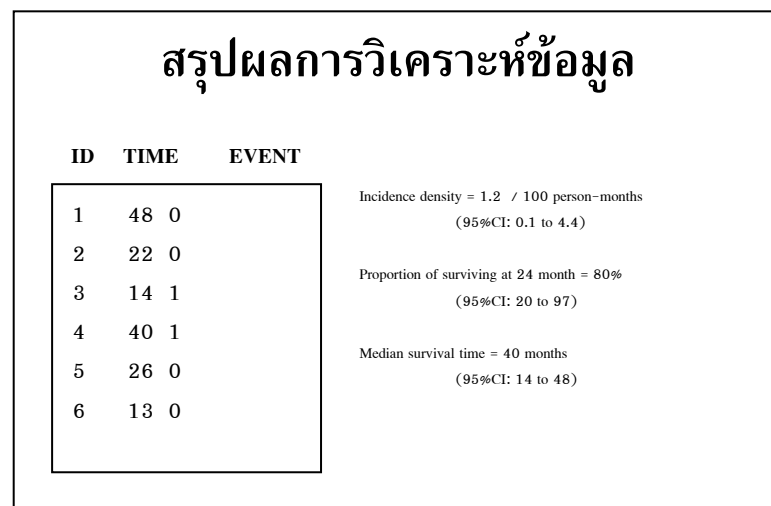
คำสั่ง `sts g` เป็นคำสั่งสำหรับสร้าง Kaplan-Meier Survival Curve การระบุ `, atrisk` ลงไปท้ายคำสั่งเพื่อให้เห็นจำนวนผู้ป่วยที่เหลือสำหรับคำนวณทางสถิติ (Number at risk) ในแต่ละช่วงเวลาด้วย ซึ่งมีประโยชน์มาก เพราะเราจะไม่แปลความหมายจนเกินจริงกับกราฟกรณีเหลือ Number at risk น้อยกว่า 5 (Altman, 1991) เป็นต้น

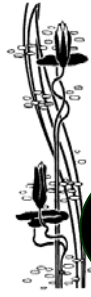
นักวิจัยสามารถนำเอากราฟข้างต้นไปใส่ในรายงานวิจัยได้โดย เพียงแต่คลิก Edit ที่ Menu bar ของ STATA แล้วคลิก Copy Graph จากนั้นเปิด Microsoft Word แล้วสร้าง Text box ก่อน จากนั้น Paste กราฟลงไปใน Text box ด้วยวิธีนี้จะง่ายต่อการย่อ ขยาย และเลื่อน ตำแหน่งของกราฟ ในรายงานวิจัย

2.3.3 สรุปผลการพรรณนา Survival experience

ที่กล่าวแล้วทั้งหมดเป็นการพรรณนา Survival experience จากที่ได้เริ่มด้วยคำถามและการวิเคราะห์ที่ไม่เหมาะสมแต่พบข้อที่แสดงในรูปที่ 2.5 นั้น สามารถสรุปใหม่ให้ถูกต้องเหมาะสมดังรูปที่ 2.14 ดังนี้

รูปที่ 2.14 สรุปผลการพรรณนา Survival experience





การวิเคราะห์ข้อมูล ระยะปลอดเหตุการณ์ 3 ที่มีมากกว่าหนึ่งกลุ่ม

การพิจารณา Survival experience ที่กล่าวแล้วข้างต้น เป็นการศึกษาในกลุ่มตัวอย่างเดียว กรณีมีมากกว่าหนึ่งกลุ่มนอกจากต้องพิจารณาในแต่ละกลุ่มแล้ว ยังต้องเปรียบเทียบระหว่างกลุ่มด้วย เพื่อแสดงให้เห็นแนวทางการวิเคราะห์ จึงสมมติข้อมูลเพิ่มเติมคือตัวแปร DRUG (1=ยาใหม่, 0=ยาเก่า) เป็นตัวแปรจำแนกกลุ่มที่เปรียบเทียบ และเพิ่มข้อมูลอีก 6 ราย ดังนี้

ID	TIME	DEAD	DRUG	
1	48	0	1	ข้อมูลเดิมที่กล่าวแล้วข้างต้น
2	22	0	1	
3	14	1	1	
4	40	1	1	
5	26	0	1	
6	13	0	1	
7	13	0	0	ข้อมูลใหม่ที่นำมาเป็นกลุ่มเปรียบเทียบ
8	6	1	0	
9	12	1	0	
10	14	1	0	
11	22	1	0	
12	13	1	0	

ต่อไปนี้จะแสดงให้เห็นแนวทางการวิเคราะห์ โดยแสดงแนวทางที่ไม่เหมาะสมก่อน (ผิดเป็นครู) จากนั้นจึงเป็นแนวทางที่ถูกต้องเหมาะสมตามลำดับดังนี้

3.1 การวิเคราะห์เปรียบเทียบระหว่างกลุ่มที่ไม่เหมาะสม

3.1.1 วิเคราะห์โดยใช้ระยะปลอดเหตุการณ์เป็นตัวแปรตาม

3.1.1.1 เปรียบเทียบค่าเฉลี่ยโดยใช้ t-test

```
. ttest time, by(drug)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	6	13.33333	2.092314	5.125102	7.954869	18.7118
1	6	27.16667	5.775908	14.14803	12.31922	42.01411
combined	12	20.25	3.595294	12.45446	12.33681	28.16319
diff		-13.83333	6.143199		-27.52123	-1.145439

Degrees of freedom: 10

Ho: mean(0) - mean(1) = diff = 0

Ha: diff < 0
t = -2.2518
P < t = 0.0240

Ha: diff ~= 0
t = -2.2518
P > |t| = 0.0480

Ha: diff > 0
t = -2.2518
P > t = 0.9760

คำสั่ง ttest ข้างต้นเป็นคำสั่งสำหรับทดสอบว่าค่าเฉลี่ย TIME ระหว่างกลุ่ม DRUG แตกต่างกันอย่างมีนัยสำคัญหรือไม่ ในวงรีคือค่าที่ควรนำไปสรุปในผลการวิเคราะห์ (ศึกษารายละเอียดใน บันทึก ถิ่นคำพ, 2543) เช่นกลุ่มได้รับยาเก่า 6 คนมีชีวิตอยู่ได้โดยเฉลี่ย 13.3 เดือน

ในขณะที่กลุ่มได้รับยาใหม่ 6 คนมีชีวิตรอดอยู่ได้โดยเฉลี่ย 27.2 เดือน กลุ่มได้รับยาใหม่มีระยะรอดชีพยาวกว่ากลุ่มได้รับยาเก่า 13.8 เดือน (95%CI: 0.1 ถึง 27.5) ซึ่งมีนัยสำคัญทางสถิติ (p-value = 0.048)

การวิเคราะห์เช่นนี้จะถูกก็ต่อเมื่อไม่มี Censored observation จึงไม่ใช่วิธีที่ถูกต้องเมื่อใช้กับข้อมูลชุดนี้

3.1.1.2 วิเคราะห์โดยใช้การถดถอยเชิงเส้น (Simple linear regression)

เป็นวิเคราะห์เพื่อดูผลของตัวแปรต้น DRUG ต่อตัวแปรตาม TIME โดยใช้ Regression model เขียนเป็นสัญลักษณ์ได้ดังนี้

$$y = a + bX$$

เมื่อ y คือ TIME และ X คือ DRUG

การวิเคราะห์ทำได้ดังนี้

```
. regress time drug
```

Source	SS	df	MS	Number of obs =	12
Model	574.083333	1	574.083333	F(1, 10) =	5.07
Residual	1132.16667	10	113.216667	Prob > F =	0.0480
Total	1706.25	11	155.113636	R-squared =	0.3365
				Adj R-squared =	0.2701
				Root MSE =	10.64

time	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
drug	13.83333	6.143199	2.252	0.048	.1454339 27.52123
_cons	13.33333	4.343897	3.069	0.012	-3.654527 23.01214

คำสั่ง regress เป็นคำสั่งสำหรับ Fit Linear regression model ในวิธีคือค่าที่ควรนำไปสรุปในผลการวิเคราะห์ (ศึกษารายละเอียดใน บันทึก)

ถึ้นค่ารพ, 2543) จะเห็นว่าค่าสัมประสิทธิ์การถดถอยเท่ากับ 13.83 (95%CI: 0.1 ถึง 27.5) ซึ่งเท่ากันกับการใช้ t-test วิเคราะห์ตามที่กล่าวแล้วข้างต้น

3.1.2 วิเคราะห์โดยใช้การเกิดเหตุการณ์เป็นตัวแปรตาม

การวิเคราะห์โดยใช้ t-test และ Regression ตามที่กล่าวข้างต้นล้วนใช้ตัวแปรตามคือระยะปลอดเหตุการณ์ TIME ซึ่งเป็นตัวแปรต่อเนื่อง บ่อยครั้งที่ผู้วิจัยเลือกเอาการเกิดเหตุการณ์ DEAD เป็นตัวแปรตามซึ่งเป็นตัวแปรแจกแจง กลุ่มนี้มักวิเคราะห์ดังนี้

3.1.2.1 เปรียบเทียบค่าสัดส่วนโดยใช้ Z-test

. prtest drug, by(dead)

```
Two-sample test of proportion                                0: Number of obs =      5
                                                            1: Number of obs =      7
```

Variable	Mean	Std. Err.	z	P> z	[95% Conf. Interval]
0	.8	.1788854	4.47214	0.0000	.449391 1.150609
1	.2857143	.1707469	1.67332	0.0943	-.0489436 .6203721
diff	.5142857	.2472944			.0295976 .9989738
under HO:	.29277	1.75662	0.0790		

```
Ho: proportion(0) - proportion(1) = diff = 0

Ha: diff < 0      Ha: diff ~= 0      Ha: diff > 0
z = 1.757         z = 1.757         z = 1.757
P < z = 0.9605   P > |z| = 0.0790     P > z = 0.0395
```

คำสั่ง prtest เป็นคำสั่งสำหรับเปรียบเทียบค่าสัดส่วน (Proportion test) ในวงรีคือค่าที่ควรนำไปสรุปในผลการวิเคราะห์ (ศึกษารายละเอียดในบันทึก ถึ้นค่ารพ, 2543) เช่นกลุ่มได้รับยาเก่า 6 คนมีอัตราตาย 80.0%

ในขณะที่กลุ่มได้รับยาใหม่ 6 คนมีอัตราตาย 28.6% กลุ่มได้รับยาเก่ามีอัตราตายสูงกว่ากลุ่มได้รับยาใหม่ 51.4% (95%CI: 29.6% ถึง 99.9%) อย่างไรก็ตามความแตกต่างดังกล่าวไม่มีนัยสำคัญทางสถิติ (p-value = 0.079)

3.1.2.2 หาขนาดความสัมพันธ์และทดสอบสถิติโดยใช้ Chi-square test

. cs dead drug, or

	drug Exposed	Unexposed	Total	
Cases	2	5	7	
Noncases	4	1	5	
Total	6	6	12	
Risk	.3333333	.8333333	.5833333	
	Point estimate		[95% Conf. Interval]	
Risk difference	- .5		-.9808315 .0191685	
Risk ratio	.4		.1220764 1.310654	
Prev. frac. ex.	.6		-.3106542 .8779236	
Prev. frac. pop	.3			
Odds ratio	.1		0	1.263859 (Cornfield)
	chi2(1) =		3.09 Pr>chi2 = 0.0790	

คำสั่ง cs เป็นคำสั่งสำหรับวิเคราะห์ข้อมูลตารางการณักร (Contingency table) กรณีทั้งตัวแปรตามและตัวแปรต้นเป็นตัวแปรทวินาม (Dichotomous) ในวงรีคือค่าที่ควรนำไปสรุปในผลการวิเคราะห์ (ศึกษารายละเอียดใน บันทึก ถิ่นคำรพ, 2543) จากผลข้างต้น สรุปได้ว่ากลุ่มที่ได้รับยาใหม่ (DRUG=1) มีโอกาสที่จะเสียชีวิต (DAED=1) เป็น 0.4 เท่าของกลุ่มที่ได้รับยาเก่า (DRUG=0) รายงานผลเป็น Relative Risk = 0.40 (95%CI: 0.12 ถึง 1.31) ซึ่งไม่มีนัยสำคัญทางสถิติ (p-value = 0.079) เพื่อเข้าใจความหมายง่ายขึ้น ผู้วิจัยอาจสลับกลุ่มเปรียบเทียบ จะ

กล่าวใหม่ได้ว่า กลุ่มที่ได้รับยาเก่ามีโอกาสที่จะเสียชีวิตเป็น 2.5 เท่าของ กลุ่มที่ได้รับยาใหม่ (ได้จาก $1/0.4$ เรียกว่าค่า Reciprocal ของ Relative Risk)

3.1.2.3 วิเคราะห์โดยใช้ Logistic regression

กรณีนี้มักใช้ Logistic regression วิเคราะห์ เพื่อให้ได้ Model ที่เขียนเป็นสัญลักษณ์ดังนี้

$$\text{Logit}(y) = a + bX$$

เมื่อ y คือ DEAD และ X คือ DRUG

การวิเคราะห์ทำได้วิเคราะห์ดังนี้

. logistic dead drug

Logit estimates

Number of obs = 12

LR chi2(1) = 3.26

Prob > chi2 = 0.0712

Pseudo R2 = 0.1997

Log likelihood = -6.5224523

dead	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
drug	.1	.1396395	-1.649	0.099	.0064772	1.54388

คำสั่ง logistic เป็นคำสั่งสำหรับ Fit Logistic regression model ในวงรีคือค่าที่ควรนำไปสรุปในผลการวิเคราะห์ (ศึกษารายละเอียดใน บันทึกถึนคำรพ, 2542)

ทั้ง t-test และ Regression แม้ใช้ระยะปลอดเหตุการณ์เป็นตัวแปรตาม แต่ใช้ไม่ได้กับข้อมูลที่มี Censored observation กรณีที่ใช้ Z-test หรือ Chi-square test หรือ Logistic regression นั้นไม่ใช้ระยะปลอดเหตุการณ์เป็นตัวแปรตาม แต่ใช้การเกิดเหตุการณ์ โดยสนใจเพียงว่าเกิดหรือไม่เกิด

เหตุการณ์โดยไม่สนใจว่าเกิดช้าหรือเร็ว อาจเหมาะสมในบางกรณีที่ไม่สนใจระยะเวลาการเกิดเหตุการณ์ดังกล่าวนั้นเป็นเหตุผลที่ยอมรับได้ แต่การวิเคราะห์โดยใช้ Survival analysis ให้ผลไม่แตกต่างกันในกรณีดังกล่าว อีกทั้งยังให้รายละเอียดที่มากกว่า เช่นความเสี่ยงต่อการเกิดเหตุการณ์ในแต่ละจุดเวลา เป็นต้น แต่ถ้าวการละเลยซึ่งข้อมูลระยะปลอดเหตุการณ์นั้นเป็นเหตุผลที่ยอมรับไม่ได้ ต้องใช้ Survival analysis เท่านั้น

พึงตระหนักเสมอว่าข้อมูลที่มี Censored observation ต้องมีวิธีการจัดการที่เหมาะสม ไม่เช่นนั้นจะได้ข้อสรุปที่ผิด อย่างไรก็ตามแม้ข้อมูลไม่มี Censored observation คือเป็น Complete observation ทั้งหมด แต่ถ้ามีตัวแปรตามเป็นระยะเวลา การวิเคราะห์โดยใช้ Survival analysis สามารถให้องค์ความรู้ที่ละเอียดกว่า เนื่องจากมีการผนวกเอาข้อมูลระยะเวลาในแต่ละช่วงเข้าร่วมการวิเคราะห์ จึงมีประสิทธิภาพสูงกว่าวิธีการทั่วไปทางสถิติ

3.2 การวิเคราะห์เปรียบเทียบระหว่างกลุ่มที่เหมาะสม

3.2.1 เปรียบเทียบ Survival experience โดยใช้ Log-rank test

- มี Assumption เกี่ยวกับ Hazard function ว่าเป็นสัดส่วนกันระหว่างกลุ่มที่เปรียบเทียบ
- สูตรสำหรับหาคำนวน

$$\chi^2 = \sum_i^{\# \text{ of groups}} \frac{(O_i - E_i)^2}{V_i}$$

- ค่า V มีกล่าวถัดไป เพื่อง่ายต่อการทำความเข้าใจ จึงใช้สูตรอย่างง่าย
- สูตรอย่างง่าย เป็นดังนี้

$$\chi^2 = \sum_i^{\# \text{ of groups}} \frac{(O_i - E_i)^2}{E_i}$$

เมื่อ $O_i - E_i = \sum (o_{ij} - e_{ij}),$

i คือกลุ่มที่เปรียบเทียบ = 1, 2, ..., k กลุ่ม

j คือผู้ป่วย = 1, 2, ..., n คน

- คำนวณโดยต้องสร้างตาราง กรณีเปรียบเทียบ 2 กลุ่ม ตารางต้องมี 10 คอลัมน์ดังนี้

t = Time

n = Number at risk for both groups ณ เวลา t

n1 = Number at risk for group 1 ณ เวลา t

n2 = Number at risk for group 2 ณ เวลา t

d = Dead for both groups ณ เวลา t

c = Censored for both groups ณ เวลา t

O1 = Number of dead for group 1 ณ เวลา t

O2 = Number of dead for group 2 ณ เวลา t

E1 = Number of expected dead for group 1 ณ เวลา t

E2 = Number of expected dead for group 2 ณ เวลา t

จากข้อมูลตัวอย่าง เขียนใหม่ได้ดังนี้

DRUG1 = 48+, 22+, 26+, 13+, 14, 40 ให้เป็นกลุ่ม 1

DRUG0 = 13+, 6, 12, 14, 22, 13 ให้เป็นกลุ่ม 2

แสดงการคำนวณตามที่ปรากฏในตารางที่ 3.1 ดังนี้

ตารางที่ 3.1 การคำนวณ Log-rank test

t	n	n ₁	n ₂	d	c	O ₁	O ₂	$E_1 = \frac{d \times n_1}{n}$	$E_2 = \frac{d \times n_2}{n}$
6	12	6	6	1	0	0	1	(1×6)/11=0.50	(1×6)/11=0.50
12	11	6	5	1	0	0	1	(1×6)/11=0.55	(1×5)/11=0.45
13	10	6	4	1	2	0	1	(1×6)/10=0.60	(1×4)/10=0.40
14	7	5	2	2*	0	1	1	(2×5)/7 =1.43	(2×2)/7 =0.57
22	5	4	1	1	1	0	1	(1×4)/5 =0.80	(1×1)/5 =0.20
26	3	3	0	0	1	0	0	(0×3)/3 =0.00	(0×0)/3 =0.00
40	2	2	0	1	0	1	0	(1×2)/2 =0.10	(1×0)/2 =0.00
48	1	1	0	0	1	0	0	(0×1)/1 =0.00	(0×0)/1 =0.00
Totals						2	5	4.87	2.13
$\chi^2 = [(2 - 4.87)^2 / 4.87] + [(5 - 2.13)^2 / 2.13] = 5.56$ ค่า Chi-square ที่ df = 1 ได้ p-value = 0.018									

* Survival time เท่ากัน 2 คน เรียกว่าเกิด Tied failure time ต้องมีวิธีการจัดการที่เหมาะสม เพื่อง่ายต่อการทำความเข้าใจ จะกล่าวถึงกรณีนี้ตอนท้าย

ผลการทดสอบสรุปได้ว่า การรอดชีพ (Survival experience) ของสองกลุ่มแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($p\text{-value} = 0.018$)

การคำนวณตามตารางข้างต้น เป็นสิ่งแนะนำเพื่อการทำความเข้าใจวิธีการทางสถิติ แต่ในทางปฏิบัติ เราไม่คำนวณด้วยมือ ต่อไปนี้เป็นผลการวิเคราะห์ Log-rank test โดยใช้ STATA ดังนี้

```
. sts test drug
```

```
      failure _d:  dead
analysis time _t:  time
```

```
Log-rank test for equality of survivor functions
```

drug	Events observed	expected
0	5	2.13
1	2	4.87
Total	7	7.00

```
      chi2(1) =      6.67
      Pr>chi2 =      0.0098
```

- ผลการทดสอบแตกต่างกันเพราะแม้ใช้สูตรเดียวกันกับที่กล่าวข้างต้น แต่ตัวหารต่างกัน กล่าวคือแทนที่จะใช้ E_i แต่ใช้ $\text{Var}(O_i - E_i)$ แทน ค่า $\text{Var}(O_i - E_i)$ ยากต่อการคำนวณด้วยมือจึงไม่นำเสนอในที่นี้ (ผู้อ่านศึกษารายละเอียดได้ใน Kleinbaum, 1996 หน้า 61)
- ค่า E_i ที่ได้จากการคำนวณด้วยมือคือ $E_1 = 4.87$ และ $E_2 = 2.13$ ซึ่งเท่ากับค่า E_i ที่ได้จากการคำนวณโดย STATA (ค่าที่อยู่ในวงรี)

กรณีไม่แน่ใจว่า Assumption เกี่ยวกับ Hazard function ว่าเป็นสัดส่วนกันระหว่างกลุ่มที่เปรียบเทียบ ให้เลือกใช้ Wilcoxon test แทน Log-rank test ดังนี้

. sts test drug, wilcoxon

```
failure _d: dead
analysis time _t: time
```

Wilcoxon (Breslow) test for equality of survivor functions

drug	Events		Sum of ranks
	observed	expected	
0	5	2.13	25
1	2	4.87	-25
Total	7	7.00	0

chi2(1) = 5.65
Pr>chi2 = 0.0175

อย่างไรก็ตาม Log-rank test บอกเพียงค่า p-value ซึ่งไม่ได้บอกขนาด หรือทิศทางของความแตกต่างระหว่างกลุ่มเปรียบเทียบแต่อย่างใด ดังนั้นการวิเคราะห์เพื่อพรรณนา Overall survival และหา Survival probability โดยวิธี Kaplan-Meier ยังคงจำเป็นในทุกการศึกษาที่ต้องใช้ Survival analysis ดังจะกล่าวในลำดับต่อไป

การทดสอบโดย Log-rank test หรือ Wilcoxon test นี้ มิได้เป็นการทดสอบความแตกต่างค่า Median survival time หรือ Survival probability ณ จุดเวลาใดเวลาหนึ่ง แต่เป็นการทดสอบความแตกต่างของ Survival experience ในภาพรวม (เรียกว่า Global test) ระหว่าง กลุ่ม

3.2.2 คำนวณ Overall incidence rate และ Median survival time

ค่า Overall incidence rate ในที่นี้คือ Incidence density ตามที่ได้กล่าวแล้วข้างต้น ค่าดังกล่าวนี้และค่า Median survival time สามารถหาได้โดยคำสั่ง `stsum` ดังนี้

. stsum, by(drug)

```

failure _d:  dead
analysis time _t:  time

```

drug	time at risk	incidence rate	no. of subjects	Survival time		
				25%	50%	75%
0	80	.0625	6	12	13	14
1	163	.0122699	6	40	40	.
total	243	.0288066	12	13	22	40

ผลข้างต้น สรุปได้ว่า จากกลุ่มที่ใช้ยาเก่า 80 คน-เดือน มีอัตราตายเท่ากับ 6.3/100 คน-เดือน ในขณะที่กลุ่มที่ใช้ยาใหม่ 163 คน-เดือน มีอัตราตายเท่ากับ 1.2/100 คน-เดือน ระยะเวลารอดชีพเฉลี่ยในกลุ่มที่ใช้ยาเก่าเท่ากับ 13 เดือนในขณะที่กลุ่มที่ใช้ยาใหม่เท่ากับ 40 เดือน

อีกคำสั่งที่มีประโยชน์คือ `sts list` ที่ให้ Survival probability ในแต่ละจุดเวลา เปรียบเทียบระหว่างสองกลุ่ม ดังนี้

. sts list, by(drug) compare

```

failure _d:  dead
analysis time _t:  time

```

		Survivor Function	
drug		0	1
time			
6	0.8333	1.0000	
11	0.8333	1.0000	
16	0.2500	0.8000	
21	0.2500	0.8000	
26	.	0.8000	
31	.	0.8000	
36	.	0.8000	
41	.	0.4000	
46	.	0.4000	
51	.	.	

3.2.3 คำนวณ Overall incidence rate พร้อมช่วงเชื่อมั่น

ในข้อ 3.2.2 ได้ค่า Incidence density แต่ยังไม่มีความเชื่อมั่น ซึ่ง
เป็นค่าที่มีความสำคัญ สามารถหาได้ดังนี้

```
. strate drug
```

```
failure _d: dead
analysis time _t: time
```

Estimated rates and lower/upper bounds of 95% confidence intervals
(12 records included in the analysis)

drug	_D	_Y	_Rate	_Lower	_Upper
0	5	80.0000	0.0625000	0.0260142	0.1501582
1	2	163.0000	0.0122699	0.0030687	0.0490606

ค่า Incidence ตามที่กล่าวข้างต้น เป็นมาตรฐานความเสี่ยง
(Risk) แม้สามารถใช้ได้กับข้อมูลที่มี Censored observation แต่เป็นเพียง
ค่าเฉลี่ยและถือว่าเท่ากันตลอดระยะเวลาที่ศึกษา ซึ่งความเป็นจริงไม่ค่อย
เป็นไปตามนั้น จึงนำมาใช้เชิงพรรณนาข้อมูล เพื่อบอกขนาดของผลจาก
ปัจจัยที่ศึกษานั้นหายาๆ ก่อนการวิเคราะห์ที่เหมาะสมกว่า

3.2.4 คำนวณช่วงเชื่อมั่นของ Median survival time

ในข้อ 3.2.2 ได้ค่า Median survival time แต่ยังไม่มีความเชื่อมั่น
อีกเช่นกัน สามารถหาได้ดังนี้

. stquant, by(drug)

drug	no. of events	no. of ties	50 percentile	95% Confidence limits Large Sample Appr.	Brookmeyer-Crowley
0	5	0	13	11.0796 14.9204	6 22
1	2	0	40	19.6503 60.3497	14 48
total	7	2	22	13.3609 30.6391	12 48

Brookmeyer-Crowley limits could be inaccurate in presence of tied survival times.

3.2.5 คำนวณ Survival probability ในแต่ละจุดเวลา พร้อมช่วงเชื่อมั่น

บ่อยครั้งที่คำถามวิจัยต้องการเปรียบเทียบ Survival ระหว่างสอง
กลุ่มในบางจุดเวลา สามารถวิเคราะห์ได้ดังนี้

. sts list, by(drug)

	failure _d:	dead					
analysis time _t:		time					
Time	Beg. Total	Fail	Net Lost	Survivor Function	Std. Error	[95% Conf. Int.]	

drug=0							
6	6	1	0	0.8333	0.1521	0.2731	0.9747
12	5	1	0	0.6667	0.1925	0.1946	0.9044
13	4	1	1	0.5000	0.2041	0.1109	0.8037
14	2	1	0	0.2500	0.2041	0.0123	0.6459
22	1	1	0	0.0000	.	.	.
drug=1							
13	6	0	1	1.0000	.	.	.
14	5	1	0	0.8000	0.1789	0.2038	0.9692
22	4	0	1	0.8000	0.1789	0.2038	0.9692
26	3	0	1	0.8000	0.1789	0.2038	0.9692
40	2	1	0	0.4000	0.2966	0.0114	0.8290
48	1	0	1	0.4000	0.2966	0.0114	0.8290

ตัวอย่างการใช้ประโยชน์ผลข้างต้น เช่น ณ เดือนที่ 14 หลังการ
รักษา (ตัวอักษรเอียง) Survival probability ของกลุ่มที่ได้รับยาเท่าเท่ากับ

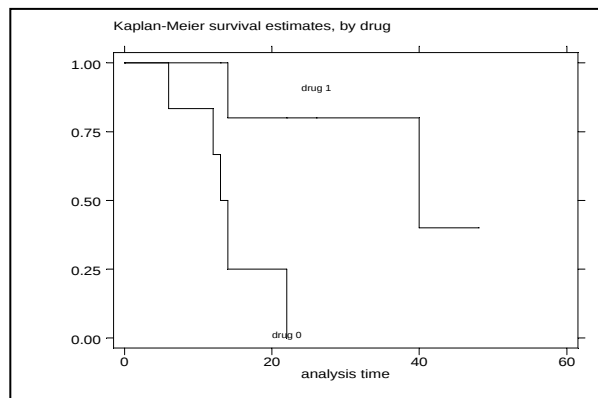
0.25 (95%CI: 0.01 ถึง 0.65) ต่ำกว่ากลุ่มที่ได้รับยาใหม่ซึ่งเท่ากับ 0.80 (95%CI: 0.20 ถึง 0.97)

3.2.6 สร้างกราฟ Kaplan-Meier survival curve

ทุกการวิจัยที่ต้องใช้ Survival analysis ควรนำเสนอ Kaplan-Meier Survival Curve ทำได้ดังนี้

`.sts graph, by(drug)`

failure _d: dead
analysis time _t:time



การวิเคราะห์ข้างต้น ได้ค่าสถิติจำแนกตามกลุ่มที่เปรียบเทียบ กรณีตัวอย่างมีสองกลุ่มก็ได้ค่าสองค่า จึงขาดค่าสถิติที่ใช้สรุปค่าจากสองกลุ่มที่เปรียบเทียบนั้น กรณี Survival analysis นี้ สามารถใช้ค่าต่อไปนี้

- Relative risk พร้อมช่วงเชื่อมั่น
- อัตราส่วน (Ratio) ของ Median survival time พร้อมช่วงเชื่อมั่น

- ผลต่าง (Difference) ของ Survival probability พร้อมช่วงเชื่อมั่น
 - Hazard ratio พร้อมช่วงเชื่อมั่น
- มีรายละเอียดการคำนวณเป็นลำดับดังต่อไปนี้

3.2.7 Relative risk พร้อมช่วงเชื่อมั่น

. stmh drug

```
failure _d: dead
analysis time _t: time
```

Maximum likelihood estimate of the rate ratio
comparing drug==1 vs. drug==0

RR estimate, and lower and upper 95% confidence limits

RR	chi2	P>chi2	[95% Conf. Interval]	
0.196	4.70	0.0302	0.038	1.012

คำสั่ง stmh นี้เหมาะที่จะใช้เมื่อมีตัวแปรที่สามที่ต้องการควบคุมผล เช่นโรงพยาบาลที่ทดลองยา (สมมติชื่อตัวแปรคือ hospital) คำสั่งจะเป็น stmh drug, by(hospital) จะยังผลให้ได้ค่า RR adjusted for hospital

. stmc drug

```
failure _d: dead
analysis time _t: time
```

Mantel-Cox comparisons

Mantel-Haenszel estimates of the rate ratio
comparing drug==1 vs. drug==0
controlling for time (by clicks)

Overall Mantel-Haenszel estimate, controlling for time

RR	chi2	P>chi2	[95% Conf. Interval]	
0.047	6.32	0.0119	0.002	1.436

คำสั่ง `stmc` นี้เป็นคำสั่งที่ให้ค่า `RR` adjusted for small time intervals กล่าวคือควบคุมผลของการเปลี่ยนแปลงอัตราตลอดช่วงที่ศึกษา

3.2.8 อัตราส่วนของ Median survival time พร้อมช่วงเชื่อมั่น

ค่านี้ไม่มีคำสั่งใน STATA แต่คำนวณด้วยมือได้ไม่ยากนัก ตามสูตรต่อไปนี้ (ผู้อ่านศึกษาเพิ่มเติมใน Altman, 1991 หน้า 385)

Ratio of the two median survival time = m_1 / m_2

95%Confidence interval อยู่ระหว่าง $\frac{m_1}{m_2} - e^{-1.96s}$ ถึง $\frac{m_1}{m_2} + e^{-1.96s}$

$$\text{เมื่อ } s = \sqrt{\frac{1}{o_1} + \frac{1}{o_2}}$$

3.2.9 ผลต่างของ Survival probability ระหว่างกลุ่ม

- เลือกเอา ณ จุดเวลาใดเวลาหนึ่งที่มีความสำคัญ ทั้งนี้อาจเลือกจุดเวลาได้มากกว่า 1 จุดเวลา แต่ยิ่งมากยิ่งมีปัญหาทางสถิติ จากนั้นคำนวณช่วงเชื่อมั่น
- การคำนวณ สามารถใช้ค่าที่ได้จากคำสั่ง `sts list` ใน STATA แล้วคำนวณจากสูตรต่อไปนี้ (Altman, 1991 หน้า 384)

ผลต่างของ Survival probability = $P_1 - P_2$

95%Confidence interval อยู่ระหว่าง

$$(P1 - P2) - [1.96SE(P1 - P2)] \text{ ถึง } [(P1 - P2) + 1.96SE(P1 - P2)]$$

$$\text{เมื่อ } SE(P1 - P2) = \sqrt{SE(P1)^2 + SE(P2)^2}$$

3.2.10 Hazard ratio พร้อมช่วงเชื่อมั่น (จาก Altman, 1991 หน้า 383)

$$\text{Hazard ratio} = (O1/E1) / (O2/E2)$$

$$95\% \text{Confidence interval อยู่ระหว่าง } K-1.96/\sqrt{V} \text{ ถึง } K+1.96/\sqrt{V}$$

$$\text{เมื่อ } K = (O1/E1) / V$$

$$V = \sum (n1 \times n2 \times d) (n-d) / [n^2 \times (n-1)]$$

คำนวณจากตารางหา Log-rank statistics (ดูรายละเอียดการคำนวณจาก Altman, 1991, หน้า 380, Table 13.5)

Hazard ratio เป็น Estimated relative risk

$$\text{จากตัวอย่าง Hazard ratio} = (2/4.87) / (5/2.13) = 0.17$$

V คำนวณไม่ได้จากข้อมูลชุดนี้เพราะข้อมูลไม่เพียงพอ กล่าวคือ เดือนที่ 48 ซึ่งเป็นเดือนสุดท้าย มี Number at risk เพียง 1 ราย ตัวหารของสูตรหาค่า V จึงเป็น 0 ทำให้คำนวณไม่ได้

อย่างไรก็ตาม Hazard ratio โดยวิธีนี้อยู่บนพื้นฐานการศึกษาที่ไม่มี Censoring ซึ่งไม่เป็นจริงตามกรณีตัวอย่าง แต่นำเสนอเพื่อช่วยให้เข้าใจ Hazard ratio จากวิธีการทางสถิติที่สลับซับซ้อนขึ้น ได้ง่ายขึ้น

ที่เหมาะสมคือ Hazard ratio ที่คำนวณจากวิธี Cox regression หาได้จากคำสั่ง `stcox` ดังต่อไปนี้ ส่วนรายละเอียดเกี่ยวกับวิธีการดังกล่าว จะกล่าวในบทต่อไป

3.2.11 Cox regression analysis

วิธีการวิเคราะห์จากข้อ 1 - 10 ที่กล่าวแล้วทั้งหมดเป็น Survival analysis ที่ไม่อาศัยโมเดลทางสถิติ วิธีการวิเคราะห์สุดท้ายที่นำเสนอคือ Cox regression analysis

ก่อนอื่นขอเสนอรูปแบบการวิเคราะห์ที่ยังคงดำเนินตามทุกข้อที่กล่าวแล้วข้างต้น คือเปรียบเทียบระหว่างสองกลุ่ม ดังนี้

คำสั่ง:

```
. stcox drug
```

```
      failure _d:  dead
analysis time _t:  time
```

```
Iteration 0:  log likelihood = -13.379792
Iteration 1:  log likelihood = -10.306472
Iteration 2:  log likelihood = -10.300562
Iteration 3:  log likelihood = -10.300554
```

```

Refining estimates:
Iteration 0: log likelihood = -10.300554

Cox regression -- Breslow method for ties

No. of subjects =      12      Number of obs   =      12
No. of failures =       7
Time at risk    =     243

Log likelihood   =  -10.300554      LR chi2(1)      =      6.16
                                      Prob > chi2      =     0.0131

```

	t	_d_	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
drug			.098265	.1093883	-2.084	0.037	.0110878 .8708715

จากผลข้างต้น สรุปได้ว่ากลุ่มที่ได้รับยาใหม่ (DRUG=1) มีโอกาสที่จะเสียชีวิต (DEAD=1) เป็น 0.1 เท่าของกลุ่มที่ได้รับยาเก่า (DRUG=0) รายงานผลเป็น Hazard ratio = 0.10 (95%CI: 0.01 ถึง 0.87) ซึ่งมีนัยสำคัญทางสถิติ (p-value = 0.037) เพื่อเข้าใจความหมายง่ายขึ้น ผู้วิจัยอาจสลับกลุ่มเปรียบเทียบ จะกล่าวใหม่ได้ว่า กลุ่มที่ได้รับยาเก่ามีโอกาสที่จะเสียชีวิตเป็น 10 เท่าของกลุ่มที่ได้รับยาใหม่ (ได้จาก $1/0.1$ เรียกว่าค่า Reciprocal ของ Hazard ratio)

จากตัวอย่างข้างต้น ตัวแปรต้นมีเพียงตัวเดียวคือ DRUG ถ้าหากการศึกษาครั้งนั้นทำใน 2 โรงพยาบาล และยังมีการเก็บข้อมูลผู้ป่วยเกี่ยวกับอายุ (AGE) เพศ (SEX) และความรุนแรงของโรค (SEVERITY) ตัวแปรเหล่านี้คือตัวแปรที่อาจส่งผลให้การสรุปผลเกี่ยวกับประสิทธิภาพของยาในการรักษาโรคไม่ถูกต้องได้ เช่นกลุ่มที่ได้รับยารุนาน有一部分มากอายุมากกว่าอีกกลุ่มหนึ่งและยาได้ผลดีกับคนอายุน้อย อัตราตายในกลุ่มนี้ย่อมต่ำกว่าโดยที่มิได้เป็นผลจากยาแต่จากอายุเป็นต้น Cox Regression เอื้อให้วิเคราะห์โดยคำนึงถึงผลจากตัวแปรต้นหรือปัจจัยได้คร่าวละหลายปัจจัยพร้อม ๆ กันได้ วิธีนี้มีชื่อว่า Proportional hazards regression analysis

Hazard function สามารถแปลความหมายเป็นความเสี่ยงต่อการเกิดเหตุการณ์ ณ เวลา t

Cox regression สามารถวิเคราะห์ข้อมูลได้คล้ายกันกับ Multiple regression เพียงแต่ตัวแปรตามใน Multiple regression เป็นตัวแปรต่อเนื่อง ส่วนใน Cox regression ตัวแปรตามเป็น ความเสี่ยงต่อการเกิดเหตุการณ์ ณ เวลา t

ถ้ามีตัวแปรต้น p ตัว แทนด้วย $X_1, X_2, X_3, \dots, X_p$ สามารถเขียน Model ได้ดังนี้

$$h(t) = h_0(t) \times \text{EXP}(b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_pX_p)$$

เมื่อ

$h_0(t)$ เป็น Baseline หรือ Underlying hazard function ได้จากเมื่อให้ค่า X ทุกตัวเป็น 0 [เพราะ $\text{EXP}(0)$ เท่ากับ 1]

$\text{EXP}(\theta)$ คือ Exponential ของค่า θ อาจเขียนใหม่เป็น e^θ ซึ่งเท่ากับ 2.71828^θ

สังกรทราบคือ $\text{Log}_e[\text{EXP}(\theta)] = \text{Ln}[\text{EXP}(\theta)] = \theta$

b_p คือค่าสัมประสิทธิ์ (Coefficient) ที่ได้จากการ Fit model

จากสมการข้างต้น จะเห็นว่า Hazard ของ X คือ Exponential ของ bX ดังนั้น อาจกล่าวอีกนัยหนึ่งได้ว่า bX ก็คือ Log Hazard ของ X นั่นเอง

ค่า b คือค่าสัมประสิทธิ์ คำนวณด้วยวิธีการที่สลับซับซ้อนโดยอาศัยคอมพิวเตอร์ เรียกว่า Fit Model วิธีการที่ใช้ประมาณค่า b เรียกว่า Partial maximum likelihood

ค่า Likelihood (L) คือค่าความน่าจะเป็นที่ค่าสังเกตจะถูกอธิบายโดย Model ค่ายิ่งสูง ยิ่งดี ปกติผลจากคอมพิวเตอร์รายงานค่านี้เป็นสเกล Log เรียก Log-likelihood เพราะง่ายที่จะนำไปคำนวณขั้นต่อไป ด้วยค่าความน่าจะเป็นอยู่ระหว่าง 0 - 1 ค่าที่สูงคือค่าที่เข้าใกล้ 1 เมื่อเป็นค่า Log แล้ว ค่าที่เข้าใกล้ 1 จะได้ค่า Log เข้าใกล้ 0 และ Log ของค่าที่น้อยกว่า 1 มีค่าติดลบเสมอ ดังนั้น ค่า Log-likelihood จึงเป็นค่าที่ติดลบเสมอ และค่ายิ่งใกล้ 0 ยิ่งดี

การได้มาซึ่งค่า b นั้นเป็นการเปรียบเทียบ Likelihood ของ Model ที่กำหนดให้ตัวแปรต้นทุกตัวไม่มีผลต่อ Model เขียนแทนด้วย $L(0)$ กับ Likelihood ของ Model ที่ตัวแปรต้นทุกตัวมีผลต่อ Model เขียนแทนด้วย $L(b)$ แล้วคำนวณหาค่า $L(b)$ ที่มากที่สุดเท่าที่จะเป็นไปได้ที่ยังผลให้ได้ค่าสัดส่วนระหว่าง $L(0)/L(b)$ หรือ Likelihood ratio นี้มีค่าต่ำสุด จึงเรียกว่า Maximum likelihood

ค่าของ Linear combination คือ $b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_pX_p$ ของคนคนหนึ่ง หาได้จากการแทนค่า X ต่างๆ ลงในสมการที่ได้จากการ Fit model ค่าที่ได้เรียกว่า Prognostic Index หรือ Risk score มีค่าเท่ากับ Log Hazard

ภายใต้ Model นี้มี Assumption เกี่ยวกับ Proportional hazard เป็นหลัก กล่าวคือความแตกต่างสัมพัทธ์หรือสัดส่วนของ Log hazard ระหว่างกลุ่มนั้น เท่ากันตลอดช่วงเวลาที่ศึกษา ไม่ว่าจะมีความถี่ใดก็ตาม จะเพิ่มหรือลดเมื่อเวลาเปลี่ยนไป หรือเพิ่มแล้วลด หรือลดแล้วเพิ่ม ขอเพียงแต่ต้องเป็นสัดส่วนคงที่ระหว่างกลุ่มหรือหน่วยศึกษาทุกหน่วย

นั่นคือ Cox proportional hazard ไม่มีการระบุลักษณะของ Baseline hazard แต่ให้คงที่ (Baseline hazard is left unparameterized but fixed) แต่ยังคงมีการประมาณค่าพารามิเตอร์ของสัมประสิทธิ์ จึงเรียกว่า เป็น Semiparametric model แบบหนึ่ง

ถ้าเป็น Parametric model ต้องมีการระบุลักษณะของ Baseline hazard ตามเวลาที่เปลี่ยนไป

$h_0(t)$ คงที่ เรียกว่า Exponential

$h_0(t)$ เป็น Monotone คือเพิ่มตามเวลา หรือลดตามเวลา เรียกว่า Weibull

$h_0(t)$ เป็น Nonmonotone คือเพิ่มแล้วลด หรือลดจากนั้นเพิ่ม เรียกว่า Log-normal

เรื่อง Parametric model สำหรับ Survival analysis มีรายละเอียดเพิ่มเติมท้ายบทนี้

Hazard บอกความเสี่ยงต่อการเกิดเหตุการณ์ ณ เวลา t ถ้าต้องการทราบความเสี่ยงต่อการเกิดเหตุการณ์ระหว่างเวลา 0 ถึง เวลา t ก็เพียงแต่รวมค่า hazard ทุกจุดเวลาจนถึงเวลา t เรียกว่า Cumulative hazard เขียนแทนด้วย $H(t)$ ดังนั้น Model สามารถเขียนใหม่เป็น

$$H(t) = H_0(t) \times \text{EXP}(b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_pX_p)$$

จากที่กล่าวข้างต้น เราทราบว่า โอกาสที่อยู่อย่างปลอดภัยเหตุการณ์จนถึงเวลา t ซึ่งเขียนแทนด้วย $S(t)$ นั้น สามารถหาได้จาก $\text{EXP}[-H(t)]$ ดังนั้นเราจึงสามารถหาโอกาสที่คนคนหนึ่งจะอยู่อย่างปลอดภัยเมื่อเราทราบค่าตัวแปรต้นต่างๆ ที่เป็นข้อมูลของเขา

การแปลความหมายจาก

$$h(t) = h_0(t) \times \text{EXP}(b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_pX_p)$$

1. สามารถดูที่ขนาดของค่า b ค่ายิ่งมาก บ่งชี้ว่ายิ่งมีโอกาสเกิดเหตุการณ์ได้มาก Exponentiate ค่า b เรียกว่า Hazard แปลความหมายคล้ายอัตรา (Rate)
2. เปรียบเทียบค่า b ระหว่างค่า X สองค่า เช่น เปรียบเทียบระหว่างชาย ($X_1=1$) กับหญิง ($X_1=0$) โดยที่ให้ค่าตัวแปรอื่นๆ ทุกตัวคงที่ (Unspecified but fixed) ได้ดังนี้

$$\frac{h_1(t)}{h_2(t)} = \frac{h_0(t) \times \text{EXP}[b_1(X_1=1) + b_2X_2 + b_3X_3 + \dots + b_pX_p]}{h_0(t) \times \text{EXP}[b_1(X_1=0) + b_2X_2 + b_3X_3 + \dots + b_pX_p]}$$

$$= \text{EXP}[b_1(1-0)] = \text{EXP}[b_1(1-0)] = \text{EXP}[b_1]$$

นี่คือค่า Hazard Ratio

โปรดสังเกตว่า Hazard ratio ไม่ได้ขึ้นอยู่กับเวลาว่า ณ จุดเวลาใด แต่ให้เป็นไปตาม Assumption คือจะเวลาใดก็ตาม Hazard ของทั้งชายและหญิงเป็นสัดส่วนกันคงที่เสมอ จะเพิ่มก็เพิ่มทั้งคู่ที่เป็นสัดส่วนสัดส่วนหนึ่ง จะลดก็ลดทั้งคู่ที่เป็นสัดส่วนหนึ่ง ทั้งนี้สัดส่วนดังกล่าวต้องเท่ากัน

นอกจากนั้นจะเห็นได้ว่า การหาค่า Hazard ratio เราไม่จำเป็นต้องทราบ Baseline hazard function หรือ $h_0(t)$ เพราะหารกันได้เท่ากับ 1

การแปลความหมาย Hazard ratio ทำคล้ายกับความเสี่ยงสัมพัทธ์ (Relative Risk) โดยแปลความหมายสัมพันธ์กับการเปลี่ยนแปลงของค่า X ถ้า X เป็นตัวแปรต่อเนื่องเช่น อายุ (เป็นปี) แล้ว

ถ้าได้ Hazard ratio เท่ากับ 1.02 หมายถึง "ทุก 1 ปีที่อายุเพิ่มขึ้น ความเสี่ยงต่อการเกิดเหตุการณ์เพิ่มขึ้น 2%"

ถ้าได้ Hazard ratio เท่ากับ 0.98 หมายถึง "ทุก 1 ปีที่อายุเพิ่มขึ้น ความเสี่ยงต่อการเกิดเหตุการณ์ลดลง 2%"

ถ้าได้ Hazard ratio เท่ากับ 4.02 หมายถึง "ทุก 1 ปีที่อายุเพิ่มขึ้น ความเสี่ยงต่อการเกิดเหตุการณ์คิดเป็น 4.02 เท่าของเมื่อก่อนหน้านี้ หรือค่าความเสี่ยงเป็นผลคูณของ 4.02 นั่นเอง"

การเปรียบเทียบระหว่างกลุ่มเช่นเพศ (0=หญิง 1=ชาย) ก็เช่นเดียวกัน ถ้าได้เท่ากับ 1.02 หมายถึง "ชายความเสี่ยงต่อการเกิดเหตุการณ์มากกว่าหญิง 2%" ถ้าได้เท่ากับ 0.98 หมายถึง "ชายความเสี่ยง"

ต่อการเกิดเหตุการณ์น้อยกว่าหนึ่ง 2%” ถ้าได้เท่ากับ 4.02 หมายถึง “ความเสี่ยงต่อการเกิดเหตุการณ์มากกว่าหนึ่งเป็น 4.02 เท่า”

ค่า Hazard ratio เท่ากับ 1 หมายถึง X ไม่มีผลต่อ Survival ถ้าน้อยกว่า 1 หมายถึงการได้รับ X มีผลป้องกันไม่ให้เกิดเหตุการณ์ หรือเอื้อให้มี Survival เรียกว่ามีผลเป็น Protective effect ตรงข้าม ถ้ามากกว่า 1 หมายถึงการได้รับ X ก่อให้เกิดเหตุการณ์ หรือเอื้อให้มี Failure เรียกว่ามีผลเป็น Risk effect

การหา Survival probability ต้องทราบค่าตัวแปรต้นทุกตัวแปรใน Model และต้องรู้ค่า $h_0(t)$ ด้วย ค่าดังกล่าวได้จากคอมพิวเตอร์

กรณีตัวแปรต้นเป็นตัวแปรต่อเนื่อง เช่นอายุ (ปี) แปลความหมายว่าอายุเพิ่มขึ้น 1 ปี ค่า Log hazard เพิ่มขึ้นเท่ากับ b และด้วยเหตุที่ Assumption เป็นความสัมพันธ์เชิงเส้นตรง (Linear relationship) อายุเพิ่มจาก 40 ไปเป็น 41 ปี จะให้ค่า Log hazard ที่เปลี่ยนไปเท่ากับอายุที่เพิ่มจาก 45 ไปเป็น 46 ปี

การเปรียบเทียบ Hazard function ในที่นี้ จะคล้ายกันกับการเปรียบเทียบ Survival probability ตามที่กล่าวแล้วข้างต้น ต่างกันตรงที่วิธีหลังนี้ได้ควบคุมผลกระทบจากปัจจัยอื่นๆที่อยู่ใน Model แล้ว

หากต้องการสร้าง Prognostic Index (PI) เพื่อนำไปทำนายเหตุการณ์ที่ศึกษาในประชากรกลุ่มอื่น สามารถเขียนเป็น Prognostic model ดังนี้

$$PI = b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_pX_p$$

ค่า PI ยิ่งมาก หมายถึง Hazard ยิ่งมาก นั่นคือ Survival ยิ่งสั้นลง ในทางตรงข้าม ถ้าค่า PI ยิ่งน้อย หมายถึง Hazard ยิ่งน้อยเช่นกัน และบอกให้รู้ว่า Survival ยิ่งยาวขึ้น แต่แปลความหมายมากนักเท่าไรจึงถือว่ามาก หรือน้อยนั้น ทำได้ยาก จึงต้องนำไปคำนวณค่าสถิติที่แปลความหมายได้ง่ายต่อไป

ถ้าต้องการรู้ Hazard ณ เวลาใดๆ ของใคร ก็นำลักษณะคนคนนั้น (ค่า X ต่างๆ) แทนค่าใน Model ได้ค่า PI

ถ้าต้องการรู้ Survival probability ณ เวลาใดๆ ของใครก็ทำเช่นเดียวกัน แต่แทนค่าลงใน Model ต่อไปนี้

$$S(t) = \text{EXP}[-H_0(t) \times \text{EXP}(PI)] \text{ หรือ } S(t) = S_0(t)^{\text{EXP}(PI)}$$

ค่า $-H_0(t)$ หรือ $S_0(t)$ ได้จากผลของการ Fit model โดยใช้คอมพิวเตอร์ (ดูบทที่ 4 ข้อ 4.9)

3.2.1.2 Parametric model for survival analysis

จากที่กล่าวแล้วข้างต้น Cox proportional hazard ไม่มีการระบุลักษณะของ Baseline hazard แต่ให้คงที่ (Baseline hazard is left unparameterized but fixed) แต่ยังคงมีการประมาณค่าพารามิเตอร์ของสัมประสิทธิ์ จึงเรียกว่าเป็น Semiparametric model แบบหนึ่ง

ด้วยเหตุที่ Cox regression ใช้ข้อมูลเฉพาะเวลาที่เกิดเหตุการณ์ (Failure) ในการเปรียบเทียบโอกาสเกิดเหตุการณ์ระหว่างหน่วยตัวอย่าง

เพื่อการคำนวณค่าต่างๆ (ได้แก่ค่าสัมประสิทธิ์ b) ในบางกรณีข้อมูลไม่ได้ถูกใช้เช่นค่าความดันโลหิต (X) ที่เปลี่ยนแปลงไปก่อนตาย (Event) หรือที่เรียกว่า X เป็น Time varying covariate นั้น Cox regression จะใช้ข้อมูลเฉพาะวันที่ตายแม้มีการวัดความดันโลหิตทุกวัน จึงมีประสิทธิภาพต่ำกว่าวิธีการวิเคราะห์ที่ใช้ข้อมูลทุกค่า ดังนั้น Parametric model จึงเข้ามามีบทบาทนี้ เพราะไม่มีการเปรียบเทียบระหว่างหน่วยตัวอย่าง หากแต่คำนวณโอกาสเกิดเหตุการณ์ในแต่ละคนเพื่อการคำนวณค่าต่างๆ จึงมีการใช้ข้อมูลทุกค่า

Parametric model ต้องมีการระบุลักษณะของ Baseline hazard ตามเวลาที่เปลี่ยนไป

$h_0(t)$ คงที่ เรียกว่า Exponential regression

$h_0(t)$ เป็น Monotone คือเพิ่มตามเวลา หรือลดตามเวลา เรียกว่า Weibull regression ซึ่งให้ Baseline hazard function ที่หลากหลาย ในลักษณะ Monotone hazard rate

$h_0(t)$ เป็น Nonmonotone คือเพิ่มแล้วลด หรือลดจากนั้นเพิ่ม เรียกว่า Log-normal regression

Parametric model ประเภทอื่นๆ ได้แก่

- Gompertz regression เป็นวิธีการเก่าแก่ ใช้สำหรับ Modeling ข้อมูลการตาย เหมาะสำหรับ Monotone hazard rate เช่นกันกับ Weibull regression แต่การเพิ่มหรือลดลงมีลักษณะเป็น Exponential ตามเวลา ค่าพารามิเตอร์ G เป็นตัวระบุลักษณะนี้ ถ้าค่าเป็นบวกคือ Hazard rate เพิ่มขึ้นตามเวลา ถ้าค่า

เป็นลบคือ Hazard rate ลดลงตามเวลา ถ้าค่าเป็นศูนย์ คือ Hazard rate คงที่ตามเวลา ก็จะกลายเป็น Exponential regression ไปทันที

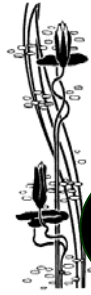
- Log-logistic regression มีวิธีการคล้าย Log-normal regression แต่มีความง่ายในเชิงคณิตศาสตร์มากกว่า
- Generalized log gamma regression เป็นวิธีการที่เอื้อให้ระบุลักษณะ Hazard rate ที่หลากหลาย จึงมักนำมาใช้วิเคราะห์เพื่อหาทางเลือก Parametric model ที่เหมาะสม

ในทางปฏิบัติ ลักษณะของ Baseline hazard มักไม่ทราบ ดังนั้นควรตรวจสอบความถูกต้องของค่า b โดยเปรียบเทียบค่า b จาก Cox proportional hazard เสมอ ถ้าไม่ตรงกัน แสดงว่าเราระบุลักษณะ Baseline hazard ไม่ถูกต้อง ต้องระบุใหม่ หรือต้องกลับมาใช้ Cox proportional hazard

หลักในการเลือก Parametric model หรือ Cox proportional hazard คือ ถ้าหากรู้ลักษณะของ Hazard ในเรื่องที่ศึกษาวิจัย และต้องการได้ค่า b ที่ได้มาจากทุกค่าข้อมูล และได้ค่า Baseline hazard ที่สะท้อนลักษณะข้อมูล ให้เลือกใช้ Parametric model ถ้าไม่เช่นนั้น ให้เลือกใช้ Cox proportional hazard

การที่จะรู้ลักษณะของ Hazard ต้องใช้ความรู้เกี่ยวกับลักษณะของ Hazard ในเรื่องที่ศึกษาวิจัย และโดยปกติจะไม่ทราบ แต่ Parametric model มีความไวมากต่อการระบุลักษณะของ Baseline hazard ถ้าระบุไม่เหมาะสมจะได้ค่าที่ผิดพลาดมาก ในขณะที่ Cox proportional hazard ไม่ต้องระบุลักษณะดังกล่าว จึงได้ค่าที่ถูกต้องในทุกกรณี เพียงแต่ Baseline

hazard ที่ได้ไม่สะท้อนลักษณะข้อมูลอย่างเต็มประสิทธิภาพเท่านี้เอง แต่ก็
ไม่ถึงกับผิด ดังนั้น Cox proportional hazard จึงเป็นทางเลือกที่ดีเสมอ



4 **แนวทางการวิเคราะห์ขั้นสูง :** **บทสรุปเนื้อหาที่สำคัญ**

ต่อไปนี้เป็นสรุปขั้นตอนการวิเคราะห์ ข้อ 4.1 ถึง 4.6 ทำคล้ายกับที่ได้กล่าวมาแล้วจึงไม่มีรายละเอียด ข้อ 4.7 เป็นต้นไปมีสรุปประเด็นสำคัญ และทั้งหมดจะนำเสนอด้วยตัวอย่างโดยละเอียด ซึ่งนำเสนอถัดจากการสรุปนี้

ขั้นตอนการวิเคราะห์ สรุปได้ดังนี้

4.1 ตรวจสอบข้อมูล

4.2 เปรียบเทียบลักษณะพื้นฐานของสองกลุ่มที่ศึกษา

4.3 ทำการ Set ข้อมูลเพื่อการวิเคราะห์ Survival analysis

4.4 พรรณนารูปแบบการศึกษา

4.5 พรรณนา Overall survival experience

4.6 ทำการวิเคราะห์อย่างหลาย (Bivariate analysis)

- หาความสัมพันธ์ระหว่างแต่ละตัวแปรกับระยะปลอดเหตุการณ์
- กรณีตัวแปรแรงนับให้วิเคราะห์ดังนี้

(1) Kaplan–Meier survival curve จำแนกตามกลุ่ม

(2) Median survival time และช่วงเชื่อมั่นจำแนกตามกลุ่ม

- (3) ค่า p-value จากการทดสอบความแตกต่างระยะปลอดเหตุการณ์ระหว่างกลุ่ม โดยเลือกใช้ Log-rank test หรือ Wilcoxon test ตามที่เห็นว่าเหมาะสม
- (4) Hazard ratio และช่วงเชื่อมั่น พร้อมกับ ค่า p-value จาก Partial likelihood ratio test
- กรณีตัวแปรต่อเนื่องให้จัดกลุ่มเป็นกลุ่มควอร์ไทล์หรือตามที่ให้ความหมายทางการแพทย์และสาธารณสุขก่อน หลังจากนั้นให้วิเคราะห์เช่นเดียวกับกรณีตัวแปรแจกแจงนับ ทั้งนี้การจัดกลุ่มดังกล่าวทำเพื่อการพรรณนาในขั้นนี้เท่านั้น กรณีไม่มี Non-linear relationship ที่ชัดเจนให้วิเคราะห์เพื่อหา Hazard ratio และช่วงเชื่อมั่น โดยไม่ต้องจัดกลุ่มร่วมด้วย

สรุปผล *Bivariate analysis*

- (1) ข้อมูล Categorical ทั้งหมด รวมทั้ง Continuous ที่จัดกลุ่มด้วย

ตัวแปร	Median time (95%CI)	Person- times	IR/100	HR (95%CI)	p-value*

หมายเหตุ: * = p-value จาก partial likelihood ratio test

(2) ข้อมูล Continuous ที่ยังไม่จัดกลุ่ม

ตัวแปร	ค่าที่เปลี่ยนแปลง	HR (95%CI)	p-value*

หมายเหตุ: * = p-value จาก partial likelihood ratio test

4.7 ทำการวิเคราะห์ตัวแปรเชิงพหุ (Multivariable analysis)

4.7.1 The initial multivariable model

- ทุกตัวแปรที่มีความสำคัญ
- ทุกตัวแปรที่ p-value < 0.25
- จำนวนตัวแปรไม่ควรมากเกินไปกว่า 1 ใน 10 เท่าของ Failure
- ตัวแปร Dichotomous ใส่เข้าโมเดลโดยตรง
- ตัวแปร Polytomous ต้องสร้าง Dummy variables ก่อนใส่เข้าโมเดล
- ตัวแปร Continuous ใส่เข้าโมเดลโดยตรงในรูปข้อมูลที่ยังไม่ถูกจัดกลุ่ม

4.7.2 หลักการตัดหรือคงไว้ซึ่งตัวแปร (เป้าหมายเพื่อ Simplify model)

- ตัดออกคราวละตัวแปร ยกเว้น Dummy variable ต้องตัดออกเป็นชุด

- ใช้ p-value จาก Wald test เป็นตัวกรองเบื้องต้น โดยตัวแปรที่ให้ค่า p-value เป็น Candidate ที่จะทดลองนำออกก่อน
- ใช้ p-value จาก Partial likelihood ratio test ทดสอบว่าการนำตัวแปรดังกล่าวออกนั้นมีนัยสำคัญทางสถิติต่อ Model หรือไม่
- ใช้ขนาดการเปลี่ยนแปลงของค่า Coefficient เป็นตัวตัดสินขั้นสุดท้าย โดยการเปรียบเทียบค่า Coefficient ของตัวแปรอื่นๆ ใน Model ที่มีกับที่ไม่มีตัวแปรที่ตัดออกไปนั้น ว่ามีค่าใดค่าหนึ่งเปลี่ยนแปลงเกิน 20% โดยประมาณหรือไม่ ถ้าไม่ก็ตัดออกไปได้ แต่ถ้าเกินก็ต้องนำตัวแปรนั้นกลับคืน Model ค่าร้อยละของการเปลี่ยนแปลงดังกล่าว หาได้จาก ค่า Coefficient จาก Model ที่ไม่มีตัวแปร ลบด้วยค่า Coefficient จาก Model ที่มีตัวแปร แล้วนำผลที่ได้หารด้วยค่า Coefficient จาก Model ที่มีตัวแปร จากนั้นคูณผลที่ได้ด้วย 100
- กรณี Dummy variable ถ้าหากบางตัวแปรมีค่า p-value จาก Wald test มากกว่า 0.05 บางตัวแปรน้อยกว่า 0.05 อาจทดลองรวมตัวแปรให้ Model ง่ายขึ้น
- ทำเช่นนี้จนครบทุกตัวแปรที่มีค่า p-value จาก Wald test มากกว่า 0.05 ใน Initial model กล่าวคือ จนกระทั่งไม่มีตัวแปรใดสามารถนำออกไปจาก Model ได้อีกแล้ว
- นำทุกตัวแปรที่ถูกตัดออกไปกลับเข้าไปใน Model อีกครั้งเพื่อตรวจสอบดูให้แน่ใจว่าตัวแปรเหล่านั้นไม่มีนัยสำคัญทางสถิติต่อ Model (ดูจาก p-value จาก Partial likelihood ratio test) และไม่เป็น Confounder (ดูจากการเปลี่ยนแปลงค่า Coefficient)

- บางตัวแปรต้องคงไว้ใน Model แม้ไม่มีผลใดๆเลยต่อ Model ถ้าหากเป็นตัวแปรที่เป็นคำถามวิจัยหรือเกี่ยวข้องกับการวางแผนการศึกษา

4.7.3 Scale selection กรณีตัวแปร Continuous

ตัวแปร Continuous ที่อยู่ใน Model เพื่อให้ได้ Assumption เกี่ยวกับความสัมพันธ์เชิงเส้นกับ Log hazards ใช้ทั้ง 3 วิธีต่อไปนี้ร่วมกันเพื่อเลือกจะทำให้ตัวแปรอยู่ใน Model ในรูป 1) Single linear term หรือ 2) Categorical variable โดยใช้ cut-off point ที่เหมาะสม หรือ 3) Best term fractional polynomial ข้อ 3) มีกล่าวถึงในที่นี้ อย่างไรก็ตามผู้อ่านสามารถศึกษารายละเอียดเพิ่มเติมใน Hosmer and Lemeshow (1999) หน้า 172 - 177

4.7.3.1 ใช้วิธี Dummy variables

- Fit model โดยแทนที่ทุกตัวแปร Continuous ด้วย Dummy variables ของแต่ละตัวแปร (ตามกลุ่มควอร์ไทล์ที่ทำใน Descriptive analysis)
- คำนวณค่า Midpoints ของแต่ละกลุ่ม
- แต่ละตัวแปร ให้นำเสนอกราฟเส้นแสดงความสัมพันธ์ระหว่างค่า Coefficient (ให้เป็นแกนตั้ง ให้ชื่อแกนว่า Log Hazard ก็ได้เพื่อช่วยต่อการทำความเข้าใจ) กับค่า Midpoints ของแต่ละกลุ่มควอร์ไทล์ (ให้เป็นแกนนอน ให้ชื่อแกนเป็นชื่อตัวแปร)
- กราฟเส้นที่เป็นเส้นตรงโดยประมาณ ถือว่ามี Linear relationship สามารถนำเข้าไปใน Model ในรูป Continuous variable ได้ ในทางตรงข้าม ถ้ามีรูปแบบไม่เป็นเชิงเส้นชัดเจน ต้องจัดกลุ่มตัวแปร และเข้าไปใน Model ในรูป Categorical variable

4.7.3.2 ใช้วิธี Fractional polynomial

- Fit Model ที่ได้ตามที่กล่าวข้างต้น แต่ไม่มีตัวแปรที่จะตรวจสอบ กล่าวอีกนัยหนึ่งคือให้ Coefficient = 0 นั่นเอง
- Fit Model ที่ได้ตามที่กล่าวข้างต้น แต่ให้มีตัวแปรที่จะตรวจสอบในรูป Continuous
- คำนวณ Partial likelihood ratio test แล้วหาค่า p-value เป็นค่าสำหรับผลการนำเข้า Model ในรูป Continuous (หรืออาจเรียกว่า Linear term)
- หา The best power โดยใส่ตัวแปรนั้นในรูปรากกำลังสอง แล้วคำนวณ Partial likelihood ratio test จากนั้นหาค่า p-value เป็นค่าสำหรับผลการนำเข้า Model ในรูป Linear term + Square root term เปรียบเทียบกับ Model ที่ใส่เป็น Linear term อย่างเดียว (Degree of freedom = 2)
- หา The best two terms โดยใส่ตัวแปรนั้นในรูป X^{-1} และ $X^{-1}\ln(X)$ แล้วคำนวณ Partial likelihood ratio test จากนั้นหาค่า p-value เป็นค่าสำหรับผลการนำเข้า Model ในรูป Linear term + Square root term + X^{-1} term + $X^{-1}\ln(X)$ term เปรียบเทียบกับ Model ที่ใส่เป็น Linear term อย่างเดียว (Degree of freedom = 4)

4.7.4 นำทุกตัวแปรที่ถูกตัดออกไปกลับเข้าไปใน **Model** อีกครั้ง

เพื่อตรวจสอบดูให้แน่ใจว่าตัวแปรเหล่านั้นไม่มีนัยสำคัญทางสถิติ ต่อ Model (ดูจาก p-value จาก Partial likelihood ratio test) และไม่เป็น

Confounder (ดูจากการเปลี่ยนแปลงค่าCoefficient) ตัวแปรเหล่านั้นจะไม่ต้องคงไว้ใน Model ก็ต่อเมื่อ

- ค่า p-value ของ Likelihood ratio test > 0.05
- ค่าการเปลี่ยนแปลงของ Coefficient ไม่มากกว่า 20%

4.7.5 การพิจารณา Interaction effect

- สร้างรายการ Interaction term ที่เป็นไปได้ (ไม่จำเป็นต้องเป็น all possible combination) จากตัวแปรเดี่ยวๆ (Main effect) ทุกตัวที่อยู่ใน Model ที่ได้ก่อนหน้านี้
- ควรมีเพียงสองตัวแปรในหนึ่ง Interaction term เว้นแต่จะมีข้อบ่งชี้ชัดเจนว่าต้องมีถึงสามตัวแปร เพราะสามารถเกิดปัญหาได้ง่ายในการ Fit model และยากต่อการแปลความหมาย
- เลือกเพียงรายการ Interaction term ที่มีความสำคัญทางการแพทย์และสาธารณสุข นำเข้า Model ที่เฉพาะ Main effect ที่ได้ก่อนหน้านี้ โดยนำเข้าคราวละ 1 Term
- Term ที่ให้ค่า p-value ของ Likelihood ratio test < 0.05 จึงให้คงไว้ใน Model

4.7.6 Assessment of model adequacy

- ตรวจสอบ 3 ประเด็นคือ 1) Assumption 2) Influential observations และ 3) Goodness-of-fit

4.7.6.1 Assessing the proportional hazard assumption

- โดย Plot ค่า Scaled Schoendfeld Residuals กับค่า Log-Time

4.7.6.2 Identification of influential and poorly fit subjects

- พิจารณา Leverage เป็นหลัก
- Leverage เป็นมาตรวัดขนาดของ "ความไม่ปกติ" ของค่าค่าหนึ่งของตัวแปรต้นตัวใดตัวหนึ่ง มีค่าเป็นสัดส่วนตรงกับผลต่างกำลังสองระหว่างค่าค่าหนึ่งของตัวแปรต้นตัวใดตัวหนึ่งนั้นกับค่าเฉลี่ยของตัวแปรนั้น หรือ $(x - \bar{x})^2$
- ค่า Leverage ที่ส่งผลต่อการเปลี่ยนแปลงค่า Coefficient มากต้องได้รับการตรวจสอบเกี่ยวกับความถูกต้องและความเป็นไปได้ของข้อมูลรายนั้น และแก้ไขให้ถูกต้อง
- ถ้าค่านั้นเป็นข้อมูลที่ถูกต้องแล้ว และเป็นไปได้ในทางการแพทย์และสาธารณสุขให้ลอง Fit model ที่ไม่มีข้อมูลของรายนั้น แล้วตรวจสอบการเปลี่ยนแปลงค่า Coefficient
- การตัดสินใจตัดข้อมูลรายนั้นทิ้งไปควรดูที่ความเป็นไปได้ในทางการแพทย์และสาธารณสุข ถ้าไม่เป็นค่าที่ผิดปกติมากนักควรให้คงไว้ เว้นแต่ Model จะไวมากต่อการมีหรือไม่มีข้อมูลรายนั้น ก็ให้ตัดทิ้ง
- หากมีการตัดข้อมูลรายที่มีผลมากต่อ Model ทิ้งไป ต้องเริ่มวิเคราะห์ข้อมูลใหม่ทั้งหมดอีกครั้ง
- สิ่งที่จะช่วยพิจารณาว่าข้อมูลรายใดส่งผลมากต่อ Model คือกราฟของ Score residuals, Scaled score residuals, Likelihood displacement statistics, และ 1-max statistics

4.7.6.3 Overall goodness-of-fit test

4.7.7 Interpretation and presentation of final model

- Model ก่อนหน้านี้ทั้งหมด ควรดูเพียงค่า Coefficient เท่านั้น จนกว่าจะผ่านขั้นตอน Assessment of model adequacy จึงเริ่มดูค่า Hazard ratio

การนำเสนอผลจาก Final model

ก. กรณีมีเป้าหมายหลักที่ Covariate adjustment

ก-1. เสนอค่าสำหรับทุกตัวแปรที่ปรากฏใน Final model แม้ว่าคำถามวิจัยจะสนใจเพียงตัวแปรเดียวที่ควบคุมผลจากตัวแปรอื่น ๆ ทั้งนี้เพื่อให้ผู้อ่านมีข้อมูลสำหรับใช้ดุลยพินิจต่อผลการวิเคราะห์ การนำเสนอเพียงบางตัวแปรแล้วหมายเหตุท้ายตารางว่าได้ควบคุมปัจจัยอะไรบ้างจึงไม่พึงทำ

ข้อมูลต่อเนื่อง อาจพิจารณานำเสนอ Hazard ratio สำหรับการเปลี่ยนแปลงข้อมูลมากกว่าหนึ่งหน่วย เช่น ให้เป็นผลการเปลี่ยนแปลงค่า 10 หน่วย ทำได้โดยคูณค่า Coefficient ด้วย 10 ก่อน Exponential ให้เป็นค่า Hazard ratio ทั้งนี้เพื่อง่ายต่อการแปลความหมายและทำความเข้าใจ

ตัวแปร	HR _{CRUDE}	HR _{ADJUSTED}	95%CI	p-value*

หมายเหตุ: * = p-value จาก partial likelihood ratio test

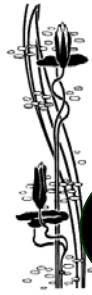
ก-2. นำเสนอ Covariate-adjusted survival curve ร่วมด้วย

ข. กรณีมีเป้าหมายหลักที่ *Prediction*

ข-1. นำเสนอ Best prediction model ที่ได้จาก Training data set โดยเสนอทั้งรูปดั้งเดิมที่ได้จากการ Fit model และรูปที่ดัดแปลงให้อยู่ในรูปที่ง่ายต่อการใช้ประโยชน์ เช่น Pocket chart ตาราง หรือกราฟ ที่เรียกว่า Nomogram เป็นต้น

ข-2. นำเสนอ Performance ของ Predictive model หรือ Risk score ที่ได้นั้นว่ามีความสามารถในการทำนาย Survival probability หรือ Hazard ได้ถูกต้องหรือไม่เมื่อนำมาใช้กับ Validating data set ซึ่งเป็นข้อมูลที่ต่างไปจากข้อมูลที่ใช้สร้าง Model

ผู้อ่านสามารถศึกษารายละเอียดเกี่ยวกับเรื่องนี้ได้ใน Kleinbaum, Kupper, Muller, and Nizam, (1998) หน้า 386-422



ตัวอย่างรวบยอด

5

แนวทางการวิเคราะห์ข้อมูล

การศึกษาเชิงทดลองเปรียบเทียบโครงการแก้ไขปัญหายาเสพติด โดยใช้ระยะเวลาต่างกันในการดำเนินการ เพื่อทราบว่าโครงการระยะสั้นหรือระยะยาวที่สามารถยืดระยะเวลาการกลับมาติดยาอีกครั้งได้นานกว่ากันต่อไปนี้เป็นรายการตัวแปร และตัวอย่างข้อมูล แฟ้มข้อมูลทั้งหมดชื่อ "UIS.DTA" สามารถดาวน์โหลดได้ที่

ftp://ftp/wiley.com/public/sci_med/survival

คำอธิบายโดยละเอียดมีใน Hosmer and Lemeshow (1998) หน้า 22 - 24 การวิเคราะห์ข้อมูลในที่นี้ดำเนินการดังนี้

- 1) แสดงการวิเคราะห์ข้อมูลอย่างเป็นระบบ
- 2) นำเสนอผลการศึกษาโดยเขียนเป็นรายงานวิจัย พร้อมกับนำเสนอตารางหรือกราฟอย่างถูกต้องเหมาะสม ทั้งนี้ควรให้อยู่ในรูปที่สามารถนำไปใส่ในรายงานวิจัย ภายใต้หัวข้อ "ผลการศึกษา" ได้

คำอธิบายตัวแปร

Variable	Description	Codes/Values
ID	Identification Code	1 - 628
AGE	Age at Enrollment	Years
BECKTOTA	Beck Depression Score	0.000 - 54.000
HERCOC	Heroin/Cocaine Use During 3 Months	1 = Heroin & Cocaine
	Prior to Admission	2 = Heroin Only
		3 = Cocaine Only
		4 = Neither Heroin nor Cocaine
IVHX	History of IV Drug Use	1 = Never; 2 = Previous;
		3 = Recent
NDRUGTX	Number of Prior Drug Treatments	0 - 40
RACE	Subject's Race	0 = White; 1 = Non-White
TREAT	Treatment Randomization Assignment	0 = Short;
		1 = Long
SITE	Treatment Site	0 = A; 1 = B
LOS	Length of Stay in Treatment (Admission	Days
	Date to Exit Date)	
TIME	Time to Drug Relapse (Measured from	Days
	Admission Date)	
CENSOR		1 = Returned to Drugs;
		0 = Otherwise

ข้อมูลตัวอย่าง

ID	AGE	BECKTOTA	HERCOC	IVHX	NDRUGTX	RACE	TREAT	SITE	LOS	TIME	CENSOR
1	39	9	4	3	1	0	1	0	123	188	1
2	33	34	4	2	8	0	1	0	25	26	1
3	33	10	2	3	3	0	1	0	7	207	1
-----ข้าม 622 records-----											
626	28	10	4	2	3	0	1	1	21	35	1
627	35	17	1	3	2	0	0	1	184	379	1
628	46	31.5	1	3	15	1	1	1	9	377	1

ตอนที่ 1 : แนวทางการวิเคราะห์ข้อมูล

คำชี้แจง:

- การวิเคราะห์ ดำเนินการเป็นลำดับที่ทำในชีวิตจริง
- คำอธิบายผลการวิเคราะห์ ได้เขียนเป็นรูปแบบการนำเสนอผลการศึกษา และนำเสนอไว้ได้ผลการวิเคราะห์ที่ได้จาก STATA และเสนอเฉพาะที่จำเป็นเพื่อให้เป็นตัวอย่าง ถ้ามีลักษณะที่ซ้ำกันจะเสนอแนะไว้เฉพาะที่ปรากฏครั้งแรกเท่านั้น
- ผลการศึกษาในรูปตารางการนำเสนอข้อมูลวิจัย ได้นำเสนอจำแนกเป็น 3 ส่วนคือ 1) Baseline characteristics; 2) Crude analysis; และ 3) Multivariable analysis และนำเสนอไว้เป็นตอน ๆ ถัดจากการวิเคราะห์ข้อมูลเพื่อเป้าหมายตามที่ระบุ
- คำสั่ง STATA ทั้งหมด ผู้อ่านสามารถคัดลอกไปใช้กับการวิเคราะห์ข้อมูลชุดอื่น ๆ ได้เลย โดยเปลี่ยนแปลงเพียงชื่อตัวแปรตามต้องการ

5.1 ตรวจสอบข้อมูล

```
. use uis.dta
```

```
. su
```

Variable	Obs	Mean	Std. Dev.	Min	Max
id	628	314.5	181.4323	1	628
age	623	32.36918	6.119534	20	56
becktota	595	17.40486	9.391132	0	54
hercoc	610	2.788525	1.100619	1	4
ivhx	610	2.047541	.9003022	1	3
ndrugtx	611	4.574468	5.494462	0	40
race	622	.2491961	.4328957	0	1
treat	628	.4904459	.5003072	0	1
site	628	.2929936	.4554983	0	1
los	628	96.14013	77.30519	1	400
time	628	234.7038	201.7011	2	1172
censor	628	.8089172	.3934675	0	1

จากข้อมูลทั้งหมด 628 ราย มีบางตัวแปรที่มี Missing data เฉพาะอย่างยิ่ง BECKTOTA มีเพียง 595 ราย นอกจากนั้นข้อมูล TIME ซึ่งเป็นตัวแปรตามในการศึกษานี้ยังมีค่าตั้งแต่ 0 ถึง 1172 วัน

5.2 เปรียบเทียบลักษณะพื้นฐานของสองกลุ่มที่ศึกษา

ข้อ 5.2.1 ถึง 5.2.8 วิเคราะห์เพื่อนำผลไปนำเสนอในการนำเสนอผลการศึกษาค้นคว้าส่วนที่ 1 คือ Baseline characteristics ของกลุ่มที่ศึกษา ในที่นี่มีสองกลุ่มเปรียบเทียบกับกันจึงวิเคราะห์ตามที่ปรากฏนี้ แต่ถ้าหากไม่มีกลุ่มเปรียบเทียบ จะวิเคราะห์เป็นภาพรวมโดยไม่แยกกลุ่ม

5.2.1 Age at Enrollment

```
. table treat, c(n age mean age sd age)
```

treat	N(age)	mean(age)	sd(age)
0	317	32.67823	5.974413
1	306	32.04902	6.260012

```
. table treat, c(n age median age min age max age)
```

treat	N(age)	med(age)	min(age)	max(age)
0	317	33	20	53
1	306	32	21	56

5.2.2 Beck Depression Score

```
. table treat, c(n becktota mean becktota sd becktota)
```

treat	N(becktota)	mean(becktota)	sd(becktota)
0	300	17.34316	9.361087
1	295	17.4676	9.43708

```
. table treat, c(n becktota median becktota min becktota max  
becktota)
```

treat	N(becktota)	med(becktota)	min(becktota)	max(becktota)
0	300	17	0	41
1	295	17	0	54

5.2.3 Heroin/Cocaine Use During 3 Months Prior to Admission

```
. tab hercoc treat, col
```

hercoc	treat		Total
	0	1	
1	65	46	111
	21.04	15.28	18.20
2	60	54	114
	19.42	17.94	18.69
3	91	87	178
	29.45	28.90	29.18
4	93	114	207
	30.10	37.87	33.93
Total	309	301	610
	100.00	100.00	100.00

5.2.4 History of IV Drug Use

```
. tab ivhx treat, col
```

ivhx	treat		Total
	0	1	
1	110	123	233
	35.60	40.86	38.20
2	52	63	115
	16.83	20.93	18.85
3	147	115	262
	47.57	38.21	42.95
Total	309	301	610
	100.00	100.00	100.00

5.2.5 Number of Prior Drug Treatments

```
. table treat, c(n ndrugtx mean ndrugtx sd ndrugtx)
```

treat	N(ndrugtx)	mean(ndrugtx)	sd(ndrugtx)
0	310	4.619355	5.243299
1	301	4.528239	5.750075

```
. table treat, c(n ndrugtx median ndrugtx min ndrugtx max  
ndrugtx)
```

treat	N(ndrugtx)	med(ndrugtx)	min(ndrugtx)	max(ndrugtx)
0	310	3	0	40
1	301	3	0	35

5.2.6 Subject's Race

```
. tab race treat, col
```

race	treat		Total
	0	1	
0	250 79.11	217 70.92	467 75.08
1	66 20.89	89 29.08	155 24.92
Total	316 100.00	306 100.00	622 100.00

5.2.7 Treatment Site

```
. tab site treat, col
```

site	treat		Total
	0	1	
0	223 69.69	221 71.75	444 70.70
1	97 30.31	87 28.25	184 29.30
Total	320 100.00	308 100.00	628 100.00

5.2.8 Length of Stay in Treatment (Admission Date to Exit Date)

```
. table treat, c(n los mean los sd los)
```

treat	N(los)	mean(los)	sd(los)
0	320	76.52812	50.2751
1	308	116.5162	93.60237

```
. table treat, c(n los median los min los max los)
```

treat	N(los)	med(los)	min(los)	max(los)
0	320	81	1	219
1	308	93.5	1	400

```
. table treat, c(n los median los min los max los)
```

treat	N(los)	med(los)	min(los)	max(los)
0	320	81	1	219
1	308	93.5	1	400

การนำเสนอผลการศึกษาส่วนที่ 1

(ใช้ผลการวิเคราะห์ข้อ 5.2.1 ถึง 5.2.8)

ลักษณะของตัวอย่างจำแนกตามกลุ่มทดลอง

ทั้งกลุ่มที่ได้รับการรักษาระยะสั้นและระยะยาว ส่วนมากมีลักษณะที่ไม่แตกต่างกัน (ตารางที่ 1)

ตารางที่ 5.1 ลักษณะของตัวอย่างจำแนกตามกลุ่มทดลอง

ตัวแปร	การรักษาระยะสั้น (n = 320)	การรักษาระยะยาว (n = 308)
Age at Enrollment		
Mean(SD)	32.7(6.0)	32.0(6.3)
Median(Min:Max)	33(20:53)	32(21:56)
Beck Depression Score		
Mean(SD)	17.3(9.4)	17.5(9.4)
Median(Min:Max)	17(0:41)	17(0:54)
Heroin/Cocaine Use During 3 Months Prior to Admission		
Heroin & Cocaine	21.0%	15.3%
Heroin Only	19.4%	17.9%
Cocaine Only	29.4%	28.9%
Neither Heroin nor Cocaine	30.1%	37.9%
History of IV Drug Use		
Never	35.6%	40.9%
Previous	16.8%	20.9%
Recent	47.6%	38.2%
Number of Prior Drug Treatments		
Mean(SD)	4.6(5.2)	4.5(5.8)
Median(Min:Max)	3(0:40)	3(0:35)
Subject's Race		
White	79.1%	70.9%
Non-White	20.9%	29.1%
Treatment Site		
Site A	69.7%	71.7%
Site B	30.3%	28.3%
Length of Stay in Treatment (Admission Date to Exit Date)		
Mean(SD)	76.5(50.3)	116.5(93.6)
Median(Min:Max)	81.0(1:219)	93.5(1:400)

5.3 ทำการ Set ข้อมูลเพื่อการวิเคราะห์ Survival analysis

```
. stset time, failure(censor)
```

```
failure event:  censor ~= 0 & censor ~= .
obs. time interval:  (0, time]
exit on or before:  failure
```

```
-----
628 total obs.
0 exclusions
-----
628 obs. remaining, representing
508 failures in single record/single failure data
147394 total analysis time at risk, at risk from t = 0
earliest observed entry t = 0
last observed exit t = 1172
```

5.4 พรรณนารูปแบบการศึกษา

```
. stdes
```

```
failure _d:  censor
analysis time _t:  time
```

Category	total	mean	min	median	max
no. of subjects	628				
no. of records	628	1	1	1	1
(first) entry time		0	0	0	0
(final) exit time		234.7038	2	166	1172
subjects with gap	0				
time on gap if gap	0				
time at risk	147394	234.7038	2	166	1172
failures	508	.8089172	0	1	1

จำนวนผู้ติดตามที่ศึกษาทั้งหมดมี 628 ราย ในระยะติดตาม 147,394 คน-วัน มีจำนวนผู้ตายซ้ำทั้งสิ้น 508 ราย

5.5 พรรณนา Overall survival experience

*Summarizing survival experience

```
. stsum, by( treat)
```

		failure _d:	censor				
		analysis time _t:	time				
treat		time at risk	incidence rate	no. of subjects	----- Survival time -----		
					25%	50%	75%
0		67752	.0039113	320	74	132	290
1		79642	.0030512	308	86	194	441
total		147394	.0034465	628	79	166	369

ในกลุ่มที่รักษาระยะสั้น จากผู้ติดตาม 320 คนในระยะติดตาม 67,752 คน-วันนั้น อัตราการติดตามซ้ำเท่ากับ 0.4 ครั้ง ต่อ 100 คน-วัน (หรือ 146.1 ครั้ง ต่อ 100 คน-ปี) ระยะเวลาปลอดการติดตามโดยเฉลี่ยเท่ากับ 132 วัน ส่วนในกลุ่มที่รักษาระยะยาว จากผู้ติดตาม 308 คนในระยะติดตาม 79,642 คน-วันนั้น อัตราการติดตามซ้ำเท่ากับ 0.3 ต่อ 100 คน-วัน (หรือ 1.1 คนต่อปี) ระยะเวลาปลอดการติดตามโดยเฉลี่ยเท่ากับ 194 วัน

*Estimating the confidence intervals of median time

```
. stquant, by( treat)
```

		no. of events	no. of ties	50 percentile	----- 95% Confidence limits ----- Large Sample Appr. Brookmeyer-Crowley		
0		265	143	132	111.3763	152.6237	113 153
1		243	100	194	169.7201	218.2799	175 222
total		508	366	166	148.1395	183.8605	144 182

Brookmeyer-Crowley limits could be inaccurate in presence of tied survival times.

ในกลุ่มที่รักษาระยะสั้น ระยะเวลาปลอดการติดตามโดยเฉลี่ยเท่ากับ 132 วัน (95%CI: 111 ถึง 152) ส่วนในกลุ่มที่รักษาระยะยาวนั้น ระยะเวลาปลอดการติดตามโดยเฉลี่ยเท่ากับ 194 วัน (95%CI: 170 ถึง 218)

*Incidence density and the 95% confidence intervals

```
. strate treat
```

```
      failure _d:  censor
analysis time _t:  time
```

Estimated rates and lower/upper bounds of 95% confidence intervals
(628 records included in the analysis)

treat	_D	_Y	_Rate	_Lower	_Upper
0	265	67752.0000	0.0039113	0.0034676	0.0044118
1	243	79642.0000	0.0030512	0.0026907	0.0034599

โดยภาพรวมแล้ว ในกลุ่มที่รักษาระยะสั้น จากผู้ติดตาม 320 คนใน
ระยะติดตาม 67,752 คน-วันนั้น อัตราการติดตามซ้ำเท่ากับ 0.39 ต่อ 100 คน-
วัน (95%CI: 0.35 ถึง 0.44) ส่วนในกลุ่มที่รักษาระยะยาว จากผู้ติดตาม 308 คน
ในระยะติดตาม 79,642 คน-วันนั้น อัตราการติดตามซ้ำเท่ากับ 0.30 ต่อ 100
คน-วัน (95%CI: 0.27 ถึง 0.35)

*Controlling for confounders using stratified analysis

```
. stmh treat, by(site)
```

```
      failure _d:  censor
analysis time _t:  time
```

Maximum likelihood estimate of the rate ratio
comparing treat==1 vs. treat==0
by site

RR estimate, and lower and upper 95% confidence limits

site	RR	Lower	Upper
0	0.73	0.59	0.90
1	0.90	0.65	1.25

Overall estimate controlling for site

RR	chi2	P>chi2	[95% Conf. Interval]
0.774	8.34	0.0039	0.651 0.922

Approx test for unequal RRs (effect modification): chi2(1) = 1.10
Pr>chi2 = 0.2934

เมื่อควบคุมผลจากพื้นที่ที่ศึกษา (SITE) แล้ว ผู้ติดยาที่ได้รับการรักษาระยะยาวมีโอกาสติดยาซ้ำเป็น 0.8 เท่าของผู้ที่ได้รับการรักษาระยะสั้น (Relative Risk = 0.8; 95%CI: 0.7 ถึง 0.9) กล่าวอีกนัยหนึ่ง ผู้ติดยาที่ได้รับการรักษาระยะสั้นมีโอกาสติดยาซ้ำเป็น 1.3 เท่าของผู้ที่ได้รับการรักษาระยะยาว (95%CI: 1.1 ถึง 1.5) ซึ่งมีนัยสำคัญทางสถิติ (p-value = 0.004) ทั้งนี้ ค่า Relative Risk ระหว่างพื้นที่ที่ศึกษานั้นแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ (p-value = 0.293)

*Controlling for variation of rate on a time scale

. stmc treat

failure _d: censor
analysis time _t: time

Mantel-Cox comparisons

Mantel-Haenszel estimates of the rate ratio
comparing treat==1 vs. treat==0
controlling for time (by clicks)

Overall Mantel-Haenszel estimate, controlling for time

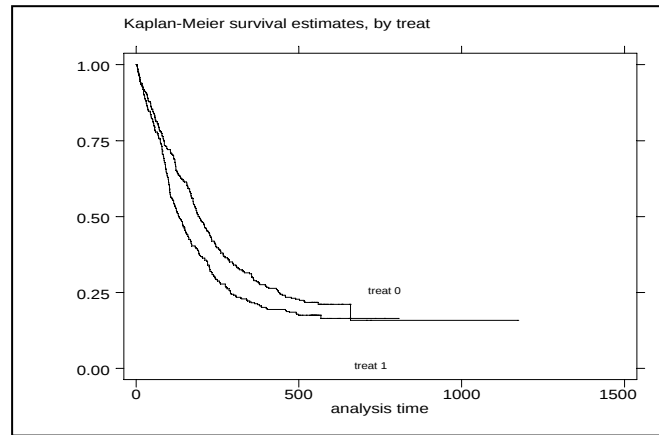
RR	chi2	P>chi2	[95% Conf. Interval]	
0.792	6.77	0.0093	0.664	0.944

เมื่อควบคุมการเปลี่ยนแปลงค่าอัตราในแต่ละช่วงเวลาสั้นๆ ตลอดระยะเวลาที่ศึกษาแล้ว ผู้ติดยาที่ได้รับการรักษาระยะยาวมีโอกาสติดยาซ้ำเป็น 0.8 เท่าของผู้ที่ได้รับการรักษาระยะสั้น (Relative Risk = 0.8; 95%CI: 0.7 ถึง 0.9) กล่าวอีกนัยหนึ่ง ผู้ติดยาที่ได้รับการรักษาระยะสั้นมีโอกาสติดยา

ซ้ำเป็น 1.3 เท่าของผู้ที่ได้รับการรักษาระยะยาว (95%CI: 1.1 ถึง 1.5) ซึ่งมีนัยสำคัญทางสถิติ (p-value = 0.009)

```
. sts g, by(treat)
```

```
failure _d: censor  
analysis time _t: time
```



5.6 ทำการวิเคราะห์อย่างหยาบ (Bivariate analysis)

ข้อ 5.6.1 ถึง 5.6.8 วิเคราะห์เพื่อนำผลไปนำเสนอในการนำเสนอผลการศึกษารound 2 คือ Crude analysis ของกลุ่มที่ศึกษา ส่วนหนึ่งของผลการวิเคราะห์นี้ยังต้องถูกนำเสนอพร้อมกับ Multivariable analysis ที่จะนำเสนอในการนำเสนอผลการศึกษารound 3 ต่อไปด้วย ส่วนที่ไม่ถูกนำเสนอเป็นตารางดังกล่าว วิเคราะห์เพื่อให้รู้ลักษณะข้อมูลในอันที่จะส่งผลต่อการวิเคราะห์ Multivariable analysis ต่อไป

5.6.1 Age at Enrollment

เพื่อการพรรณนาข้อมูล ตัวแปรต่อเนื่องจะนำมาจัดกลุ่มเป็น 4 กลุ่มที่มีจำนวนตัวอย่างพอ ๆ กัน หรือที่เรียกว่าควอร์ไทล์ (Quartile) มีขั้นตอนดังนี้

```
. su age, de
```

age					

	Percentiles	Smallest			
1%	22	20			
5%	23	21			
10%	24	21	Obs		623
25%	27	21	Sum of Wgt.		623
50%	32		Mean	32.36918	
		Largest	Std. Dev.	6.119534	
75%	37	49	Variance	37.44869	
90%	40	49	Skewness	.3692559	
95%	43	53	Kurtosis	2.810814	
99%	48	56			

ทราบค่าต่ำสุดคือ 20 ค่าสูงสุดคือ 56 ค่าควอร์ไทล์ที่ 1 (ซึ่งตรงกับ 25 เปอร์เซนต์) เท่ากับ 27 ค่าควอร์ไทล์ที่ 2 เท่ากับ 32 และควอร์ไทล์ที่ 3 เท่ากับ 37 เพื่อนำไปประกอบการจัดกลุ่มต่อไป โดยสร้างเป็นตัวแปรใหม่ชื่อ AGEGG ดังนี้

```
. gen agegg = .
(628 missing values generated)

. replace agegg = 1 if age <= 27 & age ~= .
(158 real changes made)

. replace agegg = 2 if age > 27 & age <= 32 & age ~= .
(158 real changes made)

. replace agegg = 3 if age > 32 & age <= 37 & age ~= .
(184 real changes made)

. replace agegg = 4 if age > 37 & age ~= .
(128 real changes made)
```

ตรวจสอบความถี่ในแต่ละกลุ่มดังนี้

. tab agegg

agegg	Freq.	Percent	Cum.
1	158	25.16	25.16
2	158	25.16	50.32
3	184	29.30	79.62
4	128	20.38	100.00
Total	628	100.00	

ที่ผ่านมามีขั้นตอนยาว จึงเลือกใช้คำสั่งที่สั้นกว่าดังนี้

. xtile ageg = age, nq(4)

ตรวจสอบความถี่ในแต่ละกลุ่มดังนี้

. tab ageg

4 quantiles of age	Freq.	Percent	Cum.
1	158	25.36	25.36
2	158	25.36	50.72
3	184	29.53	80.26
4	123	19.74	100.00
Total	623	100.00	

พบว่าได้ผลเช่นเดียวกัน ดังนั้นคราวต่อไปไม่ต้องจัดกลุ่มเองครว
ละคำสั่ง แต่ต้องสั่ง summarize เสมอเพื่อทราบค่าของกลุ่มว่ามีช่วงเป็น
เท่าใด

เมื่อจัดกลุ่มได้แล้ว ต่อไปเป็นการวิเคราะห์ ชั้นแรกหาจำนวน
ตัวอย่างในแต่ละกลุ่มเพื่อดูว่าเมื่อจำแนกตามกลุ่มของการทดลองที่ต้องการ
เปรียบเทียบแล้ว มีบางกลุ่มหรือไม่ที่มีจำนวนตัวอย่างน้อยมาก และอาจมี
ปัญหาในการวิเคราะห์ต่อไปดังนี้

```
. tab ageg treat, col
```

quantiles of age	4	treat		Total
		0	1	
1		70	88	158
		22.08	28.76	25.36
2		80	78	158
		25.24	25.49	25.36
3		106	78	184
		33.44	25.49	29.53
4		61	62	123
		19.24	20.26	19.74
Total		317	306	623
		100.00	100.00	100.00

พบว่าไม่น่ามีปัญหา ความถี่ในแต่ละกลุ่มมีมากพอ

```
. stsum, by(ageg)
```

```
failure _d: censor
analysis time _t: time
```

ageg	time at risk	incidence rate	no. of subjects	Survival time		
				25%	50%	75%
1	35958	.0035597	158	72	164	373
2	34423	.0039218	158	79	148	285
3	44746	.0032405	184	73	164	442
4	31689	.0030294	123	101	189	401
total	146816	.0034329	623	79	167	369

```
. stquant, by(ageg)
```

ageg	no. of events	no. of ties	50 percentile	95% Confidence limits			
				Large Sample Appr.	Brookmeyer-Crowley		
1	128	43	164	120.12	207.88	121	190
2	135	45	148	115.6674	180.3326	123	174
3	145	50	164	126.1168	201.8832	121	199
4	96	19	189	148.8714	229.1286	162	233
total	504	362	167	149.1403	184.8597	148	183

Brookmeyer-Crowley limits could be inaccurate in presence of tied survival times.

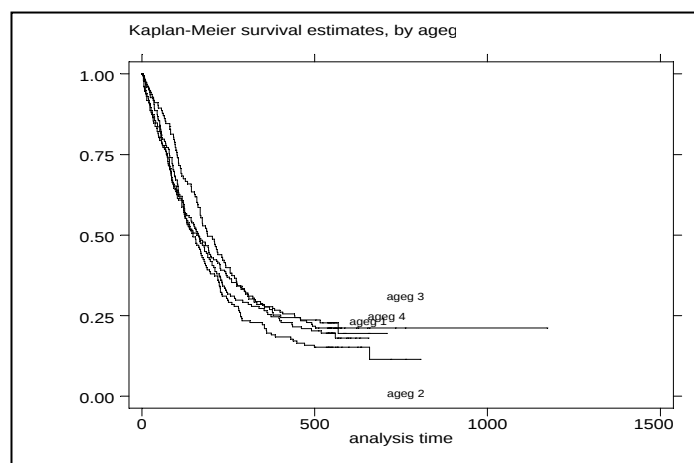
สองคำสั่งข้างต้นนี้ ค่าที่สนใจคือที่อยู่ในวงกลม ได้แก่

- 1) Time at risk คือจำนวน Person-time
(ก่อนหน้านี้ใช้คำว่า "คน-วัน")
- 2) Number of subjects คือจำนวนตัวอย่าง
- 3) Number of events คือจำนวนคนที่กลับมาติดยาซ้ำ
- 4) 50 Percentile คือค่า Median survival time พร้อมช่วงเชื่อมั่น

ผลจาก 1) ถึง 3) เพื่อการพิจารณาข้อมูล แต่บางคนก็ใช้รายงานผลด้วย ในที่นี้ ใช้เฉพาะผลในข้อ 4) เพื่อรายงานผลในตารางสรุปผลที่ท้ายของส่วนนี้

```
. sts g, by(ageg)
```

```
failure _d: censor  
analysis time _t: time
```



ดู Kaplan-Meier survival curve เปรียบเทียบ Survival ในแต่ละกลุ่มคร่าวๆ จะเห็นว่าแตกต่างกันไม่มาก

```
. xi:stcox i.ageg, nolog noshow
i.ageg          Iageg_1-4      (naturally coded; Iageg_1 omitted)

Cox regression -- Breslow method for ties

No. of subjects =          623          Number of obs =          623
No. of failures =          504
Time at risk   =         146816

Log likelihood =    -2931.2058          LR chi2(3)      =          3.81
                                      Prob > chi2      =          0.2823
```

	_t _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
Iageg_2		1.084121	.1338799	0.654	0.513	.851063 1.381002
Iageg_3		.9363168	.1136158	-0.542	0.588	.7381339 1.18771
Iageg_4		.8450363	.1141211	-1.247	0.212	.6485175 1.101106

เมื่อใช้กลุ่มอายุ 20-27 ปีเป็นกลุ่มอ้างอิงแล้ว โอกาสที่คนในกลุ่มอายุอื่นๆ จะกลับติดยาไม่แตกต่างกับ คือ Hazard ratio ประมาณ 1 ซึ่งไม่มีนัยสำคัญทางสถิติ (p-value 0.282 ผลจาก Partial likelihood ratio test)

```
. stcox age, nolog noshow

Cox regression -- Breslow method for ties

No. of subjects =          623          Number of obs =          623
No. of failures =          504
Time at risk   =         146816

Log likelihood =    -2931.4929          LR chi2(1)      =          3.24
                                      Prob > chi2      =          0.0719
```

	_t _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
age		.9872183	.0070969	-1.789	0.074	.9734062 1.001226

ถ้าไม่จัดกลุ่มตัวแปรต่อเนื่อง พบว่าทุกหนึ่งปีที่อายุเพิ่มขึ้น โอกาสกลับมาติดยาอีกนั้นลดลงประมาณ 1% (มาจาก 1 - 0.99 ได้ 0.01 แล้วคูณด้วย 100) อย่างไรก็ตาม ผลดังกล่าวไม่มีนัยสำคัญทางสถิติ (p-value = 0.072)

ด้วยการเปลี่ยนแปลงอายุเพียง 1 ปีไม่มีความสำคัญมากนัก จึงมักเปรียบเทียบการเปลี่ยนแปลงมากกว่านั้นเช่นคราวละ 5 ปี กรณีนี้ต้องหา Coefficient แล้วคูณด้วย 5 ก่อน Exponential ดังนี้

```
. stcox age, nohr nolog noshow
```

Cox regression -- Breslow method for ties

```
No. of subjects =          623          Number of obs   =          623
No. of failures =          504
Time at risk   =         146816
Log likelihood =    -2931.4929          LR chi2(1)      =          3.24
                                      Prob > chi2      =          0.0719
```

	_t _d	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age		-.0128641	.0071888	-1.789	0.074	-.0269539 .0012256

```
. disp exp(-.0128641 )      เปลี่ยนแปลงคราวละ 1 ปี
.98721829
```

```
. disp exp(5*(-.0128641))   เปลี่ยนแปลงคราวละ 5 ปี
.93770442
```

```
. disp exp(10*(-.0128641))  เปลี่ยนแปลงคราวละ 10 ปี
.87928957
```

ต่อไปนี้วิเคราะห์เพื่อใช้ผลนำเสนอในตารางสรุป โดยให้การเปลี่ยนแปลงคราวละ 5 ปี ดังนี้

```
. gen age5 = age/5
(5 missing values generated)
```

```
. stcox age5, nolog noshow
```

Cox regression -- Breslow method for ties

```
No. of subjects =          623          Number of obs   =          623
No. of failures =          504
Time at risk   =         146816
Log likelihood =    -2931.4929          LR chi2(1)      =          3.24
                                      Prob > chi2      =          0.0719
```

	_t _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
age5		.9377044	.0337048	-1.789	0.074	.8739175 1.006147

5.6.2 Beck Depression Score

```
. su becktota, de
```

becktota				

	Percentiles	Smallest		
1%	0	0		
5%	3	0		
10%	5	0	Obs	595
25%	10	0	Sum of Wgt.	595
50%	17		Mean	17.40486
		Largest	Std. Dev.	9.391132
75%	23	41		
90%	30	41	Variance	88.19335
95%	34	43	Skewness	.3721759
99%	40	54	Kurtosis	2.768088

```
. xtile beckg = becktota, nq(4)
```

```
. tab beckg treat, col
```

4			
quantiles			
of			
becktota			
	treat		
	0	1	Total
1	80	77	157
	26.67	26.10	26.39
2	78	76	154
	26.00	25.76	25.88
3	61	75	136
	20.33	25.42	22.86
4	81	67	148
	27.00	22.71	24.87
Total	300	295	595
	100.00	100.00	100.00

```
. stsum, by(beckg)
```

		failure _d: censor					
		analysis time _t: time					
beckg		time at risk	incidence rate	no. of subjects	----- Survival time -----		
					25%	50%	75%
1		42311	.0028125	157	100	215	559
2		37863	.0033014	154	81	170	353
3		29117	.0038809	136	79	143	280
4		34722	.0034848	148	71	164	403
total		144013	.0033191	595	84	170	380

. stquant, by(beckg)

beckg	no. of events	no. of ties	50 percentile	----- 95% Confidence limits -----			
				Large Sample Appr.	Brookmeyer-Crowley		
1	119	32	215	173.8296	256.1704	170	245
2	125	33	170	131.9993	208.0007	136	208
3	113	35	143	109.4325	176.5675	119	180
4	121	28	164	123.7669	204.2331	122	193
total				152.7576	187.2424	156	188

Brookmeyer-Crowley limits could be inaccurate in presence of tied survival times.

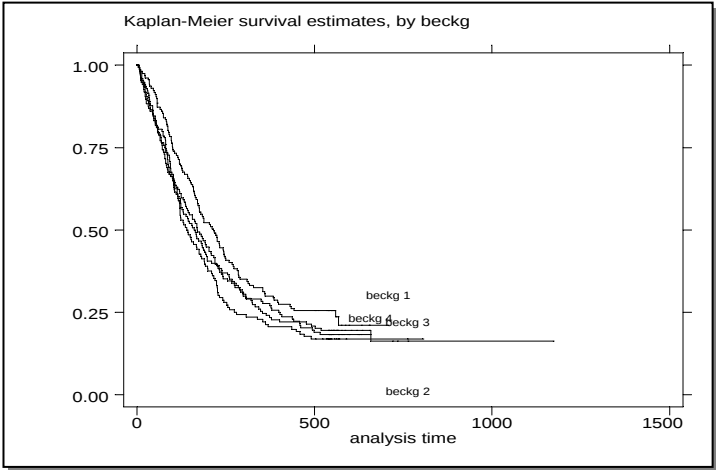
. sts g, by(beckg)

failure _d: censor
analysis time _t: time

. xi:stcox i.beckg

i.beckg Ibeckg_1-4 (naturally coded; Ibeckg_1 omitted)

failure _d: censor
analysis time _t: time



. xi:stcox i.beckg, nolog noshow

i.beckg Ibeckg_1-4 (naturally coded; Ibeckg_1 omitted)

Cox regression -- Breslow method for ties

```

No. of subjects =          595                Number of obs =          595
No. of failures =          478
Time at risk    =        144013
Log likelihood   =    -2760.0003
LR chi2(3)      =          6.08
Prob > chi2     =          0.1078

```

	_t _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
Ibeckg_2		1.202513	.1541435	1.439	0.150	.93536 1.545968
Ibeckg_3		1.363781	.1793146	2.360	0.018	1.053964 1.76467
Ibeckg_4		1.246216	.1609414	1.704	0.088	.9675331 1.60517

```
. stcox becktota, nolog noshow
```

```
Cox regression -- Breslow method for ties
```

```

No. of subjects =          595                Number of obs =          595
No. of failures =          478
Time at risk    =        144013
Log likelihood   =    -2760.3812
LR chi2(1)      =          5.32
Prob > chi2     =          0.0211

```

	_t _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
becktota		1.011022	.0047673	2.325	0.020	1.001721 1.020409

ผลจากข้างต้น เป็นค่า Hazard ratio ของการเปลี่ยนแปลงค่า BECKTOTA คราวละ 1 คะแนน ต่อไปนี้วิเคราะห์เพื่อใช้ผลนำเสนอในตารางสรุป โดยให้การเปลี่ยนแปลงคราวละ 10 คะแนน ดังนี้

```
. gen beck10 = becktota/10
(33 missing values generated)
```

```
. stcox beck10, nolog noshow
```

```
Cox regression -- Breslow method for ties
```

```

No. of subjects =          595                Number of obs =          595
No. of failures =          478
Time at risk    =        144013
Log likelihood   =    -2760.3812
LR chi2(1)      =          5.32
Prob > chi2     =          0.0211

```

	_t _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
beck10		1.115852	.0526159	2.325	0.020	1.017348 1.223892

5.6.3 Heroin/Cocaine Use During 3 Months Prior to Admission

```
. tab hercoc treat, col
```

hercoc	treat		Total
	0	1	
1	65 21.04	46 15.28	111 18.20
2	60 19.42	54 17.94	114 18.69
3	91 29.45	87 28.90	178 29.18
4	93 30.10	114 37.87	207 33.93
Total	309 100.00	301 100.00	610 100.00

```
. stsum, by( hercoc)
```

		failure _d: censor				
		analysis time _t: time				
hercoc	time at risk	incidence rate	no. of subjects	----- Survival time -----		
				25%	50%	75%
1	23653	.0038896	111	56	150	317
2	23973	.0041714	114	80	151	289
3	46943	.0028971	178	89	186	490
4	50831	.0032461	207	87	181	434
total	145400	.0033906	610	81	168	373

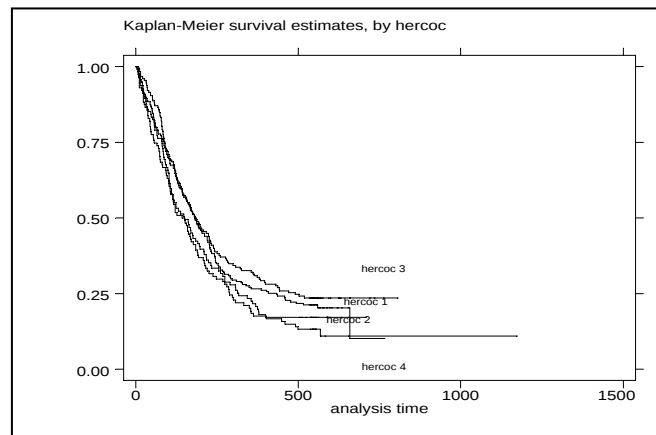
```
. stquant, by( hercoc)
```

hercoc	no. of events	no. of ties	50 percentile	----- 95% Confidence limits -----			
				Large Sample Appr.	Brookmeyer-Crowley		
1	92	20	150	102.3492	197.6508	106	190
2	100	25	151	110.0249	191.9751	110	177
3	136	46	186	143.8734	228.1266	147	224
4	165	50	181	150.7872	211.2128	153	211
total	493	349	168	149.6521	186.3479	153	184

Brookmeyer-Crowley limits could be inaccurate in presence of tied survival times.

```
. sts g, by( hercoc)
```

```
      failure _d:  censor
analysis time _t:  time
```



```
. xi:stcox i.hercoc, nolog noshow
```

```
i.hercoc          Iherco_1-4   (naturally coded; Iherco_1 omitted)
```

```
Cox regression -- Breslow method for ties
```

```
No. of subjects =          610          Number of obs   =          610
No. of failures =          493
Time at risk    =        145400
Log likelihood   =   -2855.2457          LR chi2(3)      =          7.76
                                      Prob > chi2      =          0.0513
```

_____	_____	_____	_____	_____	_____	_____
_t						
_d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
Iherco_2	1.0804	.156115	0.535	0.593	.8139319	1.434105
Iherco_3	.7753589	.1047386	-1.883	0.060	.5950028	1.010384
Iherco_4	.8506636	.1107275	-1.243	0.214	.6591137	1.097881
_____	_____	_____	_____	_____	_____	_____

5.6.4 History of IV Drug Use

```
. tab ivhx treat, col
```

ivhx	treat		Total
	0	1	
1	110 35.60	123 40.86	233 38.20
2	52 16.83	63 20.93	115 18.85
3	147 47.57	115 38.21	262 42.95
Total	309 100.00	301 100.00	610 100.00

```
. stsum, by( ivhx)
```

ivhx	failure _d: censor analysis time _t: time		no. of subjects	Survival time		
	time at risk	incidence rate		25%	50%	75%
1	64253	.0026925	233	96	194	659
2	27335	.0034022	115	81	170	354
3	53812	.0042184	262	67	147	306
total	145400	.0033906	610	81	168	373

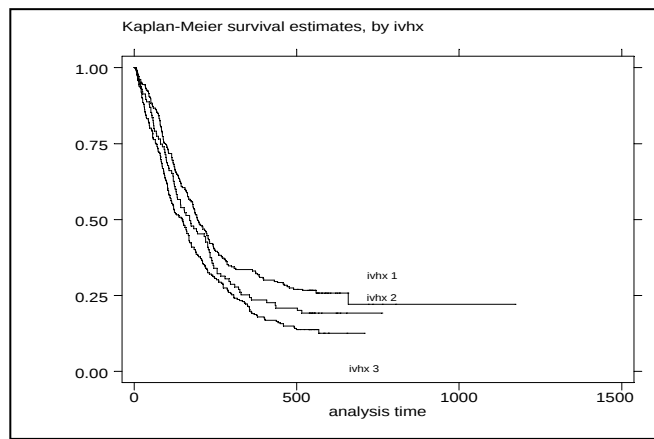
```
. stquant, by( ivhx)
```

ivhx	no. of no. of 50			95% Confidence limits			
	events	ties	percentile	Large Sample Appr.	Brookmeyer-Crowley		
1	173	69	194	165.8087	222.1913	171	227
2	93	18	170	108.4488	231.5512	129	220
3	227	109	147	118.1078	175.8922	113	166
total	493	349	168	149.6521	186.3479	153	184

Brookmeyer-Crowley limits could be inaccurate in presence of tied survival times.

```
. sts g, by( ivhx)
```

```
failure _d: censor  
analysis time _t: time
```



```
. xi:stcox i.ivhx, nolog noshow
i.ivhx          Iivhx_1-3      (naturally coded; Iivhx_1 omitted)

Cox regression -- Breslow method for ties

No. of subjects =          610          Number of obs   =          610
No. of failures =          493
Time at risk   =       145400
Log likelihood = -2851.8054          LR chi2(2)        =       14.64
                                      Prob > chi2         =       0.0007
```

	_t _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
Iivhx_2		1.215818	.1564479	1.519	0.129	.9447966 1.564583
Iivhx_3		1.46962	.1487287	3.804	0.000	1.205207 1.792043

5.6.5 Number of Prior Drug Treatments

```
. su ndrugtx, de
```

ndrugtx					
Percentiles			Smallest		
1%	0	0			
5%	0	0			
10%	0	0	Obs	611	
25%	1	0	Sum of Wgt.	611	
50%	3		Mean	4.574468	
			Std. Dev.	5.494462	
75%	6	31			
90%	10	34	Variance	30.18912	
95%	16	35	Skewness	2.746518	
99%	30	40	Kurtosis	12.73594	

```
. xtile ndrugtxg = ndrugtx, nq(4)
```

```
. tab ndrugtxg treat, col
```

quantiles of ndrugtx	treat		Total
	0	1	
1	86 27.74	97 32.23	183 29.95
2	84 27.10	81 26.91	165 27.00
3	75 24.19	63 20.93	138 22.59
4	65 20.97	60 19.93	125 20.46
Total	310 100.00	301 100.00	611 100.00

```
. stsum, by( ndrugtxg)
```

ndrugtxg	failure _d: censor analysis time _t: time		incidence rate	no. of subjects	Survival time -----		
	time at risk				25%	50%	75%
1	47066	.0029958	183	83	170	494	.
2	43149	.0028506	165	90	177	.	.
3	28905	.0041169	138	74	136	313	.
4	23882	.0047316	125	54	123	279	.
total	143002	.0034685	611	79	166	362	.

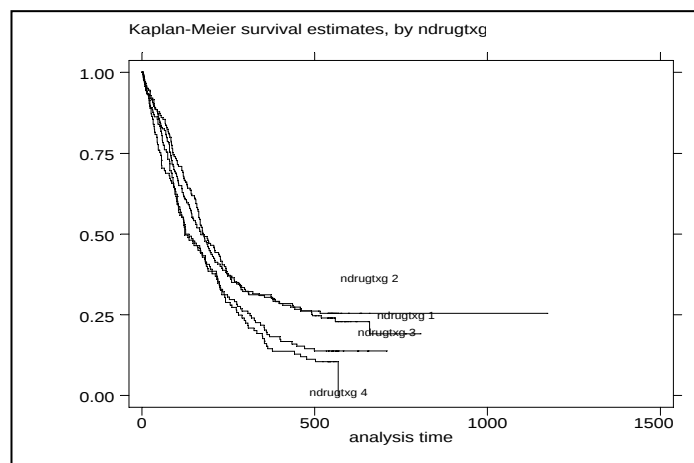
```
. stquant, by( ndrugtxg)
```

ndrugtxg	no. of events	no. of ties	50 percentile	----- 95% Confidence limits -----			
				Large Sample Appr.	Brookmeyer-Crowley		
1	141	38	170	123.9502	216.0498	140	224
2	123	32	177	154.7861	199.2139	161	204
3	119	34	136	97.356	174.644	106	181
4	113	28	123	77.4897	168.5103	106	176
total	496	354	166	148.211	183.789	144	181

Brookmeyer-Crowley limits could be inaccurate in presence of tied survival times.

```
. sts g, by( ndrugtxg)
```

```
failure _d: censor
analysis time _t: time
```



```
. xi:stcox i.ndrugtxg, nolog noshow
i.ndrugtxg          Indrug_1-4 (naturally coded; Indrug_1 omitted)

Cox regression -- Breslow method for ties

No. of subjects =          611                Number of obs   =          611
No. of failures =           496                LR chi2(3)       =          14.50
Time at risk   =        143002                Prob > chi2      =          0.0023
Log likelihood =    -2867.7232
```

	_t _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
Indrug_2		.9503908	.1173509	-0.412	0.680	.746103	1.210614
Indrug_3		1.304845	.1627081	2.134	0.033	1.021923	1.666095
Indrug_4		1.4403	.1824525	2.880	0.004	1.123636	1.846207

```
. stcox ndrugtx, nolog noshow

Cox regression -- Breslow method for ties

No. of subjects =          611                Number of obs   =          611
No. of failures =           496                LR chi2(1)       =          13.35
Time at risk   =        143002                Prob > chi2      =          0.0003
Log likelihood =    -2868.299
```

	_t _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
ndrugtx		1.029808	.0077214	3.917	0.000	1.014785	1.045053

ผลจากข้างต้น เป็นค่า Hazard ratio ของการเปลี่ยนแปลงค่า NDRUGTX คราวละ 1 ครั้ง ต่อไปนี้วิเคราะห์เพื่อใช้ทำนายเสนอในตารางสรุป โดยให้การเปลี่ยนแปลงคราวละ 5 ครั้ง ดังนี้

```
. gen ndrugg5 = ndruggtx / 5
(17 missing values generated)
```

```
. stcox ndrugg5, nolog noshow
```

Cox regression -- Breslow method for ties

```
No. of subjects =      611          Number of obs   =      611
No. of failures =      496
Time at risk    =     143002
Log likelihood   =    -2868.299
LR chi2(1)      =     13.35
Prob > chi2     =     0.0003
```

	_t					
	_d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
ndrugg5		1.158192	.04342	3.917	0.000	1.076141 1.246498

5.6.6 Subject's Race

```
. tab race treat, col
```

race	treat		Total
	0	1	
0	250 79.11	217 70.92	467 75.08
1	66 20.89	89 29.08	155 24.92
Total	316 100.00	306 100.00	622 100.00

```
. stsum, by(race)
```

```
failure _d: censor
analysis time _t: time
```

race	time at risk	incidence rate	no. of subjects	Survival time		
				25%	50%	75%
0	103665	.0037428	467	69	152	339
1	42207	.0027484	155	115	193	659
total	145872	.0034551	622	79	166	364

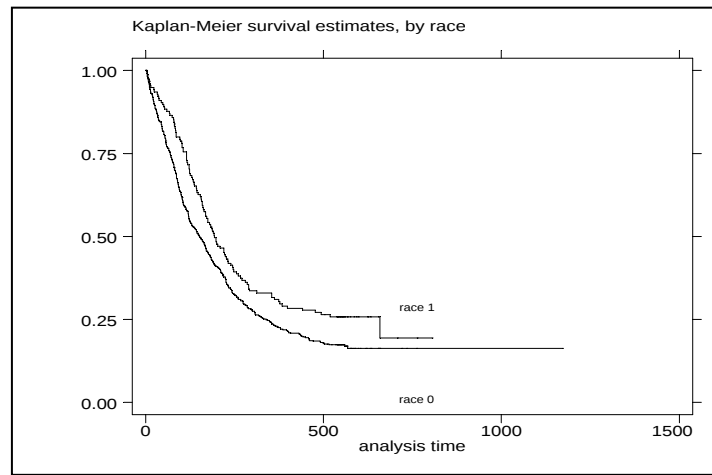
```
. stquant, by(race)
```

race	no. of			50 percentile	95% Confidence limits		
	events	ties			Large Sample Appr.	Brookmeyer-Crowley	
0	388	250	152	128.1753	175.8247	124	170
1	116	32	193	156.3987	229.6013	162	231
total	504	360	166	148.2251	183.7749	147	182

Brookmeyer-Crowley limits could be inaccurate in presence of tied survival times.

```
. sts g, by(race)
```

```
failure _d:  censor
analysis time _t:  time
```



```
. stcox race, nolog noshow
```

Cox regression -- Breslow method for ties

```
No. of subjects =      622          Number of obs =      622
No. of failures =      504
Time at risk   =     145872
Log likelihood = -2927.6972          LR chi2(1) =      7.57
                                      Prob > chi2 =     0.0059
```

_t _d	[95% Conf. Interval]					
	Haz. Ratio	Std. Err.	z	P> z		
race	.7525163	.0796946	-2.685	0.007	.6114628	.9261082

5.6.7 Assignment Treatment Site

```
. tab site treat, col
```

site	treat		Total
	0	1	
0	223 69.69	221 71.75	444 70.70
1	97 30.31	87 28.25	184 29.30
Total	320 100.00	308 100.00	628 100.00

```
. stsum, by(site)
```

		failure _d: censor		analysis time _t: time					
site		time at risk	incidence rate	no. of subjects		Survival time			
						25%	50%	75%	
0		101281	.003594	444		76	156	328	
1		46113	.0031228	184		81	200	431	
total		147394	.0034465	628		79	166	369	

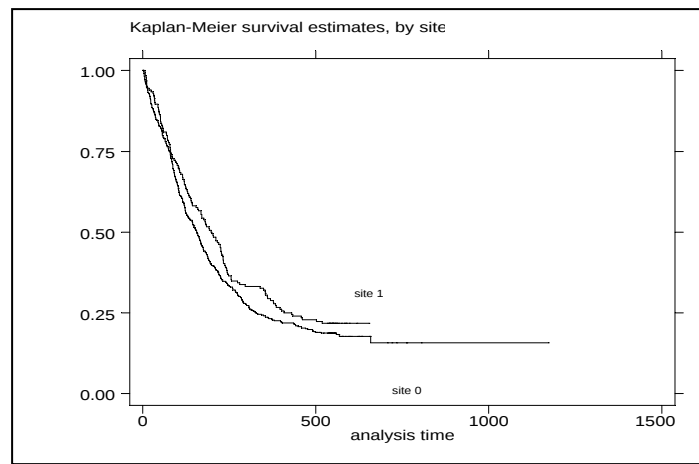
```
. stquant, by(site)
```

site	no. of events	no. of ties	50 percentile	95% Confidence limits			
				Large Sample Appr.		Brookmeyer-Crowley	
0	364	242	156	134.9114	177.0886	130	171
1	144	37	200	163.2878	236.7122	159	227
total	508	366	166	148.1395	183.8605	144	182

Brookmeyer-Crowley limits could be inaccurate in presence of tied survival times.

```
. sts g, by(site)
```

```
failure _d: censor
analysis time _t: time
```



```
. stcox site, nolog noshow
```

Cox regression -- Breslow method for ties

```
No. of subjects =      628      Number of obs =      628
No. of failures =      508
Time at risk   =     147394
Log likelihood =  -2959.3033      LR chi2(1)   =      2.40
                                   Prob > chi2   =     0.1211
```

```
-----+-----
      _t |
      _d | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      site |   .8597066   .0847279    -1.534   0.125     .7086971    1.042893
-----+-----
```

5.6.8 Treatment assignment

```
. tab treat
```

```
      treat |      Freq.   Percent   Cum.
-----+-----
          0 |         320     50.96     50.96
          1 |         308     49.04    100.00
-----+-----
       Total |         628    100.00
```

. stsum, by(treat)

		failure _d: censor					
		analysis time _t: time					
treat		time at risk	incidence rate	no. of subjects	----- Survival time -----		
					25%	50%	75%
0		67752	.0039113	320	74	132	290
1		79642	.0030512	308	86	194	441
total		147394	.0034465	628	79	166	369

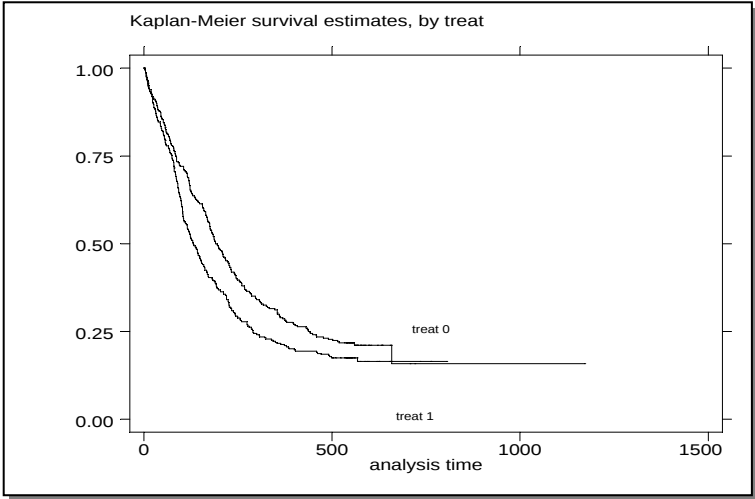
. stquant, by(treat)

		no. of events	no. of ties	50 percentile	----- 95% Confidence limits -----			
treat					Large Sample Appr.	Brookmeyer-Crowley		
0		265	143	132	111.3763	152.6237	113	153
1		243	100	194	169.7201	218.2799	175	222
total		508	366	166	148.1395	183.8605	144	182

Brookmeyer-Crowley limits could be inaccurate in presence of tied survival times.

. sts g, by(treat)

failure _d: censor
analysis time _t: time



```
. stcox treat, nolog noshw
```

```
Cox regression -- Breslow method for ties
```

```
No. of subjects =          628          Number of obs   =          628
No. of failures =          508
Time at risk    =         147394
Log likelihood   =    -2957.1293          LR chi2(1)      =          6.75
                                          Prob > chi2     =          0.0094
```

```
-----+-----
      _t |
      _d | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      treat |      .7936656   .0706311    -2.597   0.009   .6666319   .9449069
-----+-----
```

5.6.9 ตัวแปร Length of Stay in Treatment (Admission Date to Exit Date) ไม่คำนึงถึงในที่นี้เพราะเป็นเรื่องเดียวกันกับ Treatment

การนำเสนอผลการศึกษาส่วนที่ 2

(ใช้ผลการวิเคราะห์ข้อ 5.4 ข้อ 5.5 และข้อ 5.6 ซึ่งรวมข้อ 5.6.1 ถึง 5.6.8)

ผลของปัจจัยต่าง ๆ ต่อระยะปลอดการติดยาซ้ำ โดยไม่คำนึงถึงผลจากปัจจัยอื่น

ส่วนที่ 2.1 ทุกตัวแปร รวมทั้งตัวแปรต่อเนื่องที่จัดกลุ่มควอร์ไทล์

ตารางที่ 5.2 ผลของปัจจัยต่างๆ ต่อระยะปลอดการติดยาซ้ำสำหรับตัวแปรแขนง

ตัวแปร	Median time (95%CI)	Person-times	IR/100	HR (95%CI)	p-value*
Age at Enrollment					0.282
20-27	164 (120, 207)	35958	0.36	1	
28-32	148 (115, 180)	34423	0.39	1.08 (0.85, 1.38)	
33-37	164 (126, 201)	44746	0.32	0.94 (0.74, 1.19)	
38-56	189 (148, 229)	31689	0.30	0.85 (0.65, 1.10)	
Beck Depression Score					0.108
0-10	215 (173, 256)	42311	0.28	1	
11-17	170 (131, 208)	37863	0.33	1.20 (0.94, 1.55)	
18-23	143 (109, 176)	29117	0.39	1.36 (1.05, 1.76)	
24-54	164 (123, 204)	34722	0.35	1.25(0.97, 1.61)	
Heroin/Cocaine Use During 3 Months Prior to Admission					0.051
Both	150 (102, 197)	23653	0.39	1	
Heroin	151 (110, 191)	23973	0.42	1.08 (0.81, 1.43)	
Cocaine	186 (143, 228)	46943	0.29	0.77 (0.59, 1.01)	
Neither	181 (150, 211)	50831	0.32	0.85 (0.66, 1.10)	
History of IV Drug Use					< 0.001
Never	194 (165, 222)	64253	0.27	1	
Previous	170 (108, 231)	27335	0.34	1.22 (0.94, 1.56)	
Recent	147 (118, 175)	53812	0.42	1.47 (1.21, 1.79)	
Number of Prior Drug Treatments					0.002
0-1	170 (123, 216)	47066	0.30	1	
2-3	177 (154, 199)	43149	0.29	0.95 (0.75, 1.21)	
4-6	136 (97, 174)	28905	0.41	1.30 (1.02, 1.67)	
7-40	123 (77, 168)	23882	0.47	1.44 (1.12, 1.85)	
Subject's Race					0.006
White	152 (128, 175)	103665	0.37	1	
Other	193 (156, 229)	42207	0.27	0.75 (0.61, 0.93)	
Treatment Site					0.121
Site A	156 (134, 177)	101281	0.36	1	
Site B	200 (163, 236)	46113	0.31	0.86 (0.71, 1.04)	
Treatment					0.009
Short	132 (111, 152)	67752	0.39	1	
Long	194 (169, 218)	79642	0.31	0.79 (0.67, 0.94)	

หมายเหตุ: *ค่า p-value จาก Partial likelihood ratio test

ส่วนที่ 2.2 ตัวแปรต่อเนื่องโดยใช้ค่าข้อมูลดั้งเดิมคือค่าที่ไม่ได้จัดกลุ่ม
 ตารางที่ 5.3 ผลของปัจจัยต่าง ๆ ต่อระยะปลอดการติดยาซ้ำสำหรับตัว
 แปรต่อเนื่อง

ตัวแปร	ค่าที่เปลี่ยนแปลง	HR (95%CI)	p-value*
Age at Enrollment	5 ปี	0.94 (0.87, 1.01)	0.072
Beck Depression Score	10 คะแนน	1.12 (1.02, 1.22)	0.021
Number of Prior Drug Treatments	5 ครั้ง	1.16 (1.08, 1.25)	< 0.001

หมายเหตุ: *ค่า p-value จาก Partial likelihood ratio test

5.7 สร้างโมเดล (Model fitting)

5.7.1 The initial multivariable model

```
. xi:stcox age becktota ndruxt i.hercoc i.ivhx race treat site,
nolog noshw nohr
i.hercoc      Iherco_1-4 (naturally coded; Iherco_1 omitted)
i.ivhx        Iivhx_1-3 (naturally coded; Iivhx_1 omitted)
```

Cox regression -- Breslow method for ties

```
No. of subjects =      575                Number of obs   =      575
No. of failures =      464
Time at risk    =     138900
Log likelihood  =    -2640.0305
LR chi2(11)     =      47.91
Prob > chi2     =      0.0000
```

_____	_____	_____	_____	_____	_____	_____
_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_d						
age	-.0288716	.0081716	-3.533	0.000	-.0448876	-.0128556
becktota	.008342	.0049756	1.677	0.094	-.00141	.0180939
ndruxt	.0283735	.0083064	3.416	0.001	.0120934	.0446537
Iherco_2	.0653182	.1500091	0.435	0.663	-.2286942	.3593307
Iherco_3	-.0936223	.1654656	-0.566	0.572	-.4179288	.2306843
Iherco_4	.0279784	.1602768	0.175	0.861	-.2861584	.3421152
Iivhx_2	.1743865	.1386377	1.258	0.208	-.0973384	.4461115
Iivhx_3	.2807122	.1469297	1.911	0.056	-.0072648	.5686891
race	-.2028896	.1166875	-1.739	0.082	-.4315929	.0258137
treat	-.2399473	.0943704	-2.543	0.011	-.4249099	-.0549847
site	-.1024875	.1092668	-0.938	0.348	-.3166465	.1116716

```
. lrtest, saving(0)
```

คำสั่ง `stcox` มี `xi:` นำหน้าเพื่อบอกให้ STATA ทราบว่ามีตัวแปรต้นที่เป็น Polytomous โดยถ้าตัวแปรต้นใดที่เป็น Polytomous จะระบุ `i.` ก่อนชื่อตัวแปรนั้น จากนั้น STATA จะสร้าง Dummy variable ให้ตัวแปรเหล่านั้นโดยอัตโนมัติ พร้อมกับรายงานการทำงานนี้ไว้บรรทัดแรกๆ ของผล เป็นที่ที่จะทราบได้ว่า Dummy variable ใดที่ไม่อยู่ใน Model (STATA ระบุว่า “naturally coded; DummyVar omitted”) หมายถึงเป็นกลุ่มอ้างอิง (Reference group)

คำสั่ง `lrtest, saving(0)` สำหรับบันทึกค่า Log-likelihood ของ Model ที่ได้ Fit ไปนั้น ไว้ในค่า 0 สำหรับไว้เปรียบเทียบกับ Model ต่อไป เป็นการทดสอบสมมติฐานว่าตัวแปรที่ถูกตัดออกไปนั้น มีความสำคัญต่อ Model หรือไม่ โดยใช้ Likelihood ratio test

ค่า p-value ที่อยู่ในวงกลมนั้น เป็นค่าจาก Wald test ซึ่งทดสอบสมมติฐานว่าค่า Coefficient ของตัวแปรนั้นต่างจาก 0 หรือไม่ ใช้สำหรับเป็นตัวกรองขั้นต้นว่าตัวแปรใดใน Model ที่จะถูกเลือกออก

5.7.2 เลือกแบบเจาะจง (Purposeful selection) โดย Backward elimination

จากผลข้างต้น ตัวแปรแรกๆ ที่เลือกตัดออกจาก Model คือ HERCOC ซึ่งจะยังผลให้ Dummy variable ทั้งสามตัวแปรคือ `Iherco_2` `Iherco_3` และ `Iherco_4` ถูกตัดออกไปจาก Model ดังนี้

```
. xi:stcox age becktota ndrugtx i.ivhx race treat site, nolog
noshow nohr
i.ivhx          Iivhx_1-3      (naturally coded; Iivhx_1 omitted)

Cox regression -- Breslow method for ties

No. of subjects =          575          Number of obs   =          575
No. of failures =          464
Time at risk    =       138900

LR chi2(8)      =       46.51
```

Log likelihood = -2640.7278 Prob > chi2 = 0.0000

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	-.028217	.00817	-3.454	0.001	-.0442298	-.0122042
becktota	.0079385	.0049673	1.598	0.110	-.0017972	.0176741
ndrugtx	.0277628	.0082863	3.350	0.001	.011522	.0440035
Iivhx_2	.1959922	.1372134	1.428	0.153	-.0729412	.4649255
Iivhx_3	.3327979	.119912	2.775	0.006	.0977746	.5678212
race	-.2092451	.1158944	-1.805	0.071	-.4363939	.0179037
treat	-.2317746	.093713	-2.473	0.013	-.4154487	-.0481006
site	-.0994619	.1085414	-0.916	0.359	-.3121991	.1132753

. lrtest, using(0)

Cox: likelihood-ratio test

chi2(3) = 1.39
Prob > chi2 = 0.7068

. lrtest, saving(1)

ตัวแปร HERCOC ไม่มีผลต่อ Model (p-value = 0.707)
จากนั้น ตรวจสอบ % ของการเปลี่ยนแปลงค่า coefficient ดังนี้

% Changes of the coefficient = $100[(\text{Coeff}_{\text{REDUCED}} - \text{Coeff}_{\text{FULL}}) / \text{Coeff}_{\text{FULL}}]$

Variables	Coef. Reduced	Coef. Full	%Changes
age	-0.02822	-0.02887	-2.26728
becktota	0.007939	0.008342	-4.83697
ndrugtx	0.027763	0.028374	-2.15236
Iivhx_2	0.195992	0.174387	12.38955
Iivhx_3	0.332798	0.280712	18.55484
race	-0.20925	-0.20289	3.132492
treat	-0.23177	-0.23995	-3.40604
site	-0.09946	-0.10249	-2.95216

พบว่า % Changes of the coefficient สูงสุดเท่ากับ 18.6%
ดังนั้นตัวแปร HERCOC สามารถตัดออกไปจาก Model ได้
จากนี้ Candidate ที่จะนำออกจาก Model ต่อมาคือ SITE แต่เรา
ต้องคงไว้เพราะเป็นการกำหนดในวิธีการศึกษาคือเลือกทดลองสองพื้นที่

ดังนั้นตัวแปรที่เป็น Candidate ที่จะนำออกจาก Model คือ IVHX แต่พบว่า Dummy variable สองตัวนั้น มีตัวหนึ่ง Significant จาก Wald test จึงพิจารณาจัดกลุ่ม ดังนี้

```
. tab ivhx
```

ivhx	Freq.	Percent	Cum.
1	233	38.20	38.20
2	115	18.85	57.05
3	262	42.95	100.00
Total	610	100.00	

```
. gen ivhx01 = .
(628 missing values generated)
```

```
. replace ivhx01 = 0 if ivhx <= 3
(348 real changes made)
```

```
. replace ivhx01 = 1 if ivhx == 3
(262 real changes made)
```

```
. tab ivhx01
```

ivhx01	Freq.	Percent	Cum.
0	348	55.41	55.41
1	280	44.59	100.00
Total	628	100.00	

จากเดิม IVHX มีสามกลุ่ม จัดใหม่เหลือสองกลุ่ม จึงเปลี่ยนจาก Polytomous (ไม่เคยใช้ เคยใช้แต่เลิกแล้ว และกำลังใช้) เป็น Dichotomous (ไม่เคยใช้ กับเคยใช้แต่เลิกแล้วและ/หรือกำลังใช้) ซึ่งไม่ต้องสร้าง Dummy variable จึงง่ายกว่า Model ที่เป็น Polytomous พร้อมกับให้ชื่อตัวแปรใหม่เป็น IVHX01 เมื่อค่า 0 = Never และ 1 = Previous+Recent จากนั้น Fit Model ดังนี้

```
. xi:stcox age becktota ndruxt ivhx01 race treat site, nolog
noshow nohr
```

Cox regression -- Breslow method for ties

```

No. of subjects =      575
No. of failures =      464
Time at risk    =     138900

Number of obs   =      575

LR chi2(7)      =     44.51
Prob > chi2     =     0.0000
Log likelihood   =    -2641.7294

```

	_t _d	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age		-.0261517	.0080491	-3.249	0.001	-.0419276	-.0103758
becktota		.0083976	.0049516	1.696	0.090	-.0013074	.0181025
ndrugtx		.0290737	.0082126	3.540	0.000	.0129773	.04517
ivhx01		.2561209	.1062996	2.409	0.016	.0477776	.4644642
race		-.224462	.1152656	-1.947	0.051	-.4503784	.0014544
treat		-.2324266	.0937326	-2.480	0.013	-.4161392	-.048714
site		-.0866855	.1078637	-0.804	0.422	-.2980944	.1247233

```

. lrtest, using(1)
Cox: likelihood-ratio test

chi2(1)      =      2.00
Prob > chi2   =     0.1570

```

ด้วยวิธีการจัดกลุ่มตามที่กล่าวข้างต้น ยังผลให้ตัวแปร Dummy variable ของ IVHX หายไปหนึ่งตัว แต่ไม่มีผลต่อ Model แต่อย่างใด (p-value = 0.157)

ตรวจสอบ % Changes of the coefficient พบว่าค่าที่เปลี่ยนไปมากที่สุดเท่ากับ 12.8% ดังนั้นจึงสามารถดำเนินการดังกล่าวได้

Variables	Coef. Reduced	Coef. Full	%Changes
age	-0.02615	-0.02822	-7.3292
becktota	0.008398	0.007939	5.776546
ndrugtx	0.029074	0.027763	4.721032
race	-0.22446	-0.20925	7.269773
treat	-0.23243	-0.23177	0.283298
site	-0.08669	-0.09946	-12.8439

ถึงตรงนี้ พบว่าไม่มีตัวแปรที่จะเป็น Candidate ที่จะนำออกจาก Model ได้ จึงดำเนินการในขั้นต่อไปคือการเลือกว่า ตัวแปร Continuous จะเข้า Model เป็นแบบใด

5.7.3 Scale selection กรณีตัวแปร Continuous

ในที่นี้พิจารณาโดยวิธี 2 วิธีคือ วิธี Dummy variables และ Fractional polynomial ดังนี้

5.7.3.1 ใช้วิธี Dummy variables

ขั้นที่หนึ่งคือ Fit Model โดยแทนที่ตัวแปรที่เป็น Continuous ด้วยตัวแปรที่จัดกลุ่ม เช่นกลุ่มตามคอรัไต้ที่ทำก่อนหน้านี้ โดยทำครวละตัวแปรและใช้ Model ที่ได้จากขั้นตอนการตัดตัวแปรข้างต้น ขั้นนี้ทำเพื่อเอาค่า Coefficient ไปสร้างกราฟกับค่ากลางของกลุ่มตัวแปร (Midpoint)

```
. xi:stcox i.ageg becktota ndruxt ivhx01 race treat site, nolog
noshw nohr
i.ageg          Iageg_1-4      (naturally coded; Iageg_1 omitted)
```

Cox regression -- Breslow method for ties

```
No. of subjects =          575          Number of obs =          575
No. of failures =          464
Time at risk   =        138900
LR chi2(9)      =         45.04
Log likelihood  =   -2641.4644      Prob > chi2      =         0.0000
```

_____t_____d_____	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Iageg_2	.0358234	.1312602	0.273	0.785	-.2214419	.2930887
Iageg_3	-.2094037	.1300267	-1.610	0.107	-.4642513	.0454439
Iageg_4	-.3905878	.1499628	-2.605	0.009	-.6845094	-.0966662
becktota	.0088102	.0049298	1.787	0.074	-.0008852	.0184724
ndruxt	.029196	.0083175	3.510	0.000	.0128941	.045498
ivhx01	.242417	.1063426	2.280	0.023	.0339894	.4508446
race	-.2454856	.1161824	-2.113	0.035	-.4731989	-.0177724
treat	-.2293467	.0940323	-2.439	0.015	-.4136465	-.0450468
site	-.0791067	.1084877	-0.729	0.466	-.2917388	.1335253

ขั้นที่สอง หาค่ากลางของกลุ่มตัวแปร (Midpoint)

. su age, de

age					
Percentiles			Smallest		
1%	22		20		
5%	23		21		
10%	24		21	Obs	623
25%	27		21	Sum of Wgt.	623
50%	32			Mean	32.36918
	32			Std. Dev.	6.119534
75%	37		Largest		
90%	40		49	Variance	37.44869
95%	43		49	Skewness	.3692559
99%	48		53	Kurtosis	2.810814
			56		

. disp (20+28)/2, (28+33)/2, (33+38)/2, (38+56)/2

24 30.5 35.5 47

ทำเช่นเดียวกันกับตัวแปร AGE สองขั้นตอน ดังนี้

```
. xi:stcox age i.beckg ndrugtx ivhx01 race treat site, nolog
noshow nohr
```

```
i.beckg Ibeckg_1-4 (naturally coded; Ibeckg_1 omitted)
```

Cox regression -- Breslow method for ties

```
No. of subjects = 575 Number of obs = 575
No. of failures = 464
Time at risk = 138900
Log likelihood = -2641.9622 LR chi2(9) = 44.05
Prob > chi2 = 0.0000
```

_t _d	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	-.0259968	.0080818	-3.217	0.001	-.0418367	-.0101568
Ibeckg_2	.1219339	.1299184	0.939	0.348	-.1327016	.3765693
Ibeckg_3	.2012409	.1346418	1.495	0.135	-.0626521	.4651339
Ibeckg_4	.1342078	.1325452	1.013	0.311	-.125576	.3939915
ndrugtx	.0279403	.0082484	3.387	0.001	.0117738	.0441068
ivhx01	.2623327	.1060643	2.473	0.013	.0544506	.4702148
race	-.2182485	.1152013	-1.894	0.058	-.4440389	.0075418
treat	-.2365608	.0943569	-2.507	0.012	-.421497	-.0516246
site	-.0912409	.1079469	-0.845	0.398	-.3028128	.1203311

```
. su becktota, de
```

becktota					
Percentiles		Smallest			
1%	0	0			
5%	3	0			
10%	5	0	Obs	595	
25%	10	0	Sum of Wgt.	595	
50%	17		Mean	17.40486	
		Largest	Std. Dev.	9.391132	
75%	23	41			
90%	30	41	Variance	88.19335	
95%	34	43	Skewness	.3721759	
99%	40	54	Kurtosis	2.768088	

```
. disp (0+11)/2, (11+18)/2, (18+24)/2, (24+54)/2
5.5 14.5 21 39
```

ทำเช่นเดียวกันกับตัวแปร AGE สองขั้นตอน ดังนี้

```
. xi:stcox age becktota i.ndrugtxg ivhx01 race treat site, nolog
noshow nohr
i.ndrugtxg Indrug_1-4 (naturally coded; Indrug_1 omitted)
```

Cox regression -- Breslow method for ties

```
No. of subjects = 575 Number of obs = 575
No. of failures = 464
Time at risk = 138900
Log likelihood = -2640.2504 LR chi2(9) = 47.47
Prob > chi2 = 0.0000
```

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_d						
age	-.0277234	.0081495	-3.402	0.001	-.0436961	-.0117507
becktota	.0080507	.0049513	1.626	0.104	-.0016538	.0177551
Indrug_2	-.0695084	.1291112	-0.538	0.590	-.3225618	.1835449
Indrug_3	.2594868	.1336796	1.941	0.052	-.0025204	.521494
Indrug_4	.399345	.1403041	2.846	0.004	.1243541	.674336
ivhx01	.2490765	.1066439	2.336	0.020	.0400582	.4580948
race	-.2296649	.115418	-1.990	0.047	-.4558799	-.0034498
treat	-.2199096	.0935729	-2.350	0.019	-.4033092	-.0365099
site	-.0838404	.1079094	-0.777	0.437	-.295339	.1276582

```
. su ndruxt, de
```

ndruxt				

Percentiles	Smallest			
1%	0	0		
5%	0	0		
10%	0	0	Obs	611
25%	1	0	Sum of Wgt.	611
50%	3		Mean	4.574468
		Largest	Std. Dev.	5.494462
75%	6	31		
90%	10	34	Variance	30.18912
95%	16	35	Skewness	2.746518
99%	30	40	Kurtosis	12.73594

```
. disp (0+2)/2, (2+4)/2, (4+7)/2, (7+40)/2
```

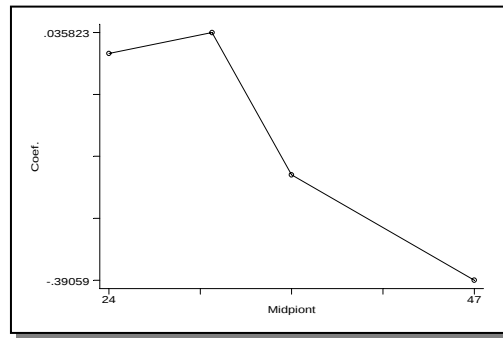
```
1 3 5.5 23.5
```

รวบรวมข้อมูลจากข้างต้นในรูปตาราง พร้อมสร้างคอลัมน์ X เพื่อ
จำแนกตัวแปรดังนี้

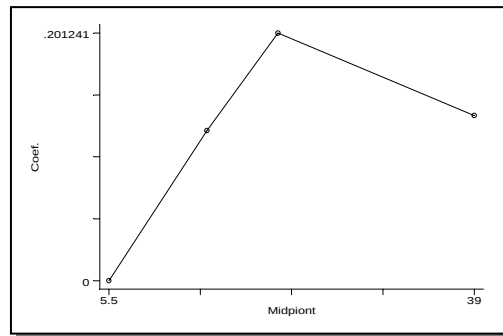
Variables	Coef.	Midpiont	x
Iageg_1	0	24.0	1
Iageg_2	0.035823	30.5	1
Iageg_3	-0.2094	35.5	1
Iageg_4	-0.39059	47.0	1
Ibeckg_1	0	5.5	2
Ibeckg_2	0.121934	14.5	2
Ibeckg_3	0.201241	21.0	2
Ibeckg_4	0.134208	39.0	2
Indrug_1	0	1.0	3
Indrug_2	-0.06951	3.0	3
Indrug_3	0.259487	5.5	3
Indrug_4	0.399345	23.5	3

Run โปรแกรม STATA ขึ้นมาเพิ่มอีกหนึ่งหน้าต่าง คลิกที่ Data
Editor ก๊อปปี้ตารางนี้ลงใน STATA Data Editor แล้วสร้างกราฟดังนี้

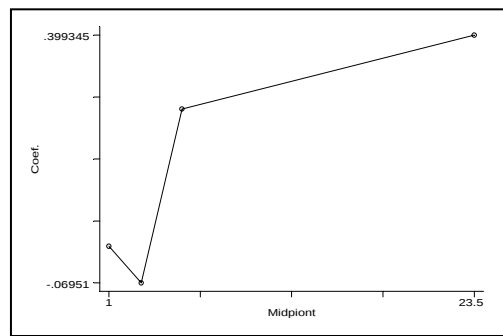
. gr coef midpiont if x == 1, c(1)



. gr coef midpiont if x == 2, c(1)

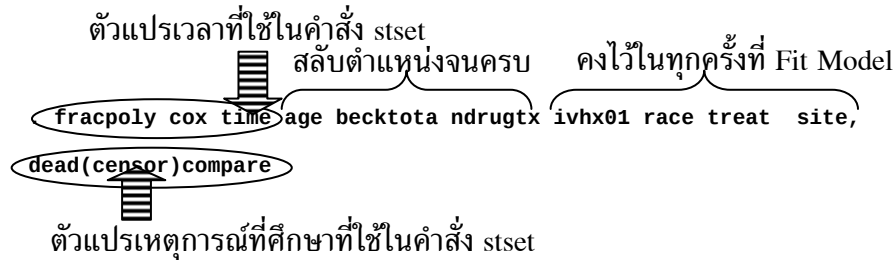


. gr coef midpiont if x == 3, c(1)



สองกราฟแรกจัดว่าเป็น Monotone ถือว่าไม่ละเมิดข้อตกลงเบื้องต้นเกี่ยวกับ Linearity แต่กราฟสุดท้ายไม่เป็น Linearity

5.7.3.2 ใช้วิธี Fractional polynomial



```
-> gen double Ibeck__1 = bectota-17.37 if e(sample)
-> gen double Indru__1 = ndruxt-4.543 if e(sample)
.....
-> gen double Iage__1 = X^3-33.96 if e(sample)
-> gen double Iage__2 = X^3*ln(X)-39.9 if e(sample)
(wherex: X = age/10)
```

Cox regression -- Breslow method for ties
Entry time 0

Number of obs = 575
LR chi2(8) = 45.59
Prob > chi2 = 0.0000
Pseudo R2 = 0.0086

Log likelihood = -2641.1921

time censor	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Iage__1	.0020128	.0323073	0.062	0.950	-.0613083	.0653339
Iage__2	-.0060879	.0201117	-0.303	0.762	-.0455062	.0333304
Ibeck__1	.0084836	.0049543	1.712	0.087	-.0012268	.0181939
Indru__1	.0292179	.0081925	3.566	0.000	.0131608	.045275
ivhx01	.2534394	.1063401	2.383	0.017	.0450166	.4618622
race	-.2298186	.1153503	-1.992	0.046	-.455901	-.0037361
treat	-.2276863	.0937621	-2.428	0.015	-.4114567	-.0439159
site	-.0771452	.1081312	-0.713	0.476	-.2890784	.134788

Deviance: 5282.384. Best powers of age among 44 models fit: 3 3.

Fractional Polynomial Model Comparisons:

age	df	Deviance	Gain	P(term)	Powers
Not in model	0	5294.224	--	--	
Linear	1	5283.459	0.000	0.001	1
m = 1	2	5282.477	0.981	0.322	3
m = 2	4	5282.384	1.075	0.954	3 3

```
. fracpoly cox time becktota age ndrugtx ivhx01 race treat site,
dead(censor)compare
```

```
-> gen double Iage__1 = age-32.38 if e(sample)
-> gen double Indru__1 = ndrugtx-4.543 if e(sample)
.....
-> gen double Ibeck__1 = X^3-5.284 if e(sample)
-> gen double Ibeck__2 = X^3*ln(X)-2.932 if e(sample)
    (where: X = (becktota+.0499992370605469)/10)
```

```
Cox regression -- Breslow method for ties
Entry time 0
```

```
Number of obs =      575
LR chi2(8)     =      46.08
Prob > chi2    =      0.0000
Pseudo R2     =      0.0086
```

```
Log likelihood = -2640.9434
```

time censor	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Ibeck__1	.0381016	.0201116	1.895	0.058	-.0013164	.0775197
Ibeck__2	-.0253573	.0147057	-1.724	0.085	-.05418	.0034654
Iage__1	-.0259231	.0080477	-3.221	0.001	-.0416964	-.0101499
Indru__1	.0284871	.0082543	3.451	0.001	.012309	.0446652
ivhx01	.24782	.106675	2.323	0.020	.0387409	.4568992
race	-.2332012	.1155157	-2.019	0.044	-.4596079	-.0067945
treat	-.2227746	.0939075	-2.372	0.018	-.40683	-.0387193
site	-.0964414	.1080121	-0.893	0.372	-.3081413	.1152585

```
Deviance: 5281.887. Best powers of becktota among 44 models fit: 3 3.
```

```
Fractional Polynomial Model Comparisons:
```

becktota	df	Deviance	Gain	P(term)	Powers
Not in model	0	5286.305	--	--	
Linear	1	5283.459	0.000	0.092	1
m = 1	2	5283.446	0.013	0.909	.5
m = 2	4	5281.887	1.572	0.459	3 3

```
. fracpoly cox time ndrugtx becktota age ivhx01 race treat site,
dead(censor)compare
```

```
-> gen double Ibeck__1 = becktota-17.37 if e(sample)
-> gen double Iage__1 = age-32.38 if e(sample)
.....
-> gen double Indru__1 = X^-1-1.804 if e(sample)
-> gen double Indru__2 = X^-1*ln(X)+1.065 if e(sample)
    (where: X = (ndrugtx+1)/10)
```

```
Cox regression -- Breslow method for ties
Entry time 0
```

```
Number of obs =      575
LR chi2(8)     =      51.43
Prob > chi2    =      0.0000
Pseudo R2     =      0.0097
```

```
Log likelihood = -2638.2717
```

time censor	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Indru__1	-.5226671	.1244122	-4.201	0.000	-.7665105	-.2788238
Indru__2	-.1947183	.0482466	-4.036	0.000	-.2892799	-.1001566
Ibeck__1	.0091589	.004987	1.837	0.066	-.0006154	.0189333
Iage__1	-.0281541	.0081328	-3.462	0.001	-.044094	-.0122141
ivhx01	.2585973	.1080215	2.394	0.017	.046879	.4703156
race	-.2421462	.1154647	-2.097	0.036	-.4684528	-.0158396
treat	-.2108924	.0936877	-2.251	0.024	-.394517	-.0272678
site	-.1053157	.109154	-0.965	0.335	-.3192536	.1086223

Deviance: 5276.543. Best powers of ndrughtx among 44 models fit: -1 -1.

Fractional Polynomial Model Comparisons:

ndrughtx	df	Deviance	Gain	P(term)	Powers
Not in model	0	5294.497	--	--	
Linear	1	5283.459	0.000	0.001	1
m = 1	2	5283.088	0.371	0.543	.5
m = 2	4	5276.543	6.915	0.038	-1 -1

▲ X^{-1} และ $(X^{-1})\ln(X)$

ผลจากสอง Model แรก Confirm ผลจากวิธี Dummy variable ว่าการนำตัวแปร AGE และ BECKTOTA เข้าใน Model แบบ Continuous สามารถกระทำได้โดยไม่ละเมิด Linearity assumption และผลจาก Model สุดท้ายก็สอดคล้องกับวิธีข้างต้น กล่าวคือ Log Hazard จากผลของ NDRUGTX ไม่เป็น Linear

ดังนั้น Scale ที่เลือกคือ AGE และ BECKTOTA เป็น Continuous ส่วน NDRUGTX ต้องจัดกลุ่มก่อนเข้าใน Model อาจใช้ Median เป็น Cut point หรือตามที่วรรณกรรมเคยระบุไว้ อย่างไรก็ตามในที่นี้ใช้ผลที่ได้จาก Fractional polynomial ข้างต้น โดยเข้าเป็น 2 ตัวแปร คือ

จากผล Fractional polynomial บ่งชี้ว่า Best power มีสอง Term คือ X^{-1} และ $(X^{-1})\ln(X)$ ทั้งนี้ผู้อ่านสามารถศึกษารายละเอียดจาก Hosmer and Lemeshow (1999) หน้า 169- 171

ดังนั้นจึงสร้างตัวแปรใหม่สองตัวแปรดังนี้

```
. gen ndruggp1 = ((ndrugtx + 1)/10)^(-1)
(17 missing values generated)
```

```
. gen ndruggp2 = ((ndrugtx + 1)/10)^(-1) *
ln((ndrugtx + 1)/10)
(17 missing values generated)
```

จากนั้น Fit Model เป็น Model ที่ได้จากการตัดตัวแปรแบบ
Backward elimination และเลือก Scale ของ Continuous variables อย่าง
เหมาะสมแล้ว Model ดังกล่าวเป็นดังนี้

```
. stcox age becktota ndruggp1 ndruggp2 ivhx01 race treat site, nolog noshw
nohr

Cox regression -- Breslow method for ties

No. of subjects =          575                Number of obs =          575
No. of failures =          464
Time at risk   =        138900

Log likelihood =   -2637.4523                LR chi2(8)      =        53.07
                                                Prob > chi2       =        0.0000
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	-.0297713	.0082121	-3.625	0.000	-.0458667	-.0136758
becktota	.0090311	.0049964	1.808	0.071	-.0007617	.0188238
ndruggp1	-.5309044	.124057	-4.280	0.000	-.7740516	-.2877572
ndruggp2	-.2006364	.0481802	-4.164	0.000	-.2950679	-.106205
ivhx01	.3057497	.1138566	2.685	0.007	.0825948	.5289046
race	-.2374182	.1153595	-2.058	0.040	-.4635186	-.0113177
treat	-.2145395	.0935829	-2.293	0.022	-.3979585	-.0311205
site	-.1528201	.1063197	-1.437	0.151	-.3612029	.0555626

5.7.4 นำทุกตัวแปรที่ถูกตัดออกไปกลับเข้าไปใน Model อีกครั้ง

```
. xi:stcox age becktota ndruggp1 ndruggp2 ivhx01 race treat site
i.hercoc, nolog noshw nohr
i.hercoc          Iherco_1-4 (naturally coded; Iherco_1 omitted)

Cox regression -- Breslow method for ties
```

No. of subjects =	575	Number of obs =	575
No. of failures =	464		
Time at risk =	138900		
		LR chi2(11) =	54.69
Log likelihood =	-2636.6391	Prob > chi2 =	0.0000

	t					
	d	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
	age	-.0308176	.0082375	-3.741	0.000	-.0469628 -.0146724
becktota		.0093385	.0050074	1.865	0.062	-.0004759 .0191529
ndrugfp1		-.5303313	.1247718	-4.250	0.000	-.7748796 -.2857833
ndrugfp2		-.2000062	.0483982	-4.133	0.000	-.294865 -.1051475
ivhx01		.2563193	.1237143	2.072	0.038	.0138438 .4987949
race		-.2254062	.1162884	-1.938	0.053	-.4533274 .0025149
treat		-.2201061	.0943234	-2.334	0.020	-.4049766 -.0352357
site		-.1430176	.1085693	-1.317	0.188	-.3558095 .0697743
Iherco_2		.0711289	.1500399	0.474	0.635	-.222944 .3652017
Iherco_3		-.1116259	.1532334	-0.728	0.466	-.4119579 .1887061
Iherco_4		-.0004887	.1463515	-0.003	0.997	-.2873323 .286355

[illegible]

```
. lrtest, saving(0)
```

```
. stcox age bectota ndruggfp1 ndruggfp2 ivhx01 race treat site,
nolog noshow nohr
```

```
Cox regression -- Breslow method for ties
```

No. of subjects =	575	Number of obs =	575
No. of failures =	464		
Time at risk =	138900		
		LR chi2(8) =	53.07
Log likelihood =	-2637.4523	Prob > chi2 =	0.0000

	t d	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age		-.0297713	.0082121	-3.625	0.000	-.0458667 - .0136758
becktota		.0090311	.0049964	1.808	0.071	-.0007617 .0188238
ndrugfp1		-.5309044	.124055	-4.280	0.000	-.7740516 -.2877572
ndrugfp2		-.2006364	.0481862	-4.164	0.000	-.2950679 -.1062055
ivhx01		.3057497	.1138566	2.685	0.007	.0825294 .5289046
race		-.2374182	.1153595	-2.058	0.040	-.4635186 -.0113717
treat		-.2145395	.0935829	-2.293	0.022	-.3979585 -.0311205
site		-.1528201	.1063197	-1.437	0.151	-.3612029 .0555626

[illegible]

[illegible]

โดยสรุป จากผลการทำ Systematic investigation of interaction effect ตามที่แสดงข้างต้น พบว่ามี Interaction effect ที่มีผลต่อ Model อย่างมีนัยสำคัญทางสถิติคือ AGE×SITE (p-value = 0.030) และ RACE×SITE (p-value < 0.001)

ดังนั้น Model ที่มี interaction terms เป็นดังนี้

```
. stcox age becktota ndrugfp1 ndrugfp2 ivhx01 race treat site
racXsite ageXsite, nohr nolog
```

```
failure _d: censor
analysis time _t: time
```

Cox regression -- Breslow method for ties

```
No. of subjects =          575          Number of obs   =          575
No. of failures =          464
Time at risk    =        138900
Log likelihood  =   -2628.5676          LR chi2(10)      =        70.83
                                      Prob > chi2       =         0.0000
```

```
-----+-----
      _t |
      _d |          Coef.   Std. Err.      z    P>|z|      [95% Conf. Interval]
-----+-----
      age |   -.0439049   .0100263    -4.379   0.000   -.0635562   -.0242537
 becktota |   .0085105   .0049666     1.714   0.087   -.0012238   .0182448
 ndrugfp1 |  -.5819714   .1246561    -4.669   0.000   -.8262928   -.33765
 ndrugfp2 |  -.2206468   .0484873    -4.551   0.000   -.31568    -.1256135
 ivhx01   |   .3233119   .1148483     2.815   0.005   .0982134   .5484105
 race     |  -.4649239   .1333076    -3.488   0.000   -.726202   -.2036459
 treat    |  -.2492435   .094199    -2.646   0.008   -.4338701   -.0646169
 site     |  -1.337526   .5316868    -2.516   0.012   -2.379613   -.2954393
 racXsite |   .9345519   .2475505     3.775   0.000   .449362    1.419742
 ageXsite |   .0314167   .0161393     1.947   0.052   -.0002158   .0630492
-----+-----
```

```
. lrtest, saving(0)
```

ทดลองนำ AGE×SITE ออกจาก Model เพราะ p-value จาก Wald test บ่งชี้ว่าไม่มีผลต่อ Model ดังนี้

```
. stcox age becktota ndrugfp1 ndrugfp2 ivhx01 race treat site
racXsite, nohr nolog
```

```
failure _d: censor
analysis time _t: time
```

Cox regression -- Breslow method for ties

```
No. of subjects =          575          Number of obs   =          575
No. of failures =          464
Time at risk    =        138900
```

```

Log likelihood = -2630.4418          LR chi2(9) = 67.09
                                     Prob > chi2 = 0.0000

```

	_t _d	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age		-.0332967	.008313	-4.005	0.000	-.0495898	-.0170036
becktota		.0087734	.0049842	1.760	0.078	-.0009954	.0185422
ndrugfp1		-.5527477	.1237794	-4.466	0.000	-.7953508	-.3101445
ndrugfp2		-.208661	.0480947	-4.339	0.000	-.302925	-.1143971
ivhx01		.330977	.1149318	2.880	0.004	.1057148	.5562392
race		-.4549021	.1330488	-3.419	0.001	-.7156729	-.1941312
treat		-.2285377	.0936393	-2.441	0.015	-.4120674	-.0450081
site		-.3354495	.1186646	-2.827	0.005	-.5680277	-.1028712
racXsite		.9694439	.2472913	3.920	0.000	.4847618	1.454126

```
. lrtest, using(0)
```

```
Cox: likelihood-ratio test
```

```

chi2(1) = 3.75
Prob > chi2 = 0.0529

```

พบว่า AGE×SITE ไม่มีผลต่อ Model (p-value 0.053) คงไว้เฉพาะ RACE×SITE ดังนั้น Final model คือ Model ที่มีตัวแปรต่อไปนี้ AGE BECKTOTA NDRUGFP1 NDRUGFP2 IVHX01 RACE TREAT SITE RACXSITE

5.7.6 Assessment of model adequacy

5.7.6.1 Assessing the proportional hazard assumption

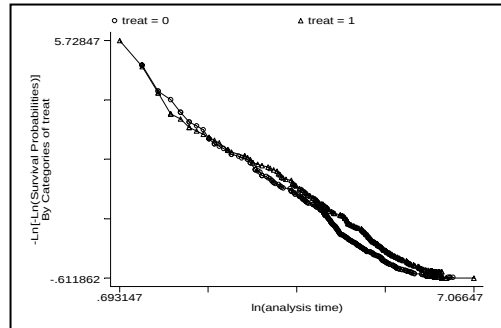
5.7.6.1.1 ดูจากกราฟ

```
. stpplot, by(treat)
```

```

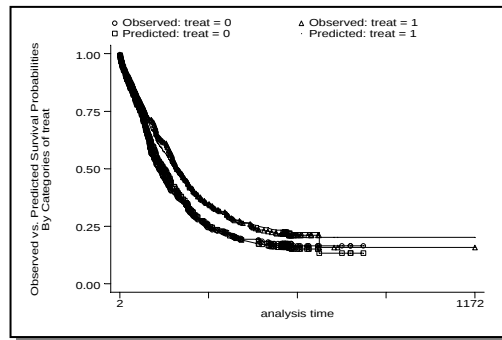
failure _d:  censor
analysis time _t:  time

```



```
. stcoxkm, by( treat)
```

```
failure _d:  censor
analysis time _t:  time
```



กราฟไม่ตัดกันอย่างชัดเจน บ่งชี้ว่า Proportional hazard assumption ไม่ถูกละเมิด จากนั้นทดสอบทางสถิติต่อไป

5.7.6.1.2 ทดสอบ Proportional hazard assumption

```
. stcox age becktota ndrugfp1 ndrugfp2 ivhx01 race treat site
racXsite, scaledsch(sca*) schoenfeld(sch*) nolog
```

```
failure _d:  censor
analysis time _t:  time
```

Cox regression -- Breslow method for ties

```
No. of subjects =          575          Number of obs   =          575
No. of failures =          464
Time at risk    =        138900
Log likelihood   = -2630.4418          LR chi2(9)       =        67.09
                                          Prob > chi2      =        0.0000
```

_____	_____	_____	_____	_____	_____	_____
_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
_d						
age	.9672515	.0080407	-4.005	0.000	.9516197	.9831401
becktota	1.008812	.0050281	1.760	0.078	.9990051	1.018715
ndrugfp1	.5753667	.0712185	-4.466	0.000	.4514228	.7333409
ndrugfp2	.8116703	.0390371	-4.339	0.000	.7386545	.8919037
ivhx01	1.392328	.1600227	2.880	0.004	1.111505	1.744101
race	.6345101	.0844208	-3.419	0.001	.4888631	.8235498
treat	.7956963	.0745084	-2.441	0.015	.6622797	.9559897
site	.7150166	.0848471	-2.827	0.005	.5666419	.9022432
racXsite	2.636478	.6519781	3.920	0.000	1.623788	4.280741
_____	_____	_____	_____	_____	_____	_____

. stphtest, rank detail

Test of proportional hazards assumption
Time: Rank(t)

	rho	chi2	df	Prob>chi2
age	0.01909	0.16	1	0.6878
becktota	-0.08564	3.24	1	0.0718
ndrugfp1	-0.00286	0.00	1	0.9512
ndrugfp2	0.00484	0.01	1	0.9174
ivhx01	0.00901	0.04	1	0.8450
race	0.05507	1.40	1	0.2362
treat	0.09062	3.81	1	0.0509
site	0.02900	0.41	1	0.5235
racXsite	-0.01632	0.12	1	0.7250
global test		10.49	9	0.3119

ผลการทดสอบ Confirm กราฟที่แสดงข้างต้น กล่าวคือ Test for departure from proportional hazard assumption ให้ค่า p-value = 0.312 ซึ่งบ่งชี้ว่าไม่ละเมิดข้อตกลงเบื้องต้นดังกล่าว

5.7.6.2 Identification of influential and poorly fit subjects

```
. stcox age becktota ndrugfp1 ndrugfp2 ivhx01 race treat site
racXsite, mgale(mg)
      failure _d: censor
      analysis time _t: time
Iteration 0: log likelihood = -2663.9851
Iteration 1: log likelihood = -2630.5633
Iteration 2: log likelihood = -2630.4418
Iteration 3: log likelihood = -2630.4418
Refining estimates:
Iteration 0: log likelihood = -2630.4418

Cox regression -- Breslow method for ties

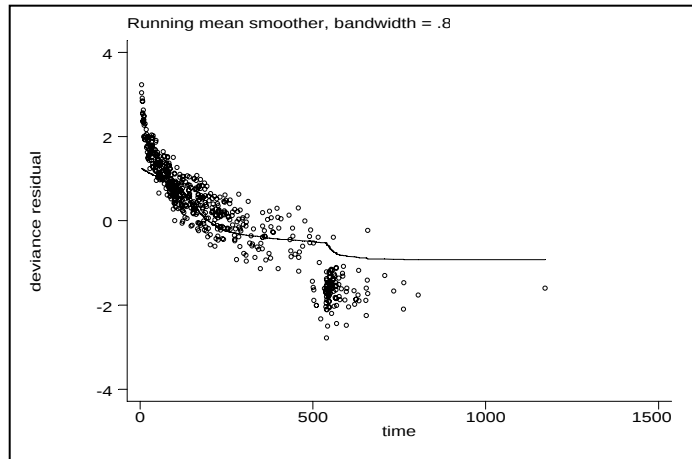
No. of subjects =          575          Number of obs   =          575
No. of failures =          464
Time at risk   =        138900

Log likelihood =   -2630.4418          LR chi2(9)       =         67.09
                                      Prob > chi2      =         0.0000
```

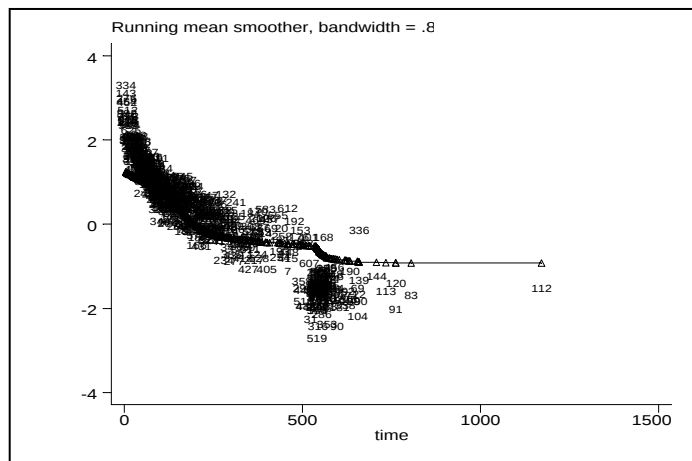
_t _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.9672515	.0080407	-4.005	0.000	.9516197	.9831401
becktota	1.008812	.0050281	1.760	0.078	.9990051	1.018715
ndrugfp1	.5753667	.0712185	-4.466	0.000	.4514228	.7333409
ndrugfp2	.8116703	.0390371	-4.339	0.000	.7386545	.8919037
ivhx01	1.392328	.1600227	2.880	0.004	1.111505	1.744101
race	.6345101	.0844208	-3.419	0.001	.4888631	.8235498
treat	.7956963	.0745084	-2.441	0.015	.6622797	.9559897
site	.7150166	.0848471	-2.827	0.005	.5666419	.9022432
racXsite	2.636478	.6519781	3.920	0.000	1.623788	4.280741

134

```
. quietly predict double dev, deviance
. ksm dev time, xla ylab
```



```
. ksm dev time, xla ylab s([id])
```



กราฟข้างต้นบ่งชี้ว่า ข้อมูลรายที่มี ID = 112 เป็น Outlier จึงลองดูว่ามีค่าอะไรเท่าไร ดังนี้

```
. list if id == 112
```

```
Observation 510
```

id	112	age	35	becktota	11
hercoc	2	ivhx	1	ndrugtx	3
race	0	treat	1	site	0
los	51	time	1172	censor	0
ageg	3	agegg	3	beckg	2
ndrugtxg	2	ivhx01	0	ndrugfp1	2.5
ndrugfp2	-2.290727	racXsite	0		

พบว่ามียาละลายเลือดการกลับติดยาช้ายาวมากถึง 1172 วัน ซึ่งเป็นไปได้จริงตามนั้นได้ (รองลงมาคือประมาณ 800 วัน) จากการพิจารณาข้อมูลอื่นๆ พบว่าเป็นไปได้ ในเบื้องต้นจึงตัดสินใจเก็บข้อมูลรายนี้ไว้ก่อน

5.7.6.3 Overall goodness-of-fit test

```
. predict cs, csnell
(53 missing values generated)
```

```
. stset cs, failure(censor)
```

```
failure event:  censor ~= 0 & censor ~= .
obs. time interval:  (0, cs]
exit on or before:  failure
```

```
-----
      628 total obs.
      53 event time missing (cs==.)
PROBABLE ERROR
-----
      575 obs. remaining, representing
      464 failures in single record/single failure data
      464 total analysis time at risk, at risk from t =
              earliest observed entry t =
```

last observed exit t = 3.883557

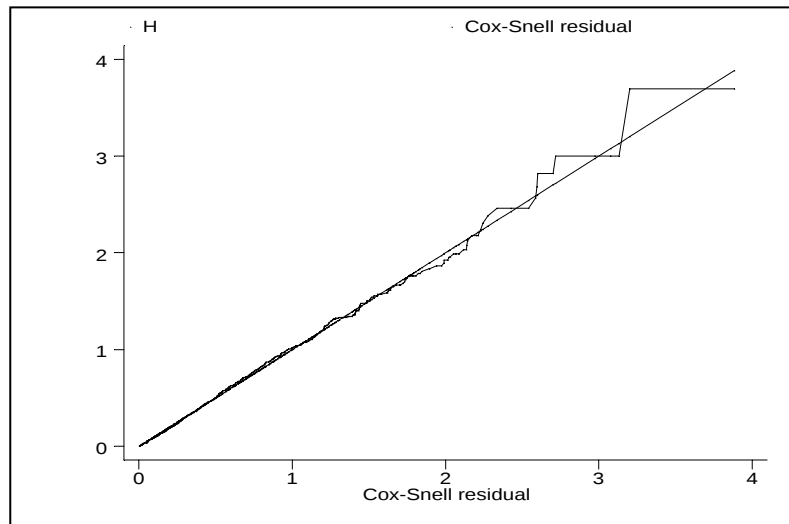
```

. sts generate km = s

. gen double H = -ln(km)
(53 missing values generated)

. graph H cs cs, c(lf) s(.) xlab ylab

```



พบว่า Model ที่ได้นี้ Fit กับข้อมูลดีพอสมควร (เส้นกราฟแนบชิดเส้นทะแยง 45 องศา) ต่อไปทำการทดสอบ Test for goodness of fit ดังนี้

คำสั่งมี 2 คำสั่ง ต้องสั่งคำสั่งแรกก่อนถ้ายังไม่สั่ง แต่ในที่นี้เราสั่งไปแล้วก่อนหน้านี้ จึงไม่ต้องสั่งอีก ให้ไปสั่งคำสั่ง stcoxgof ได้โดยตรง

```
. quietly stcox age becktota ndruggfp1 ndruggfp2 ivhx01 race treat
site racXsite mgale(mg)
```

```
. stcoxgof
```

(Table collapsed on quantiles of linear predictor)

Quantile of Risk	Observed	Expected	z	p-Norm	Observations
1	36	30.399	1.016	.31	58
2	41	37.527	.567	.571	57
3	37	48.264	-1.621	.105	58
4	44	46.199	-.323	.746	57
5	47	49.053	-.293	.769	58
6	49	40.815	1.281	.2	57
7	48	51.161	-.442	.659	58
8	54	51.836	.301	.764	57
9	55	55.308	-.041	.967	58
10	53	53.438	-.06	.952	57
Total	464	464			575

ผล Non-significance คือค่า p-value > 0.05 หมายถึง Fit ดี ซึ่งสอดคล้องกับกราฟข้างต้น

อย่างไรก็ตาม ลองหา Model ที่ดีและง่ายกว่า โดยให้ SITE เป็น Stratified variable ซึ่งจะยังผลให้ Model ไม่ต้องมี Interaction term ซึ่งมี SITE อยู่ด้วย เหตุผลคือ SITE เป็นตัวแปรที่กำหนดโดยระเบียบวิธีวิจัย (Fixed by design) ด้วยวิธีการกำหนดให้เป็น Stratified variable ทำให้มีการคำนวณ Baseline hazard function จำเพาะตามกลุ่มของ SITE ซึ่งเหมาะสมกว่าที่ใส่ไว้ใน Model

Model ที่ไม่มี Interaction term และกำหนดให้ SITE เป็น Stratified variable เป็นดังนี้

```
. stcox age becktota ndruggfp1 ndruggfp2 ivhx01 race treat,
strata(site) nolog
```

```
      failure _d:  censor
analysis time _t:  time
```

Stratified Cox regr. -- Breslow method for ties

```
No. of subjects =          575                Number of obs   =          575
No. of failures =          464
Time at risk   =        138900
Log likelihood =   -2356.9201                LR chi2(7)        =         51.89
                                                Prob > chi2       =         0.0000
```

```
-----+-----
      _t |
      _d | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      age |   .9708692   .0079707    -3.601   0.000   .9553718   .9866179
 becktota |   1.009066   .0050458     1.805   0.071   .9992249   1.019004
 ndruggfp1 |   .5886323   .0730297    -4.272   0.000   .4615707   .7506715
 ndruggfp2 |   .8187512   .0394498    -4.150   0.000   .7449696   .89984
 ivhx01   |   1.357361   .1546148     2.682   0.007   1.085767   1.696893
 race     |   .7899365   .091252     -2.041   0.041   .6298874   .9906526
 treat    |   .8058185   .07546      -2.306   0.021   .6706985   .96816
-----+-----
```

Stratified by site

ต่อไปนี้เป็นารดู Goodness-of-fit และตรวจสอบ Proportional hazard assumption

```
. quietly stcox age becktota ndruggfp1 ndruggfp2 ivhx01 race treat,
strata(site) nolog mgale(mga)
```

```
. predict csa, csnell
(53 missing values generated)
```

```
. stset csa, failure(censor)
```

```
      failure event:  censor ~= 0 & censor ~= .
obs. time interval:  (0, csa]
exit on or before:   failure
```

```

-----
628 total obs.
53 event time missing (csa==.) PROBABLY ERROR
-----
575 obs. remaining, representing
464 failures in single record/single failure data
464 total analysis time at risk, at risk from t = 0
      earliest observed entry t = 0
      last observed exit t = 4.21608

```

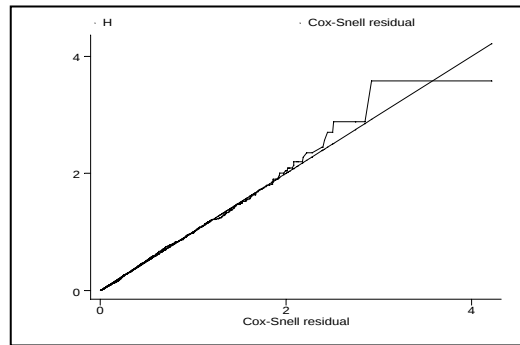
```

. sts generate kma = s

. gen double Ha = -ln(km)
(53 missing values generated)

. graph Ha csa csa, c(11) s(..) xlab ylab

```



```

. quietly stcox age becktota ndruggp1 ndruggp2 ivhx01 race treat,
strata(site) mgale(Z)

```

```

. stcoxgof

```

(Table collapsed on quantiles of linear predictor)

Quantile of Risk	Observed	Expected	z	p-Norm	Observations
1	38	33.012	.868	.385	58
2	40	40.439	-.069	.945	57
3	44	42.518	.227	.82	58
4	40	51.589	-1.613	.107	57
5	45	44.588	.062	.951	58
6	46	46.539	-.079	.937	57
7	51	47.472	.512	.609	58
8	52	51.509	.068	.945	57
9	53	55.19	-.295	.768	58
10	55	51.144	.539	.59	57

Total	464	464	575
-------	-----	-----	-----

```
. quietly stcox age bectota ndrugfp1 ndrugfp2
ivhx01 race treat, strata(site) nolog
mgale(mgal)scaledsch(sca*) schoenfeld(sch*)
```

```
. stphtest, rank detail
```

Test of proportional hazards assumption

Time: Rank(t)

	rho	chi2	df	Prob>chi2
age	0.02170	0.21	1	0.6506
becktota	-0.07821	2.71	1	0.0998
ndrugfp1	-0.00054	0.00	1	0.9908
ndrugfp2	0.00804	0.03	1	0.8636
ivhx01	0.00838	0.03	1	0.8572
race	0.05158	1.28	1	0.2580
treat	0.08817	3.62	1	0.0569
global test		10.17	7	0.1793

Model ที่ได้นี้ Fit กับข้อมูลดี และไม่ละเมิด Proportional hazard assumption

ดังนั้น Model สุดท้ายที่จะใช้ในการสรุปผลการวิจัย (The final model) เป็นดังนี้

```
. stcox age bectota ndrugfp1 ndrugfp2 ivhx01 race treat, strata(site) nolog
      failure _d:  censor
      analysis time _t:  time

Stratified Cox regr. -- Breslow method for ties

No. of subjects =          575          Number of obs   =          575
No. of failures =          464
Time at risk    =       138900
Log likelihood   = -2356.9201          LR chi2(7)        =       51.89
                                          Prob > chi2       =       0.0000
```

_t	_d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
age		.9708692	.0079707	-3.601	0.000	.9553718 .9866179
becktota		1.009066	.0050458	1.805	0.071	.9992249 1.019004
ndrugfp1		.5886323	.0730297	-4.272	0.000	.4615707 .7506715
ndrugfp2		.8187512	.0394498	-4.150	0.000	.7449696 .89984
ivhx01		1.357361	.1546148	2.682	0.007	1.085767 1.696893
race		.7899365	.091252	-2.041	0.041	.6298874 .9906526
treat		.8058185	.07546	-2.306	0.021	.6706985 .96816

การนำเสนอผลการศึกษาส่วนที่ 3

(ใช้ผลการวิเคราะห์ข้อ 5.7.1 ถึง 5.7.6)

ผลของปัจจัยต่าง ๆ ต่อระยะปลอดการติดยาซ้ำ โดยคำนึงถึงผลจากปัจจัยอื่นร่วมด้วย

เพื่อง่ายต่อการแปลความหมาย ค่า Hazard ratio ของตัวแปร AGE และ BECKTOTA ให้เป็นค่าสำหรับการเปลี่ยนแปลงคราวละ 5 หน่วย ซึ่งหาได้ดังนี้

```
. gen age5 = age/5
(5 missing values generated)

. gen beck5 = becktota/5
(33 missing values generated)

. stcox treat ivhx01 race ndrugfp1 ndrugfp2 age5 beck5,
strata(site) noshow nolog
```

Stratified Cox regr. -- Breslow method for ties

No. of subjects =	575	Number of obs =	575
No. of failures =	464		
Time at risk =	138900		
		LR chi2(7) =	51.89
Log likelihood =	-2356.9201	Prob > chi2 =	0.0000

	_____+_____					
	_t					
	_d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
treat		.8058185	.07546	-2.306	0.021	.6706985 .96816
ivhx01		1.357361	.1546148	2.682	0.007	1.085767 1.696893
race		.7899365	.091252	-2.041	0.041	.6298874 .9906526

ตัวแปร	Crude HR	Adjusted HR	95%CI	p-value*
Treatment				0.021
Short	1	1		
Long	0.79	0.81	0.67, 0.97	
History of IV Drug Use				0.007
Never	1	1		
Previous/Current	1.14	1.4	1.1, 1.7	
Subject's Race				<0.001
White	1	1		
Other	0.75	0.79	0.63, 0.99	
Number of Prior Drug Treatments				
X^{-1}	Not applicable	0.59	0.46, 0.75	<0.001
$(X^{-1})\ln(X)$	Not applicable	0.82	0.74, 0.90	<0.001
Age at Enrollment	0.94	0.86	0.80, 0.93	<0.001
Beck Depression Score	1.12	1.05	1.00, 1.10	<0.001

การคำนวณค่า p-value จาก Partial likelihood ratio test ใช้คำสั่งต่อไปนี้

```
. quietly stcox treat ivhx01 race ndruggfp1 ndruggfp2 age beckettota,
strata(site)

. lrtest, saving(0)

. quietly stcox ivhx01 race ndruggfp1 ndruggfp2 age beckettota,
strata(site)

. lrtest, using(0)
Cox: likelihood-ratio test
```

chi2(1)	=	5.32
Prob > chi2	=	0.0211

```

. quietly stcox treat race ndrugfp1 ndrugfp2 age becktota,
strata(site)

. lrtest, using(0)
Cox: likelihood-ratio test                                chi2(1) = 7.33
                                                         Prob > chi2 = 0.0068

. quietly stcox treat ivhx01 ndrugfp1 ndrugfp2 age becktota,
strata(site)

. lrtest, using(0)
Warning: observations differ: 575 vs. 579
Cox: likelihood-ratio test                                chi2(1) = 43.18
                                                         Prob > chi2 = 0.0000

. quietly stcox treat ivhx01 race ndrugfp2 age becktota,
strata(site)

. lrtest, using(0)
Cox: likelihood-ratio test                                chi2(1) = 17.84
                                                         Prob > chi2 = 0.0000

. quietly stcox treat ivhx01 race ndrugfp1 age becktota,
strata(site)

. lrtest, using(0)
Cox: likelihood-ratio test                                chi2(1) = 16.85
                                                         Prob > chi2 = 0.0000

. quietly stcox treat ivhx01 race ndrugfp1 ndrugfp2
becktota, strata(site)

. lrtest, using(0)
Warning: observations differ: 575 vs. 576
Cox: likelihood-ratio test                                chi2(1) = 25.55
                                                         Prob > chi2 = 0.0000

. quietly stcox treat ivhx01 race ndrugfp1 ndrugfp2 age
, strata(site)

. lrtest, using(0)
Warning: observations differ: 575 vs. 606
Cox: likelihood-ratio test                                chi2(1) = 341.32
                                                         Prob > chi2 = 0.0000

```

ตารางสรุปข้างต้น ได้จาก Model ที่มี Fractional polynomial terms ซึ่งยากต่อการแปลความหมาย จึงลอง Fit Model ที่ง่ายยิ่งขึ้นดังนี้

ขั้นที่ 1 จัดกลุ่ม NDRUGTX เป็น 2 กลุ่มโดยใช้ค่า Median เป็น Cut point ดังนี้

```
. su ndructx, de
```

ndrugtx				

Percentiles		Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	611
25%	1	0	Sum of Wgt.	611
50%	3		Mean	4.574468
		Largest	Std. Dev.	5.494462
75%	6	31		
90%	10	34	Variance	30.18912
95%	16	35	Skewness	2.746518
99%	30	40	Kurtosis	12.73594

```
. gen ndructmd = .
(628 missing values generated)
```

```
. replace ndructmd = .
(628 real changes made, 628 to missing)
```

```
. replace ndructmd = 0 if ndructx <=3 & ndructx ~= .
(348 real changes made)
```

```
. replace ndructmd = 1 if ndructx > 3 & ndructx ~= .
(263 real changes made)
```

```
. tab ndructmd
```

ndrugmd	Freq.	Percent	Cum.

0	348	56.96	56.96
1	263	43.04	100.00

Total	611	100.00	

```
. stcox ndructmd, nolog noshw
```

```
Cox regression -- no ties
```

```
No. of subjects =          575          Number of obs   =          575
No. of failures =          464
Time at risk    = 463.9999994
```

```

Log likelihood = -2605.3874          LR chi2(1) = 0.00
                                     Prob > chi2 = 0.9809

```

_____	_____	_____	_____	_____	_____	_____
_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
_____	_____	_____	_____	_____	_____	_____
ndrugmd	.9977492	.0937003	-0.024	0.981	.8300106	1.199387
_____	_____	_____	_____	_____	_____	_____

```

. stcox age becktota ndrugfp1 ndrugfp2 ivhx01 race treat,
strata(site) nolog noshow nohr

```

Stratified Cox regr. -- Breslow method for ties

```

No. of subjects = 575          Number of obs = 575
No. of failures = 464
Time at risk = 138900
Log likelihood = -2356.9201    LR chi2(7) = 51.89
                                Prob > chi2 = 0.0000

```

_____	_____	_____	_____	_____	_____	_____
_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_____	_____	_____	_____	_____	_____	_____
age	-.0295635	.0082099	-3.601	0.000	-.0456547	-.0134724
becktota	.0090253	.0050005	1.805	0.071	-.0007754	.018826
ndrugfp1	-.5299536	.1240668	-4.272	0.000	-.77312	-.2867872
ndrugfp2	-.199975	.0481829	-4.150	0.000	-.2944118	-.1055383
ivhx01	.3055426	.1139083	2.682	0.007	.0822864	.5287988
race	-.2358028	.1155182	-2.041	0.041	-.4622142	-.0093913
treat	-.2158967	.093644	-2.306	0.021	-.3994355	-.0323579
_____	_____	_____	_____	_____	_____	_____

Stratified by site

```

. stcox age becktota ndrugmd ivhx01 race treat, strata(site) nolog
noshow nohr

```

Stratified Cox regr. -- Breslow method for ties

```

No. of subjects = 575          Number of obs = 575
No. of failures = 464
Time at risk = 138900
Log likelihood = -2359.9438    LR chi2(6) = 45.85
                                Prob > chi2 = 0.0000

```

_____	_____	_____	_____	_____	_____	_____
_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_____	_____	_____	_____	_____	_____	_____
age	-.0280448	.008169	-3.433	0.001	-.0440558	-.0120338
becktota	.0077963	.0049566	1.573	0.116	-.0019184	.017511
ndrugmd	.3469964	.0996793	3.481	0.000	.1516285	.5423643
ivhx01	.2697392	.109117	2.472	0.013	.0558738	.4836045
race	-.2371835	.1151301	-2.060	0.039	-.4628344	-.0115326
treat	-.2305948	.0934654	-2.467	0.014	-.4137836	-.047406
_____	_____	_____	_____	_____	_____	_____

Stratified by site

Variables	Coeff. Reduced	Coeff. Full	% Changes
age	-0.02804	-0.02956	-5.13708

becktota	0.007796	0.009025	-13.6173
ivhx01	0.269739	0.305543	-11.718
race	-0.23718	-0.2358	0.585532
treat	-0.23059	-0.2159	6.807932

```
. stcox age becktota ndrugmd ivhx01 race treat, strata(site) nolog
noshow nohr
```

Stratified Cox regr. -- Breslow method for ties

```
No. of subjects =      575      Number of obs   =      575
No. of failures =      464
Time at risk    =     138900

Log likelihood   =    -2359.9438      LR chi2(6)      =     45.85
                                      Prob > chi2      =     0.0000
```

	_t _d	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age		-.0280448	.008169	-3.433	0.001	-.0440558	-.0120338
becktota		.0077963	.0049566	1.573	0.116	-.0019184	.017511
ndrugmd		.3469964	.0996793	3.481	0.000	.1516285	.5423643
ivhx01		.2697392	.109117	2.472	0.013	.0558738	.4836045
race		-.2371835	.1151301	-2.060	0.039	-.4628344	-.0115326
treat		-.2305948	.0934654	-2.467	0.014	-.4137836	-.047406

Stratified by site

```
. stcox age becktota ndrugmd ivhx01 race treat, strata(site)
scaledsch(sca*) schoenfeld(sch*) nolog
```

```
failure _d: censor
analysis time _t: time
```

Stratified Cox regr. -- Breslow method for ties

```
No. of subjects =      575      Number of obs   =      575
No. of failures =      464
Time at risk    =     138900

Log likelihood   =    -2359.9438      LR chi2(6)      =     45.85
                                      Prob > chi2      =     0.0000
```

	_t _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
age		.9723448	.0079431	-3.433	0.001	.9569005	.9880383
becktota		1.007827	.0049954	1.573	0.116	.9980834	1.017665
ndrugmd		1.414812	.1410275	3.481	0.000	1.163728	1.720069
ivhx01		1.309623	.1429021	2.472	0.013	1.057464	1.62191
race		.7888465	.09082	-2.060	0.039	.6294969	.9885336
treat		.7940612	.0742172	-2.467	0.014	.661144	.9537001

Stratified by site

```
. stphtest, rank detail
```

Test of proportional hazards assumption

Time: Rank(t)

	rho	chi2	df	Prob>chi2
--	-----	------	----	-----------

age	0.02356	0.24	1	0.6249
becktota	-0.07549	2.48	1	0.1152
ndrugmd	0.04757	1.04	1	0.3072
ivhx01	0.01360	0.08	1	0.7773
race	0.04872	1.13	1	0.2873
treat	0.08755	3.52	1	0.0605
global test		9.10	6	0.1682

. quietly predict double dev, deviance

. ksm dev time, xla ylab

. predict cs, csnell

(53 missing values generated)

. stset cs, failure(censor)

failure event: censor ~= 0 & censor ~= .
obs. time interval: (0, cs]
exit on or before: failure

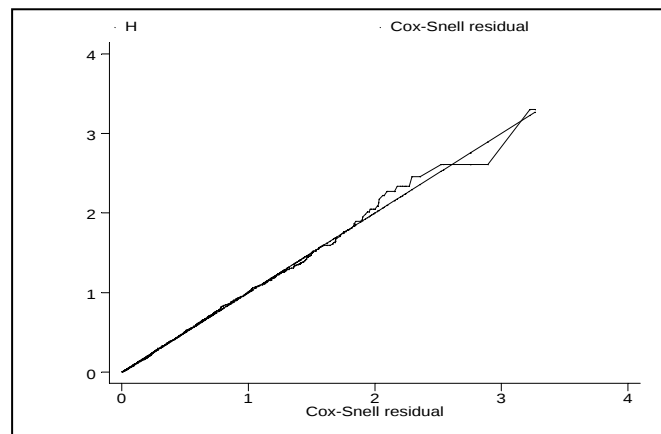
```
-----
628 total obs.
53 event time missing (cs==.) PROBABLE ERROR
-----
575 obs. remaining, representing
464 failures in single record/single failure data
464 total analysis time at risk, at risk from t = 0
earliest observed entry t = 0
last observed exit t = 3.267997
```

. sts generate km = s

. gen double H = -ln(km)

(53 missing values generated)

. graph H cs cs, c(11) s(..) xlab ylab



กราฟแสดงให้เห็นว่า Model นี้ Fit กับข้อมูลได้ดี ลองทำการทดสอบ Test for goodness of fit ดังนี้

```
. quietly stcox age becktota ndruggmd ivhx01 race treat,
strata(site) mgale(X)
```

```
. stcoxgof
```

(Table collapsed on quantiles of linear predictor)

Quantile of Risk	Observed	Expected	z	p-Norm	Observations
1	41	31.674	1.657	.097	58
2	37	43.609	-1.001	.317	57
3	37	47.612	-1.538	.124	58
4	43	45.947	-.435	.664	57
5	50	39.193	1.726	.084	58
6	52	49.279	.388	.698	58
7	47	53.02	-.827	.408	57
8	51	45.721	.781	.435	57
9	53	54.699	-.23	.818	58
10	53	53.247	-.034	.973	57
Total	464	464			575

ผล Non-significance คือค่า p-value > 0.05 หมายถึง Fit ดี ซึ่งสอดคล้องกับกราฟข้างต้น ดังนั้นเป็น Model ที่ง่ายกว่าแล้วยัง Fit กับ Data ได้ดีกว่า ไม่ละเมิด Proportional hazard assumption จึงใช้ Model นี้ในการนำเสนอผลการศึกษาดังนี้

Model:

```
. stcox treat ivhx01 race ndruggmd age5 bect5, strata(site) noshow nolog
```

Stratified Cox regr. -- Breslow method for ties

```
No. of subjects =      575          Number of obs   =      575
No. of failures =      464
Time at risk    =     138900
Log likelihood   =    -2359.9438          LR chi2(6)      =     45.85
                                          Prob > chi2     =     0.0000
```

	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
treat	.7940612	.0742172	-2.467	0.014	.661144	.9537001
ivhx01	1.309623	.1429021	2.472	0.013	1.057464	1.62191

race		.7888465	.09082	-2.060	0.039	.6294969	.9885336
ndrugmd		1.414812	.1410275	3.481	0.000	1.163728	1.720069
age5		.8691634	.0355011	-3.433	0.001	.8022949	.9416052
beck5		1.039751	.025768	1.573	0.116	.9904537	1.091502

Stratified by site							

ตารางที่ 5.5 ผลของปัจจัยต่าง ๆ ต่อระยะปลอดการติดยาซ้ำ (ไม่มี Fractional polynomial terms)

ตัวแปร	Crude HR	Adjusted HR	95%CI	p-value*
Treatment				0.014
Short	1	1		
Long	0.79	0.79	0.66, 0.95	
History of IV Drug Use				0.013
Never	1	1		
Previous/Current	1.14	1.31	1.01, 1.62	
Subject's Race				< 0.001
White	1	1		
Other	0.75	0.79	0.63, 0.99	
Number of Prior Drug Treatments				< 0.001
Median or lower (0-3)	1	1		
Higher than median (4-40)	1.10	1.41	1.16, 1.72	
Age at Enrollment	0.94	0.87	0.80, 0.94	< 0.001
Beck Depression Score	1.12	1.04	0.99, 1.09	< 0.001

หมายเหตุ: *ค่า p-value จาก Partial likelihood ratio test

การคำนวณหาค่า p-value จาก Partial likelihood ratio test ใช้คำสั่งต่อไปนี้

```
. quietly stcox treat ivhx01 race ndrugmd age becktota, strata(site)
```

```

. lrtest, saving(0)

. quietly stcox ivhx01 race ndrugmd age becktota, strata(site)

. lrtest, using(0)
Cox: likelihood-ratio test                chi2(1)    =      6.09
                                           Prob > chi2 =    0.0136

. quietly stcox treat race ndrugmd age becktota, strata(site)

. lrtest, using(0)
Cox: likelihood-ratio test                chi2(1)    =      6.19
                                           Prob > chi2 =    0.0128

. quietly stcox treat ivhx01 ndrugmd age becktota, strata(site)

. lrtest, using(0)
Warning: observations differ:  575 vs. 579
Cox: likelihood-ratio test                chi2(1)    =     43.63
                                           Prob > chi2 =    0.0000

. quietly stcox treat ivhx01 race age becktota, strata(site)

. lrtest, using(0)
Warning: observations differ:  575 vs. 585
Cox: likelihood-ratio test                chi2(1)    =    103.04
                                           Prob > chi2 =    0.0000

. quietly stcox treat ivhx01 race ndrugmd becktota, strata(site)

. lrtest, using(0)
Warning: observations differ:  575 vs. 576
Cox: likelihood-ratio test                chi2(1)    =     24.31
                                           Prob > chi2 =    0.0000

. quietly stcox treat ivhx01 race ndrugmd age, strata(site)

. lrtest, using(0)
Warning: observations differ:  575 vs. 606
Cox: likelihood-ratio test                chi2(1)    =    340.79
                                           Prob > chi2 =    0.0000

```

5.8 การสร้างกราฟของ Predicted survival curves จาก Cox

Proportional Hazards model

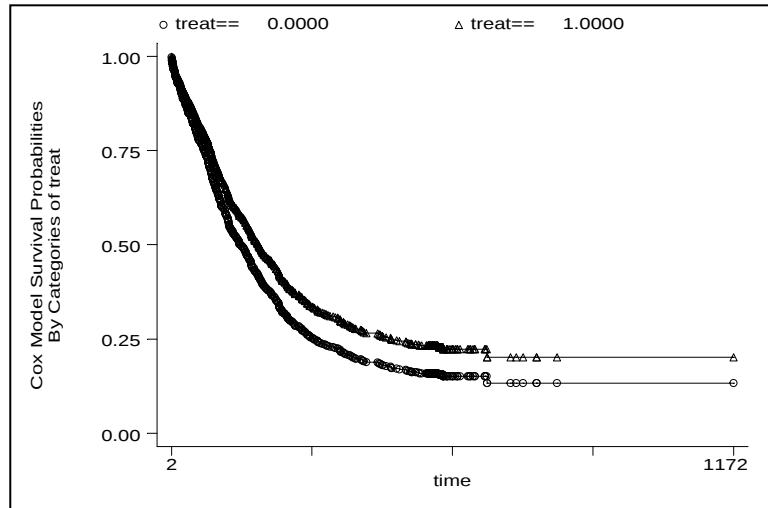
```

. stcoxplt, xvar(treat) adjust(ivhx01 race ndrugmd age5
beck5 site) c(JJ)

failure _d:  censor

```

analysis time _t: time



5.9 ส่วนเพิ่มเติมการวิเคราะห์ข้อมูล

5.9.1 การวิเคราะห์เพื่อ Prediction of survival

เพื่อช่วยต่อการทำความเข้าใจในเรื่องนี้ สมมติ Model สุดท้าย เป็น
ดังนี้

```
. stcox treat site, noshow nolog nohr
```

Cox regression -- Breslow method for ties

No. of subjects =	628	Number of obs =	628
No. of failures =	508		
Time at risk =	147394	LR chi2(2) =	9.64
Log likelihood =	-2955.6845	Prob > chi2 =	0.0081

	_t _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
treat		.7868791	.0701421	-2.689	0.007	.6607424	.9370954
site		.847137	.0836237	-1.681	0.093	.6981179	1.027966

ดังนั้น สมการคือ

$$H(t|X) = H_0(t) \times \text{EXP}(0.787\text{TREAT} + 0.847\text{SITE})$$

ค่า $H_0(t)$ เก็บที่ตัวแปร H0 จากผลของคำสั่งต่อไปนี้

```
. quietly stcox treat site, basech(H0)
```

ลองให้แสดงค่าออกมาดู

```
. sort _t
```

```
. list id treat site _t H0
```

	id	treat	site	_t	H0
1.	101	1	0	2	.00186197
2.	319	0	0	3	.00745607
3.	152	1	0	3	.00745607
4.	409	0	0	3	.00745607
5.	334	1	0	4	.01495383
6.	220	0	0	4	.01495383
7.	143	1	0	4	.01495383
8.	161	0	0	4	.01495383
9.	54	1	0	5	.02250216
10.	320	1	0	5	.02250216

----- ไม่แสดงผลต่อ -----

ถ้าต้องการรู้ Hazard ณ เวลาใดๆ ของผู้ป่วยรายใด ก็เพียงแต่แทนค่าตัวแปรที่เป็นลักษณะของผู้ป่วยรายนั้นลงใน Model ข้างต้น

ค่า $S_0(t)$ สามารถหาได้จาก $\text{EXP}[-H_0(t)]$ หรือจากคอมพิวเตอร์ โดยถูกเก็บที่ตัวแปร S0 จากผลของคำสั่งต่อไปนี้

```
. quietly stcox treat site, basesurv(S0)
```

ลองให้แสดงค่าออกมาดู

```
. sort _t
```

```
. list id treat site _t H0 S0
```

	id	treat	site	_t	H0	S0
1.	101	1	0	2	.00186197	.00186197
2.	319	0	0	3	.00745607	.00745607
3.	152	1	0	3	.00745607	.00745607
4.	409	0	0	3	.00745607	.00745607
5.	334	1	0	4	.01495383	.01495383
6.	220	0	0	4	.01495383	.01495383
7.	143	1	0	4	.01495383	.01495383
8.	161	0	0	4	.01495383	.01495383
9.	54	1	0	5	.02250216	.02250216
10.	320	1	0	5	.02250216	.02250216

----- ไม่แสดงผลต่อ -----

ถ้าต้องการรู้ Survival probability ณ เวลาใดๆ ของใคร ก็แทนค่าลงใน Model ต่อไปนี้

$$S(t) = \text{EXP}[-H_0(t) \times \text{EXP}(\text{PI})] \text{ หรือ}$$

$$S(t) = S_0(t)^{[\text{EXP}(\text{PI})]}$$

$$\text{เมื่อ PI} = 0.787\text{TREAT} + 0.847\text{SITE}$$

ค่า PI นอกจากสามารถหาได้จากสูตรข้างต้น ซึ่งเอื้อให้ผู้วิจัยกำหนดลักษณะคนที่ต้องการนำไปคาดคะเนได้แล้ว ยังสามารถหาจากคอมพิวเตอร์ ซึ่งใช้ลักษณะตามที่เป็นจริงในข้อมูล โดยถูกเก็บที่ตัวแปร XB จากผลของคำสั่งต่อไปนี้

```
. predict XB, xb
```

ลองให้แสดงค่าออกมาดู

```
. sort _t
```

```
. list id treat site _t H0 S0 XB
```

	treat	site	_t	H0	S0	XB
1.	1	0	2	.00186197	.00186197	-.2396807
2.	0	0	3	.00745607	.00745607	0
3.	1	0	3	.00745607	.00745607	-.2396807
4.	0	0	3	.00745607	.00745607	0
5.	1	0	4	.01495383	.01495383	-.2396807
6.	0	0	4	.01495383	.01495383	0
7.	1	0	4	.01495383	.01495383	-.2396807
8.	0	0	4	.01495383	.01495383	0
9.	1	0	5	.02250216	.02250216	-.2396807
10.	1	0	5	.02250216	.02250216	-.2396807

$$= 0.787 \times 1 + 0.847 \times 0$$

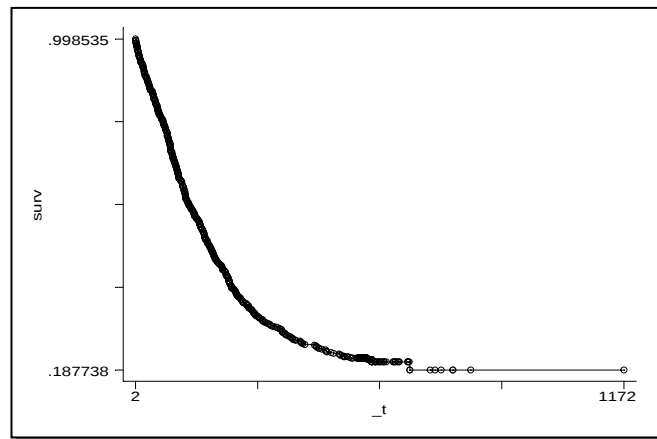
----- ไม่แสดงผลต่อ -----

เราสามารถสร้าง Adjusted survival curve จาก Cox proportional hazard model ได้จากค่า $S_0(t)$ ซึ่งคือ S0 และ PI ซึ่งคือ XB จากความสัมพันธ์ที่กล่าวแล้วข้างต้น คือ $S(t) = S_0(t)^{[EXP(PI)]}$

ตัวอย่างการสร้าง Adjusted survival curve ของ TREAT=1 และ SITE=1 ซึ่งได้ XB = PI = -0.2396807 ทำได้ดังนี้

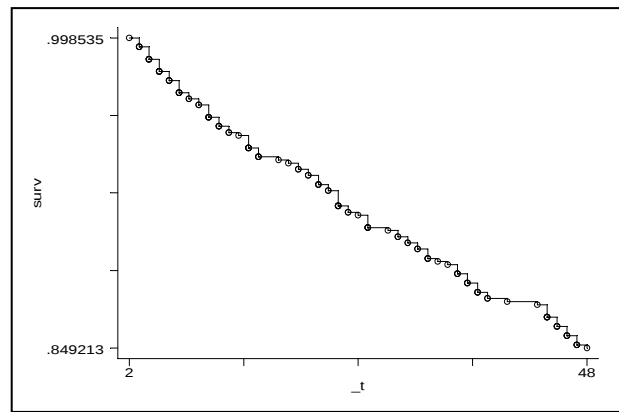
```
. gen surv = S0 ^ (exp(-0.2396807))
(628 real changes made)

. gr surv _t , c(J) sort
```



ต่อไปนี้จะแสดงกราฟเพียงข้อมูลที่มีระยะปลอดเหตุการณ์น้อยกว่า 50 วัน เพื่อให้เห็น Adjusted survival curve เป็น Step function ที่ชัดเจนขึ้น

```
. gr surv _t if _t < 50, c(J) sort
```



5.9.2 ตัวอย่างคำสั่งสำหรับการ Fit Model ที่เป็น Parametric Model

5.9.2.1 Exponential regression

```
. streg treat site, dist(exp)
```

```
      failure _d:  censor
analysis time _t:  time
```

```
Iteration 0:  log likelihood = -1043.531
Iteration 1:  log likelihood = -1038.3667
Iteration 2:  log likelihood = -1038.3392
Iteration 3:  log likelihood = -1038.3392
```

Exponential regression -- log relative-hazard form

```
No. of subjects =          628          Number of obs   =          628
No. of failures =          508
Time at risk   =         147394
Log likelihood = -1038.3392          LR chi2(2)        =          10.38
                                      Prob > chi2       =          0.0056
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
treat	.7739737	.0688443	-2.880	0.004	.6501484	.9213823
site	.8556013	.0843544	-1.582	0.114	.7052626	1.037987

5.9.2.2 Weibull regression

```
. streg treat site, dist(weibull)
```

```
      failure _d:  censor
analysis time _t:  time
```

Fitting constant-only model:

```
Iteration 0:  log likelihood = -1043.531
Iteration 1:  log likelihood = -1036.9987
Iteration 2:  log likelihood = -1036.9873
Iteration 3:  log likelihood = -1036.9873
```

Fitting full model:

```
Iteration 0:  log likelihood = -1036.9873
Iteration 1:  log likelihood = -1032.4129
Iteration 2:  log likelihood = -1032.3914
Iteration 3:  log likelihood = -1032.3914
```

Weibull regression -- log relative-hazard form

```
No. of subjects =          628          Number of obs   =          628
No. of failures =          508
Time at risk   =         147394
Log likelihood =   -1032.3914          LR chi2(2)       =          9.19
                                      Prob > chi2       =         0.0101
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
treat	.787132	.0701178	-2.687	0.007	.661032	.9372871
site	.8601256	.0848035	-1.528	0.126	.708987	1.043483
/ln_p	-.1236989	.0368963	-3.353	0.001	-.1960143	-.0513836
p	.8836458	.0326032			.8220005	.9499142
1/p	1.131675	.0417546			1.052727	1.216544

5.9.2.3 Gompertz regression

```
. streg treat site, dist(gompertz)
```

---- ไม่แสดงผลส่วนนี้ ----

5.9.2.4 Gamma regression

```
. streg treat site, dist(gamma)
```

---- ไม่แสดงผลส่วนนี้ ----

5.9.2.5 Log normal regression

```
. streg treat site, dist(lognormal)
```

```
---- ไม่แสดงผลส่วนนี้ ----
```

5.9.2.6 Log logistic regression

```
. streg treat site, dist(llogistic)
```

```
---- ไม่แสดงผลส่วนนี้ ----
```

5.9.3 การ Fit Model กรณีมี Time varying covariates

ข้อมูลตัวอย่างชุดนี้ มี sex เป็น covariate เพียงตัวเดียว ซึ่งไม่เปลี่ยนแปลงไปตามเวลา

id	time	event	sex
1	48	0	1
2	22	0	1
3	14	1	1
4	40	1	2
5	26	0	2
6	13	0	2

วิเคราะห์หาค่าผลของ sex ต่อ Survival โดยใช้ Cox proportional hazard model ได้ดังนี้

```
. stset time, failure(event)
```

```
failure event: event ~= 0 & event ~= .
obs. time interval: (0, time]
```

```
exit on or before: failure
```

6	total obs.	
0	exclusions	

6	obs. remaining, representing	
2	failures in single record/single failure data	
163	total analysis time at risk, at risk from t =	0
	earliest observed entry t =	0
	last observed exit t =	48

```
. stcox sex
```

```
failure _d: event
analysis time _t: time
```

```
Iteration 0: log likelihood = -2.3025851
Iteration 1: log likelihood = -2.2924321
Iteration 2: log likelihood = -2.2924317
Refining estimates:
Iteration 0: log likelihood = -2.2924317
```

Cox regression -- no ties

No. of subjects =	6	Number of obs =	6
No. of failures =	2		
Time at risk =	163		
		LR chi2(1) =	0.02
Log likelihood = -2.2924317		Prob > chi2 =	0.8867

<u>t</u> <u>d</u>	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
sex	1.224745	1.740957	0.143	0.887	.0755223 19.86169

เปลี่ยนแปลงโครงสร้างเพิ่มข้อมูลใหม่ ซึ่งเอื้อให้วิเคราะห์ได้หลากหลาย
ประเภทของ Modeling ดังนี้

`. stset time, failure(event) id(id)` → ส่วนนี้จำเป็นต้องมีเสมอ

```

            id: id
    failure event: event ~= 0 & event ~= .
obs. time interval: (time[_n-1], time]
  exit on or before: failure

```

6	total obs.	
0	exclusions	

6	obs. remaining, representing	
6	subjects	
2	failures in single failure-per-subject data	
163	total analysis time at risk, at risk from t =	0
	earliest observed entry t =	0
	last observed exit t =	48

```
. stsplot visit, every(6)
(24 observations=episodes created)
```

ค่าความดันโลหิตที่มาตรว

ทุก 6 เดือน

ข้อมูลเดิม แต่เปลี่ยนรูปแบบ (ข้อมูลใหม่ที่สมมติขึ้น)

id	time	event	sex	_st	_d	_t0	_t	visit	bp
1	6		1	1	0	0	6	0	90
1	12		1	1	0	6	12	6	95
1	18		1	1	0	12	18	12	92
1	24		1	1	0	18	24	18	88
1	30		1	1	0	24	30	24	86
1	36		1	1	0	30	36	30	85
1	42		1	1	0	36	42	36	84
1	48	0	1	1	0	42	48	42	83
2	6		1	1	0	0	6	0	85
2	12		1	1	0	6	12	6	82
2	18		1	1	0	12	18	12	80
2	22	0	1	1	0	18	22	18	82
3	6		1	1	0	0	6	0	100
3	12		1	1	0	6	12	6	95
3	14	1	1	1	1	12	14	12	92
4	6		2	1	0	0	6	0	85
4	12		2	1	0	6	12	6	88
4	18		2	1	0	12	18	12	83
4	24		2	1	0	18	24	18	85
4	30		2	1	0	24	30	24	84
4	36		2	1	0	30	36	30	86
4	40	1	2	1	1	36	40	36	88
5	6		2	1	0	0	6	0	83
5	12		2	1	0	6	12	6	98
5	18		2	1	0	12	18	12	95
5	24		2	1	0	18	24	18	92
5	26	0	2	1	0	24	26	24	91
6	6		2	1	0	0	6	0	102
6	12		2	1	0	6	12	6	95
6	13	0	2	1	0	12	13	12	90

ผลการวิเคราะห์เหมือนเดิมทุกประการกับที่ยังไม่เปลี่ยนแปลงข้อมูล ดังนี้

```
. stcox sex
```

---- ไม่แสดงผลส่วนนี้ ----

$\frac{-t}{-d}$	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
sex	1.224745	1.740957	0.143	0.887	.0755223 19.86169

สมมติผู้วิจัยเก็บข้อมูล bp คือค่าความดันโลหิตที่มาตรวจทุก 6 เดือน จำเป็นต้องมีแฟ้มข้อมูลตามรูปแบบข้างต้น ด้วยเหตุที่ค่า bp ของคนคนหนึ่งเปลี่ยนแปลงไปตามเวลา จึงเรียกว่า Time varying covariate การวิเคราะห์เพื่อได้ใช้ข้อมูลทุกค่า เฉพาะอย่างยิ่ง กรณีที่เรื่องที่ศึกษานั้น แต่ละค่าที่เปลี่ยนไปของตัวแปร Time varying covariate มีบทบาทมาก ต้องมีวิธีการเตรียมข้อมูลตามที่แสดงให้เห็นลำดับ จากนั้น วิเคราะห์เหมือนกับวิธีการธรรมดาตามที่ทำผ่านมาแล้วทั้งหมด เช่น วิเคราะห์ดูผลของ sex และ bp ต่อ Survival โดย Cox proportional hazard ดังนี้ (ให้ลองทำเสมือนได้ข้อมูลมาใหม่ โดยก๊อปปี้เอาเฉพาะตัวแปร id time event sex และ bp แล้วจึง stset และวิเคราะห์ต่อไป)

```
. stset time, failure(event) id(id)
```

```

      id: id
  failure event: event ~= 0 & event ~= .
obs. time interval: (time[_n-1], time]
  exit on or before: failure

```

30	total obs.	
0	exclusions	
30	obs. remaining, representing	
6	subjects	
2	failures in single failure-per-subject data	
163	total analysis time at risk, at risk from t =	0
	earliest observed entry t =	0
	last observed exit t =	48

. stcox sex bp

```

      failure _d:  event
analysis time _t:  time
            id:  id

```

```

Iteration 0:  log likelihood = -2.3025851
Iteration 1:  log likelihood = -2.0503163
Iteration 2:  log likelihood = -2.0350126
Iteration 3:  log likelihood = -2.0346005
Iteration 4:  log likelihood = -2.0346045
Refining estimates:
Iteration 0:  log likelihood = -2.0346045

```

Cox regression -- no ties

```

No. of subjects =          6              Number of obs   =          30
No. of failures =           2
Time at risk    =         163
Log likelihood   =  -2.0346045
LR chi2(2)       =           0.54
Prob > chi2      =          0.7649

```

	_t _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
sex		.7176084	1.270246	-0.187	0.851	.0223443 23.0467
bp		1.179113	.3398289	0.572	0.568	.6702442 2.07433

แบบฝึกหัด

1. การศึกษาประสิทธิผลของขมิ้นชันในการรักษาสิว

- 1.1 การศึกษาประสิทธิผลของขมิ้นชันในการรักษาสิวโดย Randomized controlled trial กลุ่มละ 10 คน กลุ่มหนึ่งให้ทาขมิ้นชันทุกวันในสามวัน ส่วนอีกกลุ่มหนึ่งล้างหน้าตามปกติโดยไม่ทาขมิ้นชัน ระยะเวลาของการศึกษาเท่ากับ 30 วัน การหายหมายถึงสิวยุบจนมองไม่เห็นร่องรอยภายใน 2 สัปดาห์

คำอธิบายตัวแปร

id	Identification number (1 - 20)
drug	Treatment (1= With Khamin ; 2 = Without Khamin)
cured	Cured within 2 weeks (1 = Yes ; 0 = No)

ข้อมูล

id	drug	cured
1	0	1
2	0	1
5	0	1
3	0	1
4	0	1
6	0	1
8	0	1
7	0	1
9	0	0
10	0	0

id	drug	cured
11	1	1
12	1	1
13	1	1
15	1	1
14	1	1
16	1	1
17	1	1
18	1	1
19	1	0
20	1	0

คำถาม

อัตราการหายของสิวภายใน 2 สัปดาห์

ทาขมิ้น = %

ไม่ทาขมิ้น = %

ความแตกต่าง = % (95%CI:% ถึง%)

ชี้แนะ : ใช้คำสั่ง STATA "*prtest cured, by(drug)*"

- 1.2 ถ้าผู้วิจัยเก็บข้อมูลเพิ่มเติมคือ time ซึ่งหมายถึง Duration of observation (days) ได้ข้อมูลตามที่แสดงไว้ท้ายนี้ จงวิเคราะห์โดยคำนึงถึงระยะเวลาที่หายด้วย

id	drug	cured	time
1	0	1	6
2	0	1	7
5	0	1	12
3	0	1	12
4	0	1	12
6	0	1	13
8	0	1	14
7	0	1	14
9	0	0	30
10	0	0	30
11	1	1	1
12	1	1	1
13	1	1	2
15	1	1	2
14	1	1	2
16	1	1	3
17	1	1	3
18	1	1	4
19	1	0	15
20	1	0	16

คำถามคือ

อัตราการหายของสิวภายใน 1 สัปดาห์

ทาขมิ้น = %

ไม่ทาขมิ้น = %

ความแตกต่าง = % (95%CI:% ถึง%)

ชี้แนะ : ใช้คำสั่ง STATA 3 คำสั่งต่อไปนี้

generate cured1 = 0

replace cured1 = 1 if cured == 1 & time <= 7

prtest cured1, by(drug)

- 1.3 ที่ผ่านมาเป็นการวิเคราะห์ที่ไม่เหมาะสมทั้งสิ้น เห็นด้วยหรือไม่ พร้อมให้เหตุผล
- 1.4 จงแสดงการวิเคราะห์ที่เหมาะสม แล้วเขียนสรุปผลที่ได้
- 1.5 ถ้าผู้วิจัยเก็บข้อมูลเพิ่มเติมคือ site ซึ่งหมายถึง Study sites (0=Hospital A; 1=Hospital B) ได้ข้อมูลตามที่แสดงไว้ท้ายนี้ จงวิเคราะห์โดยคำนึงถึงผลจาก site ด้วย

id	drug	cured	time	site
1	0	1	6	1
2	0	1	7	1
3	0	1	12	1
4	0	1	12	1
5	0	1	12	1
6	0	1	13	1
7	0	1	14	1
8	0	1	14	0
9	0	0	30	0
10	0	0	30	0
11	1	1	1	1
12	1	1	1	1
13	1	1	2	1
14	1	1	2	1
15	1	1	2	1
16	1	1	3	1
17	1	1	3	1
18	1	1	4	1
19	1	0	15	0
20	1	0	16	0

1.6 จงเขียนรายงานสรุปผลการศึกษานี้ โดยใช้ข้อมูลชุดที่มีครบทั้ง 5 ตัวแปรในข้อ 1.5 และเขียนให้อยู่ในลักษณะที่สามารถใส่ในรายงานผลการวิจัยได้

2. การทดลองวิธีการรักษามะเร็ง

จงวิเคราะห์ข้อมูลต่อไปนี้ โดย

- (1) บันทึกการวิเคราะห์ทุกขั้นตอน พร้อมทั้งคำสั่ง STATA ที่ใช้ และผลที่ได้
- (2) สรุปผลการวิเคราะห์ในรูปแบบที่ควรปรากฏในรายงานวิจัย

คำถามวิจัย: การรักษาผู้ป่วยมะเร็งโดยวิธีการใหม่ช่วยให้ผู้ป่วยมีชีวิตรยืนยาวกว่าวิธีการเดิมหรือไม่?

ตัวแปรและคำอธิบายตัวแปร:

ID	= Identification number
AGE	= Age at entry the study (years)
DIAG	= Time since the first diagnosis (months)
SCORE	= Karnofsky score
CELLTYPE	= Cell type (1=Squamous, 2=Smallcell, 3=Adeno, 4=Large)
PRIOR	= Prior therapy (0=No, 1=Yes)
TX	= Treatment (1=Standard treatment, 2=New treatment)
DEAD	= Survival status (1=Dead, 0=Alive)
DAYS	= Survival time (days)

ข้อมูลตัวอย่างจากแฟ้มข้อมูล *CANCER.DTA*

id	age	diag	score	celltype	prior	tx	dead	days
1	65	21	20	1	1	2	1	1
2	35	7	50	1	0	2	1	1
3	35	7	50	1	0	2	1	1

- - - ข้ามการแสดงผลข้อมูลผู้ป่วยไป 144 ราย - - -

148	54	12	90	1	1	1	1	577
149	54	12	90	1	1	1	1	581
150	50	7	70	1	1	1	1	590

หมายเหตุ:

1. ดาวน์โหลดข้อมูลได้ที่ <http://web.kku.ac.th/~bandit/data>
2. คำเฉลยของแบบฝึกหัด หรือคำถามอื่นใดจากการอ่านหนังสือเล่มนี้ หรือด้านอื่น ๆ ทางด้านการวิจัยทางวิทยาศาสตร์สุขภาพ และชีวิตสถิติ สามารถถามได้ใน Web Board ที่ <http://web.kku.ac.th/~bandit> ทั้งหมดในอินเทอร์เน็ตนี้เป็นบริการที่ไม่ต้องเสียค่าใช้จ่ายแต่อย่างใด

เอกสารอ้างอิง

- บัณฑิต ถิ่นคำรพ (2542). *การวิเคราะห์ข้อมูลการวิจัยทางวิทยาศาสตร์สุขภาพโดยใช้การถดถอยลอจิสติก : ตำราสำหรับศึกษาด้วยตนเอง*. ขอนแก่น: โรงพิมพ์มหาวิทยาลัยขอนแก่น.
- บัณฑิต ถิ่นคำรพ (2543). *คู่มือปฏิบัติการชีวสถิติสำหรับเรียนรู้ชีวสถิติด้วยตนเอง*. ขอนแก่น: โรงพิมพ์คลังน่านวิทยา.
- Altman, D.G. (1991). *Practical statistics for medical research*. London: Chapman and Hall.
- Christensen, E. (1987). Multivariate survival analysis using Cox's regression model. *Hepatology*. 7(6): 1346-1358
- Hosmer, D.E., and Lemeshow, S. (1999). *Applied survival analysis*. New York. John Wiley & Sons.
- Kleinbaum, D.G. (1996). *Survival analysis: A self-learning text*. New York: Springer-Verlag.
- Kleinbaum, D.G., Kupper, L.L., Muller, K.E, and Nizam, A. (1998). *Applied regression analysis and other multivariable methods*. Pacific Grove: Duxbury Press.
- Mickey, R.M., and Greenland S. (1989). The impact of confounder selection criteria on effect estimation. *Am. J. Epidemiology*. 129(1): 125-137
- Rabe-Hesketh, S. and Everitt, B. (1998). *A handbook of statistical analysis using Stata*. London: Chapman & Hall/CRC.
- StataCorp. (1999). *Stata statistical software: Release 6.0*. College Station. TX: Stata Corporation.
