

บทที่ 2

วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

การทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้องกับการประมาณค่าความชุกของเหตุการณ์ที่สนใจ เพื่อให้ได้ผลการประมาณค่าความชุกของการเกิดเหตุการณ์ที่มีความใกล้เคียงกับความ เป็นจริง โดยแบ่งการนำเสนอต่อไปนี้

1. แหล่งความผิดพลาดขององค์ความรู้ที่ได้จากการวิจัย
2. แนวทางการจัดการกับความผิดพลาดขององค์ความรู้ที่ได้จากการวิจัย
3. ผลกระทบและวิธีการแก้ไขของอคติจากการเลือกเป็นตัวอย่างในการวิจัย
4. วิธีการทางสถิติในการจัดการกับอคติจากการเลือกเป็นตัวอย่าง
5. การติดเชื้อพยาธิใบไม้ตับ
6. งานวิจัยที่เกี่ยวข้อง (นำเสนอเป็นตาราง)
7. กรอบแนวคิดการวิจัย

1. แหล่งความผิดพลาดขององค์ความรู้ที่ได้จากการวิจัย (Source of error)

การได้มาซึ่งผลลัพธ์ที่แตกต่างไปจากความเป็นจริงนั้น เกิดจากความผิดพลาดของการวิจัย ซึ่งอาจเริ่มต้นตั้งแต่การออกแบบการวิจัย การเก็บรวบรวมข้อมูล จนกระทั่งถึงขั้นตอนการวิเคราะห์ข้อมูล จนนำไปสู่การได้มาซึ่งตัวอย่างที่ไม่เหมาะสมเข้าสู่การศึกษาและทำให้ได้ผลลัพธ์ของการศึกษาบิดเบือนไปจากความเป็นจริง ซึ่งความผิดพลาดของการวิจัยที่เกิดผลกระทบต่อผลลัพธ์ของการศึกษานี้ สามารถแบ่งเป็น 2 ชนิด ได้แก่ ความผิดพลาดจากการสุ่ม (Random error) และอคติหรือความลำเอียง (Systematic Error หรือ Bias)

ความผิดพลาดจากการสุ่ม (Random error) เป็นความผิดพลาดของการสุ่มตัวอย่าง สาเหตุเกิดจากความบังเอิญแม้จะควบคุมไม่ให้เกิดอคติแล้ว อาจทำให้ได้กลุ่มตัวอย่างที่มีลักษณะไม่เหมือนกับประชากรทั้งหมด และความแตกต่างระหว่างกลุ่มตัวอย่างและกลุ่มที่ไม่ได้เป็นตัวอย่างในการศึกษา อาจไม่สามารถสรุปว่ามีความแตกต่างกันอย่างแท้จริง (Friis, & Sellers, 2004) โดยมีสาเหตุจากความผิดพลาดจากการวัดผล และความแปรผันของการสุ่มตัวอย่างที่ไม่เป็นตัวแทนประชากรที่ดี ซึ่งเกิดจากความบังเอิญจากการคัดเลือก (Aschengrau, & Seage, 2003)

อคติหรือความลำเอียง (Systematic Error หรือ Bias) เป็นผลลัพธ์จากความผิดพลาดของระบบในการออกแบบและการจัดการกับการศึกษาวิจัย โดยเฉพาะความบกพร่องในขั้นตอนของการจัดกลุ่มตัวอย่าง การเก็บรวบรวมข้อมูล การวิเคราะห์และการตีความหมายข้อมูล นำไปสู่ผลลัพธ์ที่แตกต่างไปจากความเป็นจริง อคติพบได้มากกว่าความผิดพลาดจากการสุ่ม และเมื่อเกิด

อคติขึ้น ก็ส่งผลกระทบต่อการศึกษามากกว่าเช่นกัน ซึ่งอคติที่ส่งผลต่อการวิจัยสามารถแบ่งเป็น 3 กลุ่มใหญ่ๆ (Fletcher, Fletcher SW., & Wagner, 1997; Aschengrau, & Seage, 2003; Bayona, & Olsen, 2004; Friis, & Sellers, 2004) ดังนี้

(1) อคติจากการเลือกเป็นตัวอย่าง (Selection bias) เป็นความบิดเบือนในการประมาณค่าความสัมพันธ์ระหว่างปัจจัยและการเกิดโรค ซึ่งเป็นผลมาจากการคัดเลือกตัวอย่างเข้าการศึกษาทำให้เกิดผลการประมาณค่าจากกลุ่มตัวอย่างในการศึกษาแตกต่างกับผลที่ได้จากประชากรเป้าหมาย หรือความแตกต่างของความสัมพันธ์ระหว่างปัจจัยกับการเกิดโรคของกลุ่มตัวอย่างในการศึกษากับกลุ่มที่ไม่ได้เป็นตัวอย่างในการศึกษา มีสาเหตุมาจากการไม่ตอบรับการศึกษ การปฏิเสธ หรือไม่ให้ความร่วมมือในการศึกษาของกลุ่มประชากรเป้าหมาย เป็นต้น

(2) อคติของข้อมูลที่รวบรวม (Information bias) เป็นความบิดเบือนในการประมาณค่าความสัมพันธ์ระหว่างปัจจัยกับการเกิดโรค อันเนื่องมาจากความผิดพลาดจากการวัดผล ที่ทำให้ข้อมูลด้านปัจจัยระหว่างกลุ่มเกิดความแตกต่างและทำให้ผลการประมาณค่าความสัมพันธ์ระหว่างปัจจัยกับการเกิดโรคมักจะมากกว่าหรือน้อยกว่าความจริง เช่น ความผิดพลาดที่เกิดจากการบันทึกข้อมูล การสอบถามข้อมูล การให้ข้อมูลของกลุ่มตัวอย่าง หรือการรายงานผลลัพธ์ เป็นต้น

(3) อคติจากปัจจัยร่วม (Confounding bias) เป็นอคติที่เกิดจากปัจจัยร่วมที่มีอิทธิพลต่อผลการศึกษาระหว่างปัจจัยและผลลัพธ์ ทำให้เกิดความยากลำบากในการแยกความแตกต่างผลที่เกิดจากปัจจัยร่วมและผลที่เกิดจากปัจจัยหลัก

ในการศึกษาครั้งนี้ ได้คำนึงถึงการแก้ไขปัญหาของอคติจากการเลือกเป็นตัวอย่างเพียงอย่างเดียว อันเนื่องมาจากในกระบวนการของการเลือกตัวอย่างเข้าสู่การศึกษาของโครงการวิจัย กลไกการเกิดโรคมะเร็งที่เกิดจากการติดเชื้อพยาธิใบไม้ตับในประเทศไทยในระยะที่ 1 เกิดความแตกต่างระหว่างกลุ่มตัวอย่างและกลุ่มที่ไม่เข้าร่วมการศึกษา ซึ่งอาจส่งผลกระทบต่อประมาณค่าความชุกของการติดเชื้อพยาธิใบไม้ตับ ดังจะกล่าวในหัวข้อที่ 3

2. แนวทางการจัดการกับความผิดพลาดขององค์ความรู้ที่ได้จากการวิจัย

ในการศึกษาวิจัยให้มีความถูกต้องและสมบูรณ์ จะต้องไม่มีความผิดพลาดของการวิจัยเกิดขึ้นในทุกขั้นตอนของกระบวนการวิจัย ไม่ว่าจะเป็นความผิดพลาดจากการสุ่ม ความผิดพลาดโดยสุ่ม (Random error) และความผิดพลาดอย่างเป็นระบบหรืออคติ (Systematic Error หรือ Bias) การจัดการกับความผิดพลาดของการวิจัยมีหลายวิธีการ แต่สามารถแบ่งวิธีการจัดการได้เป็น 2 ช่วงของกระบวนการวิจัย ดังนี้

2.1 ช่วงของการออกแบบการวิจัย

2.1.1 การสุ่มตัวอย่าง (Randomization) เป็นวิธีการที่ให้ประชากรได้ถูกเลือกเป็นตัวอย่างเข้าสู่การศึกษา ข้อดีคือ ช่วยป้องกันการสรุปผลการศึกษาผิดพลาด ข้อเสียคือ การทำการสุ่มอย่างเท่าเทียมกันเป็นไปได้ยากในบางงานวิจัย เช่น งานวิจัยทางคลินิกที่ทำในกลุ่มผู้ป่วยโรคที่หายากหรือมีจำนวนน้อย (Fletcher et al., 1997)

2.1.2 การตั้งข้อจำกัด (Restriction) การตั้งข้อจำกัดในการคัดเลือกประชากรเป็นการกำหนดลักษณะที่ชัดเจนให้ได้กลุ่มเป้าหมายที่สนใจเข้าสู่การศึกษา ข้อดีคือ ผลลัพธ์ที่ได้จะมีความถูกต้องมากยิ่งขึ้น ข้อเสียคือ ข้อจำกัดที่ตั้งขึ้นอาจทำให้ได้ตัวอย่างที่ไม่เป็นตัวแทนของประชากรศึกษา (Fletcher et al., 1997)

2.1.3 การจับคู่ (Matching) มักใช้ในงานที่มีการเปรียบเทียบระหว่างกลุ่มที่ได้รับปัจจัยและกลุ่มที่ไม่ได้รับปัจจัย ซึ่งภายในแต่ละกลุ่มจะต้องมีลักษณะใกล้เคียงกัน ข้อดีของการจับคู่คือ เหมาะกับงานวิจัยที่ต้องการควบคุมอคติและปัจจัยร่วมได้ ข้อเสียคือ เกิดความยากลำบากในการจับคู่ตัวอย่างให้มีลักษณะปัจจัยที่คล้ายคลึงกัน เพราะตัวอย่างแต่ละบุคคลมีลักษณะที่อาจจะแตกต่างกันในหลายปัจจัย ซึ่งจะสามารถจับคู่ได้เพียง 2-3 ปัจจัยเท่านั้น นอกจากนี้ยังพบว่า ในบางงานวิจัยที่หาความแตกต่างระหว่างกลุ่มที่ศึกษา จะไม่สามารถแยกปัจจัยร่วมในการเกิดผลลัพธ์ระหว่างกลุ่มได้ (Wunsch, Linde-Zwirble, & Angus, 2006)

2.1.4 การแบ่งกลุ่ม (Stratification) เป็นการแบ่งกลุ่มย่อยของตัวอย่างตามปัจจัยศึกษาหรือการแบ่งกลุ่มย่อยพื้นฐานประชากร ซึ่งกระจายโอกาสให้บุคคลมีโอกาสดูถูกคัดเลือกเป็นตัวอย่าง ข้อดี คือ ผลลัพธ์ที่ได้จะมีความถูกต้องมากขึ้น ข้อเสียคือ ต้องระวังเกณฑ์ในการแบ่งชั้นภูมิ ถ้ามีชั้นภูมิมาก อาจเกิดความยุ่งยาก (Wunsch et al., 2006)

2.2 ช่วงของการวิเคราะห์ข้อมูลของการวิจัย

2.2.1 การปรับข้อมูลแบบตัวแปรเดียว (Simple adjustment)

การปรับข้อมูลแบบตัวแปรเดียว เป็นการปรับข้อมูลของตัวประมาณค่าตามปัจจัยของการศึกษา ซึ่งอยู่บนพื้นฐานข้อมูลของประชากร เป็นวิธีการที่นิยมใช้ในการปรับแก้ไขความผิดพลาดจากการสุ่มตัวอย่างและอคติจากการเลือกเป็นตัวอย่าง เพื่อให้ข้อมูลมีความ

สมบูรณ์มากยิ่งขึ้นและได้ผลลัพธ์ที่ใกล้เคียงกับความจริง การปรับข้อมูลแบบอย่างง่ายมี 2 วิธีการ ได้แก่

1. การถ่วงน้ำหนักข้อมูลแบบส่วนกลับของความน่าจะเป็น (Inverse probability weighting) ได้แก่ การถ่วงน้ำหนักตามวิธีการสุ่มตัวอย่างเข้าสู่การศึกษา การถ่วงน้ำหนักข้อมูลในกลุ่มไม่ตอบรับ (Nonresponse weighting) การถ่วงน้ำหนักข้อมูลแบบโพสสแตรติฟายด์ (Poststratification weighting) วิธีการนี้ มีข้อดีคือ วิธีการนี้เป็นวิธีการที่ปรับค่าข้อมูลแบบง่าย ช่วยให้สูตรความคลาดเคลื่อนมาตรฐานของตัวประมาณค่ามีรูปแบบที่ง่ายขึ้น และการคำนวณค่าของตัวประมาณค่าจากข้อมูลตัวอย่างทำได้สะดวกขึ้น ข้อเสียคือ นำข้อมูลที่เป็นผลรวมมาใช้ในการปรับอคติจากการไม่ตอบรับการศึกษาในช่วงของการวิเคราะห์ข้อมูลของงานวิจัยที่มีข้อจำกัดและขนาดตัวอย่างใหญ่ และวิธีการนี้ไม่เหมาะสมกับงานวิจัยที่มีความซับซ้อนของวิธีการวิจัย (Little, & Rubin, 2002; Szklo, & Nieto, 2007)

2. การปรับค่ามาตรฐาน (Adjustment or Standardized Rate) โดยแบ่งเป็น การปรับค่ามาตรฐานโดยวิธีตรง (Direct standardized rate) และการปรับค่ามาตรฐานโดยวิธีอ้อม (Indirect standardized rate) วิธีการนี้ มีข้อดีคือ ช่วยในการเปรียบเทียบข้อมูลต่างๆ ที่เป็นไปอย่างถูกต้อง ปราศจากความลำเอียง เพราะได้ปรับผลความแตกต่างของ 2 กลุ่มประชากร ให้มีเหมือนกัน ข้อเสียคือ ไม่ใช่อัตราติดเชื้อที่แท้จริง เนื่องจากใช้ผลรวมของข้อมูลมาปรับความแตกต่างของประชากร และเมื่ออัตราการติดเชื้อจำเพาะของแต่ละประชากรเป็นไปในทิศทางตรงกันข้าม ก็ไม่สามารถใช้ปรับความแตกต่างของประชากรได้ (ไพบูลย์ โล่ห์สุนทร, 2538)

2.2.2 การปรับข้อมูลแบบหลายตัวแปร (Multiple adjustment)

การปรับข้อมูลแบบหลายตัวแปร เป็นวิธีการปรับข้อมูลที่คำนึงถึงตัวแปรหลายๆตัวแปรพร้อมกัน โดยเหมาะสมกับงานวิจัยที่มีความซับซ้อนและต้องการควบคุมตัวแปรทั้งหมดในการวิเคราะห์ เพื่อควบคุมอคติและความผิดพลาดของการวิจัยในขั้นตอนของการวิเคราะห์ข้อมูล โดยวิธีการปรับข้อมูลหลายตัวแปรจะใช้เทคนิคของการวิเคราะห์ถดถอยพหุ (Multiple regression) ซึ่งมีอยู่หลายวิธี อาทิเช่น การวิเคราะห์ถดถอยเชิงเส้น (Linear regression) จะใช้เมื่อตัวแปรผลลัพธ์เป็นตัวแปรต่อเนื่อง การวิเคราะห์ถดถอยลอจิสติก (Logistic regression) จะใช้เมื่อตัวแปรผลลัพธ์เป็นตัวแปรแ่งนับ การวิเคราะห์แบบคือกษาค (Cox's proportional hazards) จะใช้เมื่อตัวแปรผลลัพธ์เป็นการเกิดเหตุการณ์ตามระยะเวลา และการวิเคราะห์แบบพัวซอง (Poisson regression) จะใช้เมื่อตัวแปรผลลัพธ์คืออุบัติในการเกิดเหตุการณ์ โดยระยะเวลาเข้ามาเกี่ยวข้อง นอกจากนี้ยังมีวิธีการที่คำนึงถึงกลุ่มที่ไม่เข้าร่วมการศึกษาและค่าข้อมูลที่สูญหายไป คือ การวิเคราะห์ของตัวแบบการคัดเลือกของเฮ็กแมน (Heckman selection model) ซึ่งวิธีการนี้คำนึงถึงความน่าจะเป็นของการเกิดเหตุการณ์ในทุกค่าข้อมูล รวมทั้งข้อมูลในกลุ่ม missing

การปรับข้อมูลแบบหลายตัวแปร มีข้อดี คือ สามารถควบคุมตัวแปรหลาย ๆ ตัวแปรพร้อม ๆ กันและผลลัพธ์ที่ได้มีความน่าเชื่อถือและใกล้เคียงกับความเป็นจริง ข้อเสียคือ วิธีการนี้มีความยากลำบากที่จะเข้าถึงและเข้าใจ โดยความถูกต้องของการสร้างสมการขึ้นอยู่กับข้อตกลงเบื้องต้นของลักษณะข้อมูล ซึ่งบางครั้งอาจจะเกิดเหตุการณ์ที่ปัจจัยบางตัวที่มีความสัมพันธ์สูงกับผลลัพธ์ไม่ได้เข้าในสมการ ทำให้ผลลัพธ์ที่ได้เกิดความผิดพลาดไปจากความจริง (Szklo, & Nieto, 2007)

2.2.3 การสุ่มซ้ำ (Resampling Method)

การสุ่มซ้ำ เป็นวิธีการประมาณค่าความคลาดเคลื่อนมาตรฐาน เพื่อให้การประมาณค่าพารามิเตอร์ถูกต้องมากขึ้น โดยการสุ่มซ้ำในแต่ละครั้งจะทำให้ได้ค่าความคลาดเคลื่อนมาตรฐานมีขนาดเล็กลง เกิดความแม่นยำในการประมาณค่าพารามิเตอร์มากขึ้น (Shao, & Tu, 1995) ซึ่งวิธีการของการสุ่มซ้ำมี 2 ชนิดได้แก่ การประมาณค่าความคลาดเคลื่อนมาตรฐานแบบแจ๊คไนฟ์ (Jackknife) และการประมาณค่าความคลาดเคลื่อนมาตรฐานแบบบูทสแตรป์ (Bootstrap) วิธีการนี้มีข้อดีของการสุ่มซ้ำคือ ทำให้การประมาณค่าพารามิเตอร์ได้โดยไม่มีอคติ ทำให้ผลลัพธ์ที่ได้มีความถูกต้องมากขึ้น ข้อเสียคือ การสุ่มซ้ำจะใช้กับข้อมูลที่มีขนาดใหญ่มากกว่า 200 ขึ้นไป ถึงจะทำการสุ่มซ้ำได้ (Good, 1999; Little, & Rubin, 2002)

2.2.4 การแทนที่ค่าสูญหาย (Imputation Method)

การแทนที่ค่าสูญหาย เป็นวิธีประมาณค่าข้อมูลที่สูญหาย โดยอาศัยความสัมพันธ์ระหว่างกลุ่มข้อมูลมาแทนที่กับข้อมูลที่สูญหาย ซึ่งกระบวนการวิเคราะห์นี้จะทำให้ความคลาดเคลื่อนมาตรฐาน (Standard errors) ลดต่ำลง มีผลให้การประมาณค่ามีความแม่นยำมากขึ้น ซึ่งการแทนที่ข้อมูลที่สูญหายมี 2 วิธีการ ได้แก่ การแทนที่ข้อมูลสูญหายแบบหนึ่งค่า (Single imputation) และแบบใส่ค่าเชิงพหุ (Multiple imputation) (Allison, 2002) การแทนที่ค่าสูญหาย มีข้อดีคือ เงื่อนไขอยู่บนข้อมูลของตัวแปร สามารถจัดการสูญหายแบบสุ่มอย่างสมบูรณ์ (Missing completely at random) เพื่อช่วยลดอคติจากไม่ตอบรับ ปรับให้การประมาณค่ามีความแม่นยำมากขึ้น โดยทำให้ค่าความน่าจะเป็น (P-value) น้อยและช่วงเชื่อมั่น (Confidence interval) แคบ นอกจากนี้ยังคงไว้ซึ่งความสัมพันธ์ระหว่างข้อมูลที่สูญหายกับตัวแปรสังเกต ข้อเสียคือ เป็นวิธีการนี้มีความแม่นยำน้อยกว่า Resampling เนื่องจากวิธีการนี้ใช้เพิ่มความไม่แน่นอน (Uncertainty) ของการประมาณค่าจากข้อมูลสูญหายเพียงอย่างเดียว ดังนั้นวิธีการแทนที่ที่ถูกต้องจึงอยู่บนรากฐานของปัญหาของข้อมูลที่สูญหาย จึงง่ายต่อความผิดพลาดในการเลือกวิธีการแทนที่ ส่งผลต่อความแม่นยำในการประมาณค่า (Little, & Rubin, 2002)

ในการแก้ไขปัญหาอคติจากการเลือกเป็นตัวอย่าง ควรเลือกวิธีการให้เหมาะสมกับระเบียบวิธีการวิจัยและข้อมูลที่ต้องการของงานวิจัย ซึ่งในการศึกษาคั้งนี้มุ่งเน้นแก้ไขปัญหาจากการเกิดอคติจากการเลือกตัวอย่างในช่วงของการวิเคราะห์ข้อมูล เพื่อที่จะให้ได้ผลการประมาณค่าความชุกของการติดเชื้อพยาธิใบไม้ตับที่ถูกต้อง

3. ผลกระทบของอคติจากการเลือกเป็นตัวอย่างและวิธีการแก้ไขปัญหาของสถิติที่คำนึงถึงอคติจากการเลือกเป็นตัวอย่าง

อคติจากการเลือกเป็นตัวอย่าง (Selection bias) เป็นความบิดเบือนในการประมาณค่าความสัมพันธ์ระหว่างปัจจัยกับการเกิดโรค อันเนื่องมาจากความผิดพลาดจากการวัดผล ที่ทำให้ข้อมูลด้านปัจจัยระหว่างกลุ่มเกิดความแตกต่าง และทำให้ผลการประมาณค่าความสัมพันธ์ระหว่างปัจจัยกับการเกิดโรคอาจจะมีมากกว่าหรือน้อยกว่า เช่น ความผิดพลาดที่เกิดจากการบันทึกข้อมูล การสอบถามข้อมูล การให้ข้อมูลของกลุ่มตัวอย่าง หรือการรายงานผลลัพธ์ เป็นต้น (Kleinbaum, Morgenstern, & Kupper, 1981) ซึ่งสาเหตุเหล่านี้ส่งผลกระทบท่อผลลัพธ์ที่ถูกต้อง

วิธีการแก้ไขปัญหาของอคติจากการเลือกเป็นตัวอย่าง มีทั้งจัดการกับอคติตั้งแต่ช่วงแรกก่อนการวิเคราะห์ข้อมูล โดยการควบคุมการออกแบบการวิจัยและขั้นตอนของกระบวนการเป็นข้อมูลเพื่อไม่ให้เกิดอคติ เช่นเดียวกันกับวิธีการจัดการกับความผิดพลาดของการวิจัย แต่ไม่รวมถึงวิธีการจับคู่ แต่หากภายหลังการได้มาซึ่งข้อมูลเพื่อการวิเคราะห์ผลลัพธ์ของการศึกษาให้ถูกต้อง จะต้องใช้วิธีการทางสถิติที่คำนึงถึงอคติจากการเลือกเป็นตัวอย่างมาใช้ในการวิเคราะห์ข้อมูล โดยสามารถจำแนกได้ 3 วิธีการใหญ่ๆ ดังนี้

3.1 การปรับข้อมูลแบบตัวแปรเดียว (Simple adjustment)

เป็นวิธีการปรับข้อมูลเพื่อลดอคติจากการเลือกตัวอย่าง โดยใช้ข้อมูลของตัวอย่างที่มีในการศึกษามาปรับค่าข้อมูลในแต่ละเซลล์ ให้ข้อมูลมีน้ำหนักมากขึ้น ทำให้ผลลัพธ์จากการวิเคราะห์ข้อมูลมีความถูกต้องมากยิ่งขึ้น (Korn & Graubard, 1999; Levy, & Lemeshow, 1999; Little, & Rubin, 2002) ซึ่งวิธีการปรับค่าข้อมูลมี 2 วิธีการ ได้แก่

3.1.1 การถ่วงน้ำหนักข้อมูลแบบส่วนกลับความน่าจะเป็น (Inverse probability weighting)

เป็นวิธีการถ่วงน้ำหนักข้อมูลที่อยู่บนพื้นฐานการสุ่มตัวอย่างหรือการตอบรับการศึกษา โดยใช้หลักการของส่วนกลับความน่าจะเป็นของการคัดเลือกเป็นตัวอย่างมาคูณกับผลรวมของค่าเหตุการณ์ที่สนใจในหน่วยของตัวอย่าง และหากมีการแบ่งแบบชั้นภูมิร่วมด้วย ค่าการถ่วงน้ำหนักจะต้องกำหนดให้เป็นไปตามสัดส่วนตัวอย่างในแต่ละชั้นภูมิ (Fleiss, Levin, & Paik, 2003; สุชาติ ธีระนันท์, 2538) ซึ่งวิธีการนี้จะใช้มากในการสำรวจตัวอย่างแล้วเกิดปัญหาของการไม่ตอบรับเข้าการศึกษา เมื่อทุกการสำรวจมีข้อมูลสูญหายของตัวอย่างที่สนใจหรือการไม่เข้าร่วมการศึกษาของตัวอย่างที่สนใจ จึงนำข้อมูลจากกลุ่มที่ตอบรับการศึกษามาทำการถ่วงน้ำหนักเพื่อลดอคติของกลุ่มที่ไม่ตอบรับการศึกษา (Korn & Graubard, 1999) สามารถแบ่งออกเป็น 3 วิธีการใหญ่ๆ ดังนี้

(1) การถ่วงน้ำหนักส่วนกลับความน่าจะเป็นของการคัดเลือกสามารถแบ่งตามพื้นฐานวิธีการสุ่มตัวอย่าง ซึ่งวิธีการสุ่มตัวอย่างมี 4 วิธีการ ได้แก่ การสุ่มตัวอย่างแบบอย่างง่าย (Simple random sampling) ซึ่งวิธีการนี้นอกจากจะใช้งานวิจัยที่มีวิธีการสุ่มตัวอย่างแบบอย่างง่ายแล้ว มักใช้ในงานวิจัยที่ไม่ผ่านการสุ่มตัวอย่าง การสุ่มตัวอย่างแบบการสุ่มอย่างมีระบบ (Systematic sampling) เป็นวิธีการนี้ส่วนกลับความน่าจะเป็นที่ได้จะค้ำนึ่งถึงจำนวนกลุ่มที่แบ่งได้และจำนวนตัวอย่างในแต่ละกลุ่ม การสุ่มตัวอย่างแบบการสุ่มชั้นภูมิ (Stratified random sampling) วิธีการนี้จะค้ำนึ่งถึงชั้นภูมิที่ใช้ในการแบ่งกลุ่มและการสุ่มตัวอย่างแบบแบ่งกลุ่ม (Cluster sampling) จะมี 2 ชั้น ได้แก่ การสุ่มแบบแบ่งกลุ่มชั้นที่ 1 (Simple One-Stage Cluster Sampling) มีวิธีการถ่วงน้ำหนักคล้ายกับการสุ่มแบบ Systematic และการสุ่มแบบแบ่งกลุ่มชั้นที่ 2 (Two-Stage Cluster Sampling) จะการปรับด้วยความน่าจะเป็นของจำนวนกลุ่มและจำนวนคนในแต่ละกลุ่ม (Levy, & Lemeshow, 1999)

(2) การถ่วงน้ำหนักแบบกลุ่มไม่ตอบรับการศีกษา (Nonresponse Weighting) เป็นวิธีการถ่วงน้ำหนักแบบกลุ่มไม่ตอบรับเป็นการปรับน้ำหนักข้อมูลตามกลุ่มที่ไม่ตอบรับการศีกษาจะคล้ายกับการถ่วงน้ำหนักในกลุ่มที่ตอบรับการศีกษา แต่ประยุกต์การถ่วงน้ำหนักตามเงื่อนไขการสูญหายของการสุ่มในแต่ละกลุ่ม โดยเป็นการถ่วงน้ำหนักแบบส่วนกลับความน่าจะเป็นของกลุ่มตัวอย่างในการตอบรับของแต่ละกลุ่ม (Korn & Graubard, 1999)

(3) การถ่วงน้ำหนักแบบโพสสเตรทิฟายด์ (Poststratification Weighting) เป็นวิธีการถ่วงน้ำหนักที่แบ่งตามตัวแปรที่ใช้ในการแบ่งกลุ่ม แต่ไม่ใช่ตัวแปรที่ใช้ในการคัดเลือกตัวอย่าง โดยการถ่วงน้ำหนักจะเป็นไปตามสัดส่วนของประชากรในแต่ละกลุ่มของตัวแปรที่อยู่ภายใต้สัดส่วนของการตอบรับการศีกษาในแต่ละกลุ่มของตัวแปร การถ่วงน้ำหนักแบบนี้จะเหมาะกับงานวิจัยที่ทำการสำมะโนประชากร (Census projection) ต้องการปรับเพื่อให้ข้อมูลมีความแม่นยำมากขึ้น (Levy, & Lemeshow, 1999)

3.1.2 การปรับค่ามาตรฐาน (Adjustment or Standardized Rate)

เป็นวิธีการทางระบาดวิทยาที่มาประยุกต์ใช้ในการเปรียบเทียบอัตราตายจำเพาะตามอายุ (Age specific death rate : ASDR) โดยอาศัยประชากรมาตรฐานมาปรับความแตกต่างของสองกลุ่มประชากรตามโครงสร้างของตัวแปร และตัวแปรที่นิยมคือ อายุและเพศ (Fleiss et al., 2003; Nordness, 2006; Jekel et al., 2007) การปรับค่ามาตรฐานที่ใช้ในการปรับความแตกต่างของสองกลุ่มประชากรมี 2 วิธี ได้แก่

(1) การปรับค่ามาตรฐานโดยวิธีตรง (Direct standardized rate) เป็นวิธีการนี้จะปรับความแตกต่างของสองกลุ่มประชากรตามโครงสร้างของอายุ ซึ่งประชากรที่นำมาเปรียบเทียบจะต้องมีลักษณะที่คล้ายคลึงกัน โดยในแต่ละกลุ่มประชากร นำผลรวมของจำนวนประชากรของสองกลุ่มในแต่ละกลุ่มอายุ มาคูณกับอัตราการเกิดเหตุการณ์จำเพาะอายุของแต่ละกลุ่มอายุ ทำให้ได้จำนวนที่คาดว่าจะตายในแต่ละกลุ่มอายุ แล้วนำมาหารด้วยผลรวมของ



จำนวนประชากรของทั้งสองกลุ่ม ทำให้ได้อัตราการเกิดเหตุการณ์มาตรฐานของแต่ละกลุ่มประชากร ซึ่งการปรับค่ามาตรฐานโดยวิธีตรงจะให้อัตราเกิดเหตุการณ์จริงเท่ากับหรือใกล้เคียงกับอัตราการเกิดเหตุการณ์ของข้อมูลจริง การปรับค่าประชากรมาตรฐานด้วยวิธีทางตรง เป็นวิธีการที่นิยมมากในการจัดผลกระทบของอดีตในความแตกต่างของประชากรและสามารถประยุกต์ใช้ได้กับการหาอุบัติการณ์ของโรคเช่น การติดเชื้อ หรือการบาดเจ็บและการตาย เป็นต้น

(2) การปรับค่ามาตรฐานโดยวิธีอ้อม (Indirect standardized rate)

เป็นวิธีการที่นิยมใช้ในกรณีที่ไม่ทราบอัตราการตายจำเพาะตามอายุของประชากรแต่ละกลุ่มหรือมีจำนวนน้อยมาก เพื่อให้มีความน่าเชื่อถือมากขึ้น ในการปรับจะใช้อัตราตายจำเพาะอายุของกลุ่มประชากรที่ทราบหรือประชากรใดประชากรหนึ่ง มาคูณกับจำนวนประชากรในแต่ละกลุ่มอายุของประชากรที่ไม่ทราบอัตราการตายจำเพาะแล้วนำมาหาค่าสัดส่วนการตายมาตรฐาน (Standardized mortality ratio: SMR) โดยการนำจำนวนสัดส่วนของการตายจริงมาหารด้วยจำนวนสัดส่วนของคาดว่าจะตายในแต่ละกลุ่มประชากร

3.2 การปรับข้อมูลแบบหลายตัวแปร (Multiple adjustment)

การประมาณค่าเหตุการณ์ที่สนใจเพื่อให้ได้การประมาณค่าเหตุการณ์ที่สนใจมีความถูกต้องและใกล้เคียงกับความจริงมากขึ้น ซึ่งวิธีการทางสถิติที่ใช้วิเคราะห์ข้อมูลโดยส่วนใหญ่ใช้หลักการของการปรับข้อมูลหลายตัวแปร ซึ่งเป็นเทคนิคของการวิเคราะห์ถดถอยพหุ (Multiple regression) ซึ่งมีอยู่หลายวิธี อาทิเช่น การวิเคราะห์ถดถอยเชิงเส้น (Linear regression) การวิเคราะห์ถดถอยลอจิสติก (Logistic regression) การวิเคราะห์แบบค็อกซฮาด (Cox's proportional hazards) และการวิเคราะห์แบบพัวซอง (Poisson regression) เป็นต้น แต่มีวิธีการที่คำนึงถึงกลุ่มคนที่ไม่ได้เข้าร่วมการศึกษาและมีการสูญหายของกลุ่มเป้าหมายที่ต้องการในการคัดเลือกเข้าสู่การศึกษา ได้แก่ การวิเคราะห์ด้วยตัวแบบการคัดเลือกของเฮ็กแมน (Heckman selection model) ซึ่งเป็นวิธีการทางสถิติพื้นฐานที่ใช้ในการวิเคราะห์ตัวอย่างที่คัดเลือกที่ไม่ผ่านการสุ่ม หรือมีการละเลยตัวแปรบางตัวแปร (Error specification) ที่ทำให้เกิดอคติจากการเลือกเป็นตัวอย่าง โดยจะอยู่บนพื้นฐานประชากรหรือขนาดตัวอย่างที่มีการสูญหายของข้อมูล (Heckman, 1976, 1979)

การวิเคราะห์ด้วยตัวแบบการคัดเลือกของเฮ็กแมน (Heckman selection model) มี 2 วิธีการโดยแบ่งตามลักษณะตัวแปรผลลัพธ์ ได้แก่ วิธีการที่ 1 ตัวแบบของการคัดเลือกของเฮ็กแมน (Heckman selection model) แบบ 2 ขั้นตอน (Two step) จะใช้เมื่อตัวแปรผลลัพธ์เป็นตัวแปรต่อเนื่อง ตามหลักการแรกเริ่มของ Heckman ต่อมามีการพัฒนาให้สามารถวิเคราะห์ได้กับข้อมูลที่มีตัวแปรผลลัพธ์เป็นตัวแปรแจกแจงได้แก่ วิธีการตัวแบบโพรบิตของเฮ็กแมน (Heckman probit model) ซึ่งวิธีการนี้จะอยู่บนหลักการของ maximum likelihood และข้อมูลมีการแจกแจงปกติ ในการประมาณค่าพารามิเตอร์ที่มีความแม่นยำมากขึ้น (Heckman, 1979;



Sartori, 2003) สำหรับการศึกษาที่มีตัวแปรผลลัพธ์เป็นตัวแปรเจนนับ จึงขออธิบายขั้นตอนของการวิธีการวิเคราะห์ด้วยตัวแบบโพรบิทของเฮ็กแมน (Heckman probit model) ดังนี้

(1) ข้อมูลสังเกตที่ใช้ในการวิเคราะห์จะมีค่าการสูญหาย (missing) อันเนื่องมาจากอคติจากการเลือกเป็นตัวอย่าง

(2) สร้างสมการถดถอยแบบโพรบิท (Probit regression) ซึ่งเป็นสมการการประมาณค่าความน่าจะเป็นของการถูกคัดเลือกเป็นตัวอย่างโดยไม่ผ่านการสุ่มตัวอย่างหรือละเลยบางตัวแปร ทำให้ได้ผลลัพธ์ที่มีอคติและได้ค่าความสัมพันธ์หรือค่าที่ไม่เท่ากับศูนย์

(3) สร้างสมการการคัดเลือกชั้นใหม่จากตัวแปรที่ใช้ในการคัดเลือก โดยประยุกต์จากสมการถดถอย (regression) แรก โดยใช้วิธีการของ maximum likelihood เพื่อประมาณค่าพารามิเตอร์จากทุกค่าที่เป็นไปได้ของข้อมูล รวมทั้งข้อมูลในกลุ่ม missing ปรับค่าพารามิเตอร์ไม่ให้มีอคติจากการเลือกเป็นตัวอย่างและมีความแม่นยำมากขึ้น

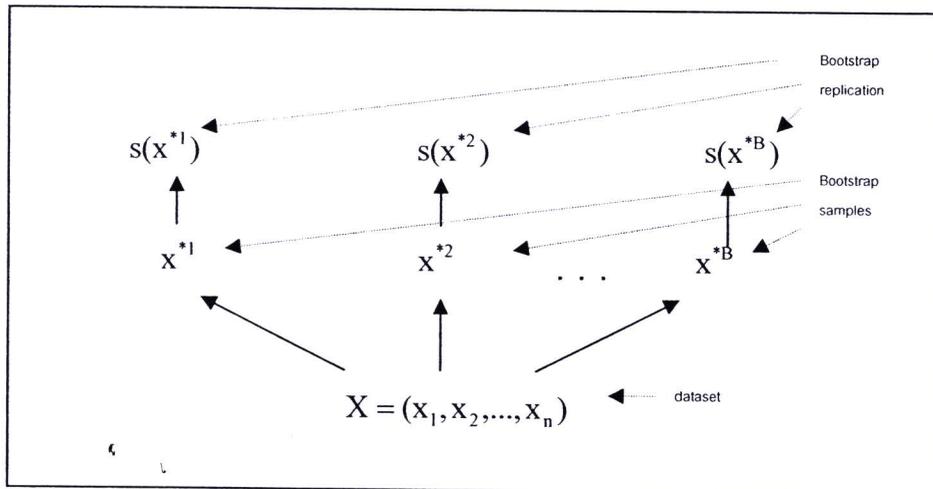
(4) นำมาทำนายค่าความน่าจะเป็นของการเกิดเหตุการณ์ของทุกค่าของข้อมูล ตามความสัมพันธ์ของปัจจัยที่เกี่ยวข้อง

3.3 การประมาณค่าแบบการสุ่มตัวอย่างซ้ำ (Resampling Method)

เป็นวิธีการประมาณค่าความคลาดเคลื่อนมาตรฐาน เพื่อให้การประมาณค่าพารามิเตอร์ถูกต้องมากขึ้น โดยการสุ่มซ้ำในแต่ละครั้งจะทำให้ได้ค่าความคลาดเคลื่อนมาตรฐาน มีขนาดเล็กลง เกิดความแม่นยำในการประมาณค่าพารามิเตอร์มากขึ้น โดยแต่ละการสุ่มตัวอย่างซ้ำของข้อมูลประชากรที่ซับซ้อน มี 2 วิธีได้ดังนี้

3.3.1 การประมาณค่าความคลาดเคลื่อนมาตรฐานแบบแจ๊คไนฟ์ (Jackknife) เป็นวิธีการแบบ Jackknife เป็นเทคนิคการสุ่มซ้ำ เพื่อการประมาณค่าพารามิเตอร์และความคลาดเคลื่อนมาตรฐาน และอคติของการประมาณค่าจากตัวอย่างที่ไม่ผ่านการคัดเลือกเป็นตัวอย่าง ซึ่งเป็นขั้นตอนแรกของการสุ่มซ้ำโดยปราศจากการแทนที่ โดยจะอยู่บนพื้นฐานการคัดเลือกตัวอย่างจากข้อมูล (Shao, & Tu, 1995; Karim, 2009)

3.3.2 การประมาณค่าความคลาดเคลื่อนมาตรฐานแบบบูทสแตรป์ (Bootstrap) เป็นวิธีการสุ่มตัวอย่างซ้ำโดยใช้ข้อมูลของตัวอย่างที่มีทำการสุ่มซ้ำหลาย ๆ ชุด เพื่อใช้ในการอนุมานเชิงสถิติ เพื่อให้ได้ผลการอนุมานที่ดีขึ้นหรือมีความคลาดเคลื่อนน้อยลง (มานพวรภักดิ์, 2550) วิธีการแบบ Bootstrap จะอนุมานที่อยู่บนพื้นฐานลักษณะทางประชากรของกลุ่มตัวอย่าง เพื่อให้เกิดความแม่นยำของข้อมูลรายละเอียดของลักษณะทางประชากรของกลุ่มตัวอย่างที่ครบถ้วนสมบูรณ์มากขึ้น (Efron, & Tibshirani, 1998; Karim, 2009) สามารถอธิบายด้วยภาพที่ 1 ดังนี้



ภาพที่ 1 แผนภาพอธิบายของวิธีการแบบ Bootstrap

จากภาพที่ 1 Bootstrap algorithm เริ่มต้นจากการ generate จำนวนมากของ Bootstrap samples ที่เป็นอิสระกัน x^1, x^2, \dots, x^B ในแต่ละขนาดตัวอย่าง n จาก original data set $X = (x_1, x_2, \dots, x_n)$ แล้วทำการสุ่มซ้ำของ Bootstrap ของ S กล่าวคือ $\hat{\theta}^*(b) = s(x^b)$, $b=1, 2, \dots, B$ ได้ $s(x^1), s(x^2), \dots, s(x^B)$ ซึ่งแต่ละตัวอย่างของ Bootstrap จะมีความคล้ายคลึงกัน (Efron, & Tibshirani, 1998)

3.4 การประมาณค่าแบบการแทนที่ค่าสูญหาย (Imputation Method)

การแทนค่าข้อมูลสูญหาย เป็นวิธีการแทนที่ข้อมูลที่สูญหายด้วยการทำนายอย่างมีเหตุผลของแต่ละค่าการสูญหาย หรืออาศัยความสัมพันธ์ของตัวแปรอิสระระหว่างกลุ่มข้อมูลที่มีกับข้อมูลสูญหาย มาทำนายการเกิดเหตุการณ์ของข้อมูลสูญหายในแต่ละรายข้อมูลที่สูญหาย และเมื่อได้ข้อมูลที่สมบูรณ์จะช่วยการประมาณค่ามีความแม่นยำมากขึ้นและค่าความคลาดเคลื่อนมาตรฐานลดต่ำลง (Allison, 2002) ซึ่งวิธีการแทนค่ามี 2 แบบ ได้แก่

3.4.1 การแทนที่แบบหนึ่งค่า (Single imputation)

เป็นแทนค่าโดยใช้ค่า ๆ หนึ่งเข้าไปแทนที่ค่าที่สูญหายในตัวแปรผลลัพธ์ เพื่อให้ข้อมูลสมบูรณ์ (Little, & Rubin, 2002) ซึ่งค่าที่นิยมใช้ในการแทนที่สามารจจำแนกดังนี้

(1) วิธีการของตัวแบบที่ชัดเจน (Explicit Modeling Method)

เป็นค่าการทำนายที่ชัดเจน จะต้องเป็นไปตามหลักเกณฑ์ทางสถิติ ได้แก่

การแทนที่ด้วยค่าเฉลี่ย (Mean Imputation) โดยนำค่าเฉลี่ยของข้อมูลที่มีอยู่โดยไม่มีค่าสูญหาย เมื่อได้ค่าเฉลี่ยแล้วนำแทนที่ข้อมูลที่สูญหาย การแทนที่ด้วยค่าเฉลี่ยนนำไปสู่การประมาณค่าที่คล้ยกับการถ่วงน้ำหนัก

การแทนที่ด้วยสมการถดถอย (Regression Imputation) โดยนำค่าการทำนายที่ได้จากการหาความสัมพันธ์ของข้อมูลที่มีอยู่โดยไม่มีค่าการสูญหาย แล้วนำค่าการทำนายที่ได้ไปแทนที่ข้อมูลที่สูญหาย แล้วทำการหาความสัมพันธ์ใหม่

การแทนที่ด้วยการถดถอยแบบสโตแคสติก (Stochastic regression Imputation) ใช้หลักการเช่นเดียวกันกับ Regression Imputation แต่จะมีค่าส่วนที่เหลือ (Residual) จะให้ค่าความสัมพันธ์มีค่าใกล้เคียงกับค่าจริง

(2) วิธีการของตัวแบบโดยนัย (Implicit Modeling Method) เป็นวิธีการที่อยู่ภายใต้เงื่อนไขของตัวแบบตามหลักการเหตุและผลของการเลือกใช้ ได้แก่

การแทนที่แบบฮอตเด็ค (Hot deck Imputation) เป็นการแทนที่ค่าการสูญหายด้วยค่าการกระจายของข้อมูล เช่น ค่าเฉลี่ย (mean) ค่ามัธยฐาน (median) เป็นต้น

การแทนที่แบบแทนค่า (Substitution Imputation) จะใช้ควบคุมการไม่ตอบรับในการสำรวจข้อมูลภาคสนาม เช่น ในกลุ่มตัวอย่างที่ไม่อยู่ อาจทำการแทนที่ข้อมูลด้วยกลุ่มตัวอย่างที่ไม่ได้ถูกเลือกแทน ซึ่งในขั้นตอนการวิเคราะห์อาจจะต้องเลือกค่าบางชนิดมาแทนที่

การแทนที่แบบโคลด์เด็ค (Cold deck Imputation) เป็นการแทนที่ค่าการสูญหายด้วยค่าคงที่จากภายนอก เช่น ค่าที่ได้จากก่อนการสำรวจ ค่าที่ตอบรับได้ในรายละเอียดหรือข้อมูลที่ต้องการศึกษา เป็นต้น

การแทนที่แบบประกอบด้วยหลายส่วน (Composite methods) เป็นการรวมแนวคิดจากวิธีการต่าง ๆ เข้ามาใช้ในการแทนที่ข้อมูลที่สูญหาย

3.4.2 การแทนที่แบบใส่ค่าเชิงพหุ (Multiple imputation)

เป็นวิธีการแทนค่าที่มากกว่า 1 ค่า จะใช้ในการหาความสัมพันธ์ การประมาณค่าข้อมูลที่มีการสูญหายจากการสุ่มตัวอย่าง เพื่อขจัดอคติจากการเลือกตัวอย่าง โดยค่าที่นำมาแทนที่จะมาจาก single imputation ซึ่งอาจมาจาก 1 ตัวแบบหรือมากกว่า แต่ส่วนใหญ่จะสร้างจาก explicit regression model และ implicit model (Little, & Rubin, 2002) ในขั้นตอนของการแทนที่ด้วยวิธีการนี้จะเพิ่มเติมจากวิธีการของ single imputation ที่การแทนค่าจะมากกว่า 1 ค่า จากตัวแบบที่ต่างกัน แล้วทำการสุ่มซ้ำภายใต้ข้อมูลการไม่ตอบรับเข้าการศึกษา การให้ได้การประมาณค่าเพียงค่าเดียว จะได้จากการนำการประมาณค่าที่ได้จากทุกตัวแบบมาหาค่าเฉลี่ย ซึ่งผลลัพธ์ที่ได้จะมีความถูกต้องและค่าความคลาดเคลื่อนมาตรฐานต่ำ

การประมาณค่าที่ได้จากการแทนที่แบบใส่ค่าเชิงพหุให้ความไวของการวิเคราะห์น้อยกว่าการใช้ maximum likelihood เพราะค่าที่ใช้ในการแทนที่ จะใช้แทนที่ข้อมูลการสูญหายเพียงอย่างเดียวไม่ได้ช่วยประมาณค่าพารามิเตอร์ตัวอื่น แต่สิ่งที่ดีกว่าคือ ไม่จำเป็นต้องมีการประยุกต์ในการใช้ตัวแบบการแจกแจงปกติของตัวแปรหลายตัว ที่มีพื้นฐานการประมาณค่าแบบ maximum likelihood โดยมีเงื่อนไขคือ ตัวแปรทุกตัวต้องมีการแจกแจงปกติและตัวแปรทุกตัวมีความสัมพันธ์เชิงเส้นกัน (Allison, 2002)

ในการศึกษาครั้งนี้ ต้องการประมาณค่าความชุกของการติดเชื้อพยาธิใบไม้ตับให้มีความถูกต้องมากขึ้น จึงพิจารณาวิเคราะห์ข้อมูลด้วยวิธีการทางสถิติที่คำนึงถึงอคติจากการเลือกเป็นตัวอย่าง หลายรูปแบบ ดังจะกล่าวในข้อที่ 4

4. วิธีการทางสถิติในการจัดการกับอคติจากการเลือกเป็นตัวอย่างของการวิจัย

การจัดการกับอคติจากการคัดเลือกเป็นตัวอย่างในการประมาณค่าความชุกการติดเชื้อพยาธิใบไม้ตับในช่วงของการวิเคราะห์ข้อมูลในการศึกษาครั้งนี้ เพื่อให้ได้ซึ่งผลการประมาณค่าความชุกการติดเชื้อพยาธิใบไม้ตับมีความถูกต้องมากขึ้น ได้ทำการศึกษาข้อมูลของโครงการวิจัยกลไกการเกิดโรคมะเร็งที่เกิดจากการติดเชื้อพยาธิใบไม้ตับในประเทศไทย ในระยะที่ 1

วิธีการจัดการกับอคติจากการเลือกเป็นตัวอย่างที่ผู้วิจัยพิจารณาแล้วว่า เหมาะสมกับการควบคุมอคติจากการเลือกเป็นตัวอย่างในการศึกษานี้ เพื่อเพิ่มความถูกต้องของการประมาณค่าความชุกในการติดเชื้อพยาธิใบไม้ตับ ได้ดังต่อไปนี้

4.1 การถ่วงน้ำหนักข้อมูลแบบส่วนกลับความน่าจะเป็น (Inverse probability weighting) เนื่องจากข้อมูลของโครงการวิจัยข้างต้น มีกลุ่มตัวอย่างในการศึกษาเป็นกลุ่มที่เข้าร่วมเป็นอาสาสมัครด้วยความสมัครใจและส่งอุจจาระตรวจให้กับโครงการวิจัย ซึ่งมีการสูญหายของกลุ่มเป้าหมายที่สนใจจากการไม่อยู่บ้านของประชากร วิธีการที่เหมาะสมในการจัดการกับอคติจากการเลือกเป็นตัวอย่างประชากร ได้แก่ การถ่วงน้ำหนักข้อมูลแบบส่วนกลับความน่าจะเป็นของการคัดเลือกเป็นตัวอย่าง (Inverse probability weighting) แบบการสุ่มอย่างง่าย (Simple random sampling) การถ่วงน้ำหนักข้อมูลแบบกลุ่มไม่ตอบรับการศึกษ (Nonresponse Weighting) ซึ่งเป็นวิธีการถ่วงน้ำหนักข้อมูลพื้นฐานที่เกิดอคติจากการคัดเลือกเป็นตัวอย่างตามเงื่อนไขการสูญหายแบบสุ่มอย่างสมบูรณ์ (Missing completely at random)

การถ่วงน้ำหนักข้อมูลแบบส่วนกลับความน่าจะเป็น มีข้อดีคือ เป็นวิธีการที่ปรับค่าข้อมูลแบบง่าย โดยใช้ส่วนกลับของความน่าจะเป็นตัวอย่างในแต่ละกลุ่มแบ่ง มาปรับความแตกต่างของข้อมูลที่เกิดอคติจากการไม่ตอบรับการศึกษในแต่ละกลุ่ม ข้อเสียคือ นำข้อมูลที่เป็น

ผลรวมมาใช้ในการปรับอคติจากการไม่ตอบรับการศึกษในช่วงของการวิเคราะห์ข้อมูลของงานวิจัย ซึ่งไม่ได้คำนึงถึงผลลัพธ์ในแต่ละรายบุคคล (Little, & Rubin, 2002)

4.2 การปรับค่ามาตรฐาน (Adjustment or Standardized Rate) เนื่องจากมีข้อมูลประชากรทั้งหมดของโครงการวิจัยข้างต้น จึงพิจารณาใช้วิธีการปรับค่ามาตรฐานโดยวิธีตรง (Direct standardized rate) โดยปรับตามโครงสร้างของตัวแปรอายุและเพศ วิธีการนี้เหมาะสมกับงานวิจัยที่ทราบข้อมูลประชากร และมีการเปรียบเทียบอัตราการเกิดเหตุการณ์ระหว่างข้อมูลประชากร เพื่อให้การเปรียบเทียบปราศจากการอคติ (Jekel et al., 2007)

การปรับค่ามาตรฐานโดยวิธีตรง มีข้อดีคือ เป็นการนำข้อมูลประชากรมาตรฐานมาปรับผลความแตกต่างของประชากร เพื่อขจัดอคติจากการเลือกเป็นตัวอย่างของกลุ่มตัวอย่าง ข้อเสียคือ ต้องทราบข้อมูลประชากรตามโครงสร้างตัวแปรของประชากรที่ต้องการปรับ เพื่อที่จะนำผลรวมของข้อมูลประชากรในแต่ละกลุ่ม มาปรับความแตกต่างของ 2 กลุ่มประชากรให้มีความเหมือนกัน หากลักษณะข้อมูลประชากรของ 2 กลุ่มประชากรแตกต่างกันมาก อาจทำให้ผลลัพธ์ไม่ถูกต้อง (Jekel et al., 2007; ไพบูลย์ โล่ห์สุนทร, 2538)

4.3 การปรับค่าข้อมูลแบบหลายตัวแปร (Multiple adjustment) เนื่องจากทราบข้อมูลประชากรทั้งหมด และมีการสูญหายของกลุ่มเป้าหมายในขั้นตอนของการคัดเลือกตัวอย่าง ทำให้กลุ่มตัวอย่างที่ไม่เป็นตัวแทนประชากร จึงพิจารณาใช้การวิเคราะห์ด้วยตัวแบบโพรบิทของเฮ็กแมน (Heckman probit model) ในการจัดการกับอคติจากการเลือกเป็นตัวอย่าง ซึ่งวิธีการนี้เหมาะสมกับงานวิจัยที่ต้องการทำนายความน่าจะเป็นของการเกิดเหตุการณ์ของข้อมูลที่สมบูรณ์ร่วมกับข้อมูลที่สูญหาย โดยไม่จำเป็นต้องตัดบุคคลที่มีข้อมูลสูญหายออกจากการศึกษา เพื่อให้ผลการประมาณค่ามีความถูกต้องมากขึ้น (Heckman, 1976)

การวิเคราะห์ด้วยตัวแบบโพรบิทของเฮ็กแมน มีข้อดี คือ สามารถทำนายความน่าจะเป็นของการเกิดเหตุการณ์ของทุกค่าในข้อมูลรวมทั้งข้อมูลในกลุ่มสูญหาย พร้อมกับควบคุมตัวแปรหลายๆตัวแปรพร้อม ๆ กัน ทำให้ได้ผลการประมาณค่าจากทุกค่าในข้อมูล ผลลัพธ์มีความถูกต้องมากขึ้น ข้อเสียคือ ความถูกต้องของการสร้างสมการการทำนายขึ้นอยู่กับข้อตกลงเบื้องต้นของลักษณะข้อมูล ซึ่งบางครั้งอาจจะเกิดเหตุการณ์ที่ปัจจัยบางตัวที่มีความสัมพันธ์สูงกับผลลัพธ์ไม่ได้เข้าในสมการ ทำให้ผลลัพธ์ที่ได้เกิดความผิดพลาดไปจากความจริง (Heckman, 1976; Allison, 2002)

4.4 การแทนที่แบบใส่ค่าเชิงพหุ (Multiple Imputation) แล้วสุ่มซ้ำ (Resampling)

เนื่องจากการศึกษาทราบข้อมูลประชากรทั้งหมด ประกอบด้วยข้อมูลกลุ่มตัวอย่างและกลุ่มที่ไม่เข้าร่วมการศึกษา ที่มีลักษณะทางประชากร ได้แก่ เพศ อายุ และหมู่บ้าน แต่ไม่ทราบผลการติดเชื้อของกลุ่มที่ไม่เข้าร่วมการศึกษา จึงพิจารณาใช้การแทนที่แบบใส่ค่าเชิงพหุ แล้วสุ่มซ้ำ โดยยึดหลักความน่าจะเป็นของการติดเชื้อพยาธิใบไม้ตับของกลุ่มที่ไม่ได้เข้าร่วมในการศึกษา ที่มีลักษณะทางประชากรคล้ายคลึงกับกลุ่มตัวอย่าง เพื่อแทนที่ค่าการติดเชื้อพยาธิใบไม้ตับในกลุ่มที่ไม่เข้าร่วมการศึกษา เพื่อให้ได้การประมาณค่าความชุกของการติดเชื้อพยาธิใบไม้ตับที่ถูกต้องมากขึ้น และลดอคติจากการเลือกเป็นตัวอย่าง เนื่องจากวิธีการนี้คำนึงถึงการแทนที่ค่าสูญหายของกลุ่มที่ไม่ได้เข้าร่วมการศึกษาในแต่ละรายบุคคล โดยอาศัยลักษณะทางประชากรระหว่างกลุ่มตัวอย่างและกลุ่มที่ไม่เข้าการศึกษาที่เหมือนกัน โดยนำค่าการติดเชื้อพยาธิใบไม้ตับของกลุ่มตัวอย่าง มาแทนค่าการติดเชื้อพยาธิใบไม้ตับของกลุ่มที่ไม่เข้าร่วมการศึกษาในแต่ละรายบุคคล ที่มีลักษณะทางประชากรเหมือนกันกลุ่มตัวอย่าง โดยเพิ่มวิธีการสุ่มซ้ำมากกว่า 1 รอบ ในการสร้างข้อมูลให้สมบูรณ์ในหลายชุดข้อมูล

การแทนที่แบบใส่ค่าเชิงพหุ แล้วสุ่มซ้ำ โดยยึดหลักความน่าจะเป็นของการติดเชื้อพยาธิใบไม้ตับของกลุ่มที่ไม่ได้เข้าร่วมในการศึกษา ที่มีลักษณะทางประชากรคล้ายคลึงกับกลุ่มตัวอย่าง เพื่อแทนที่ค่าการติดเชื้อพยาธิใบไม้ตับในกลุ่มที่ไม่เข้าร่วมการศึกษา มีข้อดีคือ เป็นการแทนที่ค่าการติดเชื้อพยาธิใบไม้ตับของกลุ่มที่ไม่เข้าร่วมการศึกษาแต่ละรายบุคคล ตามความสัมพันธ์ของปัจจัยที่เหมือนกันระหว่างกลุ่มตัวอย่างกับกลุ่มไม่เข้าร่วมการศึกษา โดยเพิ่มการสุ่มซ้ำมากกว่า 1 รอบ ซึ่งในแต่ละรอบของการสุ่มซ้ำทำการประมาณค่าความคลาดเคลื่อนมาตรฐานให้น้อยลง ส่งผลให้การประมาณค่ามีความถูกต้องและแม่นยำมากขึ้น ข้อเสียคือ หากมีข้อมูลสูญหายมากกว่า 1 ตัวแปร จำเป็นต้องเลือกใช้วิธีการแทนที่ค่าการสูญหายหลายแบบ มีความยุ่งยากของการวิเคราะห์ที่มีหลายขั้นตอน จึงจำเป็นต้องมีซอฟต์แวร์ที่ตีมาช่วยในการวิเคราะห์ข้อมูล (Allison, 2002)

5. การติดเชื้อพยาธิใบไม้ตับ

พยาธิใบไม้ตับ (*Opisthorchis viverrini*: OV) เป็นพยาธิที่รูปร่างคล้ายใบไม้ แบนยาว ส่วนท้ายป้านกว่าส่วนหัวในอุจจาระ ซึ่งพยาธิใบไม้ตับเป็นพยาธิใบไม้ตับของสุนัข และแมว ไข่จะปนมากับอุจจาระลงไปใต้น้ำ หอยน้ำจืดบางชนิดจะกินไข่เข้าไป ไมราซิเดียมจะฟักออกจากไข่ในลำไส้ของหอยแล้วไข่ออกนอกลำไส้หอย ไข่เข้าไปใต้เกล็ดและเนื้อปลา เมื่อคนรับประทานอาหารที่ปรุงจากปลาดิบ ทำให้เกิดการติดเชื้อพยาธิใบไม้ตับ (วันชัย มาลีวงษ์, ผิวพรรณ มาลีวงษ์ และ มรกต นิมิตร, 2544)

ความรุนแรงของโรคที่เกิดจากการติดเชื้อพยาธิใบไม้ตับขึ้นอยู่กับจำนวนพยาธิใบไม้ตับที่อยู่ในร่างกาย โดยทำการตรวจนับจำนวนไข่พยาธิในอุจจาระใช้ หากตรวจเจอไข่พยาธิในอุจจาระ 1-999 ไข่ อยู่ระดับไม่รุนแรง 1,000-9,999 ไข่ อยู่ระดับรุนแรงเล็กน้อย 10,000-29,999 ไข่ อยู่ระดับรุนแรงปานกลาง และมากกว่า 30,000 ไข่ อยู่ระดับรุนแรงมาก (SADUN, 1955; สว่างใจ พึ่งพักตร์, 2536; อุ่น เกียรติวุฒิ และคณะ, 2545)

พยาธิสภาพของผู้ติดเชื้อพยาธิใบไม้ตับ ในระยะแรกจะมีการอักเสบของเซลล์บุผนังท่อน้ำดีเล็ก เกิดจากการมีพยาธิจำนวนมากอุดตันที่ท่อน้ำดี (สว่างใจ พึ่งพักตร์, 2536) โดยแสดงอาการ ท้องอืด ท้องเฟ้อ เจ็บท้องบริเวณใต้ชายโครง เบื่ออาหาร อ่อนเพลีย คลื่นไส้ อาเจียน เป็นต้น หากมีสาเหตุแทรกซ้อน มีผลทำให้ท่อน้ำดีโป่งพอง เกิดถุงน้ำหรือก้อนเนื้อ กลายเป็นมะเร็งท่อน้ำดี (Cholangiocarcinoma) และมะเร็งตับ (อุ่น เกียรติวุฒิ และคณะ, 2545)

การตรวจหาไข่พยาธิตับหรือตัวพยาธิใบไม้ตับในอุจจาระหรือในตับก่อน เพื่อให้ทราบถึงความรุนแรงของโรคได้ โดยมีวิธีการตรวจหาไข่หรือตัวพยาธิใบไม้มีหลากหลายวิธีการ ยกตัวอย่าง เช่น การตรวจหาไข่พยาธิในอุจจาระ โดยนับจำนวนไข่พยาธิในอุจจาระ ซึ่งเป็นวิธีที่ดีและน่าเชื่อถือได้มากที่สุด (ประยงค์ ระดมยศ, สุวณี สุภเวทย์ และศรัชัย หล่ออารีย์สุวรรณ, 2539) แบ่ง 2 วิธี ได้แก่ 1) การตรวจโดยวิธีธรรมดา (Simple smear) 2) การตรวจโดยวิธีเข้มข้น (Concentration method) (บรรณศาสตร์ ไชยพรรณนา และอภิชาติ ชมภูบุตร, 2536) นอกจากนี้ยังมีวิธีการตรวจหาตัวพยาธิในอุจจาระ จะได้รับการตรวจภายหลังการทานยาขับพยาธิ ตลอด 24 ชั่วโมง นำมาร้อนหาตัวพยาธิและชนิดของพยาธิ การตรวจน้ำดีหรือน้ำย่อย การตรวจทางอิมมูโน และการตรวจทางรังสี เพื่อเห็นพยาธิของโรคและภาวะแทรกซ้อน (สว่างใจ พึ่งพักตร์, 2536)

การรักษาโรคพยาธิใบไม้ตับในประเทศไทย แบ่งได้ 2 วิธีการ ได้แก่ วิธีการแรกจะทำการรักษาด้วยยา โดยยาที่ใช้รักษา เช่น ยาพราซิควอนเทล (Praziquantel) ยามีเบนดาโซล (Mebendazole) ยาอัลเบนดาโซล (Albendazole) เป็นต้น (คณาจารย์ภาควิชาปรสิตวิทยา, 2536; สว่างใจ พึ่งพักตร์, 2536) วิธีการที่สอง คือ วิธีการผ่าตัดใช้แก้ไขปัญหากลุ่มหลังรักษาด้วยยาแล้วไม่ได้อผล เนื่องจากพยาธิสภาพของโรคเกิดความรุนแรง จนทำให้อุดตันในท่อน้ำดี นิ้วในท่อน้ำดี หรือมะเร็งท่อน้ำดี จึงใช้วิธีการผ่าตัดเพื่อระบายน้ำดีจากตับลงสู่ลำไส้ (สว่างใจ พึ่งพักตร์, 2536)

ในการศึกษาครั้งนี้ ใช้ข้อมูลโครงการวิจัยกลไกการเกิดโรคมะเร็งที่เกิดจากการติดเชื้อพยาธิใบไม้ตับในประเทศไทย ในระยะที่ 1 ซึ่งผู้ที่เข้าร่วมเป็นอาสาสมัครภายหลังการลงชื่อในใบยินยอมเข้าร่วมเป็นอาสาสมัครของโครงการวิจัย จะได้รับการตรวจหาไข่พยาธิใบไม้ตับในอุจจาระ (Fecal exams) ด้วยเทคนิคฟอร์มาลิน-เอทิล อะซิเตท (Formalin-ethyl acetate technique) เพื่อนับไข่พยาธิใบไม้ตับในอุจจาระ ดูการติดเชื้อพยาธิใบไม้ตับของกลุ่มที่เข้าร่วมเป็นอาสาสมัครที่ส่งอุจจาระตรวจ ผู้ที่ติดเชื้อพยาธิใบไม้ตับ คือ ผู้เข้าร่วมโครงการวิจัยที่ส่งตรวจอุจจาระและตรวจเจอไข่พยาธิใบไม้ตับตั้งแต่ 1 ใบขึ้นไป

6. งานวิจัยที่เกี่ยวข้อง (นำเสนอเป็นตาราง)

ในงานวิจัยที่ประสบปัญหาของอคติจากการเลือกเป็นตัวอย่าง ทำให้ได้ผลการประมาณค่าเหตุการณ์ที่สนใจไม่ถูกต้อง จึงมีหลายงานที่ได้เลือกใช้วิธีการทางสถิติที่คำนึงถึงอคติจากการเลือกเป็นตัวอย่าง มาใช้ในการคำนวณในช่วงของการวิเคราะห์ข้อมูล เพื่อได้ผลการประมาณค่าความชุกที่สนใจให้ถูกต้องมากยิ่งขึ้น และทราบผลกระทบของอคติจากการเลือกเป็นตัวอย่างต่อการประมาณค่าความชุกการติดเชื้อพยาธิใบไม้ตับ ดังตัวอย่างงานวิจัยในตารางที่ 1

ตารางที่ 1 การทบทวนวรรณกรรมที่เกี่ยวข้องกับการประมาณค่าที่คำนึงถึงอคติจากการเลือกเป็นตัวอย่าง

ชื่อเรื่อง	ปีที่ศึกษา	ระเบียบวิธีวิจัยและกลุ่มตัวอย่าง โดยย่อ	การเกิดอคติ จากการเลือก เป็นตัวอย่าง	ผลการวิเคราะห์ข้อมูล โดยย่อ	สถิติที่ใช้ในการประมาณ ค่าที่คำนึงถึงอคติจากการ เลือกเป็นตัวอย่าง
Multiple Imputation to Account for Estimating the Prevalence of Osteoporosis (Kmetz et al., 2002)	2002	ทำการประมาณค่าความชุกของโรคกระดูกพรุนในประชากรประเทศแคนาดา โดยมีรูปแบบการศึกษา Cohort study ทำการสุ่มคัดเลือกตัวอย่างได้เท่ากับ 9,423 คน โดยทำการส่งแบบสอบถามถึงปัจจัยเสี่ยงที่ทำให้เกิดโรคกระดูกพรุนให้กลับกลุ่มประชากรศึกษา	การศึกษานี้มีตัวอย่างในการศึกษา 9,423 คน คิดเป็น 42% ของประชากรแคนาดาทั้งหมด	ค่าความชุกของการเกิดโรคกระดูกพรุนของกลุ่มที่เข้าร่วมการศึกษามีค่าเฉลี่ยเท่ากับ 0.41 (95% CI = 0.36 ถึง 0.47) ค่าความชุกของการเกิดโรคของกลุ่มไม่เข้าร่วมการศึกษามีค่าเฉลี่ยเท่ากับ 0.39 (95% CI = 0.31 ถึง 0.47) แต่ค่าที่ได้จากเทคนิค Multiple imputation มีค่าเฉลี่ยเท่ากับ 0.39 (95% CI = 0.35 ถึง 0.45)	- เทคนิคการปรับข้อมูลด้วย Multiple imputation



ตารางที่ 1 การทบทวนวรรณกรรมที่เกี่ยวข้องกับการประมาณค่าที่คำนึงถึงอคติจากการเลือกเป็นตัวอย่าง (ต่อ)

ชื่อเรื่อง	ปีที่ศึกษา	ระเบียบวิธีวิจัยและกลุ่มตัวอย่าง โดยย่อ	การเกิดอคติจากการเลือกเป็นตัวอย่าง	ผลการวิเคราะห์ข้อมูล โดยย่อ	สถิติที่ใช้ในการประมาณค่าที่คำนึงถึงอคติจากการเลือกเป็นตัวอย่าง
Comparison of STD prevalence in the Mwanza, Rakai and Masaka trial population : the role of selection bias and diagnostic error (Orroth et al., 2003)	2003	ทำการประมาณค่าความชุกของโรคติดต่อทางเพศสัมพันธ์ (sexually transmitted diseases : STD) ของประชากรใน 3 ชุมชนที่มีอายุระหว่าง 15-54 ปี ได้แก่ Mwanza, Rakai และ Masaka โดยทำการเปรียบเทียบกับการประมาณค่าความชุกภายหลังการปรับค่าข้อมูลความชุก STD ด้วยการวิเคราะห์ sensitivity และ specificity ของ diagnostic test	ประชากรไม่ได้ผ่านการสุ่ม ซึ่งเป็นการเลือกที่เป็นการเลือกที่ยินยอมเข้าร่วมโครงการวิจัย	ค่าความชุกของการติดเชื้อทางเพศสัมพันธ์ (STD) ของทั้ง 3 ชุมชน underestimate ภายหลังการปรับค่าข้อมูล ยกตัวอย่างเช่น ในกลุ่มเพศหญิงที่มีอายุ 15-39 ปี ที่ตรวจ CT ก่อนการปรับข้อมูล พบว่า ค่าความชุกของโรคติดต่อทางเพศสัมพันธ์ของ Mwanza, Rakai และ Masaka เท่ากับ 2.7, 1.4 และ 6.4 (ตามลำดับ) แต่ภายหลังการปรับค่าข้อมูลได้ค่าความชุกเท่ากับ 3.2, 1.6 และ 13.0 (ตามลำดับ)	การปรับข้อมูลด้วย sensitivity และ specificity ของ diagnostic test

ตารางที่ 1 การทบทวนวรรณกรรมที่เกี่ยวข้องกับการประมาณค่าที่คำนึงถึงอคติจากการเลือกเป็นตัวอย่าง (ต่อ)

ชื่อเรื่อง	ปีที่ศึกษา	ระเบียบวิธีวิจัยและกลุ่มตัวอย่าง โดยย่อ	การเกิดอคติ จากการเลือก เป็นตัวอย่าง	ผลการวิเคราะห์ข้อมูล โดยย่อ	สถิติที่ใช้ในการประมาณ ค่าที่คำนึงถึงอคติจาก การเลือกเป็นตัวอย่าง
Disclosure of HIV Test Results and Non- response Bias in Seroprevalence Surveys (Reniers et al., 2007)	2006	ในการศึกษาผลการติดเชื้อ HIV ของ ผู้ป่วยเข้ารับการรักษาในโรงพยาบาล Addis Ababa กลุ่มตัวอย่างที่ ลงทะเบียนเข้าการศึกษาทั้งหมด 27,000 คน โดยประชากรศึกษาจะมี การได้รับเก็บข้อมูลทั่วไปของลักษณะ ทางประชากรและประวัติการรักษา เพื่อนำข้อมูลไปใช้ทำนายผลการ ทดสอบ HIV ในกลุ่มที่ปฏิเสธการเข้า ร่วมการศึกษา	ตัวอย่างที่ได้ ไม่ได้ผ่านการ สุ่มตัวอย่าง จึง ไม่ได้กลุ่ม ประชากร เป้าหมาย ทั้งหมด เพราะ มีกลุ่มที่ปฏิเสธ การเข้าร่วม การศึกษา	ในการประมาณค่าความ ชุกของการติดเชื้อ HIV ด้วย Heckman probit model คือ 17.7% ในขณะที่การประมาณค่า ความชุกของ logistic regression เท่ากับ 23.1% และความชุกของข้อมูล จริงเท่ากับ 22.2%	การวิเคราะห์ Heckman probit model

ตารางที่ 1 การทบทวนวรรณกรรมที่เกี่ยวข้องกับการประมาณค่าที่คำนึงถึงอคติจากการเลือกเป็นตัวอย่าง (ต่อ)

ชื่อเรื่อง	ปีที่ศึกษา	ระเบียบวิธีวิจัยและกลุ่มตัวอย่าง	การเกิดอคติจากการเลือกเป็นตัวอย่าง	ผลการวิเคราะห์ข้อมูลโดยย่อ	สถิติที่ใช้ในการประมาณค่าที่คำนึงถึงอคติจากการเลือกเป็นตัวอย่าง
Resampling methods to reduce the selection bias in genetic effect estimation in genome-wide scans (Wu, Lee, Shi, Sun, & Bull, 2004)	2004	ทำการจำลองข้อมูลขึ้นเพื่อลดปัญหา selection bias สำหรับการประมาณค่าผลทางพันธุกรรมในการตรวจสอบ genome-wide ของกลุ่มประชากร 4 กลุ่มได้แก่ Aipotu (AI), Danaca (DA), Karangar (KA) และ New York (NY) โดยทำการประมาณค่าแบบง่ายของข้อมูลทางพันธุกรรมเปรียบเทียบ three bootstrap-based resampling estimators ได้แก่ shrinkage, an out-of-sample, a weight estimator	ในการคัดเลือกรหัสของ nuclear families ของ 4 กลุ่มประชากร โดยเฉพาะ NY ที่มีหลายเชื้อสาย จึงทำให้มีการคัดออกของประชากร NY จำนวนมาก	ในการประมาณค่าแบบง่ายของข้อมูลทางพันธุกรรมจริงที่ overestimation เมื่อเปรียบเทียบกับ simulated 3 วิธีการยกตัวอย่างเช่น การประมาณค่า SNP Marker ใน Danaca (DA) และมี chromosome 5 (replicate 67) พบว่าการประมาณค่าแบบ naïve ของข้อมูลทางพันธุกรรมจริงเท่ากับ 0.42 แต่ simulated 3 วิธีได้ค่าผลทางพันธุกรรมต่ำกว่าและพบว่า การประมาณค่าแบบ weight ให้อคติน้อยที่สุด และการประมาณค่าแบบ shrinkage ให้อคติสูงสุดเมื่อเทียบกับ an out-of-sample, a weight estimator	จำลองข้อมูลแบบ three bootstrap-based resampling estimators

ตารางที่ 1 การทบทวนวรรณกรรมที่เกี่ยวข้องกับการประมาณค่าที่คำนึงถึงอคติจากการเลือกเป็นตัวอย่าง (ต่อ)

ชื่อเรื่อง	ปีที่ศึกษา	ระเบียบวิธีวิจัยและกลุ่มตัวอย่างโดยย่อ	การเกิดอคติจากการเลือกเป็นตัวอย่าง	ผลการวิเคราะห์ข้อมูลโดยย่อ	สถิติที่ใช้ในการประมาณค่าที่คำนึงถึงอคติจากการเลือกเป็นตัวอย่าง
Is the Prevalence of Overactive Bladder Overestimated? A Population-Based Study in Finland. (Tikkinen, Tammela, Rissanen, Valpas, & Huhtala, 2007)	2003 ถึง 2004	ทำการประมาณค่าความชุกของภาวะกระเพาะปัสสาวะไวต่อความรู้สึก (Overactive Bladder) และภาวะปัสสาวะรดที่นอน (nocturia) และทำการเปรียบเทียบภายหลังการหาอัตราการเกิดโรคจำเพาะตามกลุ่มอายุ ในกลุ่มคนที่มีอายุ 18-79 ปี ของประชากรที่ลงทะเบียนกับ Finnish Register Center โดยทำการส่งแบบสอบถามผ่านทางเมลล์ให้กับกลุ่มคนที่สนใจ 6000 รายที่ผ่านการสุ่ม ทำการเก็บข้อมูลใช้ระบบ Validate Danish Prostatic Symptom Score กับข้อมูลของ European standard population	การส่งแบบสอบถามที่ผ่านทางเมลล์มีผู้ตอบรับเพียง 62.4% จากผู้เข้าร่วมทั้งหมด	ภายหลังการหาอัตราการเกิดโรคจำเพาะตามกลุ่มอายุพบว่า ความชุกของภาวะ Overactive Bladder ต่ำกว่า 8% ของข้อมูล European standard population จากการศึกษาที่ไม่ได้รวมคนที่ไม่ตอบรับ สามารถบอกได้ว่า การศึกษาที่ไม่ได้รวมคนไม่ตอบรับ ปรากฏ overestimated ที่มีกฎเกณฑ์ที่ไม่ชัดเจนและการคัดเลือกประชากรศึกษาที่เกี่ยวข้องกับการแจกแจงตามอายุและการเข้าร่วมต่ำ	การเปรียบเทียบการประมาณค่าความชุกของการเกิดโรคของข้อมูลจากประชากร 2 กลุ่มได้แก่ กลุ่มทะเบียนกับ Finnish Register Center กับค่าความชุกของภาวะกระเพาะปัสสาวะไวต่อความรู้สึก ข้อมูลของ European standard population

ตารางที่ 1 การทบทวนวรรณกรรมที่เกี่ยวข้องกับการประมาณค่าที่คำนึงถึงอคติจากการเลือกเป็นตัวอย่าง (ต่อ)

ชื่อเรื่อง	ปีที่ศึกษา	ระเบียบวิธีวิจัยและกลุ่มตัวอย่าง	การเกิดอคติจากการเลือกเป็นตัวอย่าง	ผลการวิเคราะห์ข้อมูล	สถิติที่ใช้ในการประมาณค่าที่คำนึงถึงอคติจากการเลือกเป็นตัวอย่าง
Estimating the Proportion of True Null Hypotheses for Multiple Comparisons (Tiang, & Doerge, 2008)	2008	เป็นการศึกษาเพื่อตรวจสอบความสมบูรณ์ของ genome microarray โดยทำการประมาณค่าสัดส่วนของสมมติฐานหลัก ซึ่งการประมาณค่าจะมีความเที่ยง อคติเล็กน้อย และความแปรปรวนต่ำ โดยทำการเปรียบเทียบ 4 วิธีการได้แก่ Lowest slope estimate (LSL), Storey's bootstrap estimate (Storey _{boot}), Storey and Tibshiran's smoother estimate (ST _{smoother}), nonparametric maximum likelihood estimate (convert)	การตรวจสอบ genomic จะใช้ข้อมูลชีววิทยาก็ใน การทดสอบ สมมติฐาน ซึ่งเป็น ข้อมูลขนาดใหญ่ บางครั้งมีการ เปรียบเทียบหลาย แบบเช่น Family wise error rate (FWER)	เมื่อเปรียบเทียบกับข้อมูลจริง พบว่า วิธีการ LSL จะมีอคติมากที่สุด มีความแปรปรวนน้อยที่สุดแต่มีค่าสัดส่วนต่ำสุด และ overestimated จาก สมมติฐานหลักอีก 3 วิธีการมี สัดส่วน underestimated จาก สมมติฐานหลัก แต่วิธีการ Storey _{boot} อคติลดลงและความแปรปรวนขนาดเล็ก แต่วิธีการ ST _{smoother} อคติจะน้อยมากแต่ ความแปรปรวนจะกว้าง	การปรับค่าข้อมูลจาก 4 วิธีการ LSL, Storey _{boot} , ST _{smoother} , convert

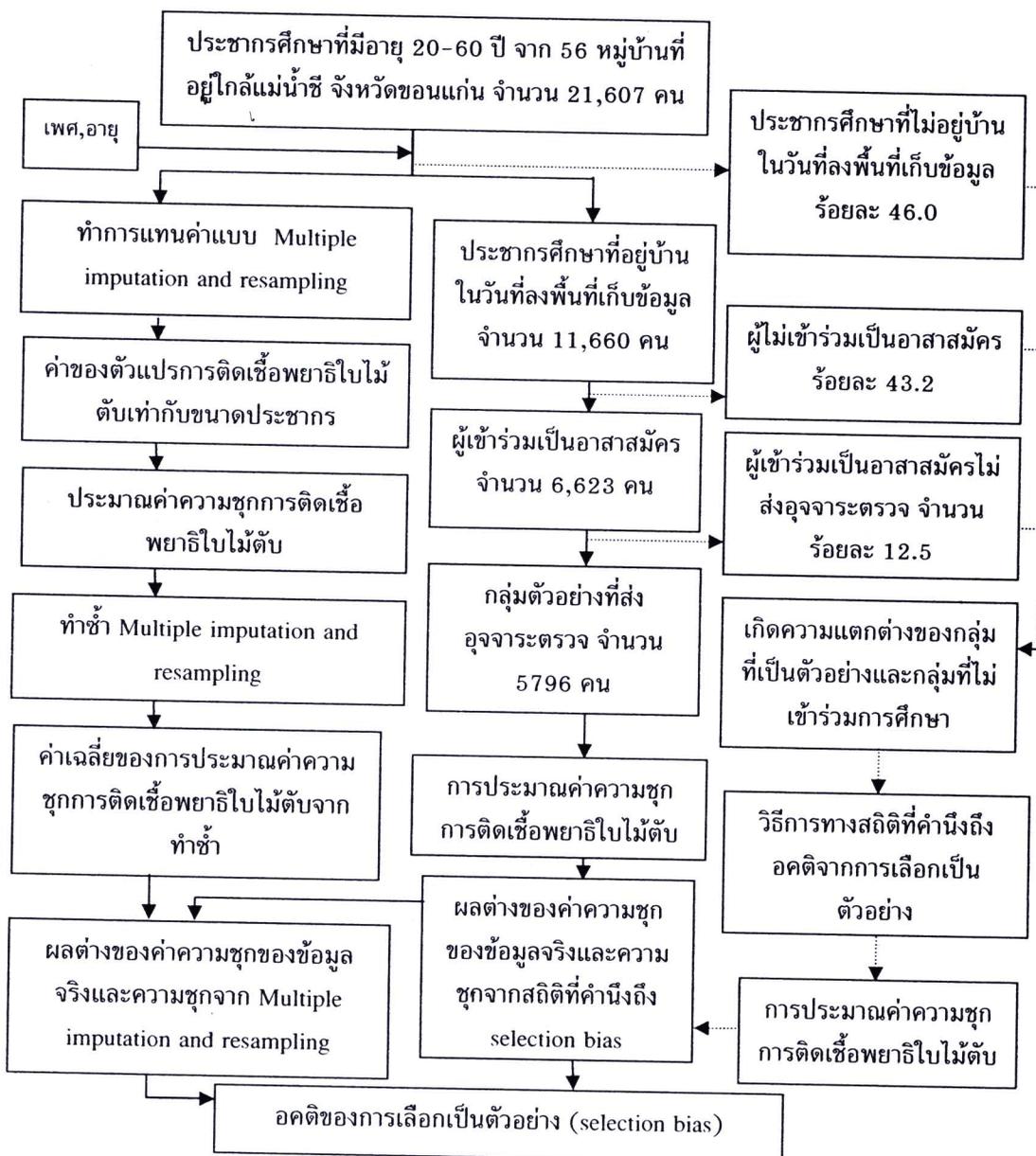
ตารางที่ 1 การทบทวนวรรณกรรมที่เกี่ยวข้องกับการประมาณค่าที่ค้ำึงถึงอดีตจากการเลือกเป็นตัวอย่าง (ต่อ)

ชื่อเรื่อง	ปีที่ศึกษา	ระเบียบวิธีวิจัยและกลุ่มตัวอย่าง	การเกิดอดีตจากการเลือกเป็นตัวอย่าง	ผลการวิเคราะห์ข้อมูลโดยย่อ	สถิติที่ใช้ในการประมาณค่าที่ค้ำึงถึงอดีตจากการเลือกเป็นตัวอย่าง
Adjustment for selection bias in observational studies with application to the analysis of autopsy data (Haneuse et al., 2009)	1992 จนถึง 2006	ทำการประมาณค่าความเสี่ยงที่จะเกิดภาวะ dementia และ Alzheimer's disease ในบุคคลที่มีอายุ 65 ปีขึ้นไปจากประชากรพื้นฐาน 23,000 ราย ที่เป็นสมาชิกของ Group Health Cooperative (GHC) ของกรุงวอชิงตัน ได้ทำการลงทะเบียนเข้าร่วมและทำการติดตามอย่างต่อเนื่อง	ประชากรที่เข้าร่วมการศึกษาจะได้รับ การทดสอบสมอง และเจาะเลือดตรวจ ทุก 2 ปี โดยมี ผู้เข้าร่วมการศึกษา ถอดตัวในระหว่าง การศึกษาและ เสียชีวิตในระหว่าง ช่วงการศึกษา	ในทุกการวิเคราะห์พบว่า การประมาณค่า ด้วย Bootstrap ให้ การประมาณค่า ดีกว่าการประมาณค่าแบบง่าย 20 ถึง 100 %	การประมาณค่าความเสี่ยงที่จะเกิดภาวะ dementia แบบ Bootstrap



7. กรอบแนวคิดการวิจัย

เนื่องจากในการศึกษาครั้งนี้ ต้องการประมาณค่าความชุกของการติดเชื้อพยาธิใบไม้ตับ ด้วยวิธีการทางสถิติที่คำนึงถึงอคติจากการเลือกเป็นตัวอย่าง เพื่อให้ได้การประมาณค่าความชุก การติดเชื้อพยาธิใบไม้ตับที่ถูกต้อง และหาผลต่างของระหว่างการประมาณค่าความชุกการติดเชื้อ พยาธิใบไม้ตับที่ไม่ได้คำนึงถึงอคติของการเลือกเป็นตัวอย่างกับการประมาณค่าความชุกของการ ติดเชื้อพยาธิใบไม้ตับที่คำนึงถึงอคติจากการเลือกเป็นตัวอย่าง เพื่อให้ทราบถึงผลกระทบจากอคติ จากการเลือกเป็นตัวอย่างต่อการประมาณค่าการติดเชื้อพยาธิใบไม้ตับ ดังรายละเอียดในภาพที่ 2



ภาพที่ 2 กรอบแนวคิดการวิจัย