



การจัดการข้อมูลและความสมบูรณ์ของข้อมูลทันตกรรม ในชุดข้อมูลมาตรฐาน 43 แฟ้ม เพื่อการศึกษาระยะยาว ติดตามการสูญเสียฟัน โดยใช้โปรแกรม R

ณัฐนิชา จิระชัยประสิทธิ์¹ จันทน์พิมพ์ หินเทา^{2,*} วรธนะ พิธพรชัยกุล³

บทความวิจัย

บทคัดย่อ

วัตถุประสงค์: 1) เพื่ออธิบายวิธีการจัดการข้อมูลและผลการจัดการข้อมูลนำเข้า 2) เพื่อวิเคราะห์ความสมบูรณ์ของข้อมูลทันตกรรมจากข้อมูล 43 แฟ้ม สำหรับนำไปวิเคราะห์การรอดชีพจากการสูญเสียฟันหลังจากการตรวจฟัน

วัสดุอุปกรณ์และวิธีการ: กลุ่มตัวอย่างเป็นผู้ใช้สิทธิหลักประกันสุขภาพถ้วนหน้าที่มาตรวจฟัน ณ หน่วยบริการสังกัดกระทรวงสาธารณสุข ของจังหวัดหนึ่งภาคใต้ ปี 2562-2564 ข้อมูลการตรวจฟัน และข้อมูลการรับบริการทันตกรรม ได้จากแฟ้มสภาวะทันตสุขภาพ และ แฟ้มหัตถการผู้ป่วยนอก เชื่อมโยงข้อมูลด้วยแฟ้มข้อมูลทั่วไป การจัดการข้อมูล และวิเคราะห์สถิติเชิงพรรณนาด้วยโปรแกรม R

ผล: พบว่ามีวิธีการจัดการข้อมูล ดังนี้ 1) นำเข้าข้อมูล 2) จัดการค่าว่าง 3) กรองกลุ่มตัวอย่างตามเกณฑ์คัดเข้าและคัดออก 4) กรองคุณภาพข้อมูล โดยตรวจสอบความถูกต้องของข้อมูล 5) ตัดข้อมูลซ้ำซ้อนโดยใช้เขตข้อมูลที่ระบุความไม่ซ้ำของข้อมูล และ 6) เชื่อมโยงข้อมูลแต่ละแฟ้มด้วยเลขบัตรประจำตัวประชาชนที่ผ่านการแสะซึ่ง ความสมบูรณ์ของข้อมูลของแฟ้มสภาวะทันตสุขภาพ มีข้อมูลสมบูรณ์ของผู้ที่ได้รับการตรวจฟันจำนวน 6,820 ระเบียบ (ร้อยละ 20.33) หลังจากจัดการข้อมูลจากข้อมูลนำเข้าของแฟ้มสภาวะทันตสุขภาพ 33,553 ระเบียบ และจากข้อมูลนำเข้าของแฟ้มหัตถการผู้ป่วยนอก 4,051,350 ระเบียบ มีข้อมูลสมบูรณ์ของผู้ที่เคยได้รับการตรวจฟันแล้วมีข้อมูลการถอนฟันหรือรับบริการทันตกรรมอื่นจำนวน 4,177 ระเบียบ (ร้อยละ 0.10) ซึ่งเป็นจำนวนกลุ่มตัวอย่างที่มีข้อมูลเพื่อนำไปวิเคราะห์การรอดชีพจากการสูญเสียฟันหลังจากการตรวจฟันได้

บทสรุป: สรุปหลังจัดการและเชื่อมโยงข้อมูลทุกแฟ้ม จำนวนกลุ่มตัวอย่างที่มีข้อมูลทันตกรรมสมบูรณ์มีเพียงพอที่สามารถใช้เพื่อการศึกษาระยะยาว ติดตามการสูญเสียฟันได้ ร้อยละ 24.56 โปรแกรม R เป็นเครื่องมือที่มีประสิทธิภาพในการจัดการระเบียบอิเล็กทรอนิกส์ทางทันตกรรมเพื่อให้ได้ชุดข้อมูลที่มีความสมบูรณ์สำหรับการศึกษารติดตามการสูญเสียฟันได้

คำไรท์ : เวชระเบียนอิเล็กทรอนิกส์/ การจัดการข้อมูล/ โปรแกรม/ ความสมบูรณ์ข้อมูล

Received: Jul 05, 2024

Revised: Oct 08, 2024

Accepted: Oct 22, 2024

บทนำ

การนำข้อมูลอิเล็กทรอนิกส์ของผู้ป่วยมาใช้ในการวิจัยทางคลินิกด้านทันตกรรมมีมากขึ้น เนื่องจากมีจำนวนกลุ่มตัวอย่างมาก ทำให้ผลวิจัยสามารถสรุปอ้างอิงไปยังประชากรได้ดี แต่เวชระเบียนอิเล็กทรอนิกส์ยังมีข้อจำกัด คือ คุณภาพข้อมูลเนื่องจากการไม่กรอกข้อมูลทำให้เป็นค่าว่าง รวมถึงการกรอก

ข้อมูลโดยบุคลากรหลายคน หลายหน่วยบริการ ส่งผลต่อความถูกต้องของข้อมูล¹

เวชระเบียนอิเล็กทรอนิกส์ที่สำคัญของประเทศไทยคือ ชุดข้อมูลมาตรฐาน 43 แฟ้ม² ส่วนใหญ่ใช้เพื่อติดตามตัวชี้วัด³ มีการนำมาใช้ศึกษาปัจจัยที่มีความสัมพันธ์กับทารกแรกเกิดน้ำหนักน้อย⁴ และการศึกษาความครบถ้วนและความ

¹ หลักสูตรการฝึกอบรมทันตแพทย์ประจำบ้าน สาขาทันตสาธารณสุข คณะทันตแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์

² หน่วยวิจัยเพื่อการพัฒนาการดูแลสุขภาพช่องปาก สาขาวิชาทันตกรรมป้องกัน คณะทันตแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์

³ หน่วยวิจัยทันตแพทยศาสตร์เชิงประจักษ์เพื่อการดูแลและส่งเสริมสุขภาพช่องปาก สาขาวิชาทันตกรรมป้องกัน คณะทันตแพทยศาสตร์

มหาวิทยาลัยสงขลานครินทร์

* ผู้ประพันธ์บทความ

ถูกต้องของข้อมูลเพื่อใช้เป็นแหล่งข้อมูลสำหรับตัวชี้วัดการดูแลสุขภาพแบบมุ่งเน้นคุณค่าในกลุ่มโรคเบาหวาน⁵ การศึกษาความชุกโรคฟันผุและโรคปริทันต์⁶ การศึกษาติดตามระยะยาวเพื่อศึกษาความสัมพันธ์ของสภาวะฟันผุ การเจริญเติบโต และพัฒนาการในเด็ก 0-5 ปีในบริบทของคลินิกเด็ก⁷

แฟ้มที่ เกี่ยวข้องกับทางทันตกรรมในชุดข้อมูลมาตรฐาน 43 แฟ้ม ได้แก่ แฟ้มสภาวะทันตสุขภาพ (Dental), แฟ้มหัตถการผู้ป่วยนอก (Procedure_OPD), แฟ้มตรวจวินิจฉัยผู้ป่วยนอก (Diagnosis_OPD) และ แฟ้มข้อมูลทั่วไป (Person) การศึกษาที่ผ่านมาพบว่าแฟ้มสภาวะทันตสุขภาพ มีความสมบูรณ์ มากกว่าร้อยละ 90 ในเกือบทุกรายการยกเว้นความจำเป็นต้องใส่ฟันเทียม และมีความแม่นยำมากกว่าร้อยละ 90 และมีความตรง มากกว่าร้อยละ 90 ในเกือบทุกรายการยกเว้นจำนวนฟันและประเภทของฟันเทียมในกลุ่มผู้สูงอายุ⁸ แต่อย่างไรก็ตาม มีการศึกษาคุณภาพข้อมูลของแฟ้มสภาวะทันตสุขภาพ แฟ้มตรวจวินิจฉัยผู้ป่วยนอก และ แฟ้มหัตถการผู้ป่วยนอก จากชุดข้อมูลมาตรฐาน 43 แฟ้ม ปีงบประมาณ 2565 ที่นำเข้ารระบบคลังข้อมูลสุขภาพระดับจังหวัด พบว่าข้อมูลมีความแตกต่างกันในแต่ละโรงพยาบาล ระเบียบข้อมูลทั้งสามแฟ้มส่วนใหญ่มีความซ้ำซ้อนมากกว่าร้อยละ 60 มีความสมบูรณ์ของแฟ้มตรวจวินิจฉัยผู้ป่วยนอก และ แฟ้มหัตถการผู้ป่วยนอก ร้อยละ 88.2-99.9 แฟ้มสภาวะทันตสุขภาพ ร้อยละ 0.3-98.3 และความถูกต้องของแฟ้มตรวจวินิจฉัยผู้ป่วยนอก และแฟ้มหัตถการผู้ป่วยนอก ร้อยละ 96.3-100 ความถูกต้องของแฟ้มสภาวะทันตสุขภาพ ร้อยละ 84.3-100⁹ ดังนั้นหากจะนำข้อมูลมาใช้ในการศึกษาปัญหาทันตสาธารณสุขด้วยรูปแบบการศึกษาที่หลากหลาย ควรมีการจัดการข้อมูลทางทันตกรรมที่เหมาะสม

เครื่องมือการจัดการและวิเคราะห์ข้อมูลในปัจจุบันมีหลากหลาย แต่โปรแกรม R เป็นโอเพนซอร์ส (Open source) ที่สามารถติดตั้งใช้งานได้อย่างไม่มีข้อจำกัด ไม่มีค่าใช้จ่าย มีชุดคำสั่ง (Package) จึงสามารถใช้คัดกรองข้อมูลที่เป็นไปได้ และข้อมูลที่ซ้ำซ้อนออก ใช้เชื่อมโยง กรองข้อมูล และวิเคราะห์ข้อมูลที่ต้องการศึกษาได้ แต่นักวิจัยต้องใช้เวลาในการเรียนรู้การใช้งานด้วยการเขียนคำสั่ง (Coding)¹⁰ สถาบันวิจัยและพัฒนาสุขภาพภาคใต้ ได้ผลิตหนังสือ “การวิเคราะห์ข้อมูลสุขภาพในระบบสาธารณสุขไทยโดยใช้โปรแกรม R” โดยใช้ข้อมูลจากฐานข้อมูลสุขภาพ 43 แฟ้ม ประกอบด้วยตัวอย่างในการวิเคราะห์ข้อมูล¹¹ แต่ยังคงขาดการวิเคราะห์ข้อมูลในระยะยาว เช่น การวิเคราะห์การรอดชีพ (Survival analysis)

การศึกษานี้จึงมีวัตถุประสงค์ 1) เพื่ออธิบายวิธีการจัดการข้อมูลและผลการจัดการข้อมูลนำเข้า 2) เพื่อวิเคราะห์ความสมบูรณ์ของข้อมูลต้นตอจากข้อมูล 43 แฟ้ม สำหรับนำไปวิเคราะห์การรอดชีพจากการสูญเสียฟัน เป็นการติดตามระยะยาวของการรับบริการทันตกรรมหลังจากการตรวจฟัน โดยใช้กรณีศึกษาในกลุ่มวัยก่อนสูงอายุ คือ ช่วงอายุ 45-59 ปี ซึ่งการศึกษาระยะยาวในประเทศไทยของกลุ่มอายุนี้นี้ยังมีน้อย¹²

วัสดุอุปกรณ์และวิธีการ

การศึกษาแบบย้อนหลัง (Retrospective cohort study) ใช้ข้อมูลทุติยภูมิจากชุดข้อมูลมาตรฐาน 43 แฟ้ม โดยขออนุญาตใช้ข้อมูลโดยส่งหนังสือขอความอนุเคราะห์ไปยังสำนักงานหลักประกันสุขภาพแห่งชาติ (สปสช.) เขต 12 สงขลา ระบุข้อมูลเฉพาะส่วนที่เกี่ยวข้องกับงานวิจัยและระบุวันสิ้นสุดการใช้ข้อมูล ประกอบด้วยแฟ้มข้อมูลทั่วไป แฟ้มสภาวะทันตสุขภาพ และ แฟ้มหัตถการผู้ป่วยนอก โดยใช้เลขที่บัตรประชาชนที่ผ่านการแฮชชิ่ง (Hashing CID) เพื่อแปลงข้อมูลที่ระบุตัวตน เพื่อเชื่อมโยงข้อมูลในแต่ละแฟ้ม โครงสร้างวิจัยผ่านการพิจารณาจากคณะกรรมการจริยธรรมการวิจัยในมนุษย์ คณะทันตแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์ รหัสโครงการ EC6607-038

กลุ่มตัวอย่าง คือ กลุ่มก่อนวัยผู้สูงอายุ 45-59 ปี จังหวัดหนึ่งในเขตสุขภาพ 12 ที่มีผลตรวจฟันในแฟ้มสภาวะทันตสุขภาพ มีเกณฑ์คัดเข้าดังนี้ 1) เป็นคนไทย 2) ใช้สิทธิหลักประกันสุขภาพถ้วนหน้า 3) รับบริการ ณ หน่วยบริการสังกัดกระทรวงสาธารณสุข 4) ได้รับการตรวจฟันโดยทันตบุคลากร ในปีพ.ศ. 2562-2564 ไม่รวมการตรวจโดยผู้ให้บริการประเภทอื่นเพื่อเป็นการควบคุมคุณภาพการเก็บข้อมูล 5) มีอายุ 45-59 ปี 11 เดือน 29 วัน ณ วันตรวจฟัน และ 6) มารับบริการทันตกรรมอย่างใดอย่างหนึ่งภายหลังการตรวจฟัน ในปีพ.ศ. 2562-2566 เกณฑ์การคัดออก คือ ผู้ที่สูญเสียฟันทั้งปาก ณ วันตรวจฟัน

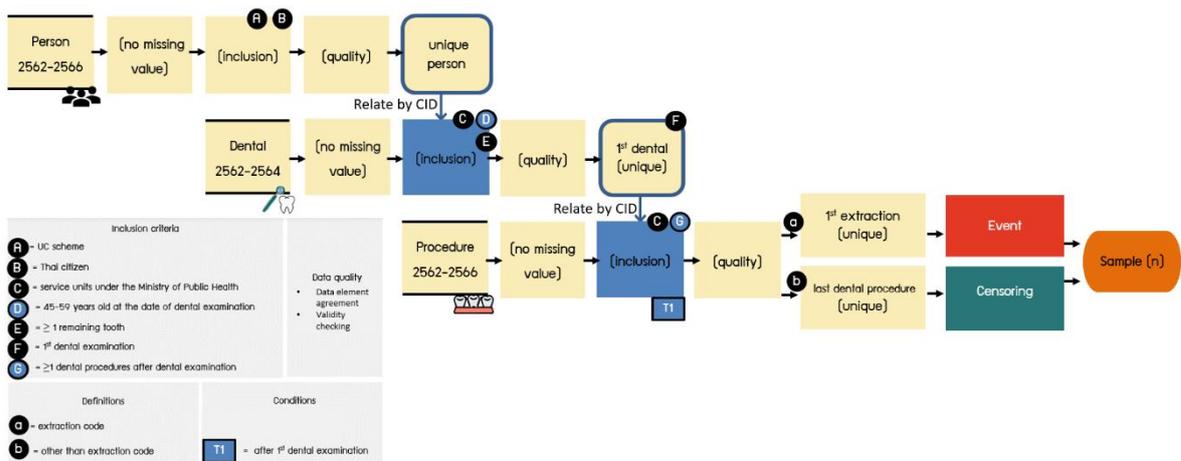
ตัวแปรที่ศึกษามีดังนี้ ตัวแปรต้น ได้แก่ 1) เพศ 2) กลุ่มอายุ แบ่งเป็น 45-49 ปี 50-54 ปี และ 55-59 ปี 3) จำนวนฟันแท้ที่มีในช่องปาก 4) จำนวนฟันแท้ผุ 5) ความจำเป็นต้องอุดหินน้ำลาย ตัวแปร 1) -2) ได้จากแฟ้มข้อมูลทั่วไป 3) -5) ได้จากแฟ้มสภาวะทันตสุขภาพ เขตข้อมูลจำนวนฟันแท้ที่มีอยู่จำนวนฟันแท้ผุที่ไม่ได้อุด และรหัสจำเป็นต้องอุดหินน้ำลาย ณ วันที่ได้รับการตรวจฟันครั้งแรก 6) ได้รับการอุดหินน้ำลาย

ได้จากแฟ้มหัตถการผู้ป่วยนอก เขตข้อมูลรหัสหัตถการและการรักษา มีค่าใดค่าหนึ่ง ดังนี้ 2277310, 2277320, 2287310, 2287320, 2367310 ตัวแปรตาม ได้แก่ 1) สถานะการสูญเสียฟัน แบ่งเป็น 1.1) การสูญเสียฟัน (Event) คือ การถอนฟันแท้หลังจากการตรวจฟันครั้งแรก ในแฟ้มหัตถการผู้ป่วยนอก เขตข้อมูลรหัสหัตถการและการรักษา มีค่าใดค่าหนึ่ง ดังนี้ 2381402, 2382600, 2392700 ซึ่งไม่รวมการผ่าฟันคุด และ 1.2) การไม่สูญเสียฟัน (Censoring) คือ การรับบริการทันตกรรมอย่างใดอย่างหนึ่งที่ไม่ใช่การถอนฟันภายหลังการตรวจฟัน ซึ่งมี 2 ลักษณะ คือรับบริการก่อนหรือตั้งแต่วันสิ้นสุดการติดตาม 2) ระยะเวลาการสูญเสียฟัน คือ ระยะเวลาเป็นเดือนนับตั้งแต่การตรวจฟันครั้งแรกในช่วงเวลาที่เข้าจนถึงเกิดการสูญเสียฟันภายในวันสิ้นสุดการศึกษา

กรอบการศึกษา ได้แก่ ช่วงเวลาตัดเข้ากลุ่มตัวอย่าง คือ 1 มกราคม 2562 - 31 ธันวาคม 2564 วันที่เริ่มการติดตามสถานะการสูญเสียฟัน คือ วันที่กลุ่มตัวอย่างได้รับการตรวจฟันครั้งแรกในช่วงตัดเข้า ถ้ามีการปรับปรุงข้อมูลใช้เขตข้อมูลวันเดือนปีที่ปรับปรุงล่าสุดของแฟ้มสภาวะทันตสุขภาพ วันสิ้นสุดการติดตามสถานะการสูญเสียฟัน คือ 30 มิถุนายน 2566 การจัดการข้อมูลใช้โปรแกรม R เวอร์ชัน 4.3.1 ร่วมกับ RStudio 2024.04.2 Build 764 เพื่อเตรียมชุดข้อมูลให้พร้อมสำหรับการวิเคราะห์การศึกษาระยะยาว คือ ระยะเวลาก่อนการสูญเสียฟันด้วยเส้นโค้งแคปแพลน-ไมเออร์ (Kaplan-

Meier curve) และศึกษาปัจจัยที่เกี่ยวข้องกับระยะเวลาก่อนการสูญเสียฟันด้วยการวิเคราะห์การถดถอย Cox (Cox regression analysis) โดยมีการจัดการข้อมูลรายแฟ้ม ดังนี้ 1) นำเข้าข้อมูล 2) จัดการค่าว่าง (Missing value) ถ้าหากตัวแปรมีค่าว่าง ไม่เกินร้อยละ 40 สามารถใช้การทดแทน (Imputation) ได้¹³ หากมีค่าว่างมากกว่านั้น พิจารณาตัดกลุ่มตัวอย่างที่ไม่มีข้อมูลตัวแปรตัวใดตัวหนึ่งที่กำหนดไว้ 3) กรองข้อมูลตามเกณฑ์คัดเข้าและคัดออก 4) กรองคุณภาพข้อมูล โดยตรวจสอบความถูกต้องของข้อมูล ทั้งความสอดคล้องของข้อมูลแต่ละเขตข้อมูล (Data element agreement) และ การตรวจสอบความเป็นไปได้ของข้อมูล (Validity checking)¹⁴ 5) ตัดข้อมูลซ้ำซ้อน (Duplicated) โดยใช้กุญแจหลัก (Primary key) ซึ่งเป็นเขตข้อมูลที่ระบุความไม่ซ้ำของข้อมูล และ 6) เชื่อมโยงข้อมูลแต่ละแฟ้ม ซึ่งขั้นตอนนี้มีลำดับที่แตกต่างกันในการจัดการแต่ละประเภท (รูปที่ 1) เพื่อควบคุมคุณภาพของวิธีการจัดการข้อมูลด้วย R มีการตรวจสอบคำสั่งพร้อมกับผู้เชี่ยวชาญ (Pair programming)¹⁵

การวิเคราะห์ผลด้วยสถิติเชิงพรรณนา คือ จำนวนและร้อยละของข้อมูลที่ถูกคัดออกในแต่ละขั้นตอนของการจัดการข้อมูล และความสมบูรณ์ของข้อมูลแต่ละแฟ้มที่ผ่านการจัดการข้อมูล พร้อมนำไปวิเคราะห์ต่อไป โดยความสมบูรณ์ของข้อมูลคือ ข้อมูลทุกเขตข้อมูล มีความถูกต้อง และความสมบูรณ์ในแนวแถว¹⁴



รูปที่ 1 การจัดการข้อมูล 6 ขั้นตอนเพื่อให้ได้ตัวแปรที่กำหนด
Figure 1 Data Management 6 Steps to Obtain Variables

ผล

การจัดการข้อมูลเริ่มจากเพิ่มข้อมูลทั่วไป เพิ่มสภาวะทันตสุขภาพ และเพิ่มหัตถการผู้ป่วยนอก ตามลำดับ ดังรูปที่ 1

1. วิธีจัดการเพิ่มข้อมูลทั่วไป 1.1) นำเข้าข้อมูลจำนวน 3,652,490 ระเบียบ 1.2) จัดการค่าว่าง พบว่ากุญแจหลัก (Primary key) และตัวแปรสำคัญไม่มีค่าว่างเลย 1.3) กรองข้อมูลตามเกณฑ์คัดเข้าและคัดออก คือ ใช้สิทธิ์หลักประกันสุขภาพถ้วนหน้า และ สัญชาติไทย พบว่าไม่มีข้อมูลที่ไมผ่านเกณฑ์ 1.4) กรองคุณภาพข้อมูล ตรวจสอบความเป็นไปได้ข้อมูล พบในเขตข้อมูลเพศ มีค่าทั้ง 1 และ 2 จำนวน 3 ระเบียบ ไม่พบข้อมูลในเขตข้อมูลวันเกิดที่ไม่ผ่านเกณฑ์ ตรวจสอบความสอดคล้องข้อมูล พบว่ามีการปรับปรุงข้อมูลเพศและอายุมากกว่า 2 ครั้ง จำนวน 569 ระเบียบ ซึ่งไม่ทราบว่าการปรับปรุงครั้งใด คือ ค่าที่ถูกต้อง จึงพิจารณาตัดข้อมูลออก 1.5) ตัดข้อมูลซ้ำซ้อน การศึกษานี้ไม่คำนึงถึงการปรับปรุงข้อมูลล่าสุด เพราะแต่ละระเบียบที่มีหลายเขตและหลายวันเกิดออกในขั้นตอนกรรณคุณภาพข้อมูลแล้ว พบว่ามีข้อมูลซ้ำจากข้อมูลบุคคลสะสมในทุกปีจำนวน 3,241,498 ระเบียบ 1.6) การเชื่อมโยงข้อมูล ทำหลังจากจัดการค่าว่างในเพิ่มสภาวะทันตสุขภาพแล้ว

เพิ่มข้อมูลทั่วไปนี้เป็นเพิ่มลำดับแรกที่ต้องจัดการเพื่อเป็นข้อมูลเชื่อมโยงกับเพิ่มสภาวะทันตสุขภาพ เพื่อใช้ในการคำนวณอายุ ณ วันตรวจฟัน

2. วิธีจัดการเพิ่มสภาวะทันตสุขภาพ 2.1) นำเข้าข้อมูลจากเพิ่มสภาวะทันตสุขภาพ จำนวน 33,553 ระเบียบ 2.2) จัดการค่าว่าง พบค่าว่างเป็นร้อยละ 51.2 จึงไม่สามารถใช้การทดแทนได้¹³ พบจำนวนค่าว่างมากที่สุดในจำนวนฟันผุ รองลงมาความจำเป็นต้องซุดหินน้ำลาย และจำนวนฟัน คือ 17,126 16,844 และ 16,666 ระเบียบ ค่าว่างคิดเป็นร้อยละ 51.04 50.20 และ 49.67 ตามลำดับ 2.3) เชื่อมโยงเพิ่มข้อมูลทั่วไป และ เพิ่มสภาวะทันตสุขภาพ ด้วยเลขที่บัตรประชาชน เพื่อนำข้อมูลวันเกิด (เขตข้อมูลวันเกิด) มาใช้คำนวณอายุ ณ วันตรวจฟัน และกรองข้อมูลสภาวะทันตสุขภาพ ตามเกณฑ์คัดเข้า 2.4) กรองข้อมูลตามเกณฑ์คัดเข้าและคัดออก ดังนี้ รับประทานที่หน่วยบริการนอกสังกัดกระทรวงสาธารณสุข จำนวน 357 ระเบียบ ผลตรวจฟันที่ไม่มีข้อมูลเพศ หรือวันเกิดในเพิ่มข้อมูลทั่วไป จำนวน 9 ระเบียบ จำนวนฟันน้อยกว่า 1 ซี่ 7,243 ระเบียบ ผลตรวจฟันทั้งหมดตรวจโดยทันตบุคลากร พบผู้ที่ไม่อยู่ในช่วงอายุ 45-59 ปี ณ วันที่ตรวจฟัน จำนวน 1,943 ระเบียบ 2.5)

กรองคุณภาพข้อมูล ด้วยการตรวจสอบความเป็นไปได้ข้อมูล พบจำนวนฟันอยู่นอกช่วง 1-32 ซี่ 1 ระเบียบ ไม่มีระเบียบที่จำนวนฟันอยู่นอกช่วง 0-32 ซี่ ไม่มีค่าข้อมูลในเขตข้อมูลความจำเป็นต้องซุดหินน้ำลาย นอกเหนือจากค่า 1 หรือ 2 ตรวจสอบความสอดคล้องข้อมูลพบจำนวนฟันเกินจำนวนฟันในช่องปาก 1 ระเบียบ พบว่ามีข้อมูลตรวจทันตสุขภาพทุกครั้ง 6,820 ระเบียบ 2.6) ตัดข้อมูลซ้ำซ้อน พบการตรวจฟันมีวันที่ปรับปรุงข้อมูลวันเดียว และไม่ใช้การตรวจฟันครั้งแรก 1,060 ระเบียบ

ทั้งนี้เพิ่มทันตสุขภาพต้องจัดการเป็นลำดับถัดมา เพราะการตรวจฟันเป็นจุดเริ่มต้นการติดตาม

3. วิธีจัดการเพิ่มหัตถการผู้ป่วยนอก 3.1) นำเข้าข้อมูลจากเพิ่มหัตถการผู้ป่วยนอก จำนวน 4,051,350 ระเบียบ 3.2) จัดการค่าว่าง โดยเพิ่มหัตถการผู้ป่วยนอก ไม่พบตัวแปรสำคัญมีค่าว่าง 3.3) เชื่อมโยงข้อมูล ระหว่างเพิ่มสภาวะทันตสุขภาพที่ผ่านการจัดการข้อมูลขั้นตอนที่ 2 เชื่อมโยงกับเพิ่มหัตถการผู้ป่วยนอก ด้วยเลขที่บัตรประชาชน 3.4) กรองข้อมูลตามเกณฑ์คัดเข้าและคัดออก ดังนี้ พบการรับบริการที่หน่วยบริการนอกสังกัดกระทรวงสาธารณสุข 21,929 ระเบียบ ไม่มีผลการตรวจฟันครั้งแรก 3,830,735 ระเบียบ ไม่ใช้หัตถการทันตกรรม 146,896 ระเบียบ วันที่มารับบริการทันตกรรมไม่เกินวันตรวจฟันครั้งแรก 21,105 ระเบียบ 3.5) กรองคุณภาพข้อมูล โดยการตรวจสอบความเป็นไปได้ข้อมูล ไม่พบวันที่บริการอยู่นอกช่วงปีพ.ศ. 2562-2566 และ 3.6) ตัดข้อมูลซ้ำซ้อนจากระเบียบในขั้นตอน 3.5) แยกจัดการเป็น 2 กลุ่ม กลุ่มสูญเสียฟันพบรหัสหัตถการและการรักษาที่ไม่ใช่การถอนฟัน 26,968 ระเบียบ พบการถอนฟันที่ไม่ใช่วันที่ปรับปรุงข้อมูลล่าสุด 60 ระเบียบ พบว่าไม่ใช่การถอนฟันครั้งแรก 1,928 ระเบียบ ในส่วนกลุ่มไม่สูญเสียฟันเริ่มจากระเบียบจากขั้นตอน 3.5) เช่นกัน ตัดข้อมูลจากกลุ่มคนสูญเสียฟัน 18,427 ระเบียบ พบบริการทันตกรรมอื่น ๆ ซึ่งไม่ใช่วันที่ปรับปรุงข้อมูลล่าสุด 5,228 ระเบียบ บริการทันตกรรมอื่น ๆ ที่ไม่ใช่ครั้งสุดท้าย 4,376 ระเบียบ

เพิ่มหัตถการผู้ป่วยนอกใช้เพื่อติดตามวันที่รับบริการทันตกรรมหลังตรวจฟัน การจัดการข้อมูลในขั้นตอนนี้จะให้ข้อมูลวันสิ้นสุดการติดตามของกลุ่มตัวอย่าง

ในขั้นตอนการจัดการข้อมูล ร้อยละข้อมูลที่ต้องคัดออกแยกแยะเพิ่ม ได้แก่ 1) เพิ่มข้อมูลทั่วไป คัดออกร้อยละ 88.76 ซึ่งเกือบทั้งหมดคัดออกเพราะข้อมูลซ้ำจากข้อมูลสะสมแต่ละปี 2) เพิ่มสภาวะทันตสุขภาพคัดออกร้อยละ 82.83

ส่วนใหญ่เพราะเป็นค่าว่างในตัวแปรจำนวนฟัน จำนวนฟันผุ ความจำเป็นต้องขูดหินน้ำลาย คิดเป็นร้อยละ 51.20 และ 3) แพ้หมดการผู้ป่วยนอก คัดออกร้อยละ 99.89 ส่วนใหญ่เพราะไม่มีผลตรวจฟันในแฟ้มสภาวะทันตสุขภาพ คิดเป็นร้อยละ 94.55 (ตารางที่ 1)

4. แฟ้มสภาวะทันตสุขภาพมีข้อมูลในขั้นตอนนำเข้า 33,553 ระเบียบ คิดเป็น 17,009 คน ก่อนเชื่อมโยงข้อมูล มีค่าว่างร้อยละ 51.20 จึงได้ข้อมูลที่สมบูรณ์ร้อยละ 48.8 หลังจากจัดการข้อมูลแฟ้มสภาวะทันตสุขภาพ พบข้อมูลสมบูรณ์ของการตรวจฟันทุกครั้ง จำนวน 6,820 ระเบียบ คิดเป็นร้อยละ 20.33 และหลังจากเชื่อมโยงข้อมูลทุกแฟ้ม พบข้อมูลที่สมบูรณ์

และพร้อมนำไปวิเคราะห์ต่อในการศึกษาการวิเคราะห์การ รอดชีพ จำนวน 4,177 คน คิดเป็นร้อยละ 24.56 แบ่งเป็นกลุ่ม ที่สูญเสียฟัน 1,582 คน และกลุ่มที่ไม่สูญเสียฟัน 2,595 คน มี ร้อยละของระเบียบสมบูรณ์จำแนกเป็นโรงพยาบาลชุมชน โรงพยาบาลส่งเสริมสุขภาพตำบล โรงพยาบาลศูนย์ และหน่วย บริการประเภทอื่น ๆ คือ 60.5, 30.1, 3.5 และ 5.9 ตามลำดับ ทั้งนี้ไม่มีข้อมูลจากโรงพยาบาลทั่วไปเข้าในการศึกษาเลย เนื่องจากข้อมูลในเขตข้อมูลรหัสประเภทบุคลากรในแฟ้ม สภาวะทันตสุขภาพปี.ศ. 2562-2564 เป็นค่าว่าง จึงระบุไม่ได้ ว่าเป็นการตรวจฟันโดยทันตบุคลากร (ตารางที่ 2)

ตารางที่ 1 ความถี่และร้อยละระเบียบที่คัดออกและที่ผ่านการจัดการข้อมูล

Table 1 Frequency and Percentage of Records Excluded and Processed

Data management	File/ n (%)		
	Person	Dental	Procedure_OPD
Importing data	3,652,490 (100.00)	33,553 (100.00)	4,051,350 (100.00)
Missing value	0 (0.00)	17,179 (51.20)	0 (0.00)
Data did not meet the inclusion criteria	0 (0.00)	9,552 (28.47)	4,020,871 (99.25)
Inaccurate data	572 (0.02)	2 (0.00)	0 (0.00)
Duplicated data	3,241,498 (88.75)	1,060 (3.16)	Event 2,135 (0.05) Censoring 24,167 (0.60)
Processed data	410,420 (11.24)	5,760 (17.17)	Event 1,582 (0.04) Censoring 2,595 (0.06)

ตารางที่ 2 ความถี่ (ร้อยละ) ระเบียบ คน และแห่งหน่วยบริการจากแฟ้มสภาวะทันตสุขภาพ จำแนกตามประเภทหน่วยบริการและขั้นตอนการจัดการข้อมูล

Table 2 Frequency (Percentage) of Records, People, and Units from Dental Files, Classified by Type of Service Unit and Data Management

Type of service unit	n (%)								
	Importing data			After processing the Dental file data			After processing and linking all the data		
	Records	People	Units	Records	People	Units	Records	People	Units
Regional hospital	17,477 (52.1)	5,944 (34.1)	1 (1.1)	307 (4.5)	227 (3.9)	1 (1.5)	147 (3.5)	147 (3.5)	1 (1.6)
Community hospital	11,824 (35.2)	7,517 (43.2)	14 (15.6)	4,170 (61.1)	3,425 (59.5)	12 (18.5)	2,527 (60.5)	2,527 (60.5)	12 (19.4)
Primary care unit	3,361 (10.1)	3,123 (17.9)	62 (68.9)	2,004 (29.4)	1,813 (31.5)	48 (73.8)	1,258 (30.1)	1,258 (30.1)	44 (72.6)
Others	534 (1.6)	515 (3.0)	8 (8.9)	339 (5.0)	295 (5.1)	4 (6.2)	245 (5.9)	245 (5.9)	4 (6.5)
Department of Local Administration	357 (1.1)	315 (1.8)	5 (5.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Total	33,553	17,414	90	6,820	5,760	65	4,177	4,177	61

บทวิจารณ์

หลักการจัดการข้อมูล โดยทั่วไปเริ่มจากขั้นตอนการดึงข้อมูลจากเวรเซียบีเยนอีเล็กทรอนิกส์ การจัดการข้อมูลซ้ำ การกรองค่านอกเกณฑ์ (Outliers) ออก การกรองข้อมูลที่สนใจศึกษา การจัดการข้อมูลว่าง การตรวจสอบคุณภาพข้อมูล การเชื่อมโยงข้อมูล การปรับโครงสร้างสำหรับข้อมูลระยะยาวเป็น Wide format หรือ Long format และการแก้ไขข้อมูลให้ถูกต้องตามโครงสร้างมาตรฐานข้อมูล^{16,17} ในการศึกษานี้ได้กล่าวครอบคลุมทุกขั้นตอน ยกเว้น การปรับโครงสร้างข้อมูลระยะยาว และการแก้ไขข้อมูลตามโครงสร้างมาตรฐาน เพราะในกรณีศึกษานี้เลือกข้อมูลตัวแปรจากจุดเวลาที่สนใจ เช่น การตรวจฟันครั้งแรก การบริการทันตกรรมอื่น ๆ ครั้งสุดท้าย เป็นต้น ในการศึกษาที่ใช้ข้อมูลภายใต้โครงสร้างมาตรฐาน 43 แฟ้มและได้จาก สปสช. ซึ่งผ่านการจัดการความซ้ำซ้อน ความถูกต้องเบื้องต้น โดยเฉพาะในแฟ้มหัตถการผู้ป่วยนอกและกรองข้อมูลเฉพาะสิทธิหลักประกันสุขภาพถ้วนหน้า ผลตรวจฟันโดยทันตบุคลากรเท่านั้น ทุกคนมีเลขที่บัตรประชาชนที่ผ่านการแปลงข้อมูล แต่อย่างไรก็ตามยังพบว่ามีข้อมูลซ้ำในแฟ้มสถานะทันตสุขภาพจำนวนมากเพราะ กลุ่มตัวอย่างคนเดียวกันได้รับบริการตรวจฟันหลายวัน แต่การศึกษานี้เลือกเฉพาะการตรวจฟันครั้งแรกเท่านั้น

หลังเชื่อมโยงข้อมูลและตัดข้อมูลตามเกณฑ์การศึกษา พบว่าข้อมูลแฟ้มสถานะทันตสุขภาพมีความสมบูรณ์ร้อยละ 24.56 ซึ่งสาเหตุข้อมูลที่ถูกต้องมากที่สุด คือ จำนวนฟัน จำนวนฟันผุ หรือความจำเป็นต้องขูดหินน้ำลายเป็นค่าว่าง รองลงมาคือ จำนวนฟันน้อยกว่า 1 ซี่ ซึ่งส่วนมากจำนวนฟันเท่ากับ 0 เมื่อเปรียบเทียบกับการศึกษาของณัฐนิช สาสีวรรณ (2567) ที่ศึกษาชุดข้อมูลมาตรฐาน 43 แฟ้มของทุกกลุ่มอายุ ในปีงบประมาณ 2565 พบว่ามีความสมบูรณ์ในแนวแถวของข้อมูลแฟ้มสถานะทันตสุขภาพใน 4 โรงพยาบาล ร้อยละ 22.3-98.3 สอดคล้องกับการศึกษานี้ เพราะความสมบูรณ์ข้อมูลมีค่าที่ผันแปรได้มาก ตามนโยบายของแต่ละหน่วยบริการข้อมูลในกลุ่มวัยเด็กอายุ 0-12 ปีมีการบันทึกข้อมูลสมบูรณ์มากที่สุดเนื่องจากเป็นตัวชี้วัดตามแผนพัฒนาระบบบริการสุขภาพ (Service Plan) สาขาสุขภาพช่องปาก และมีการติดตามโดยดูรายงานจาก HDC service⁶ นอกจากนี้ยังมีการตั้งค่าโดยปริยาย (Default value) ของจำนวนฟันและจำนวนฟันผุเป็น “0”⁹ ซึ่งเป็นวิธีการเพิ่มความสมบูรณ์ข้อมูลในเขตข้อมูลที่ต้องไม่ว่าง (Not null) ตามโครงสร้างข้อมูล¹⁹ เนื่องจาก

การไม่กรอกข้อมูลทำให้ข้อมูลไม่เป็นไปตามโครงสร้าง²⁰ และเมื่อเปรียบเทียบการศึกษาของที่ผ่านมาของณัฐนิช สาสีวรรณ ศรีทอง และปิยะดา ประเสริฐสม (2562) ที่ศึกษาแฟ้มสถานะทันตสุขภาพปีงบประมาณ 2560 พบความสมบูรณ์ในแนวคอลัมน์ของทุกเขตข้อมูลมากกว่าร้อยละ 90 ซึ่งมากกว่าการศึกษานี้ (ร้อยละ 48.96-50.33) อาจเกิดจากนโยบายที่แตกต่างกันในแต่ละช่วงเวลา โดยปีงบประมาณ 2560 มีนโยบายพัฒนาคุณภาพข้อมูลในหน่วยบริการ⁸

หลังจากตัดข้อมูลไม่มีคุณภาพ พบว่าข้อมูลแฟ้มสถานะทันตสุขภาพมีความถูกต้องสูงมาก โดยข้อมูลถูกต้องออกเพียง 2 ระเบียบจากจำนวนฟันที่มากกว่า 32 ซี่ และมีจำนวนฟันผุที่เกินจำนวนฟันในช่องปาก สอดคล้องกับการศึกษาของณัฐนิช สาสีวรรณ (2567) พบความถูกต้องร้อยละ 84.3-100⁹ และในการศึกษาของณัฐนิช สาสีวรรณ ศรีทอง และปิยะดา ประเสริฐสม (2562) พบความถูกต้องมากกว่าร้อยละ 90 ยกเว้น จำนวนฟันที่มีความถูกต้องร้อยละ 75.9⁸ ส่วนแฟ้มหัตถการผู้ป่วยนอก มีทั้งความสมบูรณ์และความถูกต้องที่สูงซึ่งสาเหตุข้อมูลที่ถูกต้องออกมาก คือ ไม่มีผลตรวจฟัน หรือไม่ใช้หัตถการทันตกรรม เมื่อเปรียบเทียบกับการศึกษาของณัฐนิช สาสีวรรณ (2567) มีความสมบูรณ์และความถูกต้องของแฟ้มหัตถการผู้ป่วยนอกของ 4 โรงพยาบาล ร้อยละ 88.2-99.9 และ ร้อยละ 96.3-100 ตามลำดับ⁹ อย่างไรก็ตามร้อยละความสมบูรณ์และความถูกต้องในการศึกษานี้ไม่สามารถเปรียบเทียบกับการศึกษาอื่นได้โดยตรง เนื่องจากการศึกษานี้มีการตัดข้อมูลออกจากแฟ้มต่าง ๆ จากนิยามของตัวแปรที่กำหนดและการเชื่อมโยงกันระหว่างแฟ้มด้วยเงื่อนไขเฉพาะของการวิเคราะห์ที่ด้วยการวิเคราะห์การรอดชีพ ดังนั้นจึงพบว่าการศึกษานี้มีความสมบูรณ์และความถูกต้องที่มากกว่าการศึกษานี้ เพราะวิเคราะห์แยกแฟ้ม

จุดแข็งของการศึกษานี้ คือ 1) แสดงการใช้โปรแกรม R เพื่อจัดการข้อมูลทันตกรรมซึ่งยังพบการศึกษาเช่นนี้ไม่มาก ซึ่ง R เป็นเครื่องมือที่ใช้ในการวิเคราะห์และแสดงผลได้หลากหลายกว่าโปรแกรมทางสถิติสำเร็จรูป¹⁰ 2) เป็นการศึกษาความเป็นไปได้ของการทำวิจัยโดยใช้ชุดข้อมูลมาตรฐาน 43 แฟ้ม ดังเช่นการศึกษาของณัฐนิช สาสีวรรณ (2567) ที่ศึกษาระยะเวลาก่อนการสูญเสียฟันในกลุ่มก่อนวัยผู้สูงอายุ²¹ แตกต่างจากการศึกษาส่วนใหญ่ที่วิเคราะห์ความสมบูรณ์จากฐานข้อมูลโดยตรง ซึ่งทราบแต่ร้อยละความสมบูรณ์ในภาพรวมแต่ยังไม่ทราบร้อยละความสมบูรณ์ในแง่การใช้ประโยชน์ในเชิง

วิชาการได้มากเพียงใด จึงเป็นแนวทางพิจารณาให้นักวิจัยที่สนใจการวิเคราะห์จากข้อมูลทุติยภูมิเช่นนี้ เกิดประโยชน์ในการวิจัยคลินิกด้านทันตกรรม จุดอ่อนของการศึกษานี้ คือ 1) การศึกษาจากทุติยภูมิเป็นการกรอกข้อมูลโดยบุคลากรหลายคน หลายหน่วยบริการ จึงส่งผลต่อความถูกต้องของข้อมูล¹ ทำให้เกิด Information bias ได้ 2) พบการไม่กรอกข้อมูลซึ่งทำให้เป็นค่าว่าง และหลังคัดข้อมูลตามเกณฑ์คัดเข้าของแฟ้มสภาวะทันตสุขภาพ พบร้อยละข้อมูลถูกต้องแตกต่างกันตามสถานพยาบาล อาจเกิดจากพันธกิจที่แตกต่างกันในแต่ละหน่วยบริการ ซึ่งทำให้เกิด Selection bias ได้

การใช้เวชระเบียนอิเล็กทรอนิกส์ในการศึกษาระยะยาวสามารถจัดการตัวแปรที่เปลี่ยนแปลงไปตามระยะเวลาตลอดของช่วงเวลาติดตามได้ เช่น จำนวนครั้งการชูดหินปูน หรือจำนวนฟันผุ ซึ่งจะเป็นตัวแปรที่ทำนายการสูญเสียฟันได้เหมาะสมมากขึ้น มีการใช้เวชระเบียนอิเล็กทรอนิกส์ทางทันตกรรมร่วมกับการเรียนรู้ของเครื่อง (Machine learning) เพื่อสร้างโมเดลทำนายสำหรับประเมินความเสี่ยงเกิดโรคปริทันต์และปัจจัยเสี่ยงในคลินิก²² และมีการใช้ขั้นตอนวิธีอัตโนมัติโดยคอมพิวเตอร์ (Automated computer algorithms) เพื่อการศึกษาระยะยาวติดตามการดำเนินของโรคปริทันต์เป็นเวลานาน 15 ปี ซึ่งสามารถนำไปศึกษาวิจัยด้านคลินิกได้²³ นอกจากนี้ การศึกษาของวรรณพงษ์ ชลนภาสถิตย์ (2566) ที่ศึกษาความสัมพันธ์ระหว่างสภาวะฟันผุ ภาวะโภชนาการ และพัฒนาการในเด็ก 0-5 ปี ที่เปลี่ยนแปลงได้ตามระยะเวลา⁷

การศึกษานี้สามารถใช้เป็นแนวทางศึกษาวิจัย พัฒนาองค์ความรู้ และร้อยละความสมบูรณ์ของข้อมูลยังเป็นข้อมูลพื้นฐานคุณภาพข้อมูลของผู้รับบริการที่ใช้สิทธิหลักประกันสุขภาพถ้วนหน้าในหน่วยบริการสังกัดกระทรวงสาธารณสุขเสนอแนะให้ทันตบุคลากรกรอกข้อมูลแฟ้มสภาวะทันตสุขภาพที่จำเป็น สำหรับใช้วิเคราะห์ต่อไป ในอนาคตควรศึกษา 1) การพัฒนาคุณภาพข้อมูลอิเล็กทรอนิกส์ของผู้ป่วย 2) การวิเคราะห์ข้อมูลการตรวจฟันจากโรงพยาบาลชุมชนซึ่งเป็นหน่วยบริการที่มีข้อมูลสมบูรณ์มากที่สุดเพื่อลดอคติ 3) การจัดการข้อมูลที่ได้จากหน่วยบริการนอกกระทรวงสาธารณสุข 4) การจัดการค่าว่างด้วยการทดแทน 5) การดึงข้อมูลจากฐานสิทธิของ สปสช. เพื่อศึกษาผู้ใช้สิทธิการรักษานอกเหนือจากสิทธิหลักประกันสุขภาพถ้วนหน้า ผลการศึกษานี้สามารถใช้สร้างคู่มือการบันทึกและวิเคราะห์เวชระเบียนอิเล็กทรอนิกส์ด้านทันตกรรมเพื่อการดูแลผู้ป่วยและวางแผนการบริการได้

บทสรุป

การจัดการข้อมูลสำหรับการศึกษาระยะยาวโดยใช้โปรแกรม R จากข้อมูลมาตรฐาน 43 แฟ้ม ซึ่งเมื่อนำข้อมูลมาวิเคราะห์จะเกิดองค์ความรู้ใหม่ และเป็นประโยชน์ทั้งในด้านคลินิกและระบบสาธารณสุข แฟ้มสภาวะทันตสุขภาพโดยรวมก่อนเชื่อมโยงข้อมูล มีความสมบูรณ์ร้อยละ 48.80 หลังเชื่อมโยงข้อมูลและมีการติดตามระยะยาว มีความสมบูรณ์ร้อยละ 24.56 และโดยเมื่อจำแนกข้อมูลสมบูรณ์ตามประเภทหน่วยบริการพบเป็นข้อมูลของโรงพยาบาลชุมชนมากที่สุด รองลงมาคือ โรงพยาบาลส่งเสริมสุขภาพตำบล หน่วยบริการอื่น ๆ และ โรงพยาบาลศูนย์ ตามลำดับ

กิตติกรรมประกาศ

ขอขอบคุณ คุณกิตติศักดิ์ ชูมาลี สาขาวิชาเวชศาสตร์ครอบครัวและเวชศาสตร์ป้องกัน คณะแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์ รวมถึงผู้ที่เกี่ยวข้องทุกท่าน

เอกสารอ้างอิง

1. Song M, Liu K, Abromitis R, Schleyer TL. Reusing electronic patient data for dental clinical research: a review of current status. *J Dent.* 2013;41(12):1148-63.
2. Yongsirivit T, Pradmal S, Mobnorin J, Mapong K, Puritakul C, Suktawi S, et al. Provincial, regional, and ministry- level health data center (HDC) systems. 2020 [cited 2022 Nov 1]. Available from: <https://itjournal.moph.go.th/page/detail/22>.
3. Puvasanti S, Suwanwong Y. Study on the opportunities for development of health information technology systems to support medical and public health operations in saraburi province. *Institute for Urban Disease Control and Prevention Journal.* 2021;6(1):125-51.
4. Chakriyanich R. Factors associated with low birth weight in health region 5. *J Pub Health Nurse.* 2020;34(3):1-17.

5. Wichanuwat S, Sriratanaball J. Completeness and accuracy of the standard 52- file dataset for use as a data source for value- based health care indicators in diabetes. *JHSR*. 2021;15(4):11.
6. Sritong N, Hemmanee P. Oral health in the working age group. 2020. Available from: https://dental.anamai.moph.go.th/web-upload/migrated/files/dental2/n4196_e6259c6e55b612675d922b9e4755004b_sur1559y.pdf.
7. Chonlaphasathit W. Health promotion services model in well-child clinics and the relationship of dental caries, growth, and development in children aged 0-5 years [dissertation]. Prince of Songkla University; 2023.
8. Sritong N, Prasertsom P. Quality of dental health records according to the ministry of public health's medical and health data standard structure. *Th Dent PH J*. 2019;24(1):27-43.
9. Saleewon N. Quality of dental data in the health data center system of community hospitals in ratchaburi province [dissertation]. Prince of Songkla University; 2024.
10. Wijitkunakorn P. Medical research, public health, and data analysis using R. 1st Bangkok: Sammitr Phatthana Printing (1992); 2024. 450 p.
11. Wijitsunthornkul K, Thitichai P, Chalorm K, Antimanon A, Kasetpiban N, Nalampang K, et al. Health data analysis in the thai public health system using R. 1st Songkhla: IQ Media; 2020.
12. Techavituravong R, Thianmontri A, Phitthapornchaikun S. Factors affecting one-year loss in type 2 diabetes patients. *J Dent Assoc Thai*. 2023;73(2):104-12.
13. Jakobsen JC, Glud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials—a practical guide with flowcharts. *BMC Med Res Methodol*. 2017;17(1):162.
14. Lewis AE, Weiskopf N, Abrams ZB, Foraker R, Lai AM, Payne PRO, et al. Electronic health record data quality assessment and tools: a systematic review. *J Am Med Inform Assoc*. 2023;30(10):1730-40.
15. Community UGA. Quality assurance of code for analysis and research (version 2023.2): Office for National Statistics, Analytical Standards and Pipelines hub; 2020. Available from: <https://best-practice-and-impact.github.io/qa-of-code-guidance/>.
16. Tableau. Guide to data cleaning: definition, benefits, components, and how to clean your data. 2024. Available from: <https://www.tableau.com/learn/articles/what-is-data-cleaning>.
17. Alexcernat. Understanding the longitudinal data workflow: a comprehensive guide. 2024. Available from: <https://longitudinalanalysis.com/understanding-the-longitudinal-data-workflow-a-comprehensive-guide/>.
18. Phansima N, Chaisakul A. Comparison of R and SPSS programs for classifying advance payment data of Khon Kaen University Under Data Mining Concept. *J Sci Tech UBU*. 2014;16(1):16-31.
19. Office of the Permanent Secretary, Ministry of Public Health. Handbook for Collecting and Submitting Data According to Health Data Standard Structure, Fiscal Year 2021, version 2. 4. Ministry of Public Health; 2021.
20. Zhang R, Indulska M, Sadiq S. Discovering data quality problems. *Bus Inf Syst Eng* 2019;61(5):575-93.
21. Jirachaiprasit N. Time until tooth loss among the pre-elderly and related factors: a study based on the 43- file standard datasets of a province in health region 12, Thailand [dissertation]. Prince of Songkla University; 2024.

22. Patel JS, Su C, Tellez M, Albandar JM, Rao R, Iyer V, et al. Developing and testing a prediction model for periodontal disease using machine learning and big electronic dental record data. *Front Artif Intell*. 2022;5:979525.
23. Patel JS, Kumar K, Zai A, Shin D, Willis L, Thyvalikakath TP. Developing automated computer algorithms to track periodontal disease change from longitudinal electronic dental records. *Diagnostics* 2023;13:1028.

ผู้ประพันธ์บรรณกิจ

จันทร์พิมพ์ หินเทาวิ

หน่วยวิจัยเพื่อการพัฒนาการดูแลสุขภาพช่องปาก
สาขาวิชาทันตกรรมป้องกัน คณะทันตแพทยศาสตร์
มหาวิทยาลัยสงขลานครินทร์

อำเภอหาดใหญ่ จังหวัดสงขลา 90112

โทรศัพท์ 074 287 601

จดหมายอิเล็กทรอนิกส์ : janpim.h@psu.ac.th



Data Management and Completeness of Dental Data in the 43-File Standard Dataset for Longitudinal Tooth Loss Study Using R Program

Nutnicha Jirachaiprasit¹ Janpim Hintao^{2,*} Wattana Pithpornchaiyakul³

Research Article

Abstract

Objective: This study aimed to 1) describe the methods of data management and the results of data processing, and 2) analyze the completeness of dental data from the 43-standard dataset files to conduct survival analysis of tooth loss after dental examination.

Materials and Methods: The sample group consisted of individuals utilizing the Universal Health Coverage Scheme who received dental check-ups at Ministry of Public Health-affiliated healthcare units in a southern province during 2019-2021. Dental examination data and dental service information were obtained from the dental health status files and outpatient procedure files. The data were linked through general data files, managed, and analyzed descriptively using the R program.

Results: The study found the following steps for data management: 1) data importation, 2) handling of missing values, 3) filtering the sample group based on inclusion and exclusion criteria, 4) data quality screening by checking data accuracy, 5) removing duplicate data using unique fields, and 6) linking files via hashed national ID numbers. The completeness of the dental health status files revealed that 6,820 records (20.33%) out of 33,553 imported records were complete for individuals who had undergone dental examinations. From the 4,051,350 records imported from the outpatient procedure files, 4,177 records (0.10%) had complete information for those who had undergone a dental examination and received either tooth extraction or other dental services, providing data to analyze tooth loss survival after dental check-ups. In conclusion.

Conclusion: after managing and linking all files, the sample group with complete dental data was sufficient for long-term studies on tooth loss, accounting for 24.56%. The R programming is effective tool for managing electronic dental health records to create a dataset suitable for conducting longitudinal studies on tooth loss.

Keywords: Electronic Health Records /Data Management/ Software/ Data Completeness

Corresponding Author

Janpim Hintao
Research Unit for Oral Health Care Development,
Department of Preventive Dentistry,
Faculty of Dentistry, Prince of Songkla University,
Hat Yai, Songkhla 90112
Tel.: +66 74 287 601
Email: janpim.h@psu.ac.th

¹ Residency Training Program in Dental public Health, Faculty of Dentistry, Prince of Songkla University.

² Improvement of Oral Health Care Research Unit, Department of Preventive Dentistry, Faculty of Dentistry, Prince of Songkla University.

³ Evidence-based Dentistry for Oral Health Care and Promotion Unit, Department of Preventive Dentistry, Faculty of Dentistry, Prince of Songkla University.

* Corresponding Author