



שם הפרויקט		מס' פרויקט
כלים אנליטיים לזיהוי פגמי איכות בנתוני טיפול נמרץ		P-2018-021
מנחה שותף	מנחים אקדמיים	
	ד"ר אדיר אבן, ד"ר יובל ביתן	
		חברת הצוות
		אורן קרמר
		orenkram@post.bgu.ac.il

תקציר

נתונים קליניים מהווים בסיס חשוב לקבלת החלטות רפואיות. פגמי איכות בנתונים אלו עלולים לפגום משמעותית באיכות קבלת ההחלטות והטיפול הרפואי. מרבית מחקרי איכות הנתונים הקליניים התמקדו עד עתה בעיקר בנתוני רשומה רפואית אלקטרונית המוזנים ידנית. מחקר זה יתמקד בבחינת איכות הנתונים הקליניים בסביבה המתאפיינת בקלט אוטומטי – למשל, בעזרת חיישנים או מכשור רפואי אחר המשמש לניטור המטופל. המחקר מבקש לבחון כיצד ניתן להגדיר ולמדוד פגמי איכות בנתונים קליניים הנאספים באופן אוטומטי, וכיצד פגמים כגון אלו עשויים להשפיע על קבלת ההחלטות הרפואיות ואיכות הטיפול.

המחקר נערך בשיתוף פעולה עם היחידה לטיפול נמרץ של המרכז הרפואי איכילוב בתל אביב. יחידה זו מטפלת בחולים במצב חמור, המצריך מעקב צמוד וניטור רציף של מדדיהם הרפואיים, במטרה לשמור על תפקוד מערכות תקין. מדדים רפואיים, כגון לחץ דם, פעימות לב, וקצב הנשימה, נקלטים באופן אוטומטי בקצב דגימה גבוה על ידי מכשור רפואי וחיישנים המוצמדים לכל חולה. ניסיון העבר מראה כי בסיס הנתונים אליו נצברים נתונים אלו סובל מפגמי איכות העלולים להיווצר עקב תקלות במכשור הרפואי, קריאות שגויות של חיישנים, חיבור רופף של המכשור למטופל וכד'. שאלת המחקר שעלתה בעקבות דיונים מקדימים עם צוות היחידה היא האם וכיצד ניתן לאתר בעיות איכות כגון אלו, למדוד את שיעורן, ולהבין את השפעתן על איכות הטיפול, המצריך קבלת ההחלטות נכונות ומדויקות של הצוות הרפואי ביחידה.

הפרויקט שנערך בשנה זו, כשלב מקדים לעבודת תזה מחקרית, התמקד במספר שלבי הכנה ואפיון מקדימים חשובים. סקר ספרות מקיף של מחקרים קודמים הצביע על פערים משמעותיים בהבנה ואפיון של פגמי איכות בנתונים הנאספים באופן אוטומטי, ובפרט בהקשר הרפואי. הסקר אף סייע לאתר שיטות אלגוריתמיות שונות שהוצעו בעבר, העשויות להתאים כבסיס לפיתוח כלים אנליטיים לזיהוי ומדידת פגמי איכות. מחקרים רבים נוטים לבחון את סוגיית איכות הנתונים על פי מימדים שונים, המשקפים סוגים שונים של פגמי איכות. לאחר בחינה ראשונית של מדגם נתונים רחב היקף שהתקבל מבית החולים, הוחלט להתמקד בשלושה מימדי איכות העשויים להיות רלוונטיים ומועילים בהקשר של איכות הנתונים הנאספים ביחידה לטיפול נמרץ: א. שלמות – מימד המשקף פגיעות לנתונים חסרים, ב. תקפות – מימד המשקף חריגות ממרחב הערכים התקני המוגדר למדד, ו-ג. עקביות – מימד המשקף תנודתיות חריגה של ערכי נתונים לאורך זמן.

הפרויקט כלל פיתוח סביבת מחקר, המבוססת על שפת R וספריית Shiny לויזואליזציה של נתונים. סביבה זו תומכת בתחקור הנתונים והבנתם, ובחינת כלים אנליטיים העשויות להתאים להערכת איכותם על פי מימדים שונים. כיוון שנתוני המדדים הרפואיים נאספים בדגימות אחידות לאורך זמן, ניתוח הנתונים מתבסס בעיקרו על טכניקות לניתוח סדרות עתיות. סביבת המחקר מאפשרת בחינה שיטתית של כלים אנליטיים שונים, תוך מדידת ביצועים ביחס לסט מבחן שתויג מראש. בחינה ראשונית של מספר כלים שפותחו מצביע על אפשרות לזהות נכונה חלק מבעיות איכות הנתונים.

המשך המחקר יתמקד בשיפור כלי מדידת האיכות, וישערך באמצעותם את איכות הנתונים עבור דגימות נתונים בהיקף נרחב בהרבה במטרה לבחון השפעות אפשריות של פגמי האיכות על ההחלטות הרפואיות שניתן לקשרן לנתונים הנאספים ביחידה לטיפול נמרץ.

מילות מפתח: יחידת טיפול נמרץ, איכות נתונים, שלמות, תקפות, עקביות.



Project No.	Project Title	
P-2018-021	Analytical Tools for Detecting Quality Anomalies in Emergency Care Data	
Academic Advisors		Co-Advisor
Dr. Adir Even, Dr. Yuval Bitan		
Team Member		
Oren Kramer		
orenkram@post.bgu.ac.il		

Abstract

Poor clinical data quality (DQ) might harm decision-making and medical treatment. So far, clinical DQ research focused on manually-entered data. This study examines the quality of clinical data in scenarios that are characterized by automatic acquisition - e.g. by sensors and medical devices uses for patient monitoring. The study explores methods for defining and measuring DQ in such scenarios, alerting on defects and examine their possible impact on medical decision making.

This study is performed in collaboration with the Intensive-Care Unity (ICU) department of Tel-Aviv Ichilov Medical Center. The ICU unit treats patients in severe and life-threatening conditions, which require minute-to-minute monitoring of clinical parameters, such as heart rate, blood pressure and respiratory rate. These parameters are recorded at a high sampling rate by devices and sensors attached to each patient. Past experience shows that the data collected by those devices suffers from quality defects, caused by device malfunctions, environmental noise, loose connection, and possibly other causes. Preliminary discussions with ICU staff raised the question of how such quality defects can be identified and measured and to what extent do they affect doctors' decision making and the quality of medical care.

This years' project aimed at establishing the base for the study. A review of past literature has indicated significant gaps in the understanding and characterization of quality defects in automatically collected data, particularly in medical contexts. The review has also identified relevant algorithmic methods, later used as a basis for developing tools for detection and measurement of DQ defects. A preliminary examination of a large real-world data sample led to the definition of three relevant DQ dimensions: A. Completeness, reflecting the extent of missing data values, B. Validity, reflecting the extent of actual values confirming to a certain pre-defined value domain, and C. Consistency, reflecting the extent of value stability over time.

The project also established a software-supported research environment, based on the R Studio and the Shiny library for data visualization. This environment permits visual exploration of data, as well as DQ assessment along different dimensions, with tools that adapt relevant time-series analysis techniques. This environment enables systematic examination of various analytical tools, while comparing performance for a pre-tagged test-set. A preliminary evaluation of several tools has yielded promising results, in terms of identifying DQ problems correctly.

The next phase of this study will aim at improving the DQ measurement tools, evaluating them with larger datasets, and examining how DQ defects might affect ICU-related decisions.

Keywords: Intensive Care Unit (ICU), Data Quality (DQ), Completeness, Validity, Consistency.