# Data Quality Oriented Efficacy Evaluation Method for Ambient Assisted Living Technologies

**Femina Hassan Aysha Beevi**
Section of ECE
Department of Engineering
Aarhus University
8200 Aarhus
Denmark
feay@eng.au.dk

**Christian Fischer Pedersen**
Section of ECE
Department of Engineering
Aarhus University
8200 Aarhus
Denmark
cfp@eng.au.dk

**Stefan Wagner**
Section of ECE
Department of Engineering
Aarhus University
8200 Aarhus
Denmark
sw@eng.au.dk

**Stefan Hallerstede**
Section of ECE
Department of Engineering
Aarhus University
8200 Aarhus
Denmark
sha@eng.au.dk

PervasiveHealth 2016, May 16-19, Cancun, Mexico

DOI 10.4108/eai.16-5-2016.2263753

## Abstract
In this paper we present a data quality oriented method for the assessment of efficacy of Ambient Assisted Living (AAL) technologies. The method evaluates the efficacy by measuring the data quality of the technologies along the dimensions of accuracy, completeness, timeliness and interpretability. The method was evaluated on four activities of daily living (ADL) products by Any Group: Door guard, Living room guard, Bed guard and Toilet guard. The evaluation revealed the data quality issues of the technologies; thus the method may be effective to evaluate the efficacy of AAL technologies.

## Author Keywords
Efficacy evaluation; Technology assessment.

## ACM Classification Keywords
C.4 Performance of Systems : Performance attributes

## Introduction
Ambient assisted technologies are intelligent technologies that provide assistance to people with disabilities at any stage of their life. They are intended to provide assistive services in the areas of prevention, compensation and support, and independent and active

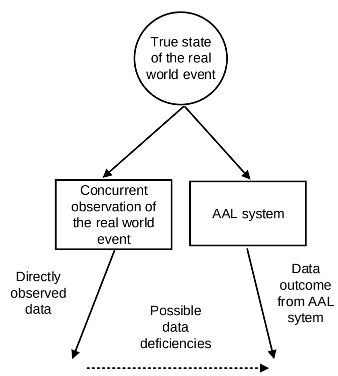


**Figure 1**. Data deficiency in an AAL system from an external black box test perspective.

aging; these areas are identified in the AALIANCE 2 roadmap [1]. The user acceptance rate of AAL technologies is low in the home settings [2]. One of the reasons for this is the lack of trust in the technology. Evaluation of the technology during the different stages of the life cycle of the technology may help to measure the functional and non-functional performance of the system. Efficacy is one of the most important evaluation dimensions as it provides data about the functional performance of the system.

AAL technologies are used for sensing, acting, reasoning, interacting and communicating. The care givers and service providers may access the data from remote locations. If the data are not fit for its purpose, it will have a negative impact on the quality of care and assistance. For example erroneous data could cause a critical event such as a fall to go unnoticed by the caregivers. Hence there is a need to assess the performance of the technologies based on the data quality. Wang and Strong define data quality as "*data that are fit for use by data consumers*". They defined the data quality dimension as "*a set of data quality attributes that represent a single aspect or construct of data quality*" [3].

Salvi et al. [7] proposed an assessment framework for AAL technologies based on the ISO/IEC 25000 standard for software quality that can be used for both technical and non-technical evaluation. It is an evaluation method based on software quality characteristics. The framework does not provide specific guidelines for the objective measurements of the characteristics of the artifacts. The activities we focus on are data capture, data maintenance and data delivery. The efficacy evaluation method presented in this paper is a sub-method of the AAL Care technology assessment method (CTAM) developed by the authors. CTAM is an HTA based method, for the assessment of AAL technologies. The aim of this study is to present an efficacy evaluation method for the assessment of AAL technologies in the home setting. This efficacy evaluation is intended for generating evidence for policy makers as part of the HTA-based CTAM method.

## Efficacy Evaluation

EUnetHTA core model defines efficacy as "the extent to which a technology does more good than harm under ideal circumstances" [8]. We propose to measure efficacy of AAL technology by measuring the quality of the data that comes out of the AAL system. The data outcome of an AAL system can be structured or semi-structured data. We measure efficacy in terms of data quality dimensions accuracy, completeness, timeliness and interpretability. Accuracy, completeness and timeliness are measured using quantitative methods and interpretability is measured using qualitative methods. The efficacy evaluation is done by looking at the AAL system in a black box perspective as presented in Figure 1. The data produced in an AAL system may have different users such as the care givers and service providers. The subjective and objective data quality measurements must be done from the different user perspectives. Each user may be acquiring the sample at a different sampling rate and the data for each user can also be different. For example, the doctor may acquire a summary of the activities of the patient, once in a month whereas a relative may monitor the activities of the patient every hour. Hence the evaluation must be initiated by acquiring the data requirements of the users. Then define the data flow through the system and decide the critical point of assessment of data

| |
|---|
| Name of the Critical data element |
| Critical assessment point user |
| Data quality thresholds |
| Type of data |
| Sampling rate |
| Source of data |
| Data quality score |

**Table 1 .** Structure of a critical data assessment element

quality. The type of data can be structured or semi-structured. The method of data quality assessment must be chosen according to the type of data.

**Accuracy**: Accuracy of data presents the closeness of the data to the values or events in the real world. Accuracy is measured by comparing the device measurements with the manual observations. The data from an AAL technology may represent data such as vital parameters or events representing activities. The accuracy measurement method depends on the type of monitoring. If the monitoring is done for identifying activity or risk conditions such as fall or physiological data of the patient crossing risk levels, true positive rates can be used to measure the accuracy. If the monitoring is done to collect the physiological data then the correlation method can be used.

**True positive rate**: The data representing events may be classified as true positive (TP), false positive (FP), true negative (TN) or false negative (FN). Fall events correctly identified by a fall detector as a fall is an example of a true positive. Identifying sitting event as fall is an example of a false positive. Identifying fall event as a non-fall event is an example of a false negative. Identifying non-fall event as non-fall event is true negative. True positive rate = TP / ( TP + FN). The accuracy of the event detectors that does a multi-class classification can be measured as the number of correct predictions across all classes, k, divided by the number of samples, n.

If the data follow a normal distribution t-test can be used to determine whether the data in the gold standard is significantly different from the data measured using the technology being assessed. Before performing the correlation test the data in the gold standard dataset must be aligned for comparison with the data in the dataset from the AAL technology being evaluated.

**Completeness**: Data may be lost due to communication failures, failure of sensors or environmental interference. Completeness of data represents the ratio of actual data in the measurements to the total number of data expected in the assessment point. Amount of data in the destination must be equal to timeframe length times the sampling frequency. The amount of expected data in the destination depends on the sampling rate of the assessed technology for different users.

**Timeliness**: Timeliness is the latency between the actual occurrence of an event (te) and the time of arrival at the data destination (ts). The destination will be the critical point where the data quality is measured. The value of minimum time ( tmin) and the maximum time (tmax ) to arrive from the source to the destination is set according to the user requirement. Timeliness=1 if $tmin \leq (ts - te) \leq tmax$ and 0 otherwise.

**Interpretability:** Interpretability represents the extent to which the meaning and properties of the data source are clear to the user. The interpretability can be measured using a subjective assessment method such as user interview. Interpretability may be rated by the user as HIGH, MEDIUM or LOW and each with a score of 1, 0.5 and 0 respectively.

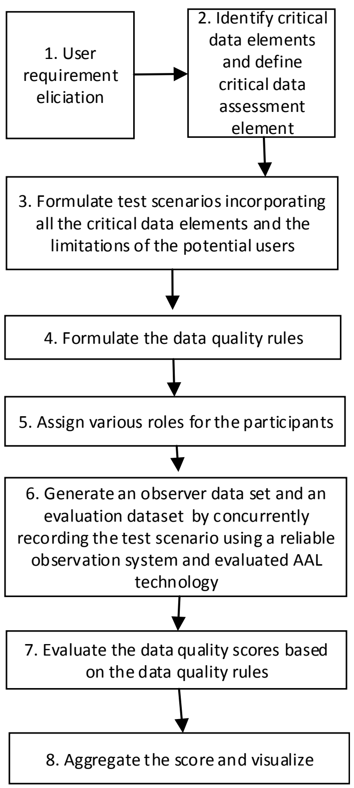


**Figure 2**. Data quality oriented efficacy evaluation process

## Efficacy evaluation process

The efficacy evaluation process is presented in Figure 2. The efficacy evaluation proposed is intended to be conducted in a laboratory environment. The efficacy evaluation commences with step 1 and step 2 in Figure 2. For step 3, test scenarios must be planned as a series of events around the critical data elements. For repeatability of the scenario, record the planned scenario and repeat the experiment to verify that the first set of experiment is representative and reproducible. Data quality rules must be formulated based on the user requirements collected in step 1. Various pre and post activities must be included around the critical event to be monitored, incorporating the possible limitations of the potential users. In order to do subjective data quality evaluation, the test participant may be assigned the roles of the various data users. The dataset for the efficacy evaluation may be made by implementing the test scenarios for different user contexts, time contexts, physical contexts and communication contexts. The test scenario may be implemented following the procedure presented below.

The test participants must be identified and recruited. The anthropometric measures of the user must be recorded. The technology being assessed must be prepared for the experiment. Conducting pilot studies will help to increase the familiarity with the technologies used in the assessment procedure. Explain the test protocol to the users. Video record the test procedure if feasible and conduct a pilot test before the actual test. The data quality assessment requires the ground truth as the "gold standard". Use an observation system to record the ground truth, meaning the actual events or data, in the context. A reliable alternative of the same technology can be used as the observation system. If no reliable alternative of the technology is available, use a manual observer and recorder. The recorder will record the details of the true event either manually or with the help of a software tool that can register the events and save it in a file.

Data quality score in step 7 can be measured by comparing the attributes logged by the evaluated system with the recorded attributes of the ground truth in the observation system. Assess the data quality along the dimensions of accuracy, completeness, timeliness and interpretability. Evaluate whether the data quality meet the requirements of the users. Define data quality rules based on the user requirements.

Calculate the data quality scores based on the definitions of data quality dimensions presented in section 2. The data quality evaluation set up is presented in Figure 3. The data quality score must be evaluated by comparing the data set from the observation system with the data set generated by the evaluated AAL technology. The data quality score must be evaluated for each critical data element. Ideally, the entities in the observed dataset and evaluated data set must have a one to one relation. But due to the data quality issues there may be a one to many relationship between the observer dataset and the evaluation dataset. Before comparing the datasets the tables must have a one to one relationship between its entities. Segment the dataset into time windows of equal size. Find the data quality scores of each window and the average of the window data scores will give the data quality score of the critical data elements. The overall data quality score can be aggregated by combining all the scores of the critical data elements. The resulting

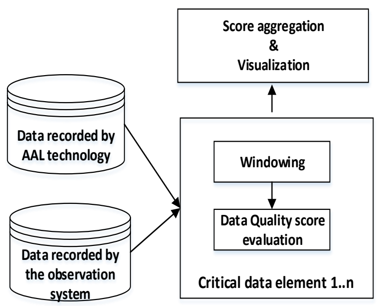


**Figure 3.** Data quality evaluation

| CDE | A | T | C | I |
|---|---|---|---|---|
| Toilet presence | 0.95 | 1 | 0.95 | 1 |
| Leave home | 0.93 | 1 | 1 | 0 |
| Leave bed | 0.87 | 1 | 2.24 | 0 |
| Fall detect | 0 | 0 | NA | NA |

**Table 2**. Data score card of the evaluation of Any group products Door Guard, Living room guard, Bed guard and Toilet guard. **A**: Accuracy, **T** : Timeliness, **C:** Completeness, **I**: Interoperability

score can be appended to the critical data assessment element.

## Evaluation of the Method

The proposed method was evaluated using it for the assessment of the ADL products: Door guard, Living room guard, Bed guard and Toilet guard from Any group. The efficacy evaluation was done as part of a CTAM based assessment of the technologies. The products are mainly intended for monitoring risky situations such as fall and nocturnal wandering in a home setting. The products may be installed as standalone components, or as part of an overall system architecture called the "Intelligent Care System", or as components that may be integrated into third party systems. The evaluation was done by four groups of four students each. Following the step 1 of the efficacy evaluation process, a user requirement collection was done. The requirements were collected through dialogue with the AAL service providers. We identified toilet presence, leave home, fall detect and leave bed as the critical data elements. The critical data elements were defined.

*Test Protocol*

The test scenarios were formulated based on the critical data element and each group implemented one scenario. The events in the scenario were carefully planned based on an experiment plan. Out of the four participants in each group, one participant enacted the events while two acted as observers and one as recorder of the ground truth. Each scenario was repeated five times by the same participant with a break of one minute between the trials. During the break the recorder recorded manually the log value in the central database. The ground truth was recorded by recording manually the event representing the ground truth and the time at which the event occurred. The reference time used was the time displayed on personal computer of the recorder. The log was constantly observed by the recorder. The manual entry was compared with the log entry to classify the events as true positive, false positive, true negative or false negative. The participants who implemented the test scenario were asked to emulate the elderly people. The scenarios in Table 3 were performed by the test participants.

## Results

The data quality measurement results, data quality score card, are presented in Table 2. The critical assessment point in the results presented is the web interface for the caregiver. The minimum value of accuracy is 0 and the maximum value is 1. If the accuracy of critical data element is 1 then it is close to the ground truth. If accuracy is zero the data is far from the ground truth. Completeness has a value more than 1 if the device detected many false positives. If the data arrive the data destination within the expected time the value of timeliness is 1 and 0 otherwise. Interpretability is measured based on a subjective analysis.

## Discussion

The evaluation of the Any Group products revealed data quality issues in laboratory scenario. The evaluation revealed data quality issues in two out of the four critical data elements. These findings indicate the need for further evaluation of the commercial technologies before they are purchased and deployed in care

| |
|---|
| Opened the door and entered the toilet and acted as though they brushed teeth or took a shower. |
| Started walking from 220 cm behind the door and left home through the door fitted with the Door guard and stopped walking at 160 cm after the door. The participant walked in a straight line with a normal pace. |
| Got up from lying on the bed and walked through the apartment and then left home. |
| Fell down while walking towards the couch and lay between the couch and the table , and sat up for 30 minutes. In the second trial the participant lay still for 30 minutes in the post fall phase. |
| Started walking from the couch at one end of the room to the bed at the other end of the room for approximately 4 minutes. Then sat on the edge of the bed and lay down on his left side. Then he twice rolled onto his right side and back to the left side. From lying on left position, he moved to sit on the edge of the bed and rose to the standing position and walked to the couch. |

**Table 3** Test Scenarios

facilities. The data quality issues were primarily associated with fall detection and leave bed detection. The fall event was not detected even after the participant lied on the floor for 30 minutes. The completeness of "leave bed" was significantly high. It reflected the high number of false positives detected by the system. It was difficult for the care giver to interpret the messages in the case of leave home and leave bed as multiple activities were represented using the same message string. During the evaluation the system failed and it took 25 minutes for recovery. This indicates the need for assessing the system from more dimensions including safety analysis.

## Conclusion

The data quality oriented efficacy evaluation may be an effective method to assess and compare the performance of the technologies as it provides a quantitative measure in terms of data quality. However, our study is limited by limited amount of participants and trials. The automation of this evaluation process will be an interesting future work.

## References


1. AALIANCE2 | European Next Generation Ambient Assistence Living Innovation Alliance: 2016. *http://www.aaliance2.eu/*. Accessed: 2016-01- 26.
2. AALIANCE2 | European Next Generation Ambient Assistence Living Innovation Alliance: 2016. *http://www.aaliance2.eu/*. Accessed: 2016-01- 26.
3. Reginatto, B. M. B. Understanding barriers to wider telehealth adoption in the home environment of older people: An exploratory study in the irish context. International Journal on Advances in Life Sciences Volume 4, Number 3 & 4, 2012, ( 2012).
4. Wang, R. Y. and Strong, D. M. Beyond accuracy: What data quality means to data consumers. J. Manage. Inf. Syst., ( 1996), 5-33.
5. Klein, A. and Lehner, W. Representing data quality in sensor data streaming environments. Journal of Data and Information Quality (JDIQ), 1, 2 ( 2009), 10.
6. Weiskopf, N. G. and Weng, C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J. Am. Med. Inform. Assoc., 20, 1 (Jan 1 2013), 144-151.
7. Pipino, L. L., Lee, Y. W. and Wang, R. Y. Data quality assessment. Commun ACM, 45, 4 ( 2002), 211-218.
8. Salvi, D., Montalvá Colomer, J. B., Arredondo, M. T., Prazak-Aram, B. and Mayer, C. A framework for evaluating Ambient Assisted Living technologies and the experience of the universAAL project. Journal of Ambient Intelligence and Smart Environments, 7, 3 ( 2015), 329-352
9. HTA Core Model Online 2015. http://meka.thl.fi/htacore/ViewApplication.aspx?id=23765 Accessed: 2016- 01- 26.