International Journal of Statistics and Probability; Vol. 8, No. 4; July 2019
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Improving the Healthcare Quality Measurement System Using Attribute Agreement Analysis Assessing the Presence and Stage of Pressure Ulcers

Sandra L. Furterer[1], Ethling Hernandez[2]

[1] University of Dayton, Department of Engineering Management, Systems & Technology, Dayton, Ohio

[2] Doral, FL

Correspondence: Sandra L. Furterer, University of Dayton, 300 College Park Avenue, Kettering Lab, Dayton, OH, 45469. E-mail: sfurterer1@udayton.edu

Received: May 29, 2019 Accepted: June 24, 2019 Online Published: June 25, 2019

doi:10.5539/ijsp.v8n4p47 URL: https://doi.org/10.5539/ijsp.v8n4p47

Abstract

Measurement Systems Analysis has recently been used in healthcare service processes mainly to assess the accuracy and use of equipment and devices. However, thirty-seven percent of emergency department healthcare malpractice claims were related to diagnosis errors. Diagnosis is heavily dependent upon human assessment and decisions. The paper describes the application of a healthcare case study that applied Measurement Systems Analysis Attribute Agreement Analysis and Gage R&R studies to assess the accuracy of the human element in a healthcare service process. The study was used to assess the accuracy of the diagnosis of pressure ulcers when patients are admitted to the hospital, either through the emergency department or directly through inpatient admitting. Creating an accurate and precise measurement system aided the hospital by standardizing the assessment of the pressure ulcer healthcare diagnosis process. Initial Attribute Agreement Analysis of whether a pressure ulcer was present resulted in a 94% assessor repeatability accuracy rate, and a 40% within assessor reproducibility accuracy. The within appraiser accuracy to the standard was 92%, and across assessors' assessment to the standard was 40%. The measurement system was poorer for assessing the pressure ulcer stages, resulting in 82% within assessor repeatability accuracy and an 8% overall accuracy to standard. This study is extremely important to 1) identify a method for healthcare providers to assess and improve the measurement system related to human diagnoses in healthcare processes; and 2) to demonstrate the usefulness of expanding gage R&R and attribute agreement analysis to human diagnosis in healthcare settings.

Keywords: measurement systems analysis, attribute agreement analysis, healthcare, pressure ulcers, attribute gage repeatability & reproducibility study, attribute data, reproducibility, repeatability

1. Introduction

Measurement Systems Analysis (MSA), including measurement assessment tools of Gage Repeatability & Reproducibility (R&R), and Attribute Agreement Analysis, has traditionally been used in manufacturing processes to ensure the accuracy of the measurement systems. There are many examples of the application of these measurement systems analysis techniques to manufacturing products and processes. These techniques have recently been used in healthcare processes mainly to assess the accuracy and use of equipment and devices. However, thirty-seven percent of emergency department healthcare malpractice claims were related to diagnosis errors. Diagnosis is heavily dependent upon human assessment and decisions. The paper describes the application of a healthcare case study that applied Measurement Systems Analysis Attribute Agreement Analysis and Gage R&R (Repeatability and Reproducibility) studies to assess the accuracy of the human element in a healthcare diagnosis service process. The study was used to assess the accuracy of the diagnosis of pressure ulcers when patients are admitted to the hospital, either through the emergency department or directly through inpatient admitting. Creating an accurate and precise measurement system aided the hospital by standardizing the assessment of the pressure ulcer healthcare diagnosis process. This study is extremely important to 1) identify a method for healthcare providers to assess and improve the measurement system related to human diagnoses in healthcare processes; and 2) to demonstrate the usefulness of expanding gage R&R and attribute agreement analysis to human diagnosis in healthcare settings.

2. Literature Review

2.1 Measurement Systems Analysis Background

Several healthcare case studies applying Gage R&R measurement systems analysis techniques have recently been

applied to the healthcare setting, mainly related to instrumentation, but not to human diagnosis processes. A repeatability and reproducibility study was performed to assess the topographic disk parameters with the retinal thickness analyzer. High variation was found in the reproducibility and repeatability measures (Hoffman and Medeiros, 2006). Another optical study performed a reproducibility and repeatability clinical evaluation of an OcuSense TearLab (TM) osmometer (Eperjesi, Maana, and Hannah, 2012). Mcalinden, Jyoti, and Konrad (2011) studied the use of appropriate statistical techniques in assessment of ocular metrology and imaging equipment. A gage repeatability and reproducibility study along with analysis of variance analysis were used to measure blood pressure gages (Dalalah, and Diabat, 2015). Yet another optical study was performed to determine the repeatability, reproducibility, and agreement of anterior chamber depth (ACD) measurements obtained with 3 Scheimpflug cameras and an anterior segment optical coherence tomography (AS-OCT) device (Wang, et al., 2015). Arani and Erdil (2017) discussed an example of applying attribute measurement analysis to residents determining if a root canal should be performed on a patient based on viewing X-rays. Arani, O. M., and Erdil also found through their literature review that most of the MSA applications in healthcare has involved measurement using instruments, not based purely on the human assessment and decisions (Arani and Erdil, 2017). Another study applied gage R&R in a hospital to assess the precision of temperature measured using an ear thermometer (Erdmann, Does and Bisgaard, 2010). A gage R&R study was re-visited with regard to thermometer temperature measurement applying a gold standard using a rectal thermometer along with an experimental design to discern significant factors' effect on measurement error (Akkerhuis, Niemeijer, Trip, Gemke, and Does, 2015). Another application of gage R&R in a hospital was used to assess the measurement system of daily patient setup in a radiation oncology department (Pawlicki and Perry, 2011).

In a study of closed emergency department claims, thirty-seven percent of the healthcare malpractice claims are related to diagnosis errors (Brown, 2010). Diagnosis is heavily dependent upon human assessment and decisions. We will describe the application of a healthcare case study that applied Measurement Systems Analysis Attribute Gage R&R, Attribute Agreement Analysis studies to assess the accuracy of the human element in the diagnosis and assessment process of pressure ulcers.

An accurate measurement system is a basic requirement for assessing the quality of any product or service process. Measurement Systems Analysis (MSA) is a quality tool for assessing the accuracy of the measurement system. Measures and metrics are vital to assessing the current state of a quality system, and for assessing whether the process improves over time. The measurement system that is used should include:

1) Metrics that are operationally defined for how they will be measured

2) Sampling method, including selection of appropriate sample size

3) Consistent and standard method for data collection, including instructions

4) Validation that the people who collect the data and make the measurement are doing so in an accurate and consistent manner.

Gage Reproducibility and Repeatability (R&R) studies are tools that enable the assessment of the accuracy of a measurement system. Gage R&R studies can measure variable or attribute data. A variable Gage R&R study measures the system for measuring variable data, such as time, temperature, or height. Attribute data is placed into categories, such as good or bad, severity of 1, 2, 3, 4, or 5. However, error can be introduced by how the person who takes the measurement follows the measurement procedure, how they use the measuring device, and from the measurement device itself. Attribute data, by the nature of a person needing to assess the category to place the measurement into can be more difficult and lend to more error in the measurement system. Attribute measurement systems assessment studies are typically called Attribute Agreement Analysis, because these studies measure how well assessors agree with each other, with themselves, and with a standard when placing the measurement into attribute categories.

Attribute Agreement Analysis / Total Gage R and R variability, is divided into:

Precision:

Repeatability – the variability from repeated measurements of the same part by the same operator

Reproducibility (which can be further divided into operator and operator-by-part components) – the variability when the same part is measured by different operators

Part to Part – the variability in measurements across different parts.

Accuracy: Comparing the assessments to a standard.

Ideally you want most of your variation in part to part (pressure ulcer to pressure ulcer) and minimal variation in how the parts (pressure ulcers) are measured – repeatability and reproducibility. This ensures that changes in the process can be detected above the measurement error.

If the Attribute Agreement Analysis fails, meaning that the precision and accuracy is too low, the following steps should be performed to improve the measurement system:

- Review operational definition accuracy to define good, bad and gray areas
- Retrain operators on how to apply the operational definition
- Conduct a Kaizen event to generate and improve the measurement system
- Perform an attribute agreement analysis again after the process has been improved

2.2 Attribute Agreement Analysis Metrics

The metrics used to measure the precision related to repeatability and reproducibility, and accuracy compared to a standard include the following metrics:

1) Within Assessor Repeatability Accuracy:

$$\frac{\sum_1^a \sum_1^t Number\ Parts\ Matched}{\sum_1^a \sum_1^p Number\ Parts\ Inspected} \text{X } 100\%$$

2) Within Assessor Reproducibility Accuracy:

$$\frac{Number\ all\ assessors' assessments\ agree\ with\ each\ other}{Number\ Parts\ Inspected} \text{X } 100\%$$

3) Across Assessor Accuracy to Standard:

$$\frac{\sum_1^a \sum_1^t Number\ Parts\ Matched\ \ to\ Standard}{\sum_1^a \sum_1^p Number\ Parts\ Inspected}$$

4) Overall Accuracy to Standard:

$$\frac{Number\ all\ assessors' assessments\ agree\ with\ known\ standard}{Number\ Parts\ Inspected} \text{X } 100\%$$

Where: t= Number of Trials Per Assessor

a = Number of Assessors

p = Number of Parts

2.3 Attribute Agreement Analysis Method

The following steps are performed to complete an Attribute Agreement Analysis.

1) Identify and operationally define the quality characteristic and metrics to be measured.
2) Define the sampling method, including selection of appropriate sample size and procedure
3) Develop a consistent and standard procedure for data collection
4) Train the operators in the data collection and measurement methods
5) Run an attribute agreement analysis as follows:

- Select the operators, which at a minimum, should be two people
- Train the operators in the data collection procedure, including the operational definition of the metric to be measured
- Select the items to be measured. At least 20 items (more is better) that represent different types of failures and gray areas. There should be roughly an equal percentage of pass versus fail items. The pass items are those that meet the quality characteristic, or the "good" items, and the fail are those that do not meet the quality characteristic. There should be a minimum of 10 items.
- An expert will measure each item once to develop the known standard information.
- Select the number of trials, there should be a minimum of two trials. Randomize the measurement order of the items across each trial.
- Have the operators measure the items in the random run order, for the total number of trials.
- Record the results of the measurement analysis.
- Assess the accuracy of the results, usually using a statistical analysis software, such as Minitab or SPSS.

- An acceptable "Accuracy" score is 80% or higher. This measures how the associates are measuring against the known standard. If the accuracy is above 80%, you may use the measurement system. If the accuracy is less than 80%, the measurement system should be improved.

2.4 Pressure Ulcer Background

Pressure ulcers are defined by the Centers for Medicare and Medicaid Services (CMS) as "A pressure ulcer is a localized injury to the skin and/or underlying tissue usually over a bony prominence, as a result of pressure, or pressure in combination with shear and/or friction." (CMS Gov A). The injury to the skin can be attributed to a patient laying on a surface for an extended period of time, without moving. When patients are immobile, the healthcare providers must turn the patient frequently, to avoid pressure ulcers from forming. The Centers for Medicare and Medicaid Services (CMS) through their payment provisions for preventable hospitable acquired conditions (HACs) are one of the recent "value-based purchasing" initiatives, with the goal of tying pay to performance in the healthcare setting (CMS Gov B). Pressure ulcers are assessed based on their stages, as shown in Table 1 (www.npuap.org, 2017)

Table 1. Pressure Ulcer Stages

Stage	Description
I	Non-blanchable erythema of intact skin
II	Partial-thickness skin loss with exposed dermis
III	Full-thickness skin loss
IV	Full-thickness skin and tissue loss
Un-stageable	Obscured full-thickness skin and tissue loss
Deep Tissue Pressure injury	Persistent non-blanchable deep red, maroon or purple discoloration

This Measurement Systems Analysis study was used to assess the accuracy of the diagnosis of pressure ulcers when patients are admitted to the hospital, either through the emergency department or directly through inpatient admitting. Specially trained nurses assess patients when they are admitted to the hospital. If a patient acquires a pressure ulcer in the hospital, the Center for Medicare/Medicaid (CMS) will typically not pay for the care related to hospital acquired pressure ulcers. It is critical that the hospitals properly assess whether pressure ulcers were present when the patient was admitted. The process analyst discovered that the specially trained nurses did not initially consistently diagnose that a pressure ulcer was present upon admission, nor consistently assess the severity of the pressure ulcers that were present. The attribute gage R&R /Attribute Agreement Analysis study identified the lack of diagnosis consistency, and allowed for additional training, that then improved the diagnosis of whether a pressure ulcer was present upon admission to the hospital. There still remained the need for additional improvement in assessing the pressure ulcer severity level after the study was complete. This study is extremely important to 1) identify a method for healthcare providers to assess and improve the measurement system related to human diagnoses in healthcare processes; and 2) to demonstrate the usefulness of expanding gage R&R and attribute agreement analysis to human diagnosis in healthcare settings.

3. Study Methodology

3.1 Motivation for and Value of the Study

The purpose of the Attribute Agreement Analysis study was to assess the accuracy of the pressure ulcer assessment measurement system within a community-based, acute care hospital. The hospital wanted to ensure that the patients being admitted to the hospital either through the Emergency Department or as an inpatient do not have pre-existing pressure ulcers. At the time of this study, CMS was starting to not reimburse hospitals with respect to certain conditions acquired while the patient is within their care, such as when a patient acquires a pressure ulcer while in the hospital. Specifically we ask two questions:

- Phase I: Do the Specially Trained Nurses (STNs) agree on what is and what isn't a pressure ulcer?
- Phase II: Do the STNs stage a pressure ulcer in the same way?

The value of this study is to apply Measurement Systems Analysis, in particular Attribute Agreement Analysis and Attribute Gage R&R studies to healthcare to improve human diagnosis of healthcare conditions. Other healthcare studies found in the literature have not yet applied MSA to attribute human diagnosis, although a dental study comparing X-rays was found (Arani and Erdil, 2017)

3.2 Current Measurement System

The current measurement system included specially trained nurses (STNs) in the identification of pressure ulcers that would assess the patients when they arrived at the Emergency Department, or when the patient was admitted to the inpatient floor. The nurses performed a two-phase assessment, as appropriate. The first phase assessed whether the

patient had any pressure ulcers, answering yes (one or more pressure ulcers were present), or no (no evidence of pressure ulcers on the patient). The second phase assessed the severity level of any identified pressure ulcers (stages I, II, III, or IV). The stages from stage I to IV represent increasing medical severity. Stage I sores are not open wounds, the skin has no breaks or tears. In Stage II sores the skin breaks away, some of the skin may be damaged beyond repair or die. In Stage III sores the sore gets worse and extends into the tissue beneath the skin forming a small crater, fat may show in the sore, but not muscle, tendon or bone. In Stage IV, the pressure ulcer is very deep, reaching into muscle and bone, damage to deeper tissues, tendons, and bones may occur (www.webmd.com, 2017)

3.3 Attribute Agreement Analysis Study Protocol

There were 21 STNs that assessed 25 pictures of skin injuries. In phase I, the STNs assessed whether the picture represented a pressure ulcer. In phase II, for the pictures where the STNs identified that it was a pressure ulcer, they then assessed the pressure ulcer stage from I to IV. The STNs repeated the assessment a second time, for each of the 25 pictures of skin injuries, resulting in two assessments being collected for each of the 25 pictures for each STN assessor. The photos were presented in a random order for each trial, as shown in Table 2.

Table 2. Pressure Ulcer Presentation Order per Trial

Trial 1 Sequence Number	**Photo Letter**	**Trial 2 Sequence Number**	**Photo Letter**
1	A	1	E
2	B	2	J
3	C	3	G
4	D	4	F
5	E	5	T
6	F	6	Y
7	G	7	C
8	H	8	R
9	I	9	P
10	J	10	D
11	K	11	L
12	L	12	S
13	M	13	V
14	N	14	N
15	O	15	A
16	P	16	W
17	Q	17	I
18	R	18	K
19	S	19	X
20	T	20	O
21	U	21	Q
22	V	22	U
23	W	23	M
24	X	24	B
25	Y	25	H

Each picture was assessed by pressure ulcer knowledge experts to provide a “gold standard” for each picture, providing whether the picture was of a pressure ulcer and the stage of the pressure ulcer.

4. Study Results

4.1 Phase I, Pass/Fail-Identification of Whether a Skin Injury Is or Is not a Pressure Ulcer

The Phase I study investigated whether the STNs would assess 25 pictures of skin injuries accurately as to whether they were a pressure ulcer or not, and within themselves (repeatability), across other STNs (reproducibility), and compared to the standard (accuracy).

4.1.1 Repeatability and Reproducibility Results

Table 3 provides the assessors’ within themselves (repeatability) summarized data for Phase I. It shows the number inspected and matched, the percent matched and the 95% confidence interval for each assessor.

Table 3. Phase I Summary of Assessor's Data for Pressure Ulcer Pass/Fail within Themselves

Assessor	Number Inspected	Number Matched	Percent Matched	95% Confidence Interval
A	25	19	76%	(54.87, 90.64%)
B	25	24	96%	(79.65, 99.90%)
C	25	25	100%	(88.71, 100.00%)
D	25	25	100%	(88.71, 100.00%)
E	25	24	96%	(79.65, 99.90%)
F	25	22	88%	(68.78, 97.45%)
G	25	24	96%	(79.65, 99.90%)
H	25	25	100%	(88.71, 100.00%)
I	25	25	100%	(88.71, 100.00%)
J	25	24	96%	(79.65, 99.90%)
K	25	25	100%	(88.71, 100.00%)
L	25	23	92%	(73.97, 99.02%)
M	25	22	88%	(68.78, 97.45%)
N	25	19	76%	(54.87, 90.64%)
O	25	24	96%	(79.65, 99.90%)
P	25	25	100%	(88.71, 100.00%)
Q	25	25	100%	(88.71, 100.00%)
R	25	23	92%	(73.97, 99.02%)
S	25	22	88%	(68.78, 97.45%)
T	25	23	92%	(73.97, 99.02%)
U	25	25	100%	(88.71, 100.00%)
Average =====➔			94%	

For the first assessment of whether a pressure ulcer was present, at a 95% confidence level, the STN A agreed 54.9% to 90.6% of the time within themselves. That means that the STN A identified correctly whether a pressure ulcer was present across both of their own assessments, between 54.9% to 90.6% of the time, with an average of 76% agreement within themselves.

The overall average Within Assessor Repeatability Accuracy of the STNs was 94%. This demonstrates that the STNs were quite consistent in assessing the skin injuries across the two trials within themselves. The Within Assessor Reproducibility Accuracy was 40% (10/25), with a 95% confidence interval of 21 to 61%. Both the repeatability and reproducibility should be greater than 80%. The repeatability is good, but the reproducibility (assessment across appraisers) was not very good.

The within STN agreement ratings graph is shown in Figure 1.

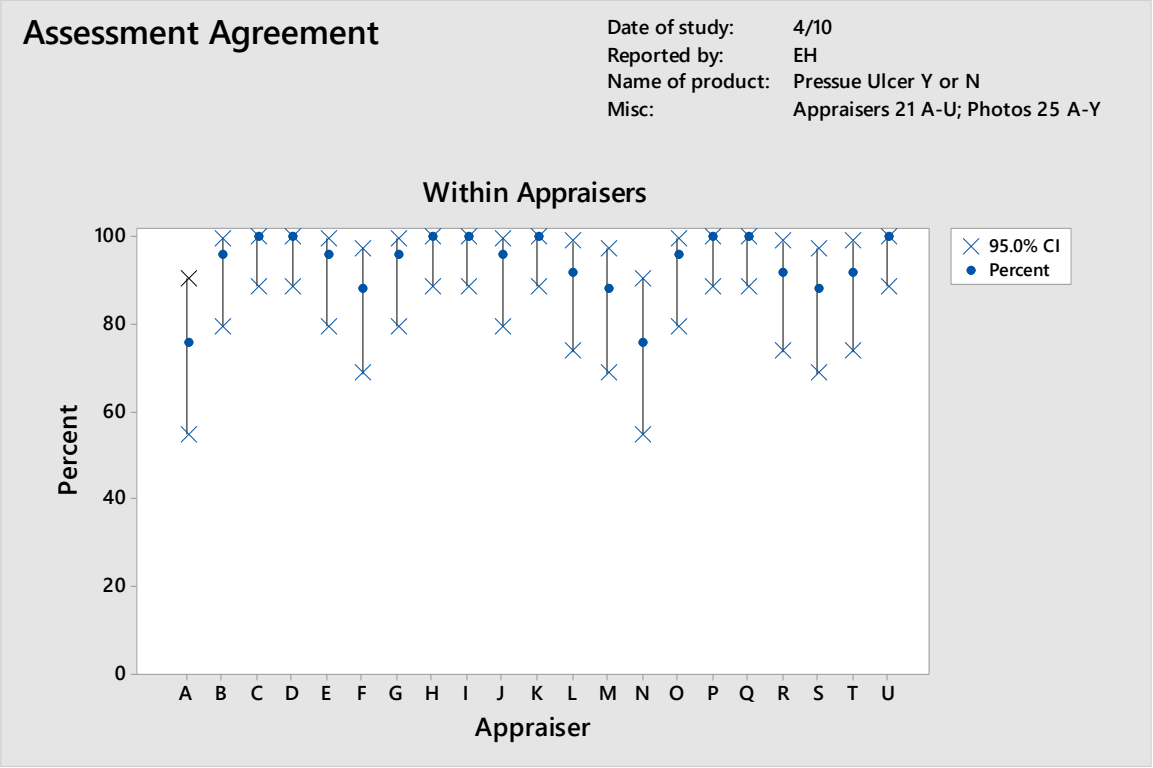

Figure 1. Phase 1 - Within Assessor Agreement of Pass/Fail Pressure Ulcers

4.1.2 Accuracy, Comparing to a Standard

Table 4 shows the summary of the assessor's data for the pressure ulcer pass or fail compared to the standard. The average *Across Assessor Accuracy to Standard* was 92%.

Table 4. Phase I Summary of Assessor's Data for Pressure Ulcer Pass/Fail to Standard

Assessor	Number Inspected	Number Matched	Percent Matched	95% Confidence Interval
A	25	19	76%	(54.87, 90.64%)
B	25	23	92%	(73.97, 99.02%)
C	25	25	100%	(88.71, 100.00%)
D	25	25	100%	(88.71, 100.00%)
E	25	24	96%	(79.65, 99.90%)
F	25	21	84%	(63.92%, 95.46%)
G	25	23	92%	(73.97%, 99.02%)
H	25	22	88%	(88.71, 100.00%)
I	25	22	88%	(88.71, 100.00%)
J	25	24	96%	(79.65, 99.90%)
K	25	25	100%	(88.71, 100.00%)
L	25	23	92%	(73.97, 99.02%)
M	25	21	84%	(68.78, 97.45%)
N	25	18	72%	(54.87, 90.64%)
O	25	24	96%	(79.65, 99.90%)
P	25	25	100%	(88.71, 100.00%)
Q	25	25	100%	(88.71, 100.00%)
R	25	23	92%	(73.97, 99.02%)
S	25	22	88%	(68.78, 97.45%)
T	25	22	88%	(73.97, 99.02%)
U	25	25	100%	(88.71, 100.00%)
Average =====➔			92%	

Although the accuracy of each assessor within themselves and to the standard was high (94% and 92% respectively), the across assessors' assessment and assessor's assessment to the standard was low. The average *Overall Accuracy to Standard* across the STNs was 40%, (10 out of 25) with a 95% confidence interval from 21 to 61%. That is, if we were to show the STNs other pressure ulcers, we are 95% confident that they would agree on whether they are a pressure ulcer anywhere from 21 to 61 percent of the time. The appraiser versus standard results are shown in Figure 2. The accuracy is not favorable, as it should be 80% or higher, resulting in the need for improving the measurement system related to the assessment of whether a skin injury is a pressure ulcer or not.

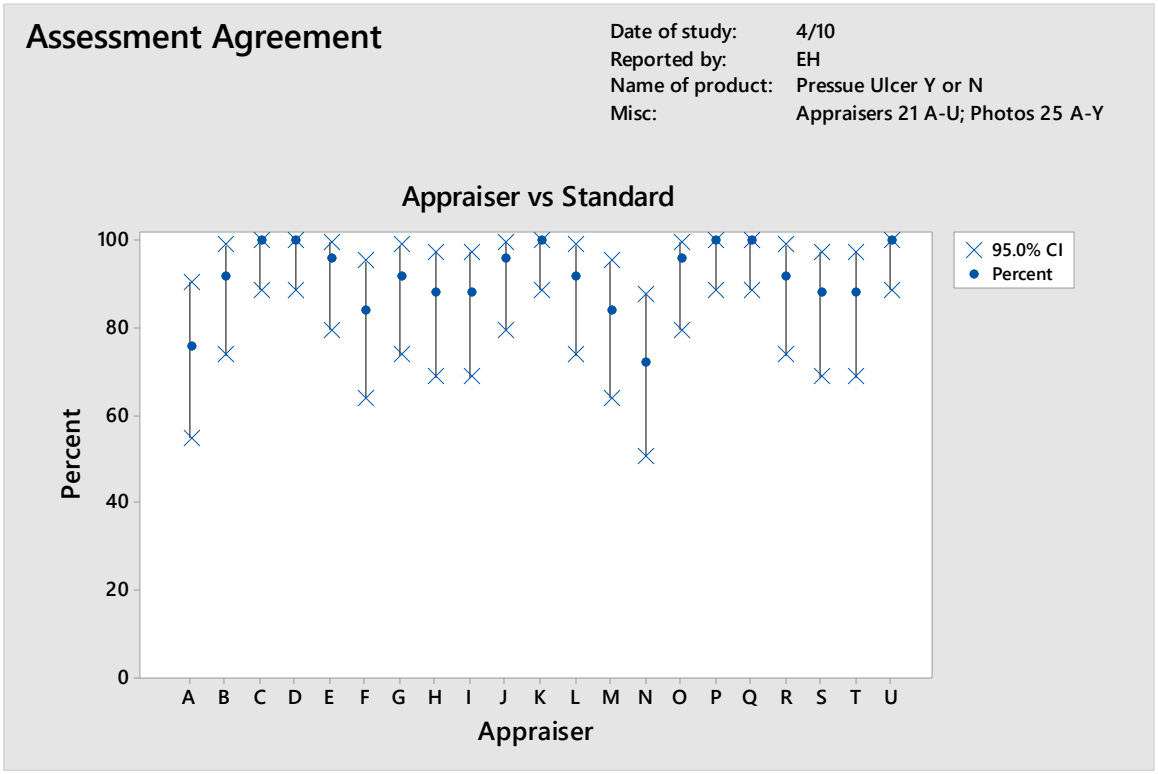

Figure 2. Phase 1 – Assessor vs. Standard of Pass/Fail Pressure Ulcers

A summary of the Attribute Agreement Analysis Metrics is shown in Table 5.

Table 5. Attribute Agreement Analysis Metrics

Metric	Results
Within Assessor Repeatability Accuracy	94%
Within Assessor Reproducibility Accuracy	40%
Across Assessor Accuracy to Standard	92%
Overall Accuracy to Standard	40%

4.2 Phase II, Staging of Pressure Ulcers

The next study assessed the accuracy of the measurement system related to the staging into four stages of the pressure ulcers by the STNs. The same 25 pictures of skin injuries were assessed twice by each of the 21 STNs. Figure 3 shows the STNs' within themselves staging results. Even though the stages were attribute categories, a Gage R&R ANOVA study was performed in Minitab® statistical software.

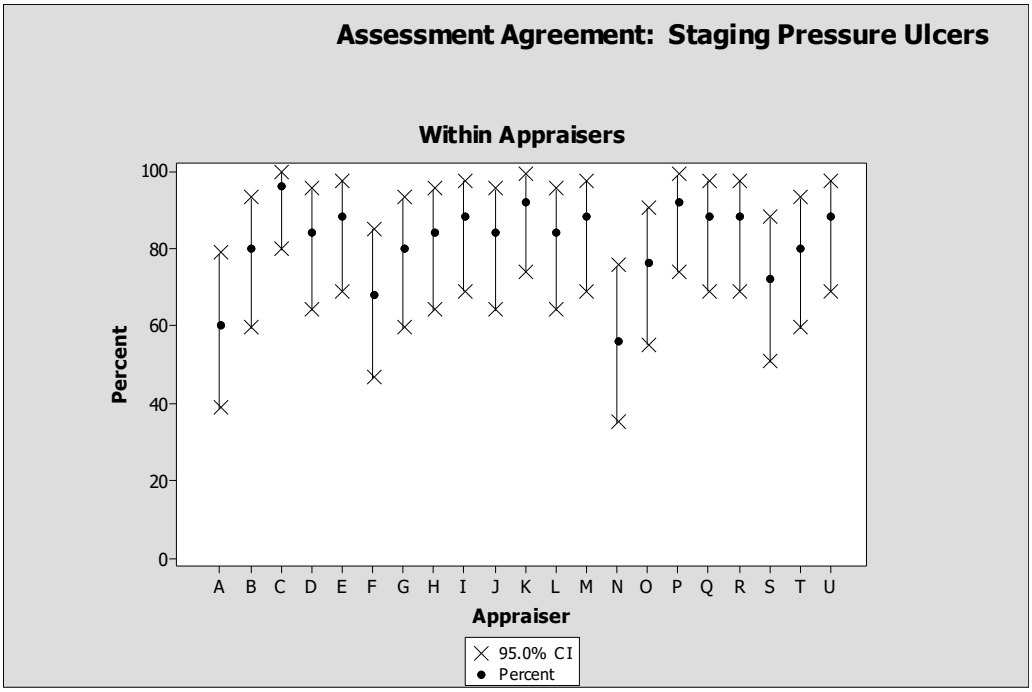

Figure 3. Phase II STNs' Within Themselves Staging Results

Figure 3 shows the 95% confidence interval of agreement for each STN. The way it can be interpreted is that "We are 95% confident that STN A will agree with his/her pressure ulcer staging anywhere from 38 to 78% of the time". The accuracy within the STNs themselves is lower when trying to stage the pressure ulcer in 4 stages versus the two categories (yes it is a pressure ulcer, or no it is not), resulting in a *Within Assessor Repeatability Accuracy* of 82% versus 94% in the phase I study. The confidence range across all of the assessors' agreement ratings was (38.67 to 78.87%).

The staging results are shown in table 6, listing the number inspected and matched, the percent matched and the 95% confidence interval for each assessor.

Table 6. Phase II Summary of Assessor's Data for Pressure Ulcer Staging within Themselves (*Within Assessor Repeatability Accuracy*)

Assessor	# Inspected	# Matched	Percent Matched	95% Confidence Interval
A	25	15	60%	(38.67, 78.87)
B	25	20	80%	(59.30, 93.17)
C	25	24	96%	(79.65, 99.90)
D	25	21	84%	(63.92, 95.46)
E	25	22	88%	(68.78, 97.45)
F	25	17	68%	(46.50, 85.05)
G	25	20	80%	(59.30, 93.17)
H	25	21	84%	(63.92, 95.46)
I	25	22	88%	(68.78, 97.45)
J	25	21	84%	(63.92, 95.46)
K	25	23	92%	(73.97, 99.02)
L	25	21	84%	(63.92, 95.46)
M	25	22	88%	(68.78, 97.45)
N	25	14	56%	(34.93, 75.60)
O	25	19	76%	(54.87, 90.64)
P	25	23	92%	(73.97, 99.02)
Q	25	22	88%	(68.78, 97.45)
R	25	22	88%	(68.78, 97.45)
S	25	18	72%	(50.61, 87.93)
T	25	20	80%	(59.30, 93.17)
U	25	22	88%	(68.78, 97.45)
Average =====➔			82%	(38.67, 78.87)

Figure 4 shows how each STN staged all of the photos. The box or bar for each STN contains 75% of the data responses, while the circle is the mean. This figure attempts to compare whether there is a trend of an assessor assessing the photos higher or lower than the other STNs.

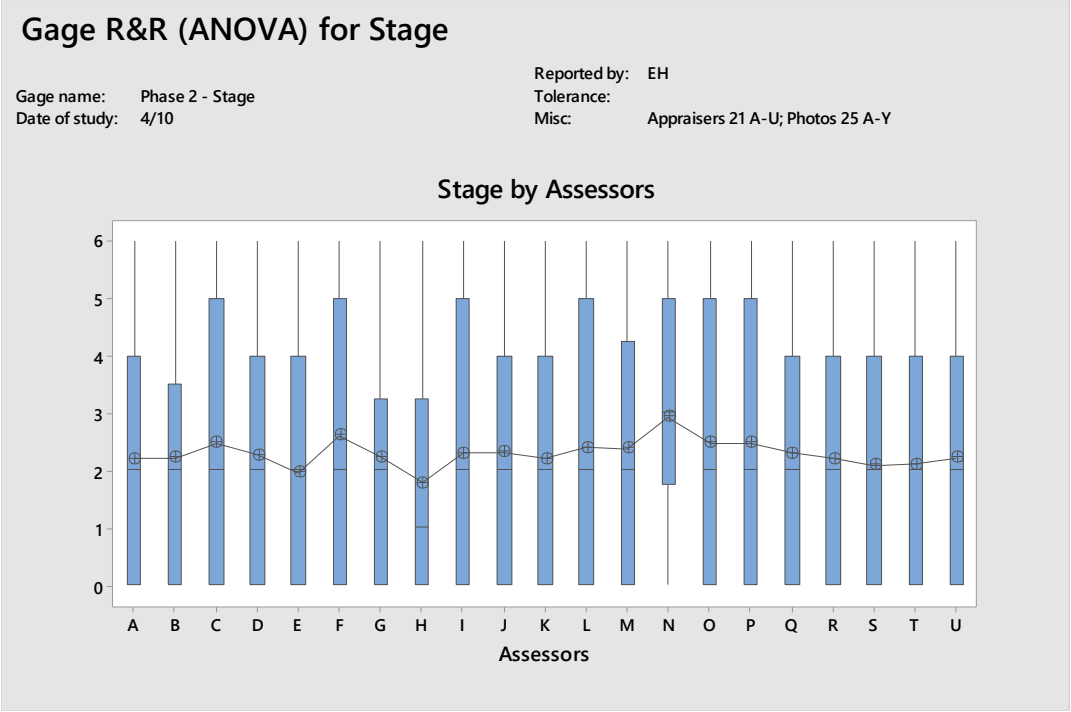

Figure 4. Pressure Ulcer Staging Ratings by Assessors

Figure 5 demonstrates that there were very few photos that everyone staged the same. The box contains 75% of all of the data for each photo. The circle is the mean, and the asterisks are outliers. If the picture was rated equally by mostly everyone, you will not see a box, you would only see a line with the circle, representing the average. This occurred only for two photos, L and R. Out of the 25 pressure ulcers, 2 of them were agreed on by all STNs on the stage from I to IV. The *Overall Accuracy to Standard* was 8%, and the 95% confidence interval is from 1 to 26%. Meaning that, if the STNs were to stage other pressure ulcers, we are 95% confident that they would stage the pressure ulcer equally anywhere from 1 to 26 percent of the time. Which clearly is extremely low from an accuracy and consistency perspective.

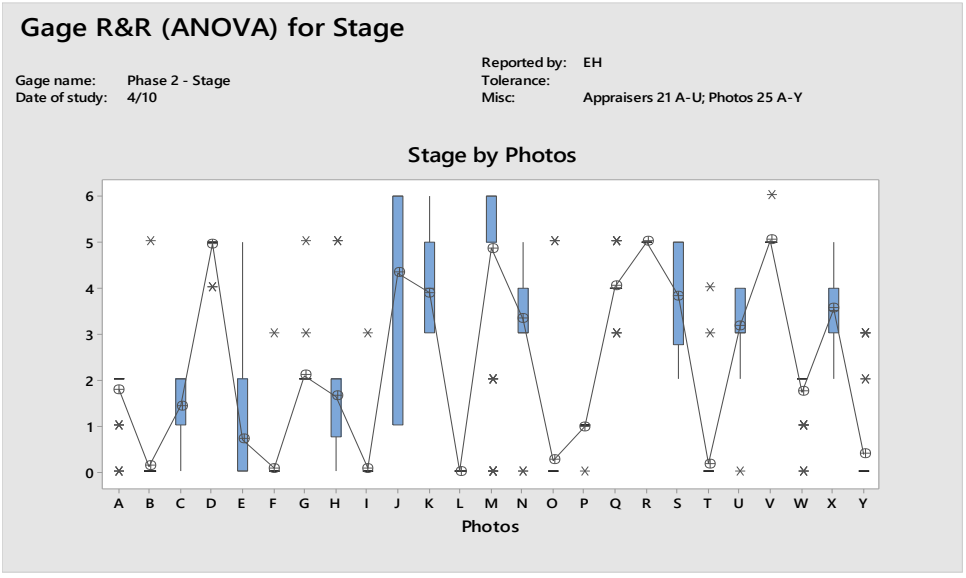

Figure 5. Pressure Ulcer Staging Ratings by Photo

Figure 6 shows the source of variation and percent of contribution for each source. Ideally you want most of your variation in part to part (because it is normal for pressure ulcers to be different stages) and minimal variation in how the parts are measured – repeatability and reproducibility. The percent of variation due to the part to part was 78.86%. The number of Distinct Categories that the measurement system was capable of distinguishing is two. This makes sense, since the accuracy was higher for distinguishing whether a pressure ulcer was present or not, but not in rating the stages from I to IV.

Variance Components

Source	VarComp	%Contribution (of VarComp)
Total Gage R&R	0.90890	21.14
Repeatability	0.60762	14.13
Reproducibility	0.30129	7.01
Operators	0.03256	0.76
Operators*Parts	0.26872	6.25
Part-To-Part	3.39013	78.86
Total Variation	4.29903	100.00

Figure 6. Source of Variation and % Contribution

According to the AIAG (AIAG, 2010), if the variation in the measurement system is less than 10% of the process variation, then it is acceptable. Table 7 shows the decision criteria for the process variation and percent contribution compared to the measurement system variation. For our results the percent contribution of the Total Gage R&R is 21.14%, which is not acceptable (greater than 9%). The percent study variation for Total Gage R&R is 45.98%, resulting is an unacceptable measurement system for assessing pressure ulcer staging. Figure 7 shows the Gage Evaluation for the percent study variation.

Table 7. Decision Criteria and Variation Resultss

Decision	**Percentage of Process Variation (% Study Variation)**	**% Contribution**
Acceptable	< 10%	< 1%
Acceptable depending upon the application, cost of fixing the measurement system, and risk	10% < x < 30%	1% < x < 9%
Not acceptable and should be improved	> 30%	> 9%
Pressure Ulcer Gage R&R Study Results	45.98% (should be improved)	21.14% (should be improved)

Gage Evaluation

Source	StdDev (SD)	Study Var (6 × SD)	%Study Var (%SV)	%Tolerance (SV/Toler)
Total Gage R&R	0.95336	5.7202	45.98	381.35
Repeatability	0.77950	4.6770	37.60	311.80
Reproducibility	0.54889	3.2934	26.47	219.56
Operators	0.18045	1.0827	8.70	72.18
Operators*Parts	0.51839	3.1103	25.00	207.35
Part-To-Part	1.84123	11.0474	88.80	736.49
Total Variation	2.07341	12.4405	100.00	829.36

Figure 7. Gage Evaluation - % Study Variation

5. Conclusions and Limitations

Based on the results of the measurement system analysis, the process analysts and the STNs agreed that the measurement system for assessing pressure ulcers needed improvement. For the Pressure Ulcers, yes or no study, the accuracy of the assessors to the standards was 40%, far below the desired 80% accuracy level that is typical for an attribute agreement analysis study. For the pressure ulcer staging the percent of process variation or percent of study variance was almost 46%, far above the desired 30% variance rejection level. The percent contribution for the pressure ulcer staging was also 21%, far above the 9% rejection percent level. A collaborative discussion was held with the STNs and the nurse experts that defined the standards to understand what prevented the STNs from assessing the pressure ulcer photos correctly. They agreed that having only pictures and not being able to actually see, touch or smell the pressure ulcer on an actual patient was limiting their ability to assess accurately. The hospital implemented additional training for the STNs, and further measurement system assessment would be done again in the future. The hospital was committed to continually improving the measurement system.

6. Future Research

This area of applying attribute agreement analysis and gage R&R to patient diagnosis by the healthcare professionals is relatively new. There can be a myriad of applications in healthcare to apply this methodology and learn what can improve assessment of patients' conditions by healthcare providers. Additionally, research could be done to assess whether the rejection levels for the Automotive Industry Action Group (AIAG) standards for measurement system accuracy are appropriate for the healthcare industry, where the human element is critical for appropriate patient diagnosis.

7. Author Bios

Sandy Furterer is an Associate Professor at the University of Dayton, in the Department of Engineering Management, Systems and Technology. Dr. Furterer received her Ph.D. in Industrial Engineering with a specialization in Quality Engineering from the University of Central Florida in 2004. She received an MBA from Xavier University, and a Bachelor and Master of Science in Industrial and Systems Engineering from The Ohio State University. Dr. Furterer has over 25 years of experience in business process and quality improvements. Dr. Furterer is an author or co-author of 4 reference textbooks on Lean Six Sigma, Design for Six Sigma and Lean Systems, including her latest book: Lean Six Sigma Case Studies in the Healthcare Enterprise by Springer publishing in 2014. Dr. Furterer is also an author or co-author of 12 refereed journal articles, and 35 refereed conference proceedings publications on process modeling, lean six sigma, process improvement and engineering education. She has over 100 conference presentations on operational excellence, process improvement, lean six sigma, business process architecture, and engineering education. She is an ASQ Certified Six Sigma Black Belt, an ASQ Certified Quality Engineer, an ASQ fellow, and a Harrington Institute certified Master Black Belt. She is a member of the ASEE, ASQ, IISE, and SWE.

Ethling Hernandez is the Director of Project Management at Mount Sinai Medical Center, in Miami Beach, Florida. Previously she was an Operational Performance Coordinator in the Enterprise Performance Excellence Center at Holy Cross Hospital, in Ft. Lauderdale, Florida. She was a Manufacturing Industrial Engineer for Beckman Coulter and a Sr. Product Supervisor at Boston Scientific. Ms. Hernandez has a Bachelor of Science in Industrial Engineering and a Masters in Engineering Management from the University of Central Florida. She is a certified Professional Engineer, and a Six Sigma Black Belt. She is fluent in English and Spanish.

References

Akkerhuis, T. S., Niemeijer, G. C. H., Trip, A., Gemke, R. J. B. J., & Does, R. J. M. M. (2015). Quality Quandaries: Precision and Accuracy of Ear Thermometry. *Quality Engineering*, *27*(4), 512. https://doi-org.libproxy.udayton.edu/10.1080/08982112.2015.1065324

Anonymous, No Date, "What are the Stages of Pressure Sores", WebMD, https://www.webmd.com/skin-problems-and-treatments/pressure-sores-4-stages#1, Accessed 5/29/2019.

Anonymous, No date, Quality-Initiatives-Patient-Assessment-Instruments https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/index.html

Arani, O. M., & Erdil, N. A. (2017). Measurement System Analysis in Healthcare: Attribute Data. Proceedings of the 2017 Industrial and Systems Engineering Conference, Pittsburgh, PA.

Ayello, E. A. (2014) Pressure Ulcer Staging, CMS Gov A, https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/IRF-Quality-Reporting/Downloads/IRF-QRP-Training-%E2%80%93-PrU-Staging-May-12-2014-.pdf, Presented at the IRF Provider Training Baltimore, Maryland May 12 , 2014.

Brown, T. W., McCarthy, M. L., Kelen, G. D., & Levy, F. (n.d.). An Epidemiologic Study of Closed Emergency Department Malpractice Claims in a National Database of Physician Malpractice Insurers. *ACADEMIC EMERGENCY MEDICINE*, *17*(5), 553-560. https://doi-org.libproxy.udayton.edu/10.1111/j.1553-2712.2010.00729.x

Dalalah, D., & Diabat, A. (n.d.). Repeatability and reproducibility in med labs: a procedure to measurement system analysis. *IET SCIENCE MEASUREMENT & TECHNOLOGY*, *9*(7), 826-835. https://doi-org.libproxy.udayton.edu/10.1049/iet-smt.2014.0363

Eperjesi, F., Aujla, M., & Bartlett, H. (n.d.). Reproducibility and repeatability of the OcuSense TearLab (TM) osmometer. *GRAEFES ARCHIVE FOR CLINICAL AND EXPERIMENTAL OPHTHALMOLOGY*, *250*(8), 1201-1205. https://doi-org.libproxy.udayton.edu/10.1007/s00417-012-1961-4

Erdmann, T., Does, R. M. M., & Bisgaard, S. (2010). Quality Quandaries*: A Gage R&R Study in a Hospital. *Quality Engineering*, *22*(1), 46-53. https://doi-org.libproxy.udayton.edu/10.1080/08982110903412924

Hoffmann, E., Medeiros, F., Kramann, C., Pfeiffer, N., & Grus, F. (n.d.). Repeatability and reproducibility of optic nerve head topography using the retinal thickness analyzer. GRAEFES ARCHIVE FOR CLINICAL AND *EXPERIMENTAL OPHTHALMOLOGY*, *244*(2), 192-198. https://doi-org.libproxy.udayton.edu/10.1007/s00417-005-0027-2

McAlinden, C., Khadka, J., & Pesudovs, K. (2011). Statistical methods for conducting agreement (comparison of clinical tests) and precision (repeatability or reproducibility) studies in optometry and ophthalmology. *Ophthalmic & Physiological Optics: The Journal of the British College of Ophthalmic Opticians (Optometrists)*, *31*(4), 330-338. https://doi-org.libproxy.udayton.edu/10.1111/j.1475-1313.2011.00851.x

Measurement Systems Analysis Reference Manual, 4th edition. Automotive Industry Action Group (AIAG) (2010). Chrysler, Ford, General Motors Supplier Quality Requirements Task Force.

Pawlicki, T., & Perry, L. (n.d.). Using Gauge R&R for Analyzing the Process of Daily IGRT. *MEDICAL PHYSICS*, *38*(6). https://doi-org.libproxy.udayton.edu/10.1118/1.3612016

Wang, Q., Ding, X., Savini, G., Chen, H., Feng, Y., Pan, C., & Huang, J. (2015). Anterior chamber depth measurements using Scheimpflug imaging and optical coherence tomography: repeatability, reproducibility, and agreement. *Journal of Cataract And Refractive Surgery*, *41*(1), 178-185. https://doi-org.libproxy.udayton.edu/10.1016/j.jcrs.2014.04.038

Copyrights