CERIAS Tech Report 2010-29

Evaluation of the Indiana Department of Correction Mug Shot Capture Process

by Gregory Hales
Center for Education and Research
Information Assurance and Security
Purdue University, West Lafayette, IN 47907-2086

Graduate School ETD Form 9
(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Greg T. Hales

Entitled
Evaluation of the Indiana Department of Corrections Mug Shot Capture Process

For the degree of Master of Science

Is approved by the final examining committee:

Dr. Stephen Elliott Chair	Dr. Stephen Elliott
Dr. Mathias Sutton	Dr. Mathias Sutton
Patrick Grother	Patrick Grother

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): Dr. Stephen Elliott

Approved by: Dr. Gary Bertoline (Head of the Graduate Program) 12/02/2010 (Date)

Graduate School Form 20
(Revised 9/10)

PURDUE UNIVERSITY
GRADUATE SCHOOL

Research Integrity and Copyright Disclaimer

Title of Thesis/Dissertation:

Evaluation of the Indiana Department of Corrections Mug Shot Capture Process

For the degree of Master of Science

I certify that in the preparation of this thesis, I have observed the provisions of *Purdue University Executive Memorandum No. C-22,* September 6, 1991, *Policy on Integrity in Research.**

Further, I certify that this work is free of plagiarism and all materials appearing in this thesis/dissertation have been properly quoted and attributed.

I certify that all copyrighted material incorporated into this thesis/dissertation is in compliance with the United States' copyright law and that I have received written permission from the copyright owners for my use of their work, which is beyond the scope of the law. I agree to indemnify and save harmless Purdue University from any and all claims that may be asserted or that may arise from any copyright violation.

Greg T. Hales

Printed Name and Signature of Candidate

12/02/2010

Date (month/day/year)

*Located at http://www.purdue.edu/policies/pages/teach_res_outreach/c_22.html

EVALUATION OF THE INDIANA DEPARTMENT OF CORRECTION MUG SHOT CAPTURE PROCESS

A Thesis

Submitted to the Faculty

of

Purdue University

by

Gregory T. Hales

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

December 2010

Purdue University

West Lafayette, Indiana

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

ABSTRACT

Hales, Gregory T. M.S., Purdue University, December 2010. Evaluation of the Indiana Department of Correction Mug Shot Capture Process. Major Professor: Stephen Elliott.

The tracking and monitoring of fugitives and persons of interest is of significant concern for the Indiana Department of Corrections (IDOC) Fugitive Detection Unit. The research conducted was to help determine the benefits of implementing a face recognition technology solution. Images were analyzed for standard compliance to help determine their suitability for input into a face recognition matcher. Results from this analysis showed the images were not in compliance with the NIST Mug Shot Best Practices, nor could the software optimize the images to make them compliant. A visit to the intake facility indicated that the process by which these mug shots were collected needed to be addressed before face recognition technology could be implemented. Consequently, the IDOC main prisoner intake facility's current mug shot image capture process was assessed. Using the analysis from the images, along with observations from the mug shot capture process, an optimized capture process was implemented for a trial period of two weeks to determine its effectiveness. Results show that the capture process improved the standard compliance of the mug shot images, determining that the images collected would be usable with face recognition technology. Another finding was that the centerline location ratio variable, which has a precise threshold, was not compliant for any images in either dataset leading to the need for further study to determine if this variable should utilize a range of values for an operational environment such as at the IDOC.

CHAPTER 1. INTRODUCTION

Recognition of a person's face is the most natural and common form of human identification used in day-to-day life, though it has limitations. Its use (prior to being an identifier) in the law enforcement environment can be seen as early as the 1850s by Pinkerton's National Detective Agency (Yeatts, 2001). Mug shot images have been taken since 1840 and are still used by law enforcement officers and eye witnesses to identify suspects. The current method relies on humans to verify an identity based on how closely the person in custody resembles the mug shot photo.

Some of the procedures used in eyewitness identification include the sequential double-blind procedure, blind sequential procedure, and the sequential procedure, which were all approved to establish unbiased eyewitness instruction in 1999 by the U.S. Department of Justice (Klobuchar, et. al, 2006). These methods are not infallible and could lead to human errors. Adler and Schuckers (2007) compared the results of humans versus machines for automatic face recognition. Over a period of eight years, from 1999-2006 they were able to see the development of the automatic face recognition algorithms over time. The results show that, in 2006, the best automatic face recognition system performed better than 37.5% of human subjects while 29.2% of human subjects performed better than the automated system (Adler & Schuckers, 2007).

The quality of a mug shot image is also an issue; if the image quality of a mug shot is poor, performance of the human or automated system is also likely to be poor.

Conversely, if the image quality is good, performance will be good. Shah and Martin (2007) conducted a workshop to explore this relationship and found that of all biometric modalities, face had the largest correlation, 0.75, between image quality and system performance.

1.1. Statement of the Problem

The Indiana Department of Correction's (IDOC) does not follow the NIST Best Practices Recommendations for the capture of facial images in their facilities, resulting in images that do not conform to the standard.

1.2. Significance of the Problem

The Fugitive Detection Unit, of the Indiana Department of Correction uses mug shot images to identify "persons of interest" and fugitives. However, IDOC would like to expand their capabilities and automate various processes using face recognition.

The Indiana DOC facilities currently have no standardized method for capturing mug shot images, so different IDOC facilities provide images that have different backgrounds, lighting, and angle/position of the head. Since the process is not standardized, the resulting images would be inconsistent in quality (many could be of poor quality) and may not perform optimally with current (2010) face recognition technology. Assessing current practices and developing a document to make improvements to the setup and process will provide the IDOC a guide to use in their facilities to capture good quality, standard-compliant images that will be usable with face recognition technology (FRT).

Figure 1.1 Sample IDOC Mug Shot Images

1.3. Statement of Purpose

The research investigates whether the mug shot photos are standard compliant, and if not, implement an optimized method so they will be. The project will use a standard compliant process based on NIST Mug Shot Best Practices Version 2.0 and was conducted using the following steps:

a) Collect data from the IDOC
    a. Calculate baseline quality and standard compliance information
    b. Distinguish problematic variables of images
    c. Randomly select 1,000 images to visually inspect
b) Observe the current capture process at the IDOC facility
c) Determine if problematic variables in images correlate to capture process observed

d) Write a process improvement document for the IDOC facility

e) Implement the recommendations in the improvement document

f) Collect images captured using improvement document

g) Determine if improvement document result in a significant difference in the amount of images standard-compliant

1.4. Definitions of Terms

Background Type - "indicates the type of background the image has" (PreFace SDK Manual, 2009, p.31).

Biometric - is "a measurable, physical characteristic or personal behavioral trait used to recognize the identity, or verify the claimed identity, of an enrollee" (Association for Biometrics, 1999 p.2).

Brightness Score - "how well the useful dynamic range of the facial region of the image is centered within the full dynamic range of the image" (PreFace SDK Manual, 2009, p.32).

Centerline Location Ratio – "is the location of the centerline as a fraction of the image width measured from the left side of the image" (PreFace SDK Manual, 2009, p.33).

Covert - Unawareness that biometric data is being measured and collected (Mansfield & Wayman, 2002).

Degree of Blur	- "indicates how much focus and/or motion blur is present in the image" (PreFace SDK Manual, 2009, p.34).
Degree of Clutter	- "indicates how much background clutter occurs in the image" (PreFace SDK Manual, 2009, p.32).
Eye Axis Angle	- "Eye Axis Angle is the slope of the eye-axis measured in degrees clockwise (positive) from the horizontal. This metric is roughly proportional to the "roll" angle specified in many standards" (PreFace SDK Manual, 2009, p.33).
Eye Axis Location Ratio	- "is the location of the eye axis as a fraction of the image height up from the bottom" (PreFace SDK Manual, 2009, p.33).
Eye Contrast	- "indicates how well the dynamic range is spread in the eye regions of the image" (PreFace SDK Manual, 2009, p.32).
Eye Separation	- "is the number of pixels between the left and right eye centers" (PreFace SDK Manual, 2009, p.34).
Facial Image	- Electronically stored portraits of an individual's face in accordance with the INCITS M1 data interchange face recognition format.

Facial Dynamic Range	- "indicates the number of bits in the dynamic range of the facial region of the input image" (PreFace SDK Manual, 2009, p.32).
Facial Recognition	- is "a physical biometric that analyzes facial features" (Association for Biometrics, 1999 p.9).
Failure to Acquire	- is "the expected proportion of transactions for which the system is unable to capture or locate an image or signal of sufficient quality" (Mansfield & Wayman, 2002).
Failure to Enroll	- is "the expected proportion of the population for whom the system is unable to generate repeatable templates" (Mansfield & Wayman, 2002).
False Match Rate	- is "the expected probability that a sample will be falsely declared to match a single randomly-selected "non-self" template. FMR is also referred to as a "false positive"" (Mansfield & Wayman, 2002 p.5).
False Non-Match Rate	- is "the expected probability that a sample will be falsely declared not to match a template of the same measure from the same user supplying the sample. FNMR is also referred to as a "false negative" (Mansfield & Wayman, 2002 p.5).

False Accept Rate	- is “the expected proportion of transactions with wrongful claims to identity (in a positive ID system) that are incorrectly confirmed” (Mansfield & Wayman, 2002 p.5).
False Reject Rate	- is “the expected proportion of transactions with truthful claims of identity (in a positive ID system) that are incorrectly denied” (Mansfield & Wayman, 2002 p.5).
File Size	- is the image “size in bytes” (PreFace SDK Manual, 2009, p.36).
Head Height to Image Height Ratio	- “is the ratio of the head height to image height” (PreFace SDK Manual, 2009, p.34).
Height to Width Ratio	- “is the ratio of image height to image width” (PreFace SDK Manual, 2009, p.34).
Identification	- is when “the user makes either no claim or an implicit “negative” claim to an enrolled identity, and a “one-to-many” search of the entire enrolled database is required” (Mansfield & Wayman, 2002 p.4).
Image Format	- is the “format for the image” (PreFace SDK Manual, 2009, p.36).

Image Height - "is the vertical dimension of the image in pixels" (PreFace SDK Manual, 2009, p.34).

Image Width - "is the horizontal dimension of the image in pixels" (PreFace SDK Manual, 2009, p.34).

Image Width to Head Width Ratio - "is the ratio of image width to head width" (PreFace SDK Manual, 2009, p.34).

Legacy Image - "are mug shot images captured in the past before the standardized capture process has been established."

Overt - is "an undisguised and candid use of a biometric system" (Association for Biometrics, 1999 p.2).

Percent Background Gray - "reflects the level of gray in the background" (PreFace SDK Manual, 2009, p.31).

Percent Background Uniformity - "reflects the variation of color throughout the background of the image" (PreFace SDK Manual, 2009, p.32).

Percent Facial Brightness – "is the average luminance of the facial region as a percent" (PreFace SDK Manual, 2009, p.33).

Percent Facial Saturation – "is the percent fraction of pixels saturated in the facial region" (PreFace SDK Manual, 2009, p.33).

Sample	- Biometric information obtained from a sensor or device.
Three-Point Lighting	- is “balanced lighting arrangement consisting of two points of illumination should be placed at approximately 45 degrees on either side of the subject’s face, the third point should be placed so as to illuminate the background uniformly” (U.S. Department of State, 2009).
Verification	- is when “the user makes a “positive” claim to an identity, requiring a “one-to-one” comparison of the submitted “sample” biometric measure to the enrolled template for the claimed identity” (Mansfield & Wayman, 2002 p.4).

1.5. Assumptions

The assumptions for this research include:

- The data collected will be limited to the electronic mug shot images made available by the Indiana Department of Correction.
- The layout and environmental conditions of the experimental setup will be constrained by Indiana Department of Corrections equipment and environment.
- Images captured using a standard-compliant capture process should result in better performance.

1.6. Limitations

The limitations for this research include:

- The results of this study are limited by the data that will be collected from the Indiana Department of Correction.

1.7. Delimitations

The delimitations for this research include:

- This study only tests the image quality and percent standard compliant of the Indiana Department of Correction mug shot database.
- Testing of other State's mug shot databases is beyond the scope of this study.
- Testing the performance of the mug shot images is beyond the scope of this study.
- The specific photography equipment used by the Indiana Department of Correction is beyond the scope of the study.

CHAPTER 2. REVIEW OF LITERATURE

The following review of literature contains four sections which deal with biometrics, face recognition, and the mug shot capture process. The first section addresses the use of biometrics technology in the law enforcement field; the next section the relationship between image quality and the performance of face recognition; the third section reviewed facial image standards; and the final section reviewed previous research related to face image capture process.

2.1. Biometrics in Law Enforcement

Human characteristics in the law enforcement and criminal justice fields can be traced back to the 1870s, and according to Jain, Ross, and Prabhakar (2004) fingerprint use began in the booking process of criminals as early as the late 19th century. The prints were stored in a card file system which acted as a manual-type database for subsequent identification. According to the Encyclopedia of Nineteenth-Century Photography, “Alphonse Bertillon developed the first scientific prisoner identification system using photography” towards the end of the 19th century (p. 150). As a clerk with the Prefecture of Police of Paris, he invented the mug shot similar to how it is still known, with a pair of photographs taken with a standardized pose and angle under standardized lighting conditions. (Hannavy, J., 2008). During this same time Allen Pinkerton began using face photographs on Wanted posters in the United States (Petersen, J., 2007).

In recent years, biometric technology's presence has been expanding in the law enforcement environment. Fingerprints are still used during the booking process today, and fingerprints collected at the state and local level get sent electronically to the Federal Bureau of Investigation's Criminal Justice Information Services (CJIS) Division, which maintains the Integrated Automated Fingerprint Identification System, IAFIS, "the world's largest repository of fingerprint data" (Spaun, 2007, p. 1). Along with fingerprints, DNA is becoming a more productive means of identification in terms of forensics and criminal justice (Spaun, 2007).

The FBI, along with state and local law enforcement agencies across the country have been using the IAFIS system, which provides the law enforcement community access to the largest fingerprint database, for latent searching, electronic searching, and electronic storage of fingerprints. The IAFIS system and the FBI process on average approximately 162,000 ten-print submissions daily (FBI.gov, 2010). While fingerprints have a long history in the law enforcement community, there is one flaw with the system: it will not work if the accused is not enrolled in the database. Enrollment in IAFIS is limited to individuals who have previously been arrested and have a criminal history and those serving in the U.S. military or have been or are employed by the federal government (FBI.gov, 2010). Any person who commits an offense for the first time will have never had their fingerprints taken before, thus leaving the largest and one of the most effective methods of criminal identification useless (Woodward, 2005).

While fingerprinting is a widely used and accepted practice, Gonzalez-Castillo (2006) insists that there are many advantages to using facial images:

> It is already socially and culturally accepted internationally, photographs do not disclose information that the person does not routinely disclose to the general public, the facial image is non intrusive and does not require

> new and costly enrollment procedures to be introduced, and many states have a legacy database of facial images captured as part of a digitized production of passport photographs which can be encoded into facial templates (p.1).

By incorporating face recognition technology in the law enforcement booking procedure, forensic examiners and criminal justice practitioners would not only be able to utilize mug shot databases of criminals, but also any results from a face recognition matcher would be more reliable in a courtroom than using only human identification.

2.2. Face Recognition Image Quality and Performance

There have been various studies related to the quality of facial images and the resulting performance of facial recognition systems. The following evaluations were conducted to assess the performance of face recognition algorithms: the DOD's Face Recognition Technology (FERET) program conducted studies from 1993 to 1997, the 2002 Face Recognition Vendor Test (FRVT), the 2005 Face Recognition Grand Challenge (FRGC), the 2006 FRVT, and the 2010 Multiple Biometric Grand Challenge (MBGC). Figure 2.1 shows the evolution of the performance of face recognition algorithms over those years.

Figure 2.1 indicates the performance rate has significantly increased since periodic evaluations began in 1993 (Phillips, et. al, 2007). One reason for this development are better algorithms, but another contributing factor is the improvements in photo capture technology. The progression from 35mm film cameras to today's digital, high resolution cameras is allowing the algorithms to work with higher quality images, resulting in performance improvements (Phillips, et. al, 2007). In 2007, NIST conducted a workshop on biometric quality that showed the correlation (Figure 2.2) of image quality to performance for different biometric modalities (Shah & Martin, 2007).

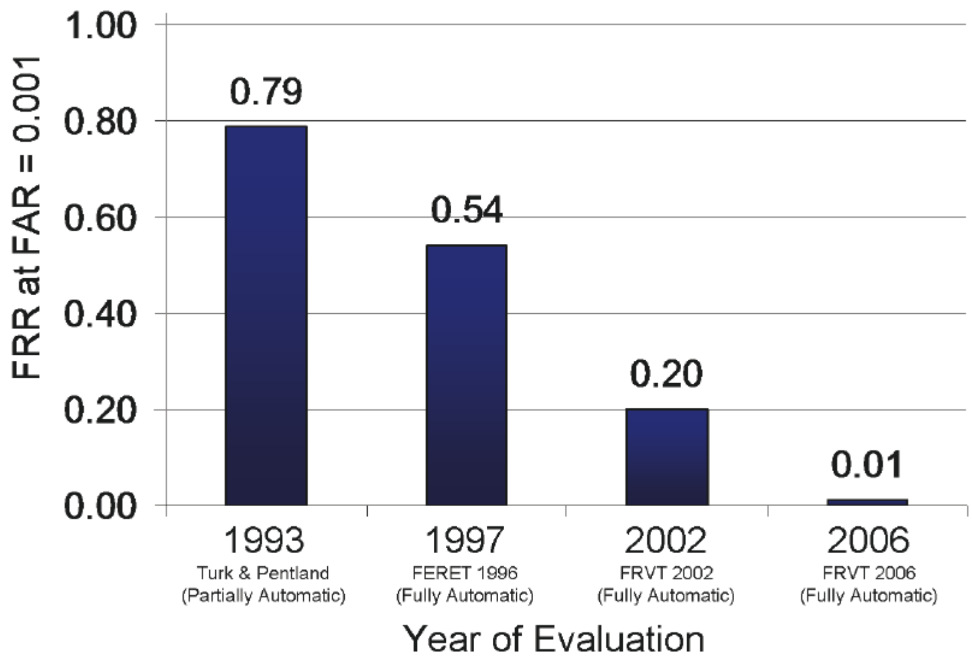


Figure 2.1 Face Recognition Algorithm Improvement (1993 - 2006), (Phillips, et. al, 2007)

Figure 2.2 shows the correlation of image quality with the matching performance of a biometric recognition system (Shah & Martin, 2007). The correlation indicates how much performance relied on the quality of the images enrolled and matched in the system. Figure 2.2 shows that both iris and finger have a relatively high correlation between match score and image quality score. However, face recognition has the highest correlation and this is an indication that the image quality of face images used in a face recognition system should be of particularly good quality in order for the system to operate at an acceptable level.

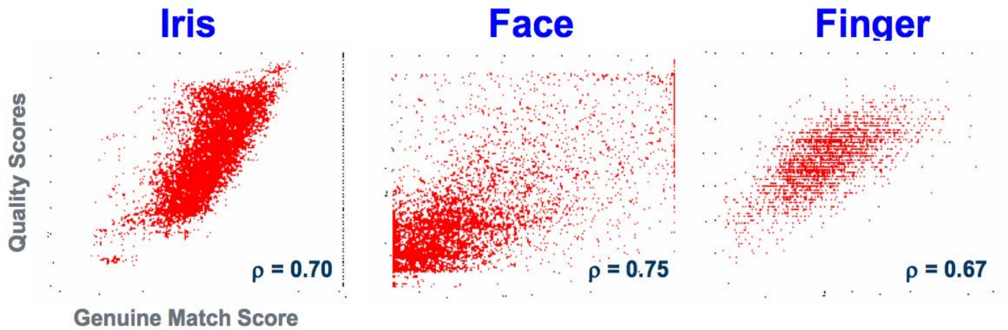


Figure 2.2 Quality and Match Performance Correlation for Iris, Face, and Finger (Shah & Martin, 2007)

Current research in the biometric community indicates that compared to other modalities, such as fingerprint, iris, and retina, face recognition does not perform as well, Table 2.1. However, with the universality and acceptability of face recognition technology, face recognition provides a promising method to identifying criminals and persons of interest (Jain, et. al, 2004). Some of the reasons for the poor performance of face recognitions systems include lighting, background, and obstructions such as hats and glasses.

Table 2.1 *False Accept Rates for Face, Fingerprint, Iris, and Voice Modalities*

Biometric Trait	Test	Test Parameters	FRR at FAR 0.001
Face	FRVT 2006	Controlled illumination	0.01
Fingerprint	FVC 2004	Exaggerated skin distortion, rotation	0.02
Fingerprint	FpVTE 2003	US Government operational data	0.001
Iris	ITIRT 2004	Indoor environment, multiple visits	0.0099
Voice	NIST 2004	Text independent, multi-lingual	0.05 – 0.10

Kukula and Elliott (2004) conducted a study to assess the effects of illumination levels on face recognition algorithm performance and found that there were significant challenges regarding face recognition under differing illumination levels, especially at low-lighting levels. Their research also indicated a need for further assessment of environmental issues regarding face recognition, such as lighting and background effects.

Part of the FRVT 2006 was a test of uncontrolled illumination. The experiment compared enrolled frontal face images taken under controlled illumination against frontal images taken under uncontrolled illumination settings. The best results from this test on the dataset of high-resolution images had a FRR interquartile range, “the range of the center half of the data” of 0.119 to 0.146 (Moor & McCabe, 2006, p.47). While these results are not as strong as seen in Table 2.1 for the controlled illumination experiment, they show improvement in algorithm development in regards to the illumination problem (Phillips, et. al, 2007).

The results of FRVT 2006 also confirm the progress in algorithms overall. The best performing dataset was the very-high resolution dataset with a False Reject Rate (FRR) of 0.01 as the set False Accept Rate (FAR) of 0.001 (Table 2.1). The high-resolution dataset did well also with a FRR interquartile range of 0.21 – 0.023 at the set FAR 0.001. These results show that the center half of the data analyzed was within the FRR range 0.021 and 0.023. The low-resolution also showed much improvement with a FRR interquartile range of 0.024 – 0.027, however, it was still the lowest performing dataset in the group further confirming the correlation to image quality with performance from Figure 2.1 (Phillips, et. al, 2007).

The latest biometric evaluation that included face recognition was the Multiple-Biometric Evaluation (MBE). The section most relevant to this research was the Evaluation of 2D Still-Image Face Recognition Algorithms. The 2D Still-Image evaluation attempted to evaluate state-of-the-art face recognition algorithms, by using large datasets of face images with populations reaching into the millions. The goals of the 2D Still-Image evaluation from MBE 2010 that were most relevant to the research being conducted in this study include:

- Leveraging massive operational corpora. The availability of images from large populations (in the millions) ensures statistical significance of all studies, particularly across demographic groups. The use of operational images brings greater operational relevance to the test results.
- Evaluating face recognition technologies in a proper one-to-many identification mode. This departs from many prior evaluations in which 1:N search accuracy was simulated via computation of N 1:1 comparisons (Grother, et. al, 2010, p.10).

Four datasets were used in MBE 2010, however the law enforcement (LEO) dataset was most relevant to this study since it consisted of facial images collected by law enforcement and transmitted to the FBI. The facial images were mug shots taken using similar method as the IDOC mug shots used in this study. Both photograph sets were taken with similar traits as shown in Figures 1.1 and 2.3. The LEO dataset was collected from the 1960s to 2008 with over 2 million images total, though not all investigations conducted utilized the entire dataset. From a visual inspection, the issues appear to be consistent with the IDOC as lighting, background, and capturing consistent images all being apparent with both IDOC and LEO mug shot images.

Figure 2.3 Examples from the LEO Dataset (Grother, et. al, 2010)

The first relevant study in MBE 2010 determined the one-to-many search accuracy of vendor algorithms while in investigation mode using a subset of the LEO photos. In investigation mode, a candidate list is populated for an expert examiner to analyze for the proper match, if it exists (Grother, P., et. al, 2010). The population size of this study was a 1.6 million subset of the LEO dataset. Results show that the best performing algorithms had a rank 1 cumulative match characteristic (CMC) score of 0.92 and 0.91 respectively. The cumulative match characteristic shows the probability that a given subject appears is different sized candidate lists (Higgins, P., 2005). These results are rank based therefore, score is ignored. If an examiner is available to investigate the candidate lists this is beneficial, however if an entirely automated process is preferred, the results of this study are not as relevant.

The identification mode one-to-many search accuracy study is relevant if the IDOC determines it would prefer an automated process instead of the investigation mode above, which requires an examiner. This study was conducted using subset samples of the LEO population at 10,000, 80,000, 320,000, and 1.6 million images. The ROC of the 1.6 million sample size reveals some very interesting findings. First, when selectivity is high (SEL > 1), medium (SEL < 0.1), and low (SEL < 0.005) a different algorithm performed best. Next, the best performing algorithm at the low selectivity produced false matches for only one in 10,000 searchers, but relability dropped so that more than one in two mated searches gave a miss. Also, it was shown that when operating threshold is high, the best algorithms essentially fail (reliability approaches zero). The algorithms solicited for the 2D Still-Image evaluation were for use in investigation mode, so since identification mode was not explicitely solicited the results may not represent the full potential for “lights out” identification with LEO type data (Grother, et. al, 2010).

The investigation of algorithms’ dependence on pose accuracy is closely related to the research being conducted in this study in that it is necessary to determine if the accuracy is dependent on the frontal pose and how much rotation from that pose is acceptable. In this pose investigation both yaw and roll angles were examined, the number of images within certain angles were calculated to determine the effect on accuracy. The histograms of yaw and roll angles showed that many images are outside the +/- 5 degree threshold in the ISO/IEC and ANIS/NIST standards (Grother, et. al, 2010).

Results indicate that error rates increased when the yaw angle was between 6 and 16 degrees from frontal, but were catastrophic when yaw angle was greater than 20 degrees. A high roll angle indicated other image quality issues were in play. Most algorithms tested in MBE 2010 gave increased error rates when non-frontal images are acquired and enrolled into recognition systems. However, some algorithms were less sensitive to pose angles than others (Grother, et. al, 2010).

These investigations provide a better understanding of how the IDOC mug shots might perform when used in a face recognition system. The LEO dataset is quite similar in makeup and much larger in quantity. MBE 2010 results, while not directly associated with this research, provide insight into how law enforcement operational mug shots perform. MBE 2010 also shows the progress in face recognition technology overall. The FNMR of 0.003 at a FMR of 0.001 was achieved for one submission. The decrease in performance over time went from the FNMR at a FMR = 0.001 from 0.79 in 1993 to 0.003 in 2010 (Grother et. al, 2010). The performance decrease is roughly three order of magnitude for error rate.

2.3. Face Recognition Standards

Many facial image standards exist to ensure interoperability, image quality, and recognition performance. These standards include ISO/IEC 19794-5 Biometric Data Interchange Formats – Part 5: Face Image Data, ANSI/NIST 385-2004 Face Recognition Format for Data Interchange, and ANSI/NIST-ITL 1-2007 Data Format for the Interchange of Fingerprint, Facial, & Other Biometric Information. McCabe (1997) wrote a best practices recommendation for the capture of mug shot images, which is incorporated in the three standards mentioned. McCabe’s document establishes a minimum set of parameters that need to be met in order to obtain standard-compliant images. The issues surrounding performance of face recognition have been linked to varied image quality and the lack of a repeatable image capture process acorss all image capture applications, (i.e. at license branches, correctional facilities, and passport photographs).

Tables 2.2 through 2.5 show face image attribute definitions in the national standard NIST/INCITS 385-2004, the international standard ISO/IEC 19794-5, the NIST Best Practices for the Capture of Mug Shots Recommendations, and the commercially off the shelf (COTS) software used in this study to analyze for standard compliance, Aware PreFace. The NIST/INCITS and ISO/IEC standards are essentially the same, just residing at different levels of the standards community so they are listed together in the table. Not all of the attributes are incorporated in the NIST Best Practices document or the COTS program since the profile used to analyze the images is based on the NIST Best Practices so some of the entries are blank in those columns.

Table 2.2 *Comparison of Face Attribute Definitions between Standards, Best Practice Recommendations, and PreFace Software*

Attribute	**NIST Mug shot Best Practices Recommedations 2.0**	**PreFace Metrics Definitions**	**ANSI/INCITS ISO/IEC Definitions**
Pose	"Frontal pose or full-face pose"	"Eye axis angle must not be less than or more than 5 degrees from center"	"The full-frontal pose shall be used and the rotation of the head should be less than +/- degrees from fontal in every direction"
Depth of Field	"The subject's captured facial image shall always be in focus from nose to ears"	Not specified	"The subject's captured image shall always be in focus from nose to ears and chin to crown"
Centering	"The approximate horizontal mid-points of the mouth and of the bridge of the nose shall lie on an imaginary vertical straight line positioned at the horizontal center of the image"	"Eye axis location ratio must be between 0.5 and 0.6" "Centerline location ration must equal "0.5"	"The approximate horizontal midpoints of the mouth and of the bridge of the nose shall lie on an imaginary vertal line AA", seen in Figure 2.3.1"

Table 2.3 *Comparison of Face Attribute Definitions between Standards, Best Practice Recommendations, and PreFace Software (Continued)*

Lighting	"Minimum of three-point balanced illumination" "Proper lighting shall also contribute to the uniformity of illumination of the background"	"Percent facial brightness is the average luminance of the facial region and shall be between 25-75%"	"Lighting shall be equally distributed on each side of the face and from top to bottom"
Background	"The subject whose image is being captured shall be positioned in front of a background which is 18% gray with a plain smooth flat surface"	"A 0 indicated a simple background, a 1 indicates a complex background" "Optimal percent background gray is 18%, though lighter is always better than darker but may not exceed 30%"	"The subject whose image is being captured shall be positioned in front of a background which is 18% gray with a plain smooth flat surface"
Aspect Ratio	"The width to height aspect ration of the capture image shall be 1:1/25"	"The width to height aspect ratio must equal 1.25"	"The aspect ratio should be between 1:1.25 and 1:1.34"
Minimum Number of Pixels	"480 pixels X 600 pixels"	"Image width must be at least 480 pixels"	Not specified
Colorspace	"24-bit RGB pixels"	"Number of color channels in the image, 3 indicates RGB"	"The 24-bit RGB colorspace shall be used"
Pixel Aspect Ration	"Square pixels with 1:1 pixel aspect ratio"	Not specified	"Pixel aspect ratio of 1:1 shall be used"

Table 2.4 *Comparison of Face Attribute Definitions between Standards, Best Practice Recommendations, and PreFace Software (Continued)*

Compression Algorithm	“JPEG Squencial Baseline mode of operation and target size of 25,000 to 45,000 bytes”	Not specified	“JPEG Sequential baseline mode of operation”
File Format	“JPEG File Interchange Format”	“Image Format must be 4, which is JPEG”	“JPEG File Interchange Format”
Expression	Not specified	“Smile indicates the likelihood of a smile, 0 indicates a smile is very unlikely”	“Expression shall be neutral (nonsmiling) with both eyes open normally and mouth closed”
Shadows	Not specified	“Background uniformity reflects the variation of color throughout the background and must not be lower than 70%”	“There shall be no shadows over the face from base of the chin to crown of head and from ear to ear. Also there shall be no dark shadows in the eye-sockets due to the brow”
Hot Spots	Not specified	“Percent facial saturation shall be between 0-3%”	“Care shall be taken to avoid hot spots on the face”
Width of Head	“The width of the subject’s head shall occupy approximately 50% of the width of the captured image”	“Image width to headth width ratio must be between 1.995 and 2.0004”	“The width of the head can be seen as CC”, in Figure 2.3.1, “also the minimum image width to head width ration is 7:4”

Table 2.5 *Comparison of Face Attribute Definitions between Standards, Best Practice Recommendations, and PreFace Software (Continued)*

Length of Head	"An imaginary horizontal line through the center of the subject's eyes shall be located at approximately the 55% point of the vertical distance up from the bottom edge of the captured image"	Not specified	"The crown to chin portion (DD) shall be no more than 80% of the vertical length of the image (B), as in Figure 2.3.1
Resolution	Not specified	Not specified	"Resolution should be roughly 90 pixels from eye center to eye center"

Figure 2.4 is the depiction of how a full frontal facial image should be captured (ANSI/INCITS, 2004). Visually comparing Figure 2.4 with the mug shot images captured from the IDOC facilities, it is apparent that the IDOC mug shots do not meet these constraints.

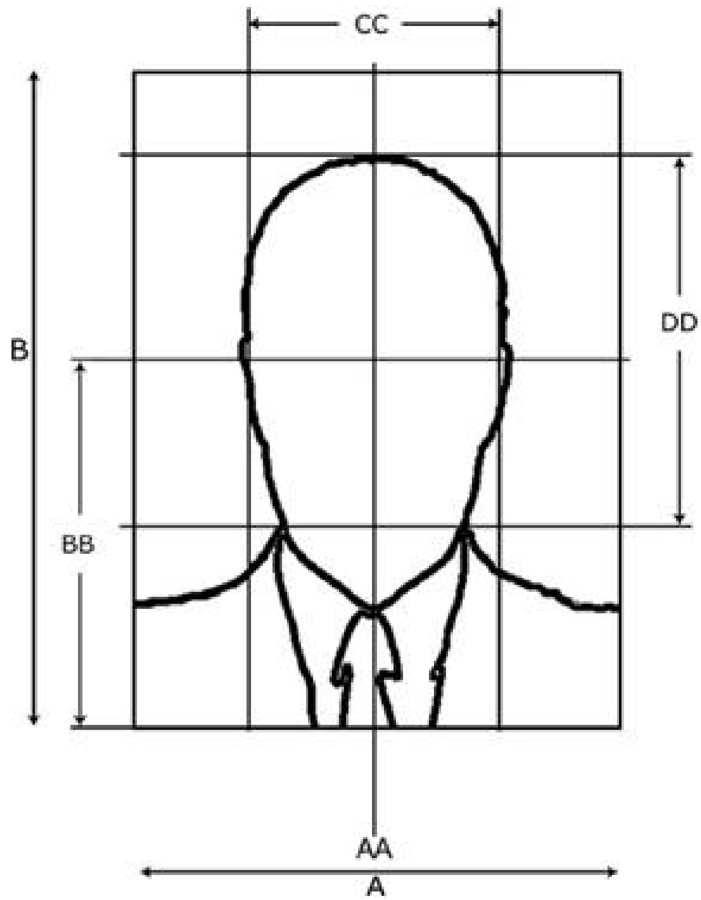


Figure 2.4 Geometric Characteristics of the Full Frontal Face Image (ISO/IEC, 2005)

2.4. Previous Work in Face Image Capture

In 2008, Theofanos et. al, conducted a study that involved assessing face image acquisition. The Thefanos study was conducted for the United States Visitor and Immigrant Status Indicator Technology (US-VISIT) program to examine the capture process to identify "any usability and human factors that may improve the existing face image capture process." Although it assessed the capture process at US ports of entry, some of the findings should be applicable to the IDOC setting. The report presents five usability improvements that can be made to the US-VISIT capture process, which include:

1. The camera should resemble a traditional camera;
2. The camera should click when the picture is taken to provide feedback to the traveler that the picture is being taken;
3. The camera should be used in portrait mode;
4. The operator should be facing the traveler and the monitor while positioning the camera and;
5. Provide some marking on the floor (such as footprints) to indicate to the traveler where to stand for the photograph (Theofanos, et. al, p. 5, 2008)

All five of these suggestions could help solve issues with the mug shots taken by the IDOC such as eyes being open and looking at the camera, subject being centered in the image, and the ratio of the head within the frame of mug shot.

Theofanos, et. al, (2009), conducted another study to test the usability of an overlay on the computer screen to improve the face image capture. A face overlay, “a visual reticule that may be superimposed onto a live video feed to facilitate the face image capture process”, was used to address if the participants could use it to effectively center the face when capturing images. The overlay they created was based on the ANSI/INCITS 385-2004 and ISO/IEC 19794-5:2005 standards mentioned above with Figure 2.4 being the optimal image.

Some of the goals of Theofanos’s study included not increasing the amount of work to capture an image, improve image quality, improve the efficiency of capturing face iamges, and improve the satisfaction of the users capturing the images. The results show that the overlay was “easy to use without instruction or training and it was obvious to participants” (Theofanos, et. al, 2009, p.6.). The overlay showed a significant improvement in the effectiveness of capturing an image. Without the overlay only 1.4% of faces were centered within the images, whereas with the overaly 53.2% of faces were perfectly centered in the image, and 45.4% were largely within the oval of the overlay (Theofanos, et. al, 2009). There was no significant difference in the time to capture face images. Although using the face overlay did not improve the efficiency of the process,it also did not negatively impact the time to capture face images either. The users of the overlay reported satisfaction with knowing when the image was centered as satisfactory and easy to use.

Since the IDOC needs to improve the process of capturing their mug shot images, the results of these studies should be beneficial when addressing the issues faced by the IDOC. In particular, using an overlay, having the camera in portrait mode, and providing markings on the floor should help improve the process.

CHAPTER 3. METHODOLOGY

The purpose of this study was to examine and identify deficiencies in the Indiana Department of Corrections mug shot image capture process. Then design a process improvement document for the IDOC to implement so that images captured using the improved process should be closer to standard-compliance and thus perform well in a facial recognition system. The following sections will discuss the design of the study and the sampling that was used as well as a timeline of events from beginning of study to its conclusion.

3.1. Study Design

The study assessed the extent to which the Indiana Department of Corrections followed the NIST Mug Shot Best Practices Recommendations. Capturing images in accordance with these recommendations should allow for good quality images that are usable in a face recognition system. Because the IDOC would like to incorporate facial recognition technology into their current processing and tracking system, it was necessary to check the current images to determine usability in a face recognition system. It was also necessary to assess the current capture process to determine if they were capturing standard-compliant images. The procedures used throughout this study are shown in Figure 3.1.

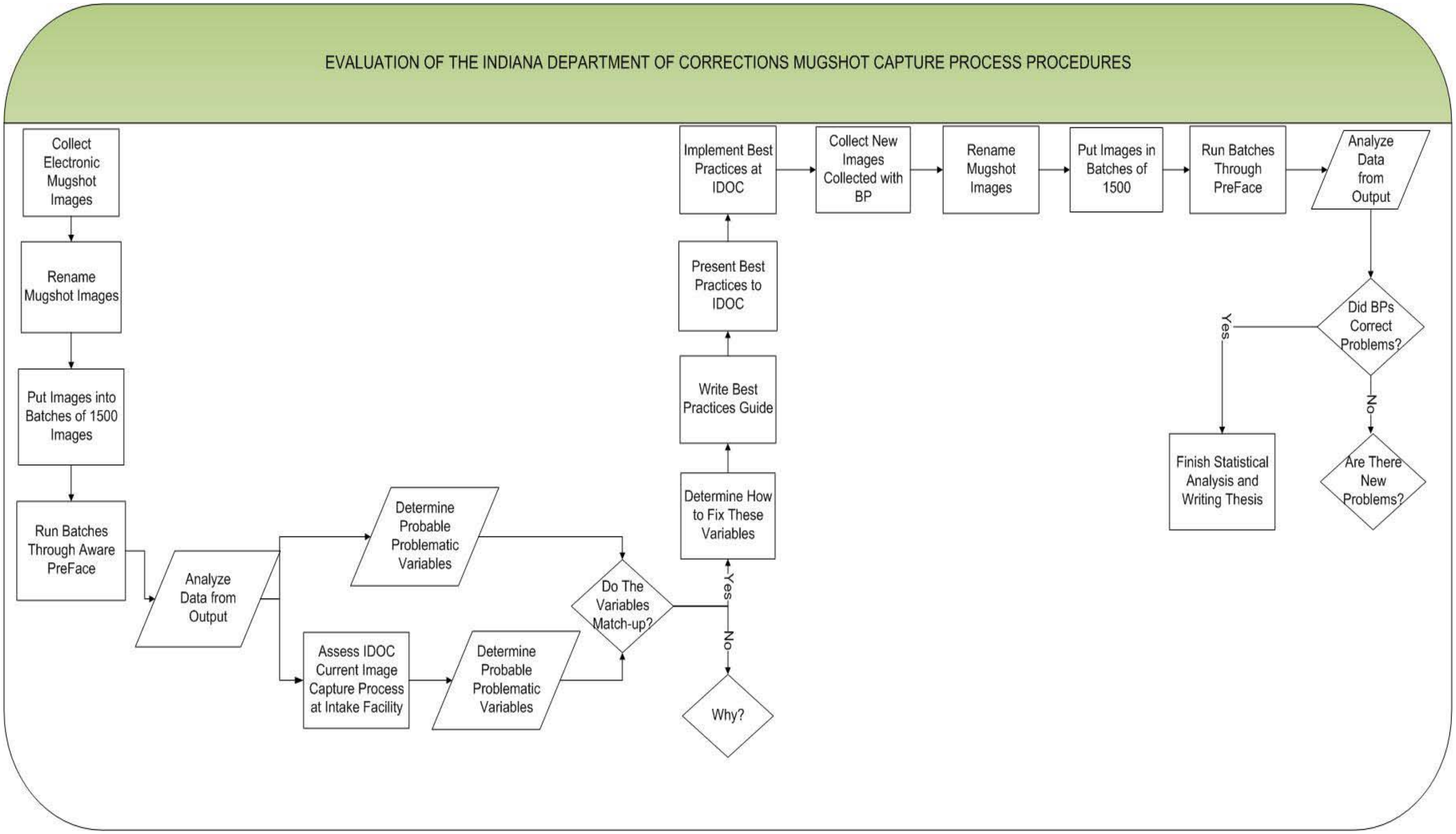


Figure 3.1 Evaluation Methodology

First, the IDOC's initial mug shot images were assessed for NIST Best Practices compliance. Using image quality and standard compliance software (Aware PreFace), which provided data on numerous variables as outlined in Figure 3.2. From this output, an assessment of standard compliance optimization, and to pinpoint flaws in the current mug shot capture process.

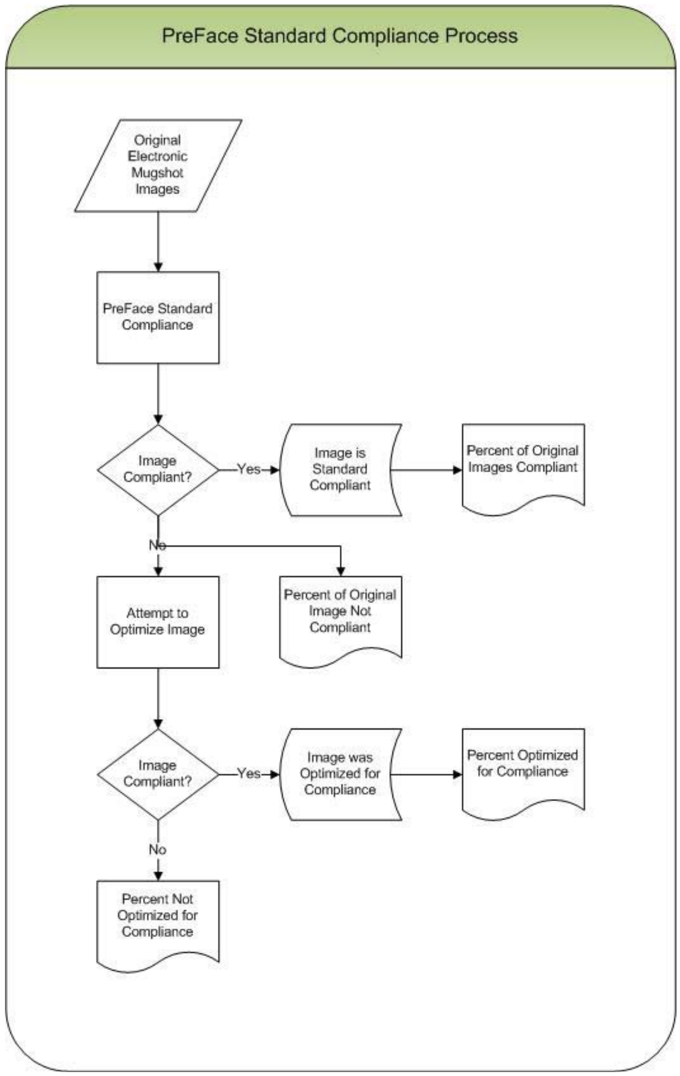


Figure 3.2 PreFace Standard Compliance Process

After the results were analyzed, a visual inspection of the dataset determined if any profile images were inadvertently included. Since the dataset was provided by the IDOC, there is the possibility of error in their process of moving images collected into their database. If any profile images were included, this would impact the overall characteristics of the dataset, since the NIST Best Practices only allows for frontal images.

Next, one thousand mug shots were randomly selected for visual examination to determine if the results from the software could be seen visually. Determining how consistent the software was at analyzing the mug shot images and if the results could be confirmed visually.

After the initial assessment of the IDOC images, and flaws in the capture process pinpointed, the capture process at the IDOC was observed. During the observation, the layout of the capture station, the equipment being used, and the mug shot capture process were documented.

From these observations, and the flaws from the mug shot image data analysis, a list of recommendations to optimize the capture process at the IDOC intake facility was created. The recommendations were implemented to determine if there was an increase in image quality and standard compliance.

The recommendations for the optimized process were based on the NIST Mug Shots Best Practices document. Once the process recommendations were completed they were implemented in the IDOC intake facility. After a period of three weeks, the mug shot images captured using the recommendations for the optimized process were collected.

The images were assessed using the same process as outlined in Figure 3.2. A statistical analysis determined if there were significantly more images standard compliant taken using the optimized process compared the old process.

The study was mixed in nature with the initial mug shot image assessment being exploratory because no prior research had been conducted analyzing the IDOC mug shots or the capture process. The image capture optimization process was experimental since the problematic variables in the capture process were manipulated to take these variables into account and attempt to improve them.

The dependent variable was standard compliance; the independent variables included the 15 face image quality characteristics measured by the software including: eye separation, eye axis angle, eye axis location ratio, centerline location ratio, image width/head width ratio, head height/image height ratio, facial dynamic range, percent background uniformity, percent background gray, percent facial brightness, brightness score, eye contrast, degree of background clutter, degree of blur, and background type.

3.2. Unit and Sampling

The sample used in the first part of the study was a collection of mug shot images taken by the IDOC. The images were taken and stored electronically in the JPEG format. A total of 49,694 mug shot images was included in the dataset. The images were assessed using the Aware PreFace software to check for standard compliance and image quality. The software provided output for the 15 image quality characteristics. The thresholds for standard compliance were determined by using the NIST Mug Shot Best Practices profile in the Aware PreFace program. During the data run, Aware PreFace automatically attempted to optimize each image to be compliant with the profile. The output from the software was then assessed to find any trends or problematic characteristics in the mug shot images. The initial hypothesis for the study was that the number of standard-compliant images would be significantly greater after the proposed process was implemented. After analyzing the initial data, it was apparent that eight variables were problematic. The proposed process attempted to improve the standard compliance of these variables

The sample used in the second part of the analysis included the images captured at the IDOC intake facility using the optimized capture process during a period of three weeks from the end of May 2010 to June 2010. According to Kurt Bensheimer (Personal Communication, December 18, 2009), the IDOC intake facility processed ~250 – 300 people per week. A period of three weeks would yield a sample of ~750 – 900 images. These images went through the same process as described above.

3.3. Threats to Internal Validity

There were seven threats to internal validity, including history, maturation, testing, instrumentation, selection bias, statistical regression, and mortality effects (Sekaran, 2003). Of these threats, instrumentation and statistical regression caused the most concern. Statistical regression was minimized due to the large sample size decreasing any one member of the sample's effect on the dependent variable. Instrumentation could have affected the study if the face image quality and standard compliance software did not work properly. In such a case, a new program would need to be chosen after the study had already begun.

3.4. Threats to External Validity

"External validity raises issues about the generalizability of the findings to other settings" (Sekaran, 2003, p. 158). The study had a sample that represented the operational, electronically stored mug shot images for the State of Indiana only. The study is generalizable only for the images captured in the State of Indiana.

CHAPTER 4. RESULTS AND ANALYSIS

There were three parts to this study; to analyze the IDOC images previously collected, analyze the current IDOC mug shot capture process, and implement an optimized process to determine if it improvemented standard compliance. This section will present the results of the data analysis for each of these parts. The final section of this chapter will present the hypotheses from above and the results of the tests conducted.

4.1. Electronic IDOC Mug Shot Image Results

A sample of 49,694 images were collected from the IDOC, no meta data were included with these images, due to IRB restrictions, so there was no way to determine when the images were captured. Due to the IRB restrictions it was impossible to link a mug shot to an individual, so no matching was undertaken.

The IDOC images were first extracted from a .zip file, converted to JPEG format and renamed to remove any identifiable information from the filename. The images were processed through Aware PreFace software to determine standard compliance with the NIST Mug shot Best Practices standard profile. The output data were converted into a Microsoft Excel® 2010 spreadsheet for initial analysis and then transferred to Minitab 16 for more extensive analysis and the hypothesis testing. Of the 49,694 images there were 617 (~1.24%) with input errors, Table 4.1.

Table 4.1 *Electronic Mug Shot Input Image Errors*

Input Error	**Number with Error**
Error 931 – Face too close to the edge of the image	14
Error 934 – Blank image	3
Error 938 – Pose is invalid	335
Error 940 – Extreme pose or not a face image	174
Error 941 – Multiple faces in field of view	23
Error 942 – Bad pose or face is too small	68

In Table 4.2, the threshold of each variable is presented along with the number of images that did not meet the threshold and were determined to be non-compliant). The eight variables of concern were chosen based on this data. The variables were determined to be problematic enough if the number not compliant was greater than 10,000 images and if, based on funding and environmental constraints from the IDOC, the variable could possibly be manipulated.

Table 4.2 *PreFace Variable Thresholds for NIST Mug Shot Best Practice Profile*

PreFace Variable	**Threshold**	**Number Not Compliant**
Facial Dynamic Range	7 - 8 bits	1780
Percent Background Uniformity	70 - 100%	1168
Percent Background Gray	0 – 30%	17856
Percent Facial Brightness	35 - 100%	8804
Brightness Score	2 – 5	1945
Eye Contrast	2 – 5	303
Degree of Clutter	= 0	2406
Degree of Blur	= 0	17345
Background Type	Simple	2508
Eye Separation	>= 90 pixels	42926
Eye Axis Angle	-5 - +5 degrees	9676
Eye Axis Location Ratio	0.5 – 0.6	22311
Centerline Location Ratio	= 0.5	49690
Head Height/Image Height Ratio	0.6 – 0.8	23246
Image Width/Head Width Ratio	1.995 – 2.004	49380

Figures 4.1 and 4.2, shows histograms of each variable, highlighting the variations of distributions.

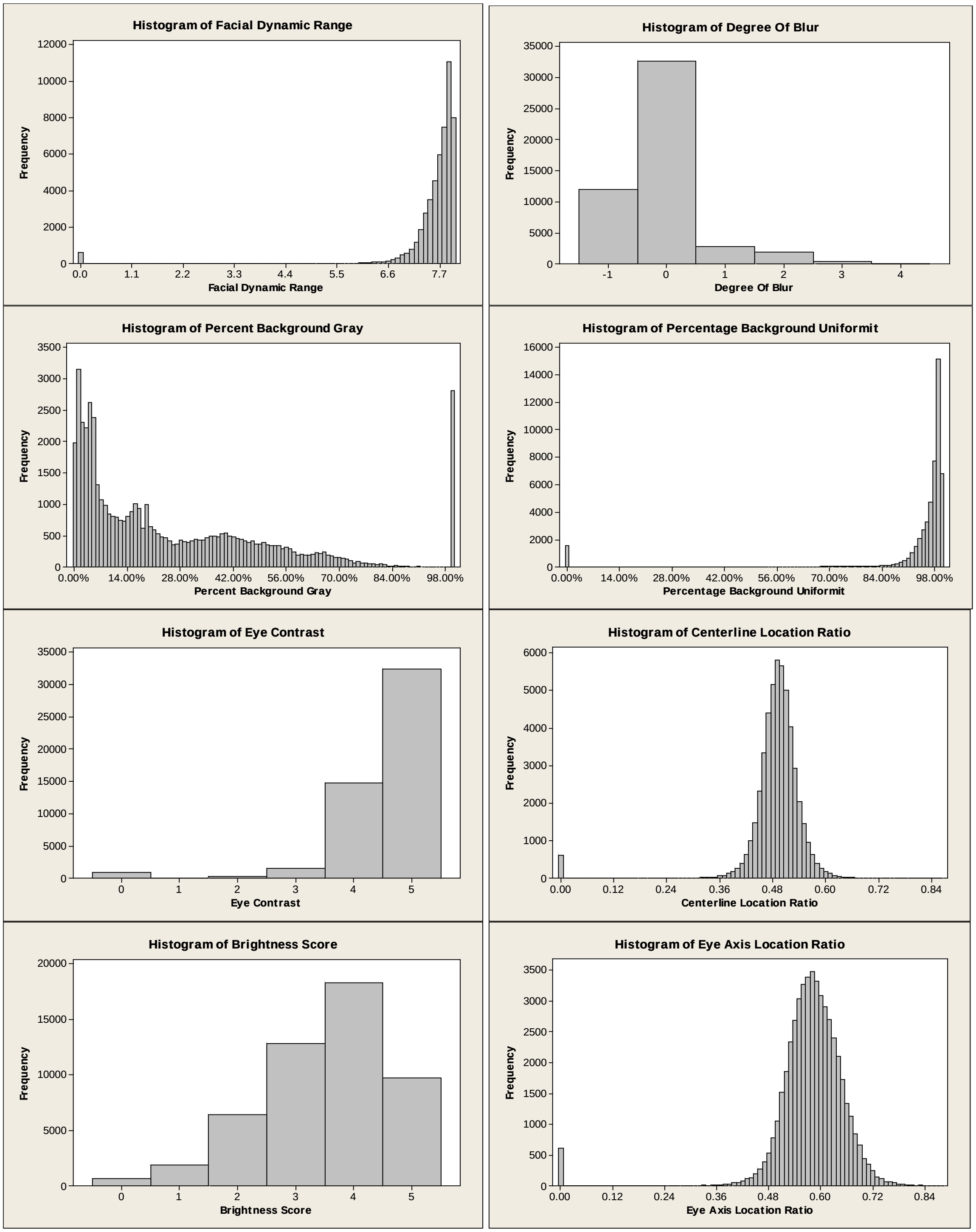


Figure 4.1 Histogram of PreFace Variables for Electronic Mug Shots

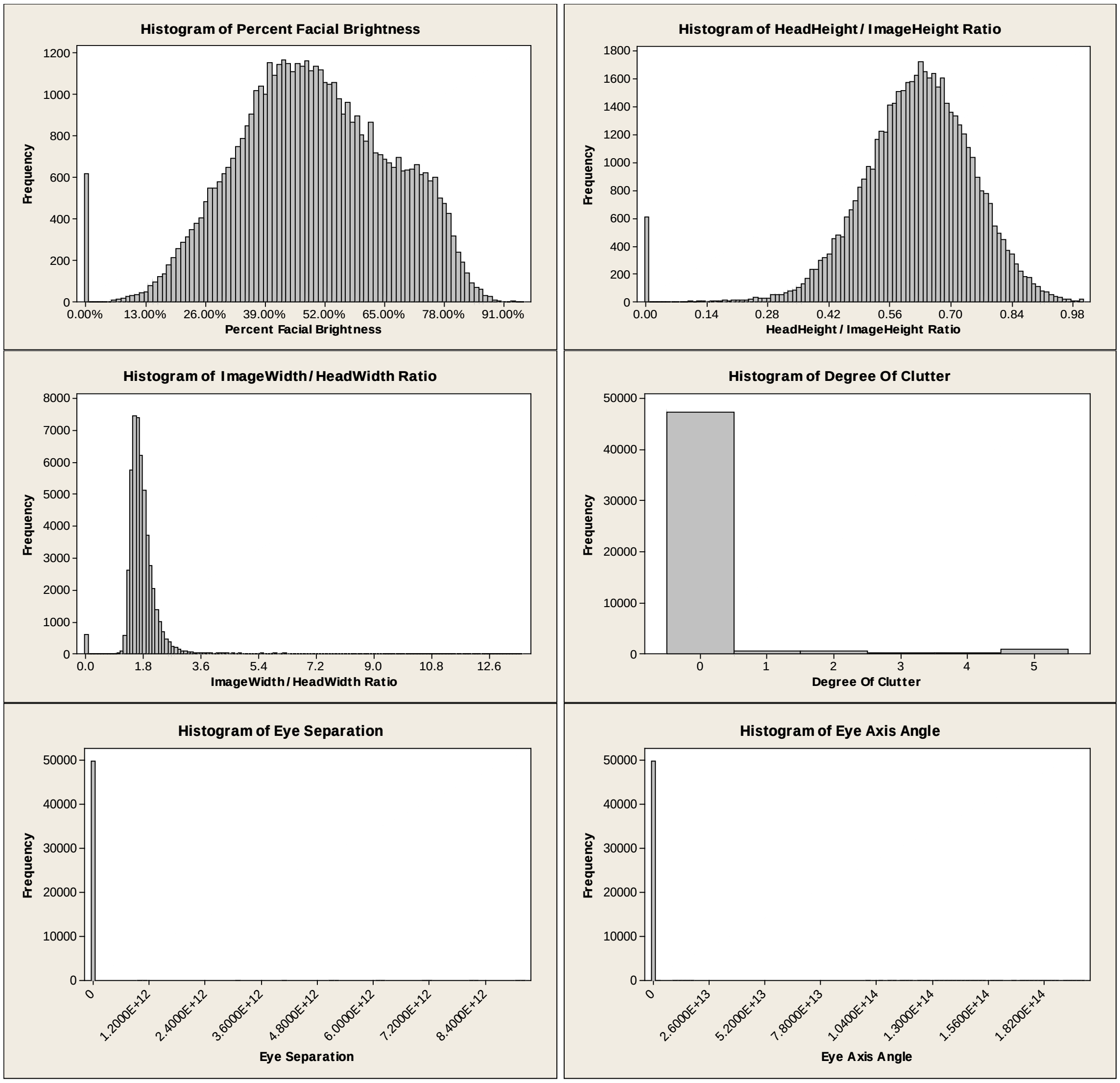


Figure 4.2 Histogram of PreFace Variables for Electronic Mug Shots (Continued)

Table 4.3 shows the problematic Pre-Face variables from the analysis of the IDOC mug shot images, It should be noted that many of the IDOC mughsot photographs were deficient in more than one of the Pre-Face variables.

Table 4.3 *Problematic PreFace Variables for Input IDOC Mug Shots*

PreFace Variable	Percent Not Compliant
Percent Background Gray	35.93%
Degree of Blur	34.91%
Eye Separation	86.38%
Eye Axis Angle	80.53%
Eye Axis Location Ratio	44.90%
Centerline Location Ratio	**100%**
Head Height/Image Height Ratio	46.78%
Image Width/Head Width Ratio	99.37%

One hundred percent of the centerline location ratio variable was non-compliant. The centerline location ratio is a static value of 0.5 for a threshold, instead of a range as is typical for the other variables. Since this is not a range it would be very difficult for any image, even taken under perfect conditions in a laboratory environment, to be compliant to the exact ratio of 0.5. The NIST Best Practices for Capture of Mug Shots states that:

> “The width of the subject's head shall occupy approximately 50% of the width of the captured image. This width shall be the horizontal distance between the mid-points of two imaginary vertical lines. Each imaginary line shall be drawn between the upper and lower lobes of each ear and shall be positioned where the external ear connects to the head” (McCabe, 1997, p.1).

The variable centerline location ratio in PreFace is defined as “the location of the centerline as a fraction of the image width measured from the left side of the image” (Aware, Inc., 2007, p.33). PreFace’s definition was taken from the ISO standard requiring the images to have “the approximate horizontal midpoints of the mouth and of the bridge of the nose shall lie on an imaginary vertical line AA positioned at the horizontal center of the image”, shown in Figure 4.3 (ISO/IEC, 2005, p. 51).

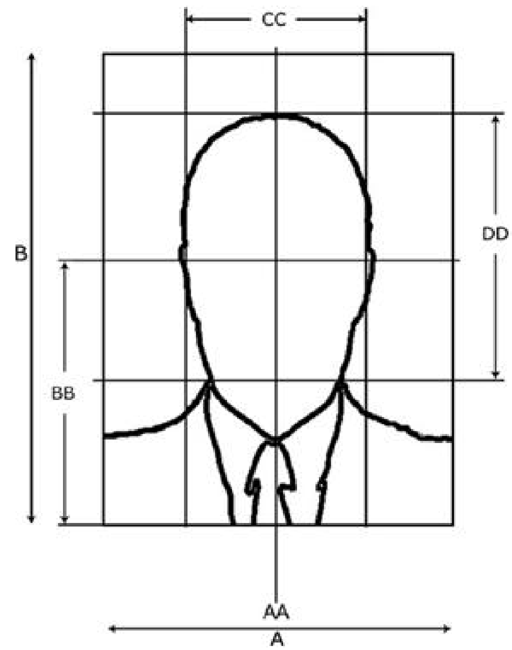


Figure 4.3 ISO Standard Geometric Characteristics for Full-Frontal Images (ISO/IEC, 2005)

The centerline location ratio was not compliant for every image in the original electronic mug shot dataset, Table 4.3. Thus, showing how an operational dataset will not meet the standards recommendation of having the head centered within the image and the value of 0.5 as the PreFace threshold is too precise for use with operational data.

The other variables that had a particularly high percentage of non-compliance; eye separation, eye axis angle, and image width-to-head width ratio, indicated the process by which these images were captured is flawed. The PreFace software attempted to optimize the input images for compliance. The software optimized the image characteristics to be within the given profile thresholds, if possible. The data showed that only 20 of the 49,694 (0.04%) images could be optimized for compliance. Of the remaining 49,674 images, 55 (~0.11%) had the optimization operation performed unsuccessfully, due to PreFace errors. The breakdown of the five most frequent errors were “Error 962: Invalid Image Width to Head Width Ratio”, which included 48,843 images, (98.29%). Having this error indicated that the original mug shot images were captured incorrectly, causing a poor image width-to-head width ratio, (Tables 4.2 and 4.3).

The variables shown in Table 4.3 were the most non-compliant variables for the electronic images. These variables were the focus when visiting the IDOC facility to view their current mug shot image capture process. The purpose of this visit was to determine if the data provided by PreFace was indicative of the deficiencies in the image capture process. The IDOC intake facility was visited with these variables in mind but wanted to take an overall approach when observing the process. This approach allowed for the process to be viewed holistically and not focus too acutely on specific parts that may have shown up in the data to determine if there were other factors not seen by PreFace or to confirm those indeed are the correct problems that needed to be addressed.

4.2. Current IDOC Mug Shot Capture Process

The process by which mug shots were taken was as follows:

1. The inmate steps into a small room where they stood against a white cinder block wall
2. The mug shot was captured, manually cropped, sent to the database, and printed onto the inmates identification card
3. The inmate stepped out of this room into the larger room
4. The inmate stepped up to the biometric collection station
5. Fingerprints from both hands were collected using a 10-print device
6. Frontal and profile mug shots were captured

The overall environment of this process included a large open room with benches lining the walls where all of the inmates being processed sat awaiting their turn. On a given day there could be between 50 – 100 inmates to go through the intake process, with only two or three guards on duty. The guards attempted to complete this process as quickly as possible, but with so many inmates to get through this process may take their whole shift (8 hours).

The original electronic mug shots collected from the IDOC analyzed in Section 4.1 were consistent with the images captured in step 1 above. The images were collected with an non-compliant background with poor lighting. The camera used was an old digital camera, which produced poor quality and sometimes blurry images, thus the high amount of images non-compliant with the degree of blur variable. Regarding the other variables of concern from Section 4.1, images were captured in a very wide view and subsequently cropped by the IDOC guard collecting the image. During the cropping process, the guards used their best judgment to determine what looked best. The guard des not give the idsubject any verbal instruction other than to stand by the wall, so subjects were prone to tilting their head, not looking at the camera, and not being centered within the frame of reference. All of these issues are consistent with the variables of concern from Section 4.1 above.

The mug shots collected at the biometric capture station, along with the fingerprints, were stored on a separate computer in the facility, sent to the Indiana State Police to be stored in their system, and subsequently sent to the FBI. Figure 4.4 shows the biometric collection station being used. The camera enclosed in the top part of the capture station is a Canon PowerShot SX110 IS 8.0 megapixel digital camera. These images however are not stored in the IDOC database. Once they were submitted to the Indiana State Police they were stored on a machine stored within the IDOC intake facility and every morning photos that were captured 90 days prior were deleted.

Figure 4.4 IDOC Multimodal Biometric Capture Station

4.3. <u>Recommendations to Improve IDOC Capture Process</u>

The results from Section 4.1 and observations from Section 4.2, were used when determining the recommendations for IDOC. The set of proposed recommendations were limited to take into account IDOC business processes. Considering these constraints the following recommendations were made:

1. The images stored in the database should be the images captured at the biometric capture station
    a. This station has the proper camera (Canon PowerShot SX110 IS 8.0 megapixel digital camera)
    b. This station has the proper background (flat uniform gray)
2. Tape will be placed on the ground centered in relation to the width of the background, in the shape of a foot. The subject (inmate) shall be asked to stand with their feet within the tape.
3. The subject should be asked to place the back of their head against the background.

4. The subject should be asked to look directly into the camera with eyes open.
5. When cropping the image within the system, make sure the bottom of the crop box is closely aligned with the top of the neck collar and the face is centered in the box (see Figure 4.5 below).

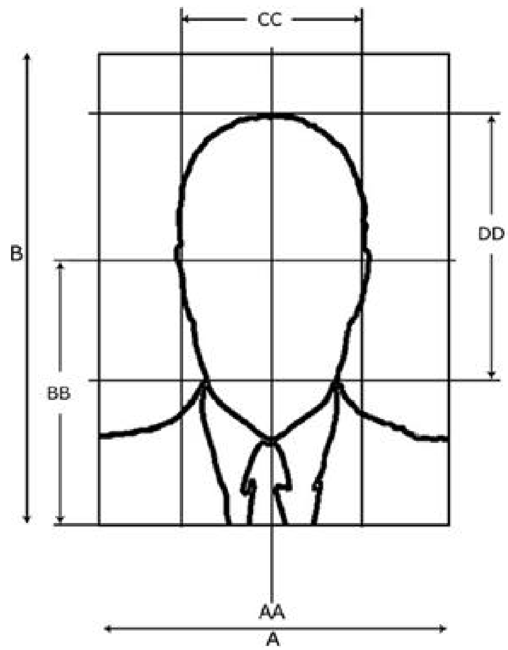


Figure 4.5 ISO Standard Geometric Characteristics for Full-Frontal Images (ISO/IEC, 2005)

Instruction was given to the personnel at the IDOC intake facility and Figure 4.5 was used in the guidance. All five recommendations were aimed at addressing the issue of centerline location ratio (which had 100% non-compliance), attempting to have the subjects' body and head centered or cropped within the image. Recommendations 1 through 4 were to address the degree of blur. Recommendation 1 was used to improve compliance with the percent of background gray variable, since the biometric capture station had the gray background setup. Recommendations 2 through 4 were developed to address the problematic variables of eye axis angle, eye axis location ratio, and eye separation. These recommendations were implemented on May 25, 2010 until June 15, 2010. This timeframe was chosen to allow for a large enough sample size to be collected.

4.4. Results IDOC Mug Shots Captured using Recommendations

The last section provides results from the IDOC images received on June 15, 2010 which were captured using the proposed recommendations. There were 953 IDOC images collected. These were collected over a three-week period with an average 318 images per week. The images were analyzed using the Aware PreFace program. The data were analyzed following the procedure used in Section 4.1. As before, the images were cleared of meta-data and renamed so no information was linked to the person in the image.

The data showed seven images returned an error (~0.73%), four were pose invalid, two were extreme pose or not a face image and one was multiple faces in the field of view. This showed a decrease from the ~1.24% of input errors for the original mug shots, (Table 4.4). Upon visual inspection these images all had similar issues. In all images the subject's head had an extreme head tilt causing improper yaw or roll angles. For three of the images the face was not centered and one had the halo effect.

Table 4.4 *Comparison of Input Errors across the Datasets*

Dataset	**Number of Input Errors**	**Percentage of Input Errors**
Electronic Images	617	1.24%
Images Captured using Recommendations	7	0.73%

The thresholds remain the same using the same NIST Best Practices profile in PreFace as before. Table 4.5 illustrates the number of images not compliant based on those thresholds.

Table 4.5 *PreFace Variable Thresholds for NIST Mug Shot Best Practice Profile*

PreFace Variable	Threshold	Number Not Compliant
Facial Dynamic Range	7 - 8 bits	315
Percent Background Uniformity	70 - 100%	43
Percent Background Gray	0 – 30%	87
Percent Facial Brightness	35 - 100%	291
Brightness Score	2 – 5	0
Eye Contrast	2 – 5	0
Degree of Clutter	= 0	228
Degree of Blur	= 0	0
Background Type	Simple	228
Eye Separation	>= 90 pixels	10
Eye Axis Angle	-5 - +5 degrees	62
Eye Axis Location Ratio	0.5 – 0.6	147
Centerline Location Ratio	= 0.5	953
Head Height/Image Height Ratio	0.6 – 0.8	140
Image Width/Head Width Ratio	1.995 – 2.004	951

From Table 4.5, not all images are compliant for every variable as was the case with the data from Section 4.1. However, there were three variables that did have all images compliant; brightness score, eye contrast, and degree of blur, whereas in Section 4.1 every variable had many images that were not compliant. To get a better picture of the overall distribution of the data for the mug shot captured post implementation histograms were produced, Figures 4.6 and 4.7.

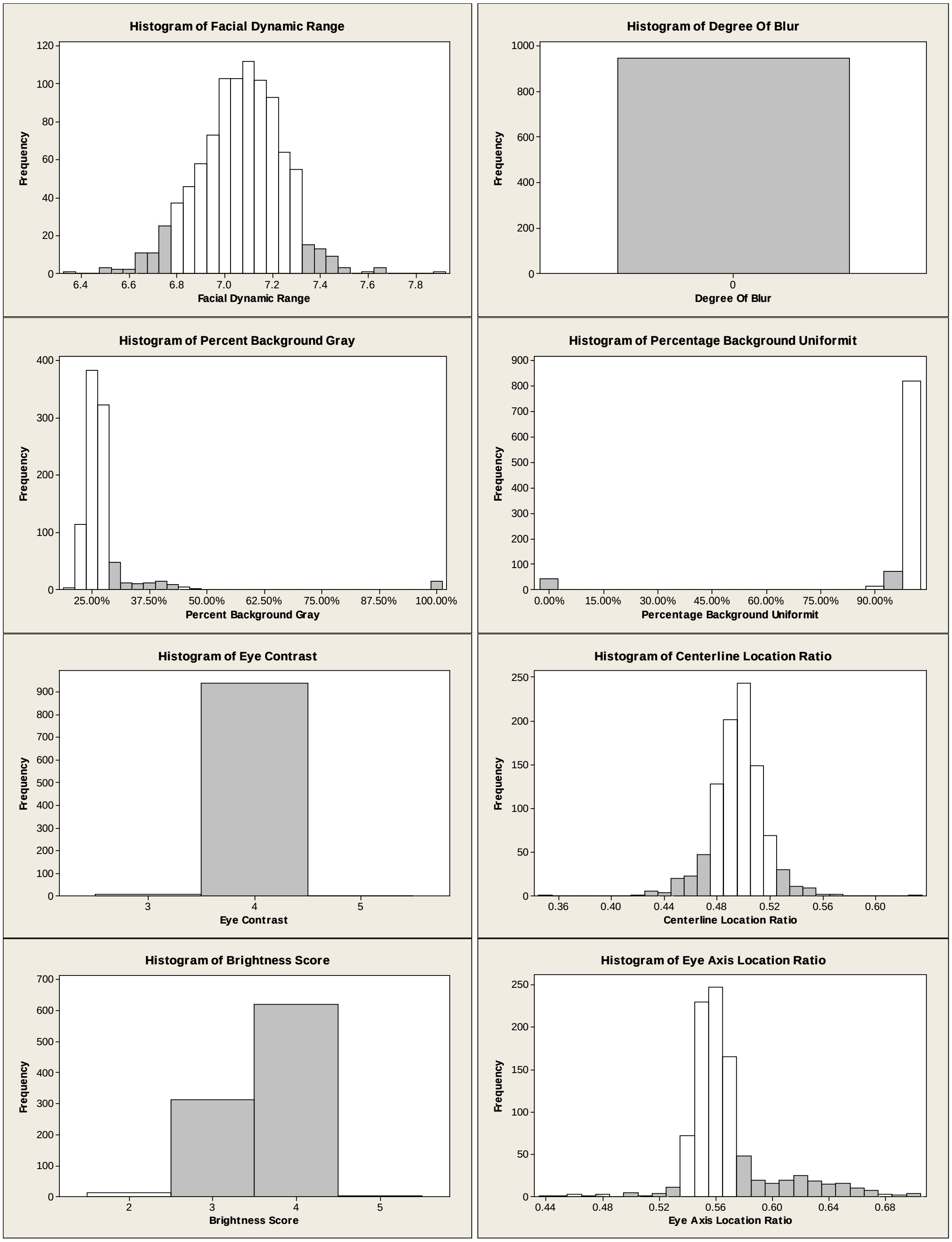


Figure 4.6 Histograms of PreFace Variables for Post Implementation Mug Shots

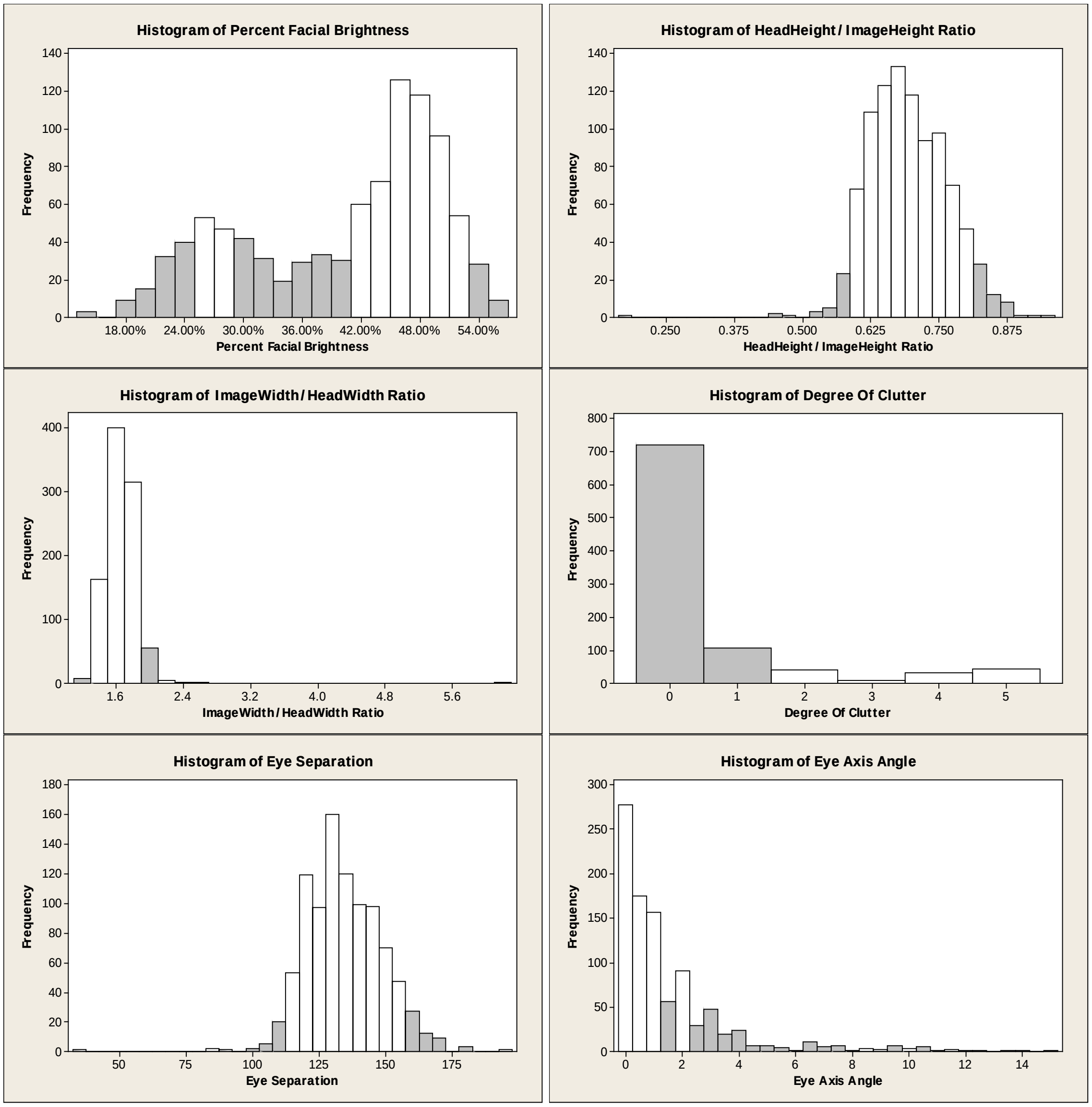


Figure 4.7 Histograms of PreFace Variables for Post Implementation Mug Shots (Continued)

From Section 4.1, the problematic variables of focus for this study included percent background gray, degree of blur, eye axis angle, eye location ratio, centerline location ratio, head height to image height ratio, image width to head width ratio, and eye separation. Looking closer at these variables, Table 4.6 illustrates the percent of non-compliant for the post implementation mug shots.

Table 4.6 *Problematic Variables Compliance for Post Implemenation Mug Shots*

PreFace Variable	Percent Not Compliant
Percent Background Gray	9.13%
Degree of Blur	0.00%
Eye Separation	1.05%
Eye Axis Angle	6.51%
Eye Axis Location Ratio	15.42%
Centerline Location Ratio	100%
Head Height/Image Height Ratio	14.69%
Image Width/Head Width Ratio	99.79%

While only one variable was reduced to zero percent, the number of images not compliant appears to be lower than the results from Section 4.1, with the exception of image width-to-head width ratio.

Again, the PreFace software attempted to optimize these images since none were already completely compliant with the NIST Best Practice profile. The software was unable to optimize any of these images (0.00%) compared to the ~0.04% in Section 4.1. This is not a great difference especially when taking into consideration the sample size of each dataset.

The post implementation images did have a better success rate for images where the optimization process was able to be completed (unsuccessfully). Of the 953 images, 770 (~80.80%) were processed by the software but the process was unable to optimize the image to be compliant. So while these images were not optimized for compliance, the fact that ~80.80% were able to be processed shows improvement when compared to the ~0.11% from Section 4.1. The remaining 183 images had errors associated with them resulting in an unsuccessful optimization attempt. The errors included 141 “Error 400: Input image not present”, 34 “Error 962: Invalid image width to head width ratio”, and one “Error 975: Image width and image-width-to-head-width ratio are inconsistent”.

4.5. Hypothesis Tests

Minitab® 16 software package was utilized. Since all nine tests are comparing percent of defectives essentially, the two-sample percent defective test was conducted for each. The test is based on a normal approximation and computes a confidence interval of the difference between the two proportions of defectives. In order for the test to be valid there should be at least five defective and five non-defective items in each sample. In addition the sample sizes should be sufficiently large enough to detect a difference between the percent defectives. In the event these constraints are not met to conduct the test based on normal approximation, the Fisher's exact test is conducted as it is accurate for all sample sizes.

4.5.1. Hypothesis 1 Comparing Overall Compliance

The first hypothesis establishes if there is a significant difference in compliance to the NIST Best Practice compliance and the original IDOC mug shots and the post implementation mug shots. In particular, the goal is to determine if the percent of non-compliant images is greater for the original IDOC mug shots versus the post implementation images. The null and alternate hypothesis can be shown as:

H_0: $p_{originalOverall} <= p_{post\text{-}implementationOverall}$

H_a: $p_{originalOverall} > p_{post\text{-}implementationOverall}$

(Where p = the percent of defective, or non-compliant images for the respective dataset)

Table 4.7 *Input Images Percent Non-Compliant for each Dataset*

Dataset	**Total Number of Images**	**Percent Non-Compliant**
Electronic Images	49,694	100.00%
Images Captured using Recommendations	953	100.00%

Table 4.7 shows the percent of non-compliant images for each dataset. As reported in the sections above, neither dataset had any images that were completely standard compliant with the NIST Best Practices PreFace profile. Figure B.1, in Appendix B, is a report showing that the test is not valid since there are not at least five defective and non-defective items for each dataset. Figure B.2 provides further emphasis on this point and reports on the power of the test. Since the two-proportion test based on the normal approximation is not valid, Figure 4.8 reports the results from the Fisher's exact test; it shows no significant difference between the two datasets. Due to this result, we fail to reject the null hypothesis and conclude that the percent non-compliant for post implementation dataset is not significantly less than the percent non-compliant for the original dataset.

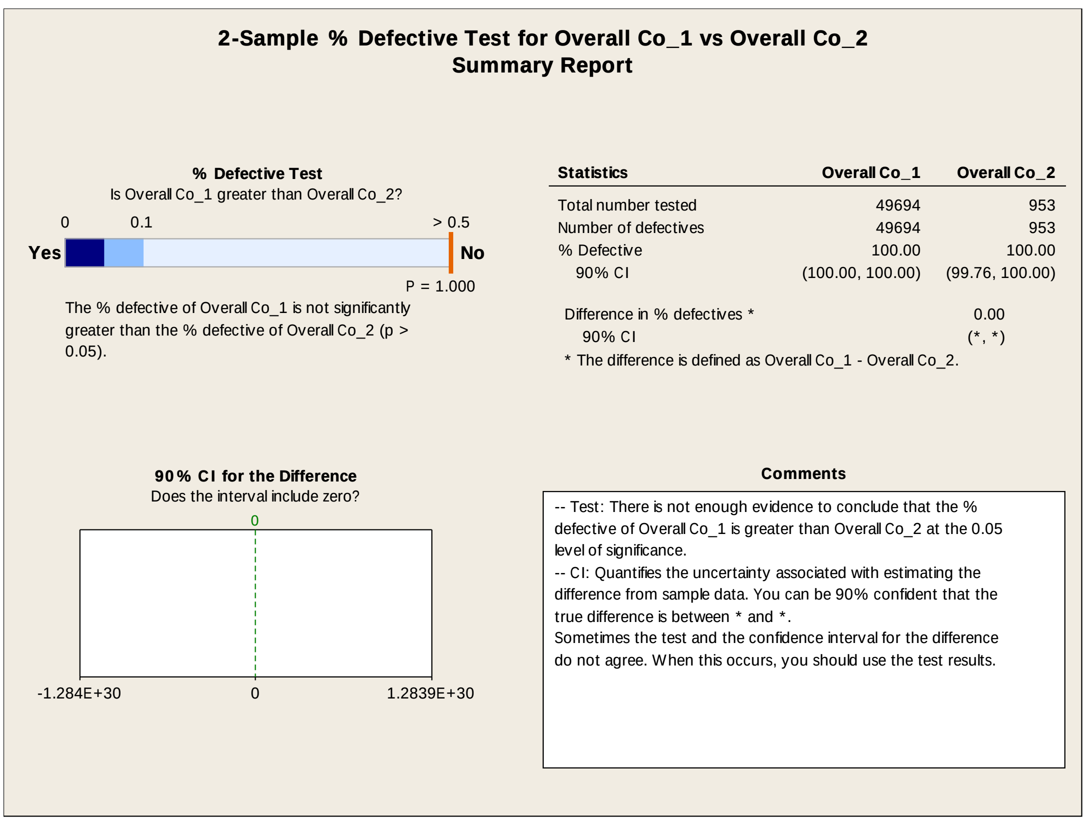


Figure 4.8 Minitab Output for Test of Hypothesis #1

4.5.2. Hypothesis 2 Comparing Percent Background Gray Compliance

Hypothesis #2 compares the percent background gray variable for the original IDOC mug shots and the post implementation mug shots. The hypothesis can be expressed by:

H_0: $p_{original\%BGgray} <= p_{post\text{-}implementation\%BGgray}$

H_a: $p_{original\%BGgray} > p_{post\text{-}implementation\%BGgray}$

(Where p = the percent of defective, or non-compliant images for the respective dataset)

Table 4.8 *Non-Compliant for Percent Background Gray for each Dataset*

Dataset	Total Number of Images	Percent Non-Compliant
Electronic Images	49,694	35.94%
Images Captured using Recommendations	953	9.13%

Table 4.8 shows the percent of non-compliance for each dataset, as can be seen the images captured post implementation of the process recommendations has a lower percent of non-compliant images by 26.81%. To test if this difference is significant the two-sample percent defective test was conducted.

Figure B.3 provides the report on the validity of the test. As can be seen the confidence intervals calculated can be considered valid since both samples had at least 5 defective and 5 non-defective images, and the samples sizes are sufficient to detect a difference.

Figure B.4 is the diagnostic report further depicting the validity of the test since the number of defective and non-defective images for each dataset was greater than five. This report also shows the power of the test, with α = 0.05 and

samples sizes of 49,694 and 953 there was a 90% chance of detecting of a difference greater than 4.51%.

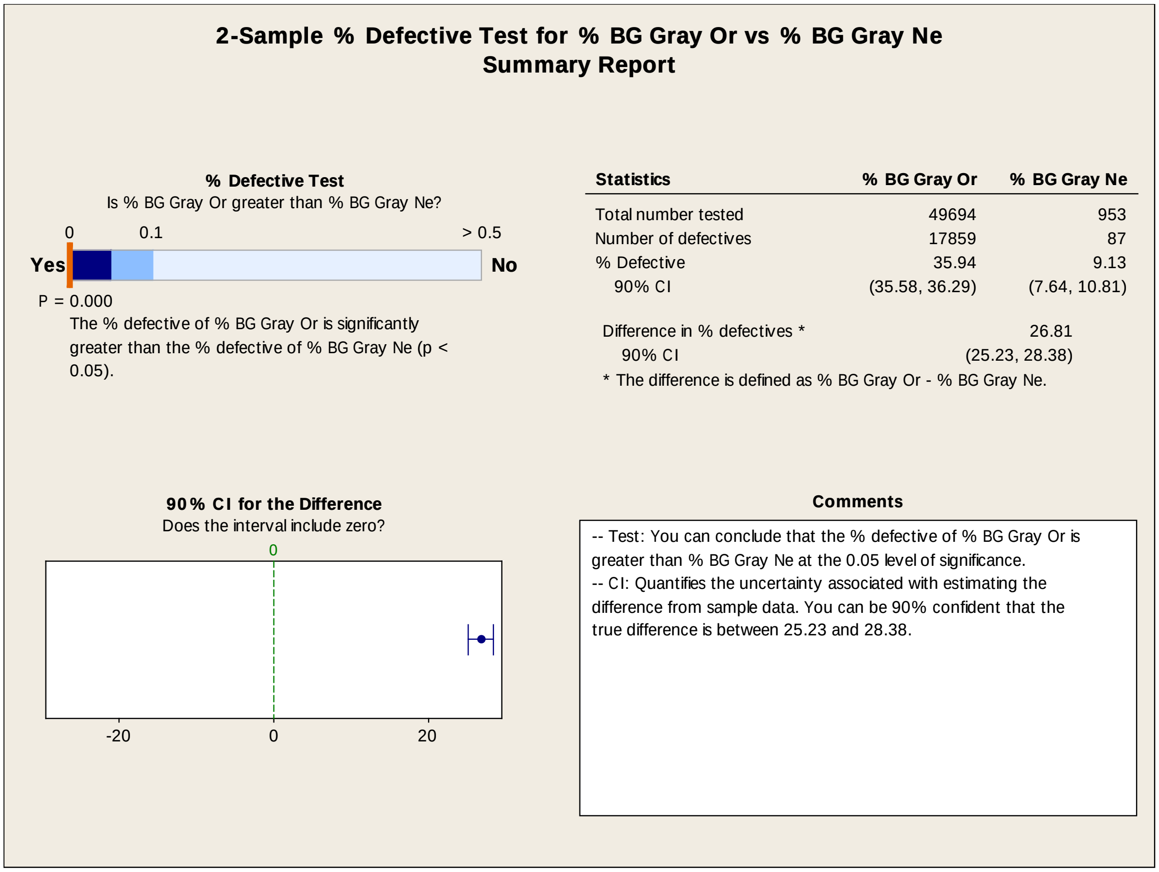


Figure 4.9 Minitab Output for Test of Hypothesis #2

Figure 4.9 is the output for the results of the two-sample percent defective test for Hypothesis #2. The difference in percent of defectives was 26.81% with the 90% confidence interval (25.53, 28.38). The *p*-value calculated equals 0.000 indicating that we can reject the null hypothesis and determine that the percent of non-compliant images in the post implementation dataset is significantly less than that of the original IDOC dataset.

4.5.3. Hypothesis 3 Comparing Degree of Blur Compliance

Hypothesis #3 compares degree of blur variable for the original IDOC mug shots and the post implementation mug shots. The hypothesis can be expressed by:

H_0: $p_{originalDegreeBlur} <= p_{post\text{-}implementationDegreeBlur}$

H_a: $p_{originalDegreeBlur} > p_{post\text{-}implementationDegreeBlur}$

(Where p = the percent of defective, or non-compliant images for the respective dataset)

Table 4.9 shows the percent of non-compliance for each dataset. The images captured post implementation of the process recommendations has a lower percent of non-compliant images by 34.91% To test if this difference is significant the two-sample percent defective test was conducted.

Table 4.9 *Non-Compliant for Degree of Blur for each Dataset*

Dataset	**Total Number of Images**	**Percent Non-Compliant**
Electronic Images	49,694	34.91%
Images Captured using Recommendations	953	0.00%

Figure B.5 provides the report on the validity of the test. As can be seen the confidence intervals calculated cannot be considered fully valid since the post implementation dataset did not have any non-compliant images. The check for validity indicates that at least five defective and non-defective images should be present, which is not the case for this variable.

Figure B.6 is the diagnostic report further depicting the non-validity of the test since the number of defective and non-defective images for each dataset was less than five. Since this is the case the software conducted Fisher's exact to test the hypothesis.

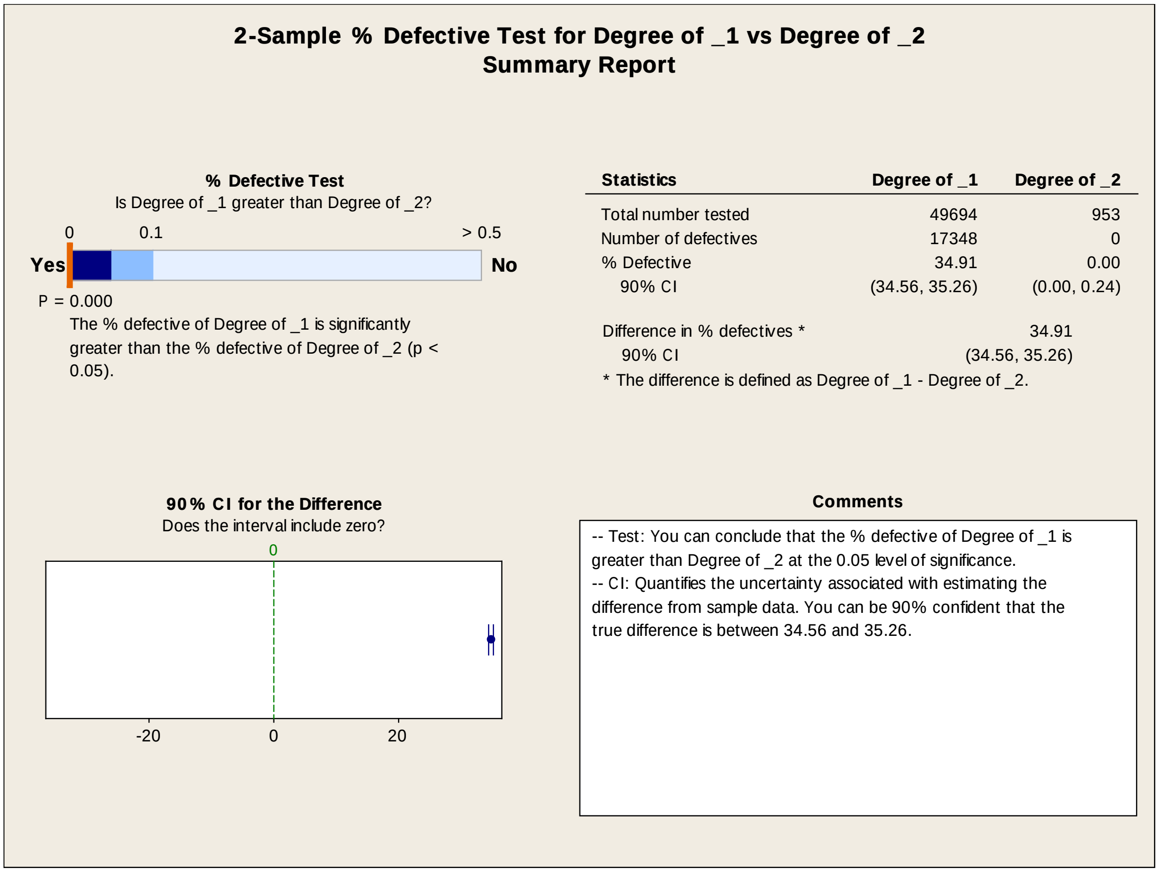


Figure 4.10 Minitab Output for Test of Hypothesis #3

Figure 4.10 is the output for the results of the Fisher’s exact test for Hypothesis #3 since the two-sample percent defective test was invalid. The difference in percent of defectives was 34.91% with the 90% confidence interval (34.56, 35.26). The *p*-value calculated equals 0.000 indicating that we can reject the null hypothesis and determine that the percent of non-compliant images in the post implementation dataset is significantly less than that of the original IDOC dataset.

4.5.4. Hypothesis 4 Comparing Eye Separation Compliance

Hypothesis #4 compares the eye separation variable for the original IDOC mug shots and the post implementation mug shots. The hypothesis can be expressed by:

H_0: $p_{originalEyeSeparation} <= p_{post\text{-}implementationEyeSeparation}$

H_a: $p_{originalEyeSeparation} > p_{post\text{-}implementationEyeSeparation}$

(Where p = the percent of defective, or non-compliant images for the respective dataset)

Table 4.10 shows the percent of non-compliance for each dataset, as can be seen the images captured post implementation of the process recommendations has a lower percent of non-compliant images by 53.90% To test if this difference is significant the two-sample percent defective test was conducted.

Table 4.10 *Non-Compliant for Eye Separation for each Dataset*

Dataset	Total Number of Images	Percent Non-Compliant
Electronic Images	49,694	54.95%
Images Captured using Recommendations	953	1.05%

Figure B.7 provides the report on the validity of the test. As can be seen the confidence intervals calculated can be considered valid since both samples had at least 5 defective and 5 non-defective images, and the samples sizes are sufficient to detect a difference.

Figure B.8 is the diagnostic report further depicting the validity of the test since the number of defective and non-defective images for each dataset was greater than five. This report also shows the power of the test, with α = 0.05 and

samples sizes of 49,694 and 953 there was a 90% chance of detecting of a difference greater than 4.78%.

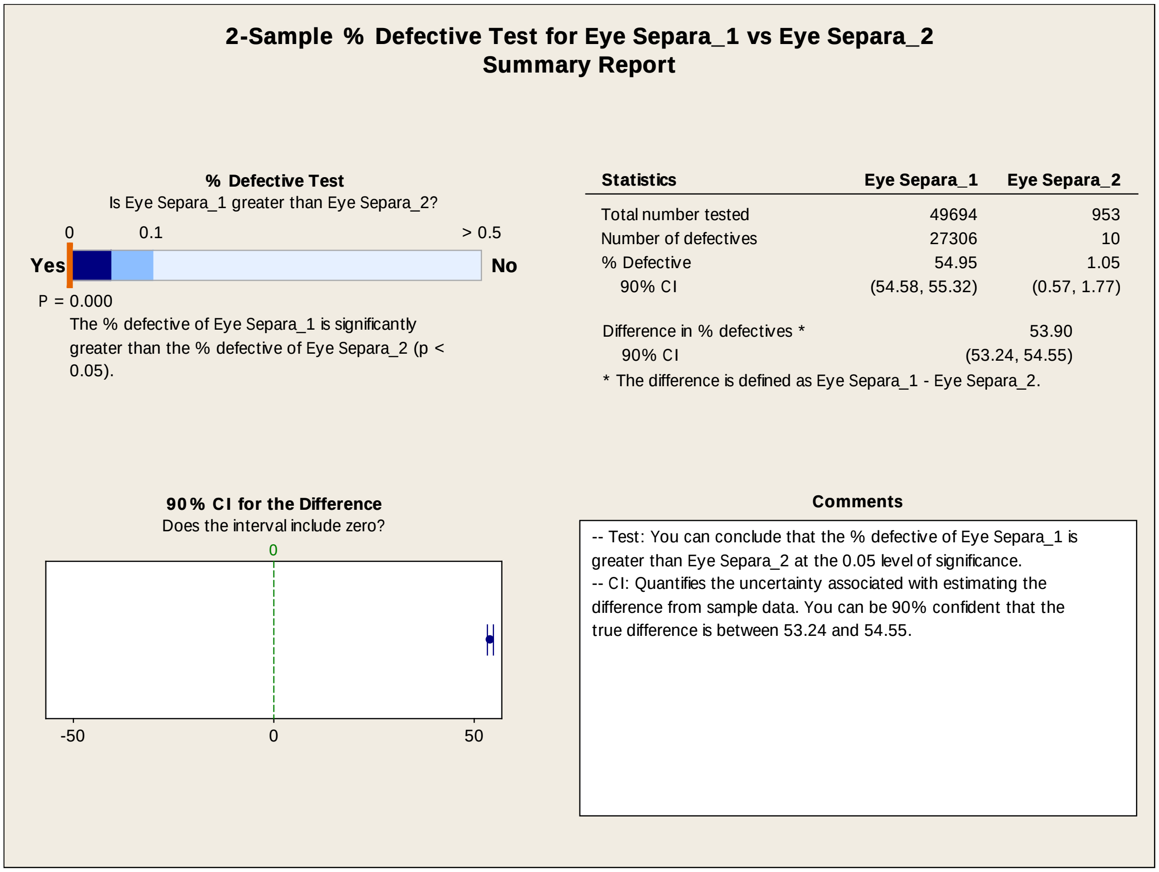


Figure 4.11 Minitab Output for Test of Hypothesis #4

Figure 4.11 is the output for the results of the two-sample percent defective test for Hypothesis #4. The difference in percent of defectives was 53.90% with the 90% confidence interval (53.24, 54.55). The *p*-value calculated equals 0.000 indicating that we can reject the null hypothesis and determine that the percent of non-compliant images in the post implementation dataset is significantly less than that of the original IDOC dataset.

4.5.5. Hypothesis 5 Comparing Eye Axis Angle Compliance

Hypothesis #5 compares eye axis angle variable for the original IDOC mug shots and the post implementation mug shots. The hypothesis can be expressed by:

H_0: $p_{originalEyeAxisAngle} <= p_{post\text{-}implementationEyeAxisAngle}$

H_a: $p_{originalEyeAxisAngle} > p_{post\text{-}implementationEyeAxisAngle}$

(Where p = the percent of defective, or non-compliant images for the respective dataset)

Table 4.11 shows the percent of non-compliance for each dataset, as can be seen the images captured post implementation of the process recommendations has a lower percent of non-compliant images by 12.97% To test if this difference is significant the two-sample percent defective test was conducted.

Table 4.11 *Non-Compliant for Eye Axis Angle for each Dataset*

Dataset	Total Number of Images	Percent Non-Compliant
Electronic Images	49,694	19.48%
Images Captured using Recommendations	953	6.51%

Figure B.9 provides the report on the validity of the test. As can be seen the confidence intervals calculated can be considered valid since both samples had at least 5 defective and 5 non-defective images, and the samples sizes are sufficient to detect a difference.

Figure B.10 is the diagnostic report further depicting the validity of the test since the number of defective and non-defective images for each dataset was greater than five. This report also shows the power of the test, with the α = 0.05

and samples sizes of 49,694 and 953 there was a 90% chance of detecting of a difference greater than 3.65%.

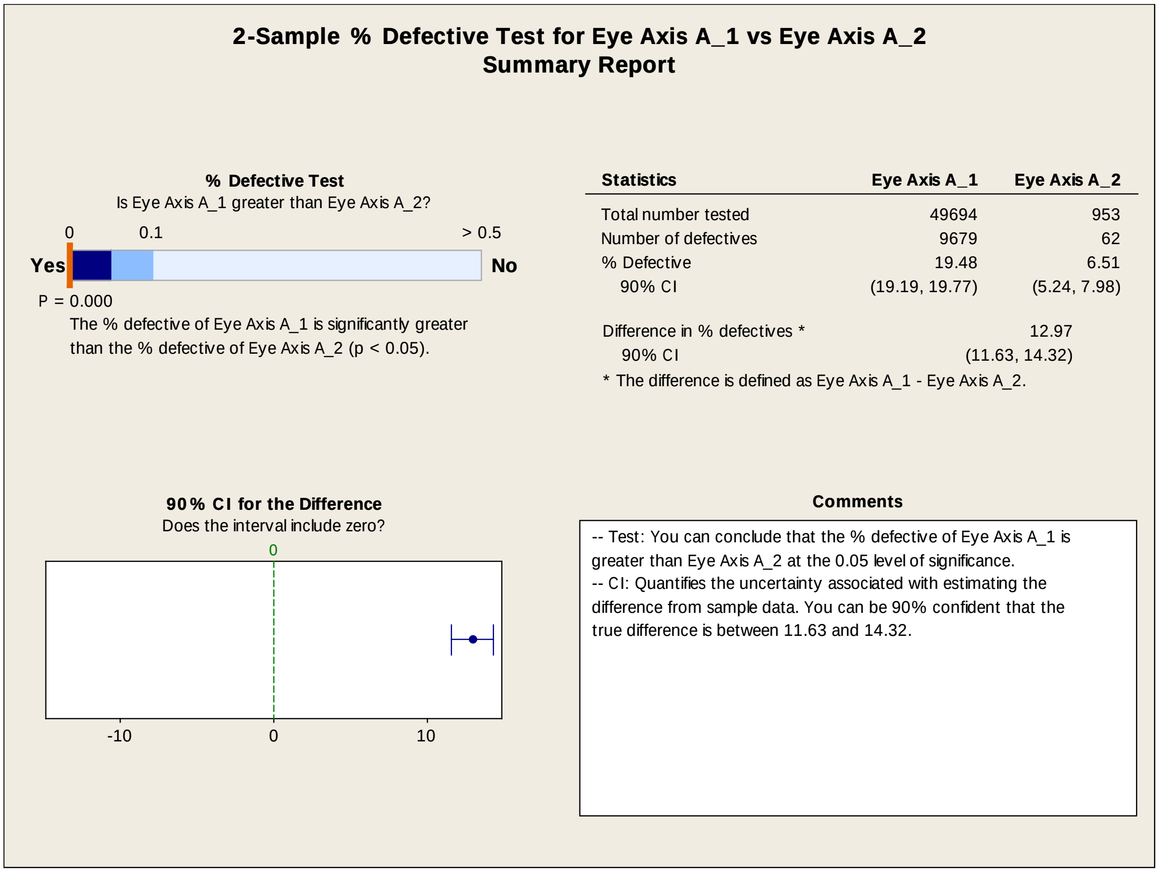


Figure 4.12 Minitab Output for Test of Hypothesis #5

Figure 4.12 is the output for the results of the two-sample percent defective test of Hypothesis #5. The difference in percent of defectives was 12.97% with the 90% confidence interval (11.63, 14.32). The *p*-value calculated equals 0.000 indicating that we can reject the null hypothesis and determine that the percent of non-compliant images in the post implementation dataset is significantly less than that of the original IDOC dataset.

4.5.6. Hypothesis 6 Comparing Eye Axis Location Ratio Compliance

Hypothesis #6 compares the eye axis location ratio variable for the original IDOC mug shots and the post implementation mug shots. The hypothesis can be expressed by:

H_0: $p_{originalEyeAxisLocRatio} <= p_{post\text{-}implementationEyeAxisLocRatio}$

H_a: $p_{originalEyeAxisLocRatio} > p_{post\text{-}implementationEyeAxisLocRatio}$

(Where p = the percent of defective, or non-compliant images for the respective dataset)

Table 4.12 shows the percent of non-compliance for each dataset, as can be seen the images captured post implementation of the process recommendations has a lower percent of non-compliant images by 29.48% To test if this difference is significant the two-sample percent defective test was conducted.

Table 4.12 *Non-Compliant for Eye Axis Location Ratio for each Dataset*

Dataset	**Total Number of Images**	**Percent Non-Compliant**
Electronic Images	49,694	44.90%
Images Captured using Recommendations	953	15.42%

Figure B.11 provides the report on the validity of the test. As can be seen the confidence intervals calculated can be considered valid since both samples had at least 5 defective and 5 non-defective images, and the samples sizes are sufficient to detect a difference.

Figure B.12 is the diagnostic report further depicting the validity of the test since the number of defective and non-defective images for each dataset was greater than five. This report also shows the power of the test, with $\alpha = 0.05$ and

samples sizes of 49,694 and 953 there was a 90% chance of detecting of a difference greater than 4.72%.

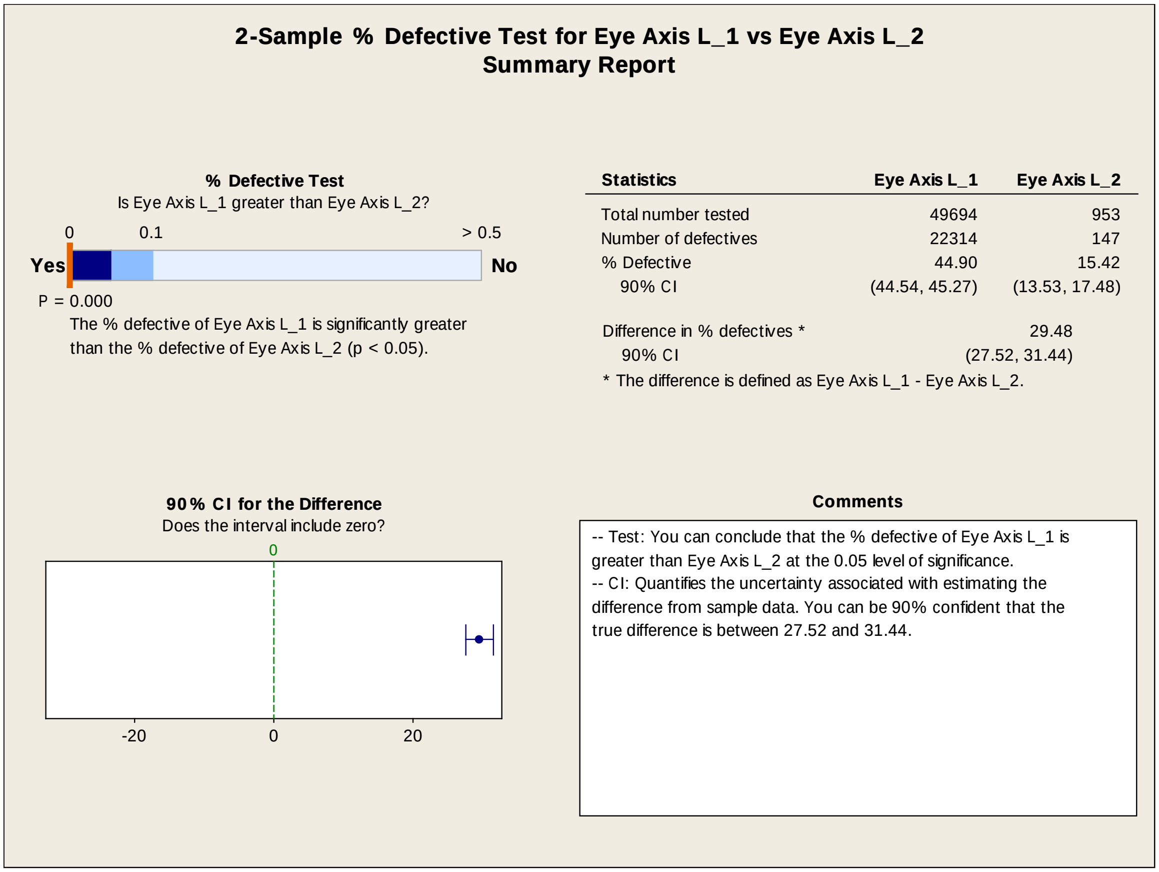


Figure 4.13 Minitab Output for Test of Hypothesis #6

Figure 4.13 is the output for the results of the two-sample percent defective test for Hypothesis #6. The difference in percent of defectives was 29.48% with the 90% confidence interval (27.52, 31.44). The *p*-value calculated equals 0.000 indicating that we can reject the null hypothesis and determine that the percent of non-compliant images in the post implementation dataset is significantly less than that of the original IDOC dataset.

4.5.7. Hypothesis 7 Comparing Centerline Location Ratio Compliance

Hypothesis #7 compares the centerline location ratio variable for the original IDOC mug shots and the post implementation mug shots. The hypothesis can be expressed by:

H_0: $p_{originalCenterlineLocRatio} <= p_{post\text{-}implementationCenterlineLocRatio}$

H_a: $p_{originalCenterlineLocRatio} > p_{post\text{-}implementationCenterlineLocRatio}$

(Where p = the percent of defective, or non-compliant images for the respective dataset)

Table 4.13 shows the percent of non-compliance for each dataset, as can be seen both datasets had 100.00% non-compliance for the centerline location ratio variable. The two-sample percent defective test was conducted to test the hypothesis.

Table 4.13 *Non-Compliant for Centerline Location Ratio for each Dataset*

Dataset	**Total Number of Images**	**Percent Non-Compliant**
Electronic Images	49,694	100.00%
Images Captured using Recommendations	953	100.00%

Figure B.13 provides the report on the validity of the test. As can be seen the confidence intervals calculated cannot be considered valid since neither samples had at least 5 defective and 5 non-defective images.

Figure B.14 is the diagnostic report further depicting the non-validity of the test since the number of defective and non-defective images for each dataset was less than five. Since this is the case the software conducted Fisher's exact to test the hypothesis.

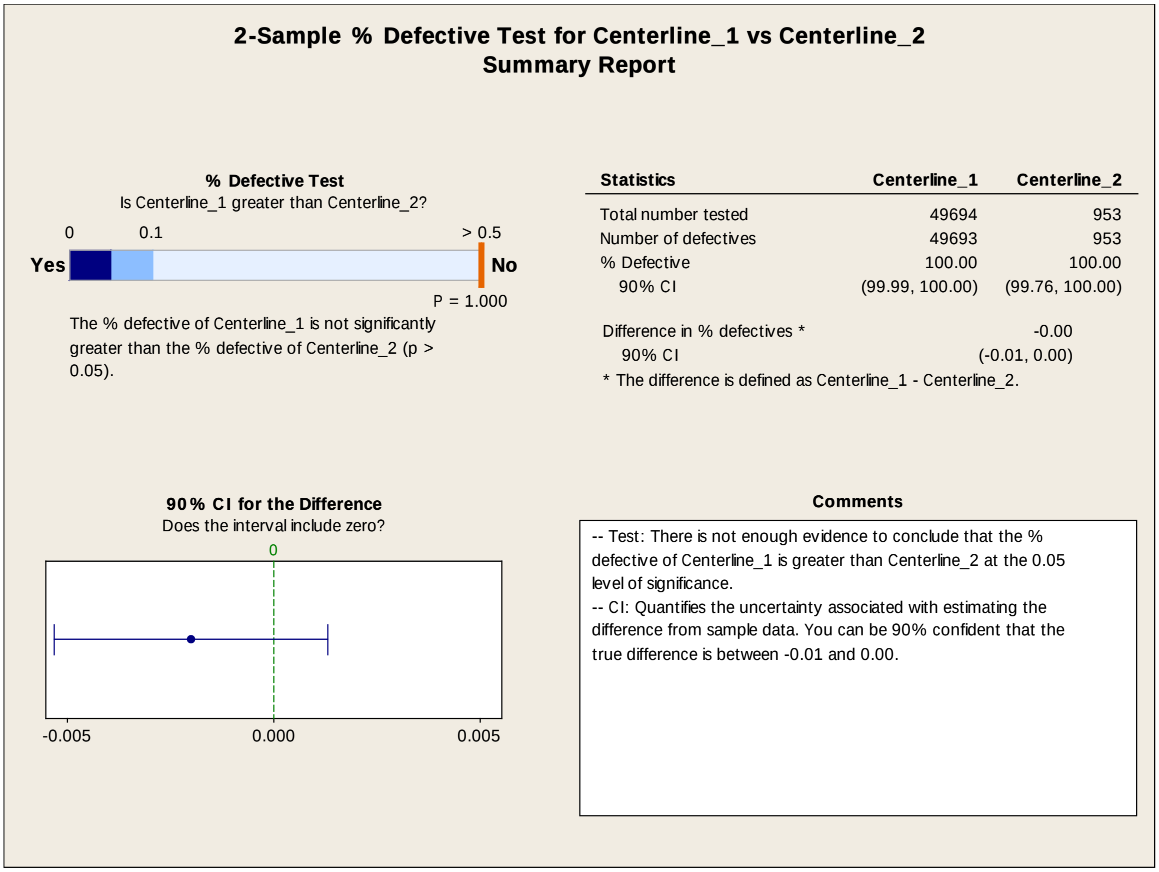


Figure 4.14 Minitab Output for Test of Hypothesis #7

Figure 4.14 is the output for the results of the Fisher's exact test for Hypothesis #7 since the two-sample percent defective test was invalid. The difference in percent of defectives was 0.00% with the 90% confidence interval (-0.01, 0.00). The *p*-value calculated was greater than 0.05 indicating that we fail to reject the null hypothesis and determine that the percent of non-compliant images in the post implementation dataset is not significantly less than that of the original IDOC dataset.

4.5.8. Hypothesis 8 Comparing Head Height to Image Height Ratio Compliance

Hypothesis #8 compares head height to image height ratio variable for the original IDOC mug shots and the post implementation mug shots. The hypothesis can be expressed by:

H_0: $p_{originalHeadHeight/ImageHeightRatio} <= p_{post\text{-}implementationHeadHeight/ImageHeightRatio}$

H_a: $p_{originalHeadHeight/ImageHeightRatio} > p_{post\text{-}implementationHeadHeight/ImageHeightRatio}$

(Where p = the percent of defective, or non-compliant images for the respective dataset)

Table 4.14 shows the percent of non-compliance for each dataset, as can be seen the images captured post implementation of the process recommendations has a lower percent of non-compliant images by 32.09% To test if this difference is significant the two-sample percent defective test was conducted.

Table 4.14 *Non-Compliant for Head Height-Image Height Ratio for each Dataset*

Dataset	**Total Number of Images**	**Percent Non-Compliant**
Electronic Images	49,694	46.78%
Images Captured using Recommendations	953	14.69%

Figure B.15 provides the report on the validity of the test. The confidence intervals calculated can be considered valid since both samples had at least 5 defective and 5 non-defective images, and the samples sizes are sufficient to detect a difference.

Figure B.16 is the diagnostic report further depicting the validity of the test since the number of defective and non-defective images for each dataset was greater than five. This report also shows the power of the test, with the $\alpha = 0.05$ and samples sizes of 49,694 and 953 there was a 90% chance of detecting of a difference greater than 4.75%.

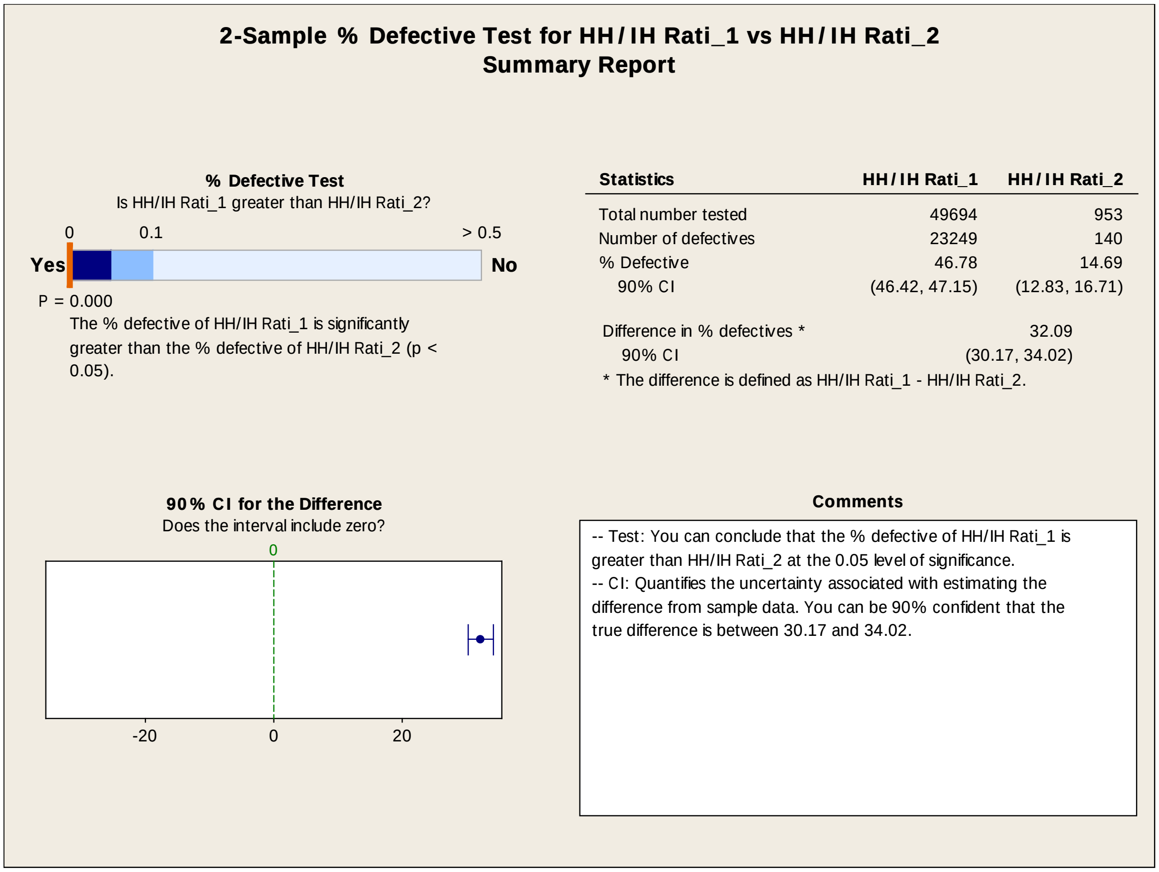


Figure 4.15 Minitab Output for Test of Hypothesis #8

Figure 4.15 is the output for the results of the two-sample percent defective test for Hypothesis #8. The difference in percent of defectives was 32.09% with the 90% confidence interval (30.17, 34.02). The *p*-value calculated equals 0.000 indicating that we can reject the null hypothesis and determine that the percent of non-compliant images in the post implementation dataset is significantly less than that of the original IDOC dataset.

4.5.9. Hypothesis 9 Comparing Image Width to Head Width Ratio Compliance

Hypothesis #9 compares the image width to head width ratio variable for the original IDOC mug shots and the post implementation mug shots. The hypothesis can be expressed by:

H_0: $p_{originalImageWidth/HeadWidthRatio} <= p_{post\text{-}implementationImageWidth/HeadWidthRatio}$

H_a: $p_{originalImageWidth/HeadWidthRatio} > p_{post\text{-}implementationImageWidth/HeadWidthRatio}$
(Where p = the percent of defective, or non-compliant images for the respective dataset)

Table 4.15 shows the percent of non-compliance for each dataset, as can be seen the both datasets had a very high percentage of the images non-compliant, and the original dataset was less than the post implementation dataset. The two-sample percent defective test was conducted to test the difference.

Table 4.15 *Non-Compliant for Image Width-Head Width Ratio for each Dataset*

Dataset	**Total Number of Images**	**Percent Non-Compliant**
Electronic Images	49,694	99.37%
Images Captured using Recommendations	953	99.79%

Figure B.17 provides the report on the validity of the test. As can be seen the confidence intervals calculated cannot be considered valid since the post implementation data only had two non-defective images.

Figure B.18 is the diagnostic report further depicting the non-validity of the test since the number of non-defective images for the post implementation dataset was less than five. Since this is the case the software conducted Fisher's exact to test the hypothesis.

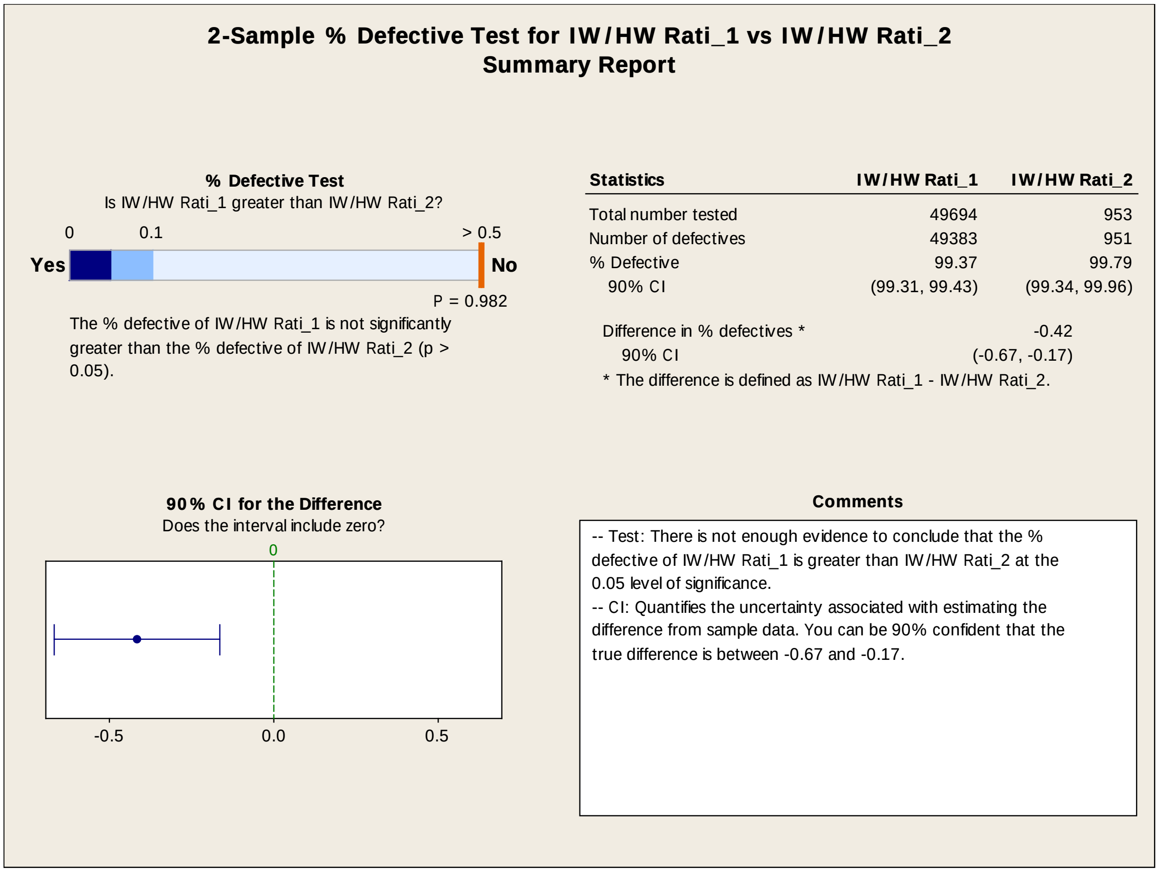


Figure 4.16 Minitab Output for Test of Hypothesis #9

Figure 4.16 is the output for the results of the Fisher's exact test for Hypothesis #9 since the two-sample percent defective test was invalid. The difference in percent of defectives was -0.42% with the 90% confidence interval (-0.67, -0.17). The *p*-value calculated was greater than 0.05 indicating that we fail to reject the null hypothesis and determine that the percent of non-compliant images in the post implementation dataset is not significantly less than that of the original IDOC dataset.

Table 4.16 shows the summary of all the hypotheses tested and the results from each test respectively.

Table 4.16 *Summary of Results for all Hypotheses*

Hypothesis	**Original % Non-Compliant**	**Post % Non-Compliant**	**Difference**	**p-value**
Hypothesis #1	100.00%	100.00%	0.00%	> 0.05
Hypothesis #2	35.94%	9.13%	26.81%	0.000
Hypothesis #3	34.91%	0.00%	34.91%	0.000
Hypothesis #4	54.95%	1.05%	53.90%	0.000
Hypothesis #5	19.48%	6.51%	12.97%	0.000
Hypothesis #6	44.90%	15.42%	29.48%	0.000
Hypothesis #7	100.00%	100.00%	0.00%	> 0.05
Hypothesis #8	46.78%	14.69%	32.09%	0.000
Hypothesis #9	99.37%	99.79%	-0.42%	> 0.05

CHAPTER 5. FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

This study examined the standard compliance of an operational mug shot dataset, and the process by which the images are captured. Many tests have been conducted (Phillps, P., et. al, 2004, Phillips, P., et. al, 2007, Theofanos, M., et. al, 2008, Theofanos, M., et. al, 2009 and Grother, P., et. al, 2010) analyzing face recognition performance, quality, and image capture process. The only tests that used operational data were the 2008 NISTIR 7540 - Assessing Face Acquisition, the National Institute of Standards and Technology test and the 2009 Usability Testing of an Overlay to Improve Face Capture conducted by Theofanos (2009).

5.1.1. Conclusions

The results of the study, Table 5.1, indicate the Indiana Department of Correction's mug shot capture process, while not overall more standard compliant, many of the individual characteristics that were pinpointed as flaws were significantly improved. The results show that the images captured, even with the process improvement recommendations, are not standard compliant with the NIST Best Practice recommendations in ISO/IEC Biometric Interchange Formats standard Part 5. The results however do show that the process did improve the individual variables that were highly problematic. The variables that showed significant improvement included percent background gray, degree of blur, eye separation, eye axis angle, eye axis location ratio, and head height to image height ratio.

One characteristic that was not compliant at all for either dataset was the centerline location ratio, the precise threshold set for this variable may be too exact to be relevant in an operational environment. This finding should be explored more in the standards community to determine if this threshold should be changed to a range instead of a precise value to allow for more flexibility that is needed in an operational environment. Other variables still not standard compliant are out the scope for this particular study due to environmental and funding constraints on the IDOC. However, these improvements along with results from Grother (2010) it can still be recommended that the IDOC continue to improve their image capture process, in accordance with the findings outlined in this paper.

5.1.2. Future Work

During the course of this study, observations emerged in areas in which this study could be improved upon. The discussion of these observations and recommendations is as follows:

1. This study only took into account the results from one commercially available software package that analyzes standard compliance of face images. Further studies could include other software packages to determine accuracy of results.
2. Due to the IRB restrictions no matching performance rates could be calculated. Future studies could attempt to collect data using the operational constraints but in a lab environment so that matching performance can be calculated.
3. During this study due to financial restrictions no lighting techniques were taken into consideration to improve the compliance. In the future some lighting techniques could be used to attempt to further improve the capture process.

4. During this study it was determined that the centerline location ratio value in the standards may be too precise for practical use with operational data. Further research should evaluate this characteristic and determine if this value is useful or should be expanded to a range.

Table 5.1 *Summary of Results for all Hypotheses*

Hypothesis	**Original % Non-Compliant**	**Post % Non-Compliant**	**Difference**	**p-value**
Hypothesis #1	100.00%	100.00%	0.00%	> 0.05
Hypothesis #2	35.94%	9.13%	26.81%	0.000
Hypothesis #3	34.91%	0.00%	34.91%	0.000
Hypothesis #4	54.95%	1.05%	53.90%	0.000
Hypothesis #5	19.48%	6.51%	12.97%	0.000
Hypothesis #6	44.90%	15.42%	29.48%	0.000
Hypothesis #7	100.00%	100.00%	0.00%	> 0.05
Hypothesis #8	46.78%	14.69%	32.09%	0.000
Hypothesis #9	99.37%	99.79%	-0.42%	> 0.05

REFERENCES

REFERENCES

Adler, A., & Schuckers, M. E. (2007). Comparing Human and Automatic Face Recognition Performance. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, *37*(5), 1248-1255.

ANSI/INCITS. (2004). Face Recognition Format for Data Interchange. ANSI.

ANSI/NIST. (2007). Data Format for the Interchange of Fingerprint Facial, and other Biometric Information - Part 1. ANSI.

Association for Biometrics. (1999). 1999 Glossary of Biomtric Terms. Association for Biometrics and International Computer Security Association. Retrieved March 10, 2009, from www.biometrieonline.net/dossiers/generalites/glossary.pdf.

Aware, Inc. (2007). PreFace by Aware: Version 4.2 for Windows Developer's Guide. Bedford, MA. Aware, Inc.

Bensheimer, K. (December 19, 2009). Personal Communication.

Gonzalez-Castillo, O. Y. (2006). Survey About Facial Image Quality. Fraunhofer Institite for Computer Graphics Research. Retrieved February 24, 2009, from http://www.face-rec.org/interestingpapers/standards/2dfiqsurveyreport_oygc.pdf.

Grother, P.J., Quinn, G.W., & Phillips, P.J. (2010) Report on the Evalution of 2D Still-Image Face Recognition Algorithms. NIST. Retrieved June 22, 2010, from http://biometrics.nist.gov/cs_links/face/mbe/mbe_2D_face_report_NISTIR_7709.pdf.

FBI.gov. (2010). Integrated Automated Fingerprint Identification System. Retrieved October 26, 2010, from http://www.fbi.gov/aboutus/cjis/fingerprints_biometrics/iafis/iais.

Hannavy, J. (Ed.). (2008). Bertillion, Alphonse. In *Encyclopedia of Nineteenth-Century Photography* (Vol. 1, pp. 150-152). New York: Routledge.

Higgins, P. T. (2005). Introduction to Biometrics. Presented at the Biometrics Consortium Conference, Arlington, VA.

Internation Biometric Group. (2005). Independent Testing of Iris Recognition Technology. Final Report. Retrieved November 20, 2009, from http://www.biometricscatalog.org/ITIRT/ITIRT-FinalReport.pdf.

ISO/IEC. (2005). Part 5: Face image data, in 19794 Information Technology - Biometric data interchange formats. ISO/IEC. Retrieved February 22, 2009, from http://webstore.iec.ch/preview/info_isoiec19794-5%7bed1.0%7den.pdf.

International Organization for Standardization/International Electrotechnical Commission. (2005). Information Technology – Biometric Data Interchange Formats – Part 5: Face Image Data. Geneva, Switzerland. ISO/IEC.

Jain, A. K., Ross, A. A., & Prabhakar, S. (2004). An Introduction to Biometric Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, *14*(1), 4-20.

Klobuchar, A., Mehrkens Steblay, N. K., & Lindell Caligiuri, H. (2006). Reforming Eyewitness Identification: Convicting the Guilty, Protecting the Innocent. *Cardozo Public Law, Policy and Ethics Journal*. Retrieved from http://www.ccfaj.org/documents/reports/eyewitness/federal/Hennepin%20County.pdf. Kukula, E. P., & Elliott, S. J. (2004). Evaluation of a Facial Recognition Algorithm across three Illumination Conditions. *IEEE Aerospace and Electronic Systems Magazine*, *19*(9), 19-23.

Maio, D., Maltoni, D., Wayman, J., & Jain, A. (2004). FVC2004: Third Fingerprint Verification Competition (pp. 1-7). Presented at the International Conference on Biometric Authentication (ICBA), Hong Kong, China. Retrieved November 20, 2009, from http://biometrics.cse.msu.edu/Publications/Fingerprint/FVC/Maioetal_FVC2004_ICBA04.pdf.

Mansfield, A., & Wayman, J. (2002). Best Practices in Testing and Reporting Performance of Biometric Devices. Biometric Working Group. Retrieved March 10, 2009, from http://www.google.com/search?hl=en&q=best+practices+in+testing+and+reporting+performance+of+biometric+devices&btnG=Google+Search&aq=1&oq=Best+Practices+in+Testing+.

McCabe, M. R. (1997, September 23). Best Practice Recommendation for the Capture of Mug shots. National Institute of Standards and Technology. Retrieved February 21, 2009, from http://www.itl.nist.gov/iaui/894.03/face/bpr_mug3.html.

Mecklenburg, S. H., Larson, M. R., & Bailey, P. J. (2008). Eyewitness Identification: What Chiefs Need to Know Now. The Police Chief, 75(10), 68-81.

Moore, D.S., & McCabe, G.P. (2006). *Introduction to the Practice of Statistics* (5th ed.).. New York, NY: W. H. Freeman and Company.

Petersen, J. K. (2001). *Understanding Surveillance Technologies: Spy Devices, Their Origins, & Applications*. Boca Raton, FL: CRC Press.

Phillips, P., Flynn, P., Scruggs, T., Bowyer, K., & Worek, W. (2006). Preliminary Face Recognition Grand Challenge Results. Southhampton, UK.

Phillips, P. J., Scruggs, W. T., O'Toole, A. J., Flynn, P. J., Bowyer, K. W., Schott, C. L., et al. (2007). FRVT 2006 and ICE 2006 Large-Scale Results. NIST. Retrieved February 22, 2009, from http://iris.nist.gov/ICE/FRVT2006andICE2006LargeScaleReport.pdf.

Przybocki, M., & Martin, A. (2004). NIST Speaker Recognition Evaluation Chronicles (pp. 12-22). Presented at the Odyssey: The Speaker and Language Recognition Workshop, Toledo, Spain. Retrieved November 20, 2009, from http://www.itl.nist.gov/iad/IADpapers/2004/ABSFBE1562.pdf.

Sekaran, U. (2003) *Research methods for business: A skill building approach.* Hoboken, NJ: John Wiley & Sons, Inc.

Shah, J., & Martin, B. (2007) Image Enhancements, Biometric (Sample) Quality and Recognition Accuracy, in *NIST Biometric Quality Workshop II*. Gaithersburg:NIST.Spaun, N. A. (2007). Forensic Biometrics from Images and Video at the Federal Bureau of Investigation. In *BTAS 2007* (pp. 1-3).

Spaun, N. A. (2007). Forensic Biometrics from Images and Video at the Federal Bureau of Investigation. In *IEEE* (pp. 1-3). Presented at the First IEEE Internation Conference on Biometrics: Theory, Applications, and Systems, 2007, IEEE. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4401932&isnumber=4401902.

Theofanos, M., Stanton, B., Sheppard, C., Micheals, R., Libert, J., & Orandi, S. (2008, September). NISTIR 7540 - Assessing Face Acquisition. National Institute of Standards and Technology. Retrieved September 12, 2009, from http://zing.ncsl.nist.gov/biousa/docs/face_IR-7540.pdf.

Theofanos, M. F., Stanton, B., Choong, Y., & Michaels, R. (2009). Usability Testing of an Overlay to Improve Face Capture. In *Third Internation Conference on Biotmetrics: Theory, Appliactions, and Systems*. Presented at the BTAS 2009, Washington, D.C.: IEEE.

Wilson, C., Hicklin, R. A., Bone, M., Korves, H., Grother, P., Zoepfl, M., et al. (2004, June). Fingerprint Vendor Technology Evaluation 2003: Summary of Results and Analysis Report. NIST Technical Report NISTIR 7123. National Institute of Standards and Technology. Retrieved November 20, 2009, from http://fpvte.nist.gov/report/ir_7123_summary.pdf.

Woodward, Jr., J. D. (2005). Using Biometrics to Achieve Identity Dominance in the Global War on Terrorism. *Military Review*, 30-34.

Yeatts, T. (2001). Forensics: Solving the Crime. The Oliver Press Inc.

Zhao, W., Chappella, R., Phillips, P., & Rosenfeld, A. (n.d.). Face recognition: A literature survey. Retrieved January 27, 2009, from http://delivery.acm.org/10.1145/960000/954342/p399-zhao.pdf?key1=954342&key2=2689703321&coll=GUIDE&dl=GUIDE&CFID=19880538&CFTOKEN=94062826.

APPENDICES

Appendix A.

PURDUE
UNIVERSITY

HUMAN RESEARCH PROTECTION PROGRAM
INSTITUTIONAL REVIEW BOARDS

To:	STEPHEN ELLIOTT KNOY 379C
From:	RICHARD MATTES, Chair Biomedical IRB
Date:	12/21/2009
Committee Action:	**Approval**
IRB Action Date:	10/08/2009
IRB Protocol #:	0909008371
Study Title:	Evaluation of Indiana Department of Corrections Mugshot Capture Process
Expiration Date:	12/07/2010

Following review by the Institutional Review Board (IRB), the above-referenced protocol has been approved. This approval permits you to recruit subjects up to the number indicated on the application form and to conduct the research as it is approved. The IRB-stamped and dated consent, assent, and/or information form(s) approved for this protocol are enclosed. Please make copies from these document(s) both for subjects to sign should they choose to enroll in your study and for subjects to keep for their records. Information forms should not be signed. Researchers should keep all consent/assent forms for a period no less than three (3) years following closure of the protocol.

Revisions/Amendments: If you wish to change any aspect of this study, please submit the requested changes to the IRB using the appropriate form. IRB approval must be obtained before implementing any changes unless the change is to remove an immediate hazard to subjects in which case the IRB should be immediately informed following the change.

Continuing Review: It is the Principal Investigator's responsibility to obtain continuing review and approval for this protocol prior to the expiration date noted above. Please allow sufficient time for continued review and approval. No research activity of any sort may continue beyond the expiration date. Failure to receive approval for continuation before the expiration date will result in the approval's expiration on the expiration date. Data collected following the expiration date is unapproved research and cannot be used for research purposes including reporting or publishing a research data.

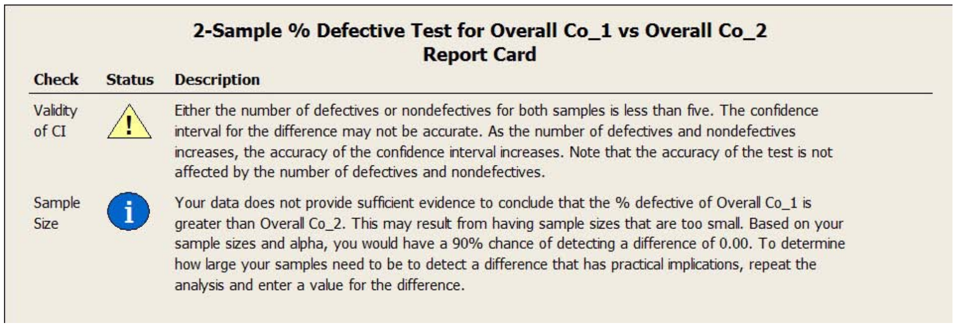
Unanticipated Problems/Adverse Events: Researchers must report unanticipated problems and/or adverse events to the IRB. If the problem/adverse event is serious, or is expected but occurs with unexpected severity or frequency, or the problem/even is unanticipated, it must be reported to the IRB within 48 hours of learning of the event and a writte report submitted within five (5) business days. All other problems/events should be reported at the time of Continuing Review.

We wish you good luck with your work. Please retain copy of this letter for your records.

Ernest C. Young Hall, 10th Floor - 155 S. Grant St. - West Lafayette, IN 47907-2114 - (765) 494-5942 - Fax: (765) 494-9911

Figure A.1 IRB Approval Form

Appendix B.


2-Sample % Defective Test for Overall Co_1 vs Overall Co_2
Report Card

Check	Status	Description
Validity of CI	!	Either the number of defectives or nondefectives for both samples is less than five. The confidence interval for the difference may not be accurate. As the number of defectives and nondefectives increases, the accuracy of the confidence interval increases. Note that the accuracy of the test is not affected by the number of defectives and nondefectives.
Sample Size	i	Your data does not provide sufficient evidence to conclude that the % defective of Overall Co_1 is greater than Overall Co_2. This may result from having sample sizes that are too small. Based on your sample sizes and alpha, you would have a 90% chance of detecting a difference of 0.00. To determine how large your samples need to be to detect a difference that has practical implications, repeat the analysis and enter a value for the difference.

Figure B.1 Minitab Check for Validity and Sample Size for Hypothesis #1

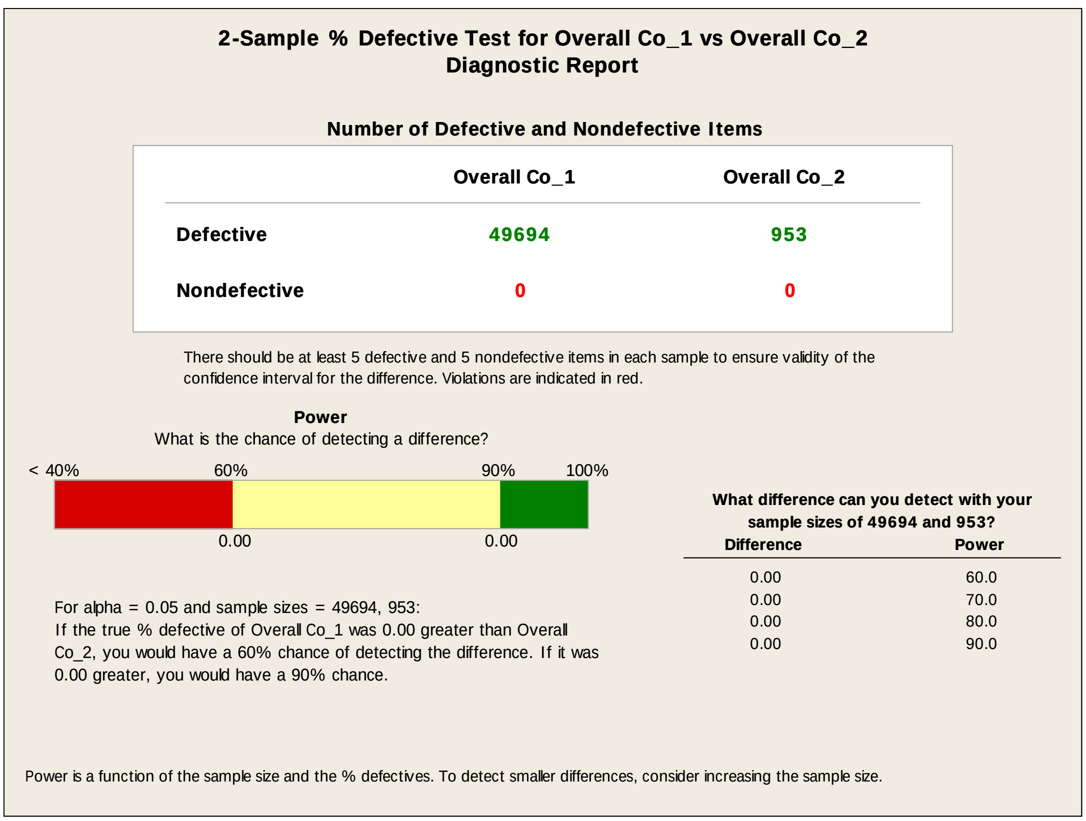


Figure B.2 Minitab Check for Validity and Power for Hypothesis #1

2-Sample % Defective Test for % BG Gray Or vs % BG Gray Ne
Report Card

Check	Status	Description
Validity of CI	✓	Because both samples have at least 5 defectives and 5 nondefectives, the confidence interval for the difference should be accurate.
Sample Size	✓	The sample is sufficient to detect a difference between the % defectives.

Figure B.3 Minitab Check for Validity and Sample Size for Hypothesis #2

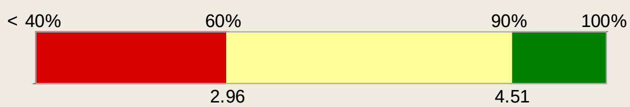
2-Sample % Defective Test for % BG Gray Or vs % BG Gray Ne
Diagnostic Report

Number of Defective and Nondefective Items

	% BG Gray Or	% BG Gray Ne
Defective	17859	87
Nondefective	31835	866

There should be at least 5 defective and 5 nondefective items in each sample to ensure validity of the confidence interval for the difference. Violations are indicated in red.

Power
What is the chance of detecting a difference?



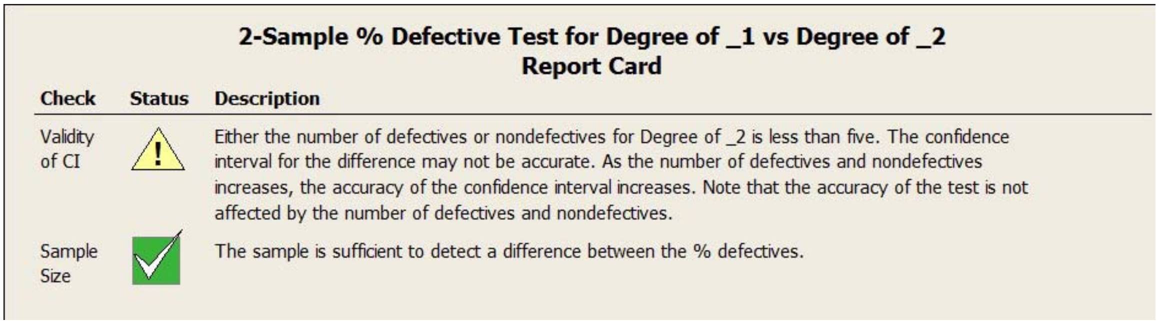
For alpha = 0.05 and sample sizes = 49694, 953:
If the true % defective of % BG Gray Or was 2.96 greater than % BG Gray Ne, you would have a 60% chance of detecting the difference. If it was 4.51 greater, you would have a 90% chance.

What difference can you detect with your sample sizes of 49694 and 953?

Difference	Power
2.96	60.0
3.37	70.0
3.85	80.0
4.51	90.0

Power is a function of the sample size and the % defectives. To detect smaller differences, consider increasing the sample size.

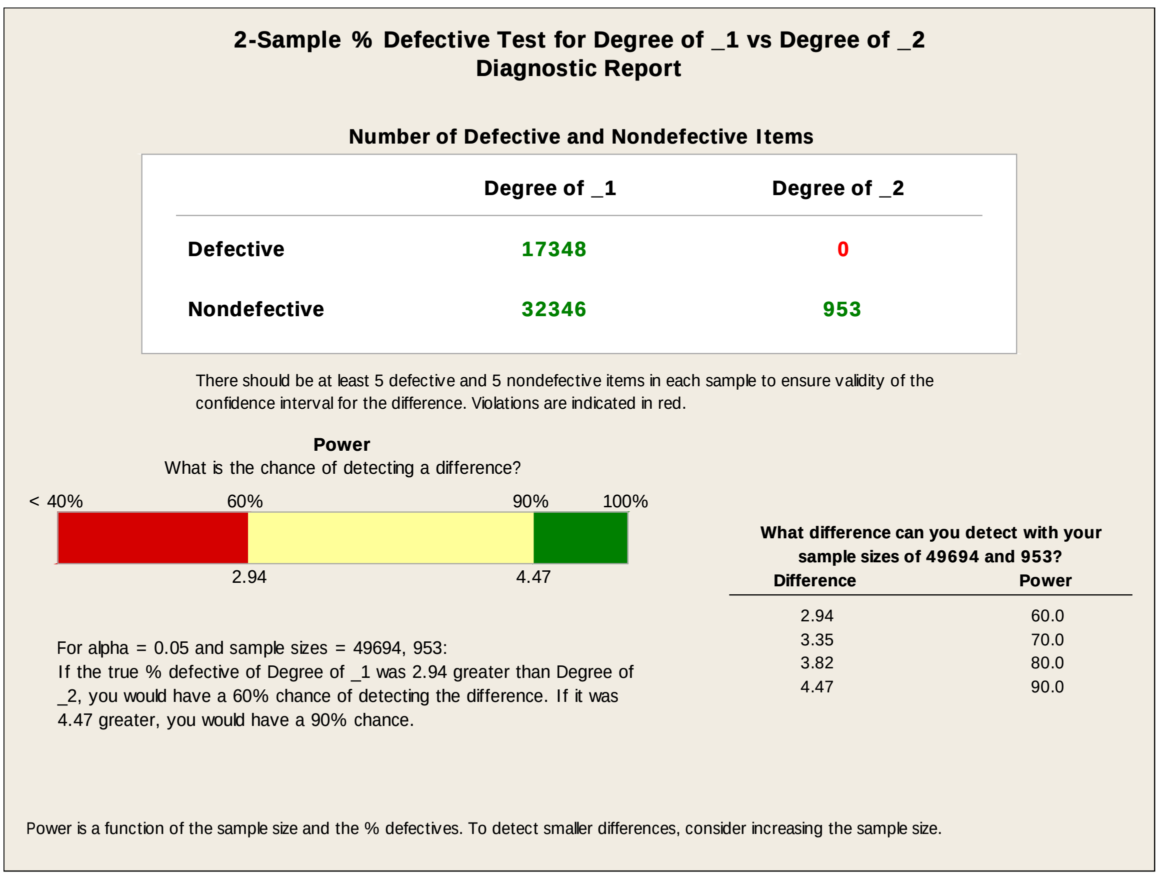
Figure B.4 Minitab Check for Validity and Power for Hypothesis #2



2-Sample % Defective Test for Degree of _1 vs Degree of _2
Report Card

Check	Status	Description
Validity of CI	!	Either the number of defectives or nondefectives for Degree of _2 is less than five. The confidence interval for the difference may not be accurate. As the number of defectives and nondefectives increases, the accuracy of the confidence interval increases. Note that the accuracy of the test is not affected by the number of defectives and nondefectives.
Sample Size	✓	The sample is sufficient to detect a difference between the % defectives.

Figure B.5 Minitab Check for Validity and Sample Size for Hypothesis #3



2-Sample % Defective Test for Degree of _1 vs Degree of _2
Diagnostic Report

Number of Defective and Nondefective Items

	Degree of _1	Degree of _2
Defective	17348	0
Nondefective	32346	953

There should be at least 5 defective and 5 nondefective items in each sample to ensure validity of the confidence interval for the difference. Violations are indicated in red.

For alpha = 0.05 and sample sizes = 49694, 953:
If the true % defective of Degree of _1 was 2.94 greater than Degree of _2, you would have a 60% chance of detecting the difference. If it was 4.47 greater, you would have a 90% chance.

What difference can you detect with your sample sizes of 49694 and 953?

Difference	Power
2.94	60.0
3.35	70.0
3.82	80.0
4.47	90.0

Power is a function of the sample size and the % defectives. To detect smaller differences, consider increasing the sample size.

Figure B.6 Minitab Check for Validity and Power for Hypothesis #3

2-Sample % Defective Test for Eye Separa_1 vs Eye Separa_2
Report Card

Check	Status	Description
Validity of CI	✓	Because both samples have at least 5 defectives and 5 nondefectives, the confidence interval for the difference should be accurate.
Sample Size	✓	The sample is sufficient to detect a difference between the % defectives.

Figure B.7 Minitab Check for Validity and Sample Size for Hypothesis #4

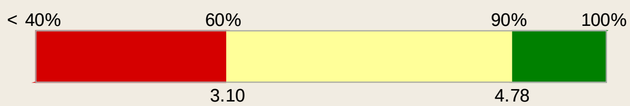
2-Sample % Defective Test for Eye Separa_1 vs Eye Separa_2
Diagnostic Report

Number of Defective and Nondefective Items

	Eye Separa_1	Eye Separa_2
Defective	27306	10
Nondefective	22388	943

There should be at least 5 defective and 5 nondefective items in each sample to ensure validity of the confidence interval for the difference. Violations are indicated in red.

Power
What is the chance of detecting a difference?



For alpha = 0.05 and sample sizes = 49694, 953:
If the true % defective of Eye Separa_1 was 3.10 greater than Eye Separa_2, you would have a 60% chance of detecting the difference. If it was 4.78 greater, you would have a 90% chance.

What difference can you detect with your sample sizes of 49694 and 953?

Difference	Power
3.10	60.0
3.54	70.0
4.06	80.0
4.78	90.0

Power is a function of the sample size and the % defectives. To detect smaller differences, consider increasing the sample size.

Figure B.8 Minitab Check for Validity and Power for Hypothesis #4

2-Sample % Defective Test for Eye Axis A_1 vs Eye Axis A_2
Report Card

Check	Status	Description
Validity of CI	✓	Because both samples have at least 5 defectives and 5 nondefectives, the confidence interval for the difference should be accurate.
Sample Size	✓	The sample is sufficient to detect a difference between the % defectives.

Figure B.9 Minitab Check for Validity and Sample Size for Hypothesis #5

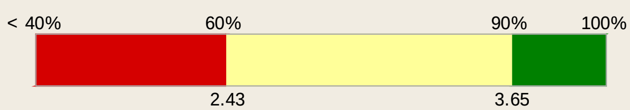
2-Sample % Defective Test for Eye Axis A_1 vs Eye Axis A_2
Diagnostic Report

Number of Defective and Nondefective Items

	Eye Axis A_1	Eye Axis A_2
Defective	9679	62
Nondefective	40015	891

There should be at least 5 defective and 5 nondefective items in each sample to ensure validity of the confidence interval for the difference. Violations are indicated in red.

Power
What is the chance of detecting a difference?



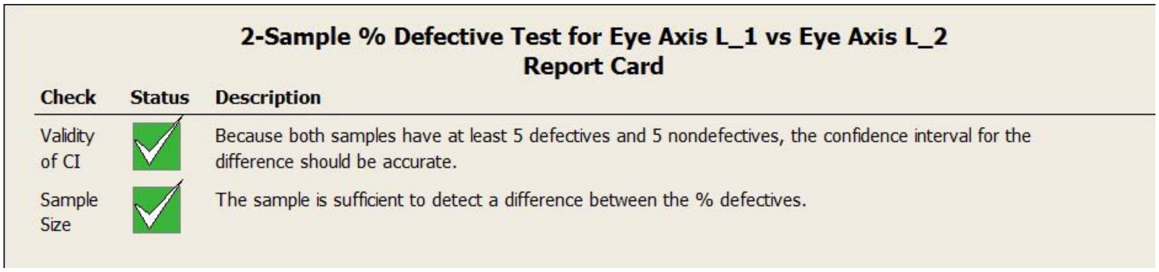
For alpha = 0.05 and sample sizes = 49694, 953:
If the true % defective of Eye Axis A_1 was 2.43 greater than Eye Axis A_2, you would have a 60% chance of detecting the difference. If it was 3.65 greater, you would have a 90% chance.

What difference can you detect with your sample sizes of 49694 and 953?

Difference	Power
2.43	60.0
2.76	70.0
3.13	80.0
3.65	90.0

Power is a function of the sample size and the % defectives. To detect smaller differences, consider increasing the sample size.

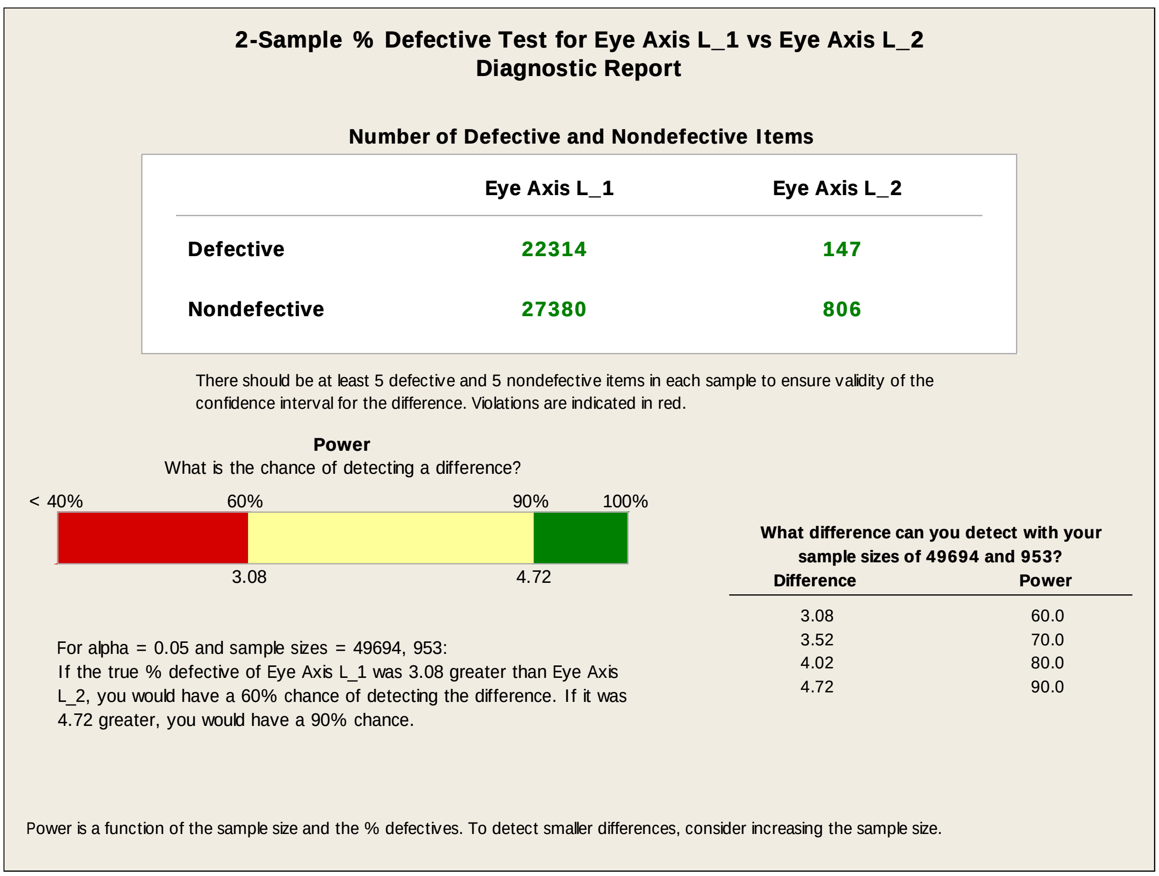
Figure B.10 Minitab Check for Validity and Sample Size for Hypothesis #5



2-Sample % Defective Test for Eye Axis L_1 vs Eye Axis L_2
Report Card

Check	Status	Description
Validity of CI	✓	Because both samples have at least 5 defectives and 5 nondefectives, the confidence interval for the difference should be accurate.
Sample Size	✓	The sample is sufficient to detect a difference between the % defectives.

Figure B.11 Minitab Check for Validity and Sample Size for Hypothesis #6



2-Sample % Defective Test for Eye Axis L_1 vs Eye Axis L_2
Diagnostic Report

Number of Defective and Nondefective Items

	Eye Axis L_1	Eye Axis L_2
Defective	22314	147
Nondefective	27380	806

There should be at least 5 defective and 5 nondefective items in each sample to ensure validity of the confidence interval for the difference. Violations are indicated in red.

Power
What is the chance of detecting a difference?

For alpha = 0.05 and sample sizes = 49694, 953:
If the true % defective of Eye Axis L_1 was 3.08 greater than Eye Axis L_2, you would have a 60% chance of detecting the difference. If it was 4.72 greater, you would have a 90% chance.

What difference can you detect with your sample sizes of 49694 and 953?

Difference	Power
3.08	60.0
3.52	70.0
4.02	80.0
4.72	90.0

Power is a function of the sample size and the % defectives. To detect smaller differences, consider increasing the sample size.

Figure B.12 Minitab Check for Validity and Sample Size for Hypothesis #6

2-Sample % Defective Test for Centerline_1 vs Centerline_2
Report Card

Check	Status	Description
Validity of CI	!	Either the number of defectives or nondefectives for both samples is less than five. The confidence interval for the difference may not be accurate. As the number of defectives and nondefectives increases, the accuracy of the confidence interval increases. Note that the accuracy of the test is not affected by the number of defectives and nondefectives.
Sample Size	i	Your data does not provide sufficient evidence to conclude that the % defective of Centerline_1 is greater than Centerline_2. This may result from having sample sizes that are too small. Based on your sample sizes and alpha, you would have a 90% chance of detecting a difference of *. To determine how large your samples need to be to detect a difference that has practical implications, repeat the analysis and enter a value for the difference.

Figure B.13 Minitab Check for Validity and Sample Size for Hypothesis #7

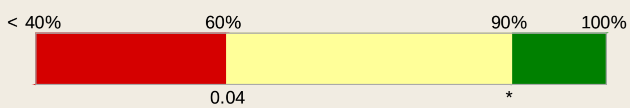
2-Sample % Defective Test for Centerline_1 vs Centerline_2
Diagnostic Report

Number of Defective and Nondefective Items

	Centerline_1	Centerline_2
Defective	49693	953
Nondefective	1	0

There should be at least 5 defective and 5 nondefective items in each sample to ensure validity of the confidence interval for the difference. Violations are indicated in red.

Power
What is the chance of detecting a difference?



For alpha = 0.05 and sample sizes = 49694, 953:
If the true % defective of Centerline_1 was 0.04 greater than Centerline_2, you would have a 60% chance of detecting the difference. If it was * greater, you would have a 90% chance.

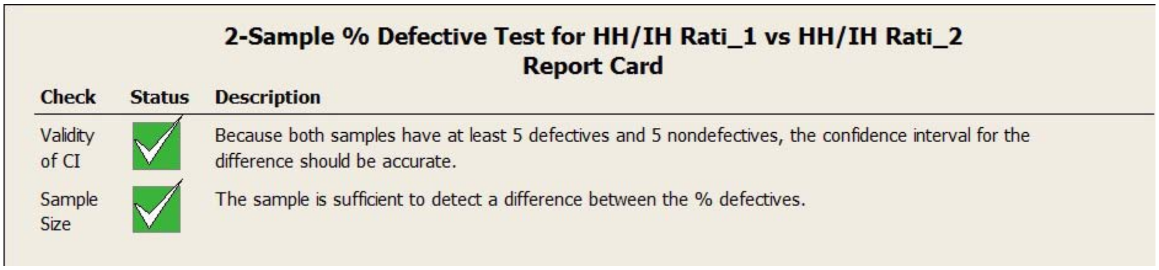
What difference can you detect with your sample sizes of 49694 and 953?

Difference	Power
0.04	60.0
0.07	70.0
*	80.0
*	90.0

* Your sample size is too small to calculate this difference.

Power is a function of the sample size and the % defectives. To detect smaller differences, consider increasing the sample size.

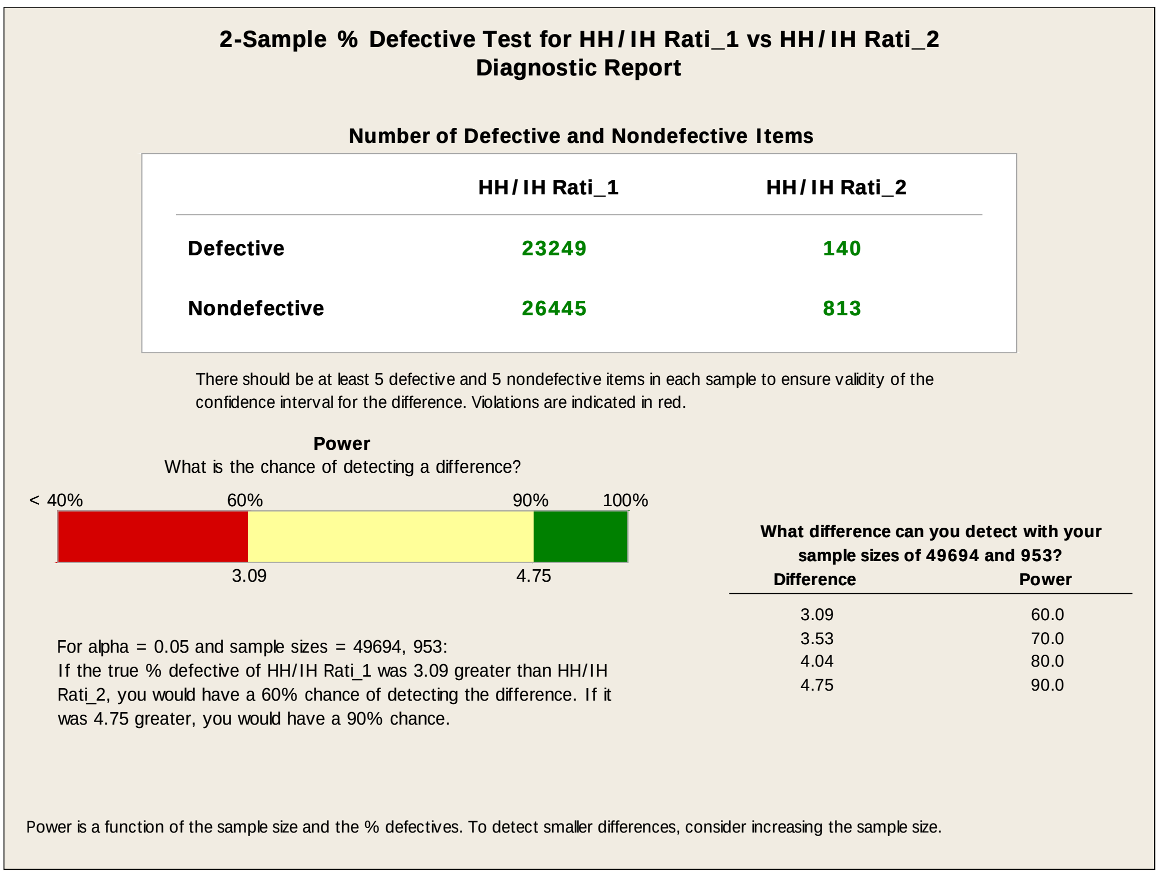
Figure B.14 Minitab Check for Validity and Sample Size for Hypothesis #7



2-Sample % Defective Test for HH/IH Rati_1 vs HH/IH Rati_2
Report Card

Check	Status	Description
Validity of CI	✓	Because both samples have at least 5 defectives and 5 nondefectives, the confidence interval for the difference should be accurate.
Sample Size	✓	The sample is sufficient to detect a difference between the % defectives.

Figure B.15 Minitab Check for Validity and Sample Size for Hypothesis #8



2-Sample % Defective Test for HH/IH Rati_1 vs HH/IH Rati_2
Diagnostic Report

Number of Defective and Nondefective Items

	HH/IH Rati_1	HH/IH Rati_2
Defective	23249	140
Nondefective	26445	813

There should be at least 5 defective and 5 nondefective items in each sample to ensure validity of the confidence interval for the difference. Violations are indicated in red.

Power
What is the chance of detecting a difference?

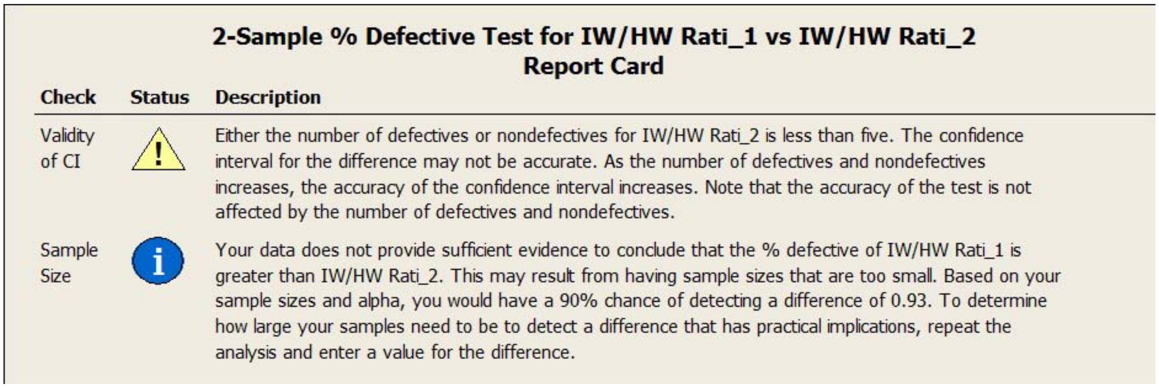
For alpha = 0.05 and sample sizes = 49694, 953:
If the true % defective of HH/IH Rati_1 was 3.09 greater than HH/IH Rati_2, you would have a 60% chance of detecting the difference. If it was 4.75 greater, you would have a 90% chance.

What difference can you detect with your sample sizes of 49694 and 953?

Difference	Power
3.09	60.0
3.53	70.0
4.04	80.0
4.75	90.0

Power is a function of the sample size and the % defectives. To detect smaller differences, consider increasing the sample size.

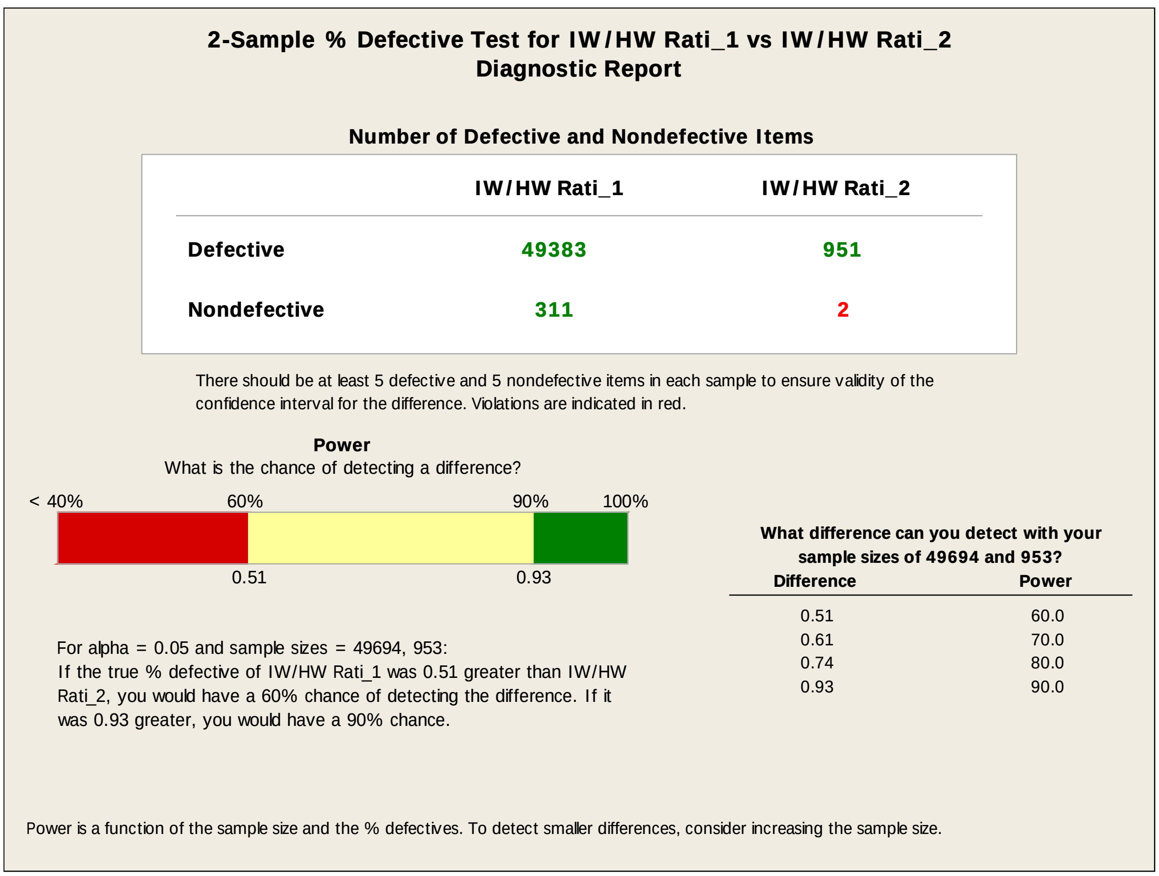
Figure B.16 Minitab Check for Validity and Sample Size for Hypothesis #8



2-Sample % Defective Test for IW/HW Rati_1 vs IW/HW Rati_2
Report Card

Check	Status	Description
Validity of CI	⚠	Either the number of defectives or nondefectives for IW/HW Rati_2 is less than five. The confidence interval for the difference may not be accurate. As the number of defectives and nondefectives increases, the accuracy of the confidence interval increases. Note that the accuracy of the test is not affected by the number of defectives and nondefectives.
Sample Size	i	Your data does not provide sufficient evidence to conclude that the % defective of IW/HW Rati_1 is greater than IW/HW Rati_2. This may result from having sample sizes that are too small. Based on your sample sizes and alpha, you would have a 90% chance of detecting a difference of 0.93. To determine how large your samples need to be to detect a difference that has practical implications, repeat the analysis and enter a value for the difference.

Figure B.17 Minitab Check for Validity and Sample Size for Hypothesis #9



2-Sample % Defective Test for IW/HW Rati_1 vs IW/HW Rati_2
Diagnostic Report

Number of Defective and Nondefective Items

	IW/HW Rati_1	IW/HW Rati_2
Defective	49383	951
Nondefective	311	2

There should be at least 5 defective and 5 nondefective items in each sample to ensure validity of the confidence interval for the difference. Violations are indicated in red.

What difference can you detect with your sample sizes of 49694 and 953?

Difference	Power
0.51	60.0
0.61	70.0
0.74	80.0
0.93	90.0

For alpha = 0.05 and sample sizes = 49694, 953:
If the true % defective of IW/HW Rati_1 was 0.51 greater than IW/HW Rati_2, you would have a 60% chance of detecting the difference. If it was 0.93 greater, you would have a 90% chance.

Power is a function of the sample size and the % defectives. To detect smaller differences, consider increasing the sample size.

Figure B.18 Minitab Check for Validity and Sample Size for Hypothesis #9