

Letter to the Editor Re: PanCanadian Evaluation of Irreversible Compression Ratios (“Lossy” Compression) for Development of National Guidelines

Geoff Norman,¹ Peter Bak,² Andre Matos,² and David A. Koff²

Because of the important implications of the findings of the Pan-Canadian Evaluation of Irreversible Compression Ratios study,¹ we have been approached to provide power estimates, or some other indication of the overall magnitude of the effect of compression on test characteristics. That is the intent of the present submission.

The concerns can be summarized into two broad categories: (1) is it possible that the few differences that were reported were simply statistical fluctuations, and can be safely discounted (a type I error)? And (2) is it possible that, at the level of individual comparisons, the sample size was so small that important differences were not detected (a type II error)? In this follow-up, we have decided to focus on indices of diagnostic performance (sensitivity, specificity, and accuracy) and not pursue measures of image quality, since performance is more central to an examination of the effect of compression.

With respect to the type I error concern, that is, the finding of a few statistically significant results, we can determine the likelihood that these may have arisen by chance. There were a total of 42 comparisons in the original paper. Of these, eight were significant at the 0.05 level and three were significant at the 0.005 level. We have calculated, using a standard formula¹ that with 42 comparisons there is a 99% likelihood that we would have observed at least one significant comparison at the 0.05 level, and a 34% likelihood that we would have observed at least one significant difference at the 0.005 level.

Further, the nature of the analysis (analysis of variance; ANOVA) is such that trends were not examined. For example, it is reasonable to presume

that if compression negatively affects image quality compressed images would have lower sensitivity and specificity than uncompressed images, and increased compression would lead to lower values. In the one significant comparison reported in the paper (sensitivity—MSK CR, Table 1), these trends were not observed; JPEG2000 at a compression ratio of 20, JPEG at a ratio of 20, and JPEG at a compression ratio of 30 all had higher sensitivity than the uncompressed images. Lowest values of sensitivity were at intermediate compression ratios (JPEG2000 at 25 and JPEG at 25). Thus, although there was a significant difference in this subtype, the observed differences were not consistent with the expected trend.

Three of the eight statistically significant comparisons are seen in computed tomography (CT) imaging. We have explored this further and analyzed sensitivity and specificity for CT images across all regions for uncompressed and JPEG and JPEG2000 compressions. Average sensitivity was 0.85 for uncompressed, 0.88 for JPEG, and 0.84 for JPEG2000 ($p=0.49$); for specificity, values were 0.70 for uncompressed, 0.66 for JPEG and 0.69 for JPEG2000 ($p=0.91$). The differences,

¹From the Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Canada.

²From the Department of Radiology, McMaster University, McMaster Innovation Park, 175 Longwood Rd South, Suite 101, Hamilton, ON L8P 0A1, Canada.

Correspondence to: David A. Koff, Department of Radiology, McMaster University, McMaster Innovation Park, 175 Longwood Rd South, Suite 101, Hamilton, ON L8P 0A1, Canada; tel: +1-905-5212100; fax: +1-905-5771443; e-mail: dkoff@mcmaster.ca

Copyright © 2010 by Society for Imaging Informatics in Medicine

Online publication 23 June 2010

doi: 10.1007/s10278-010-9308-3

¹Likelihood= $1-(1-\alpha)^k$, where k is the number of comparisons and α is the p value.

Table 1. Overall Sensitivity

	Compression			
	None	Low	Med	High
JPEG		87.9% (84-92)	86.1% (82-90)	88.6% (84-93)
UNCOMP	86.9% (83-91)			
JPEG2000		85.7% (81-90)	85.5% (81-90)	86.2% (82-90)

Effect of compression, $F(2, 501) = 0.29$, $p = 0.74$

Effect of type, $F(1,501) = 1.07$, $p = 0.30$

Interaction between type and compression $F(2,501) = 0.11$, $p = 0.89$

therefore, are very small and are not statistically significant. Thus, while it may be the case that some of the observed significant differences in the original report are real, no firm conclusion can be drawn without replication.

With respect to the Type II error concern, we decided to obtain a better overall estimate of the effect of compression by aggregating across all body regions and modalities. In doing so, we have ignored potential differences at an individual domain level and instead treat each domain as a single observation of an overall effect. While this may seem a heroic assumption, an aggregation of the data will show that, in fact, the actual magnitude of the effect of compression is extremely small, so it is unlikely that this would mask large effects within specific modalities.

Data were aggregated to determine overall sensitivity, specificity and accuracy by type of compression and amount of compression (low, medium, high). Data were analyzed with two-way ANOVA, to look for overall difference by amount of com-

Table 2. Overall Specificity

	Compression			
	None	Low	Med	High
JPEG		77.1% (70-84)	82.4% (75-90)	75.4% (67-83)
UNCOMP	77.4% (70-85)			
JPEG2000		76.3% (68-84)	76.1% (69-84)	77.3% (70-84)

Effect of compression $F(2, 497) = 0.34$, $p = 0.71$

Effect of type, $F(1,468) = 0.31$, $p = 0.58$

Interaction between type and compression $F(2,468) = 0.60$, $p = 0.55$

Table 3. Overall Sensitivity

		Sensitivity			
Effect size		0.20	0.10	0.05	0.025
Absolute diff		0.035	0.018	0.010	0.005
Power	COMP	0.99	0.97	0.43	0.13
	Type	0.99	0.99	0.50	0.14
		Specificity			
Effect size		0.20	0.10	0.05	0.025
Absolute diff		0.062	0.031	0.016	0.008
Power	COMP	0.99	0.36	0.13	-
	Type	0.99	0.46	0.18	-

pression, type of compression and interaction. The overall results are shown below in Tables 1 and 2.

Thus, there was no statistically significant effect of type or compression on any of sensitivity, specificity or accuracy. Examining individual values, the effect of compression on sensitivity ranged from -1.4% to $+1.7\%$, with a mean of -0.057% , and on specificity from -2.0% to $+5.0\%$, mean 0.056% . The effect of compression was less than 1%.

Although the estimated effects of compression are consistently small, there still may be an issue of statistical power to detect clinically meaningful differences. Accordingly we conducted a power analysis, based on the overall difference across modalities, with a sample size of 155^2 . The analysis is based on the effect size (difference/standard deviation), for sizes of $0.025-0.20$, and computed separately for detecting an effect of compression type and compression ratio (Table 3).

Thus, we had 99% power to detect differences of 3.5% in sensitivity and 6.2% in specificity and about 50% power to detect differences as small as 1% in sensitivity and 3% in specificity.

Based on these summary statistics, a reasonable conclusion is that, with the compression algorithms and ratios used in the study, digital compression does not lead to any clinically or statistically significant loss in diagnostic performance.

REFERENCE

1. Koff D, Bak P, Brownrigg P, Hosseinzadeh D, Khademi A, Kiss A, Lepanto L, Michalak T, Shulman H, Volkening A: Pan-Canadian evaluation of irreversible compression ratios ("lossy"compression) for development of national guidelines. *J Digit Imaging* 22:569-78, 2009. Epub 2008 Oct 18

²The analysis methods were based on Cohen J. *Statistical Power Analysis for the Behavioural Sciences*, 1987.