1. Background

Chronic wounds are a major health problem, not only because of their incidence, but also because of their time- and resource-consuming management. The study presented in the current paper was undertaken to investigate the possible use of colorimetric imaging during the assessment of human wound repair. The outline design of the current study is based on system requirements for colorimetric diagnostic tools as previously published.[1,2]

Today, digital photography is considered an acceptable and affordable tool in many clinical domains such as wound care and dermatology [3-12], forensics [13,14], pathology [15], traumatology, orthodontics [16,17] etc. Although technical features of most digital cameras are pretty impressive, they are unable to perform reproducible and accurate (with regards to spectrophotometry) imaging.[18-22, 28] Indeed, simply take 2 pictures of the same wound under the same lighting conditions, with the same camera and settings, immediately after each other, the pictures will most likely come out differently. No need to mention that the results are even worse when the lighting, the camera or its settings are different. As such, reproducibility is poor or even non-existent. This seems less important when photographs are taken purely for documentation, but when digital photography becomes part of the medical evaluation or is used to perform measurements, it becomes of critical importance.[5,6,18,23-26] In our view, the quality of medical photography is principally defined by its reproducibility and accuracy.[21] Without reproducibility and accuracy of images any attempt of measuring colour or geometric properties is useless.[27] A simple, practical and validated algorithm to solve this problem is called for.(Figure 1) In order to elaborate on this we must introduce some essential concepts of colorimetry first. [29]

The eye has 4 types of light sensors: rods which are not sensitive to colour but function even in low light, and 3 types of cones which are most sensitive to red, green-yellow and blue respectively. As a consequence of this almost all colours can be reconstructed using a combination of 3 base colours, usually red, green and blue (RGB) [30, 31] Together, these 3 base colours define a 3-dimensional colour space which can be used to describe colours. The accurate handling of colour characteristics of digital images is a non-trivial task because RGB signals generated by digital cameras are 'device-dependent', i.e. different cameras produce different RGB signals for the same scene, and on top of that these signals will even change over time because they are dependent on the camera settings (some of which maybe scene dependent like the shutter speed and aperture). Another way to put this is to say that generally each camera defines a custom device-dependent RGB colour space for each picture it takes. As a consequence, the term RGB (as in RGB-image) is clearly ill-defined and rather meaningless for any but trivial purposes. Furthermore, the RGB output signals of digital cameras, other than being based on red, green and blue, are not related to those of the human visual system. As the measurement of colours and colour differences are based on a standard colorimetric observer as defined by the CIE (Commission Internationale de l'Eclairage, the international standardizing body in the field of colour science), any such measurements on RGB images are impossible as long as the relationship between the varying camera RGB colour spaces with the colorimetric colour spaces (colour spaces based on said human observer) are not determined.

Luckily, there is a standard RGB colour space called sRGB that is fixed (device-independent) and has a known relationship with the CIE colorimetric colour spaces. This colour space has an added bonus that it will display realistically on most modern display devices without extra manipulation (look for a ‘sRGB’ or ‘6500K’ setting). For the moment, the downside is a rather restricted gamut (it can not represent all the colours a human can see). Our problem can now be reformulated to finding the relationship between the varying and unknown camera RGB and the sRGB colour space. Achieving this will eliminate most of the variability introduced by the camera and the lighting conditions.
The transformation between the input RGB colour space and the sRGB colour space is achieved via a colour target based calibration using a 'reference chart', namely the MacBeth Colour Checker Chart Mini [MBCCC] (GretagMacBeth AG, Regensdorf, Switzerland) (Figure 2.). This chart provides a checkerboard array of 24 scientifically prepared coloured squares or patches in a wide range of colours with known colorimetric properties under a CIE D65, noon daylight illuminant (6504K). Many of these squares represent natural objects of special interest, such as human skin, foliage and blue sky. These squares are not only the same colour as their counterparts, but also reflect light the same way in all parts of the visible spectrum.

Different calibration algorithms defining the relationship between the input RGB colour space of the camera and the sRGB colour space have been published using e.g. 3D look-up tables and neural networks. The algorithm in this study however is based on 3 1D look-up tables and polynomial modelling as previously published by Vander Haeghen et al. [32]

2. Methods
The paper has been carried out in compliance with the Helsinki Declaration; ethical committee approval is available for the methods used (B32220083450 Commissie voor Medische Ethiek Faculteit Geneeskunde Leuven Belgium).

2.1 Experiment 1
The purpose of the first experiment was to check whether camera settings or lighting conditions negatively affect the quality of the colorimetric calibration.[33] Considering that chronic wounds are assessed in different locations and environments, it was decided to challenge the calibration algorithm under extreme lighting conditions and with inappropriate camera settings.

2.1.1 Image Acquisition
Digital images of the MBCCC on a grey coloured background in a Colour Assessment Cabinet CAC 120-5 (VeriVide Leicester, UK) were taken using 2 digital cameras (Nikon D200 SLR camera (10.2 effective mega pixels) with a 60 mm AF Micro Nikkor lens and Canon Eos D10 camera (6.3 million effective pixels) with a 50 mm Canon EF lens).

Parameter Settings Number
Camera type Canon Eos D10, Nikon D200 2
Scene lighting (cabinet) D651, TL842, A3 3
Camera sensitivity 100 ISO, 400 ISO 2
Camera exposure -1EV, 0EV, +1EV 3
Camera white balance Auto, Manual, A, D65 4

Table 1: Lighting and camera settings
Table 1 gives all the parameters that were varied during image acquisition. Note that these include a couple of combinations that are really inappropriate, e.g. setting the camera white balance to D65 with an A scene illuminant. This produces very off-colour images and was done on purpose to challenge the calibration algorithm. (1 Illuminant D65: 'Artificial Daylight' fluorescent lamps conforming to Standard Illuminant D6500 (6500 K), 2 Illuminant TL84: Philips Triphosphor fluorescent lamps, often chosen as a 'Point of Sale' illuminant, (5200K), 3 Illuminant A: 'Filament (domestic) lighting' (3000K))

2.1.2 Calibration Procedure
A basic assumption for the calibration procedure is the presence of uniform illumination and a reference chart as part of the image of interest. The calibration provides a way to transform the acquired images (defined in an unknown colour space (usually RGB)), to a standard, well defined colour space called sRGB.[34] sRGB has a known relation to the CIE L*a*b* colorimetric space which allows computation of perceptual colour differences. The CIE L*a*b* colorimetric space or CIELAB space with coordinates L*, a* and b* refers to the colour-opponent space with L*
referring to Luminance and a* and b* to the colour-opponent dimensions.[34-36] The 'detection of the MBCCC' in the digital image can be done manually or automatically. The algorithm behind the 'MBCCC detection' is based on the initial detection of all the bright areas in the image, and is followed by a shape analysis. Shapes that are not rectangular, too small or too large compared to the image dimensions are discarded. The remaining areas are candidates for the white patch. For each of these white patch candidates, the corresponding black patch is searched for, taking into account the typical layout of the colour chart and the dimensions of the white patch candidate. If this succeeds the patches are checked for saturation (average pixel value $> 255-\delta$ or $<\delta$ with $\delta$ a small number, e.g. 3) in each colour channel separately. If the number of saturated patches is acceptable, (typically less than 6 out of 24 patches), the calibration proceeds and its quality is assessed. This quality assessment consists of various conditions on the colour differences between the known spectrophotometric and the computed sRGB values, both on the accepted and rejected patches. If one of these tests fails, the algorithm rejects the calibration and continues the search.

2.1.3 Analysis
Precision in this experiment is defined as a measure of how close consecutive colour measurements on an image of the same subject are to each other. This is also known as reproducibility. The precision of the MBCCC chart detection together with the calibration process were evaluated by computing the perceptual colour differences between all the possible pairs of measurements of each colour square of the MBCCC chart. These perceptual colour differences are expressed in CIE units, and are computed using the Euclidean metric in the CIE L*a*b* colour space. Theoretically 1 unit is the 'just noticeable colour difference' and anything above 5 units is clearly noticeable.

The accuracy of a procedure is a measure of how close its results are to the 'real' values, i.e. those obtained using the 'standard' procedure or measurement device. For colour measurements this would be a spectrophotometer. Consequently, the accuracy of the chart detection and colour calibration can be assessed by computing the perceptual colour differences between the measurements of the colour squares of the MBCCC chart and the spectrophotometric values of the same squares. Note that for this assessment the calibration was performed using only half of the colour patches of the MBCCC chart, while the accuracy evaluation was performed on the other half of the squares not used in the calibration. As such the accuracy is very likely to be a little higher when using the whole chart to perform the calibration. Both precision and accuracy thus result in a probability distribution for the $dE_{ab}$ errors. Tukey’s five-number summary of the $dE_{ab}$ colour differences of each patch is calculated and visualized using a box plot (the minimum, the lower quartile, the median, the upper quartile and the maximum). Wilcoxon rank-sum statistics are used to test the calibration. It compares the locations of two populations to determine if one population is shifted with respect to another. The method employed is a sum of ranks comparison which works by ranking the combined data sets and summing the ranks for each $dE_{ab}$. The sum of the ranks is then compared to significance values based on the decision alpha (p<0.05).

2.2. Experiment 2
The second experiment was designed to quantitatively assess the impact of the automatic calibration procedure, (i.e. the chart detection) on measurements. This may be important in clinical settings where large batches of images need to be calibrated (automatically) in a single run. To examine this, 40 different images of real wounds were acquired, and a region of interest (ROI) was selected in each image. 3 rotated versions (at 90, 180 and 270 degrees) of each image with its ROI were created and automatically calibrated (Figure 3.). Comparison between the
colour measurements of the ROI's of the rotated versions of each image gives us an idea of the errors introduced by the automatic chart detection part of the calibration procedure.

2.2.1 Image Acquisition

Digital images (n = 40) of the chronic wounds were taken (Sony Cybershot DSC-F828 digital camera (8.0 million effective pixels) and Carl Zeiss 28 – 200 mm equivalent lens) with fully automatic settings at different indoor locations, as is usually the case in daily clinical practice.

2.2.2 Calibration Procedure & Analysis

The calibration procedure consists of the same steps described in experiment 1. The \( dE_{ab} \) colour differences between the average colour of the ROI of the 4 rotated versions of each image are computed and visualized using a probability distribution graph.

3 Results

3.1 Experiment 1

Figures 4 and 6, show examples of realistic sample images under different illuminants and with different cameras with the corresponding calibrated images (Figures 5 and 7). The image contains a lot of patches that were saturated (see the 'x's on the patches) and were not used for the calibration, resulting in a lower quality calibration. The accuracy and reproducibility of the colour calibration using different cameras, camera settings and illumination conditions are presented using a probability distribution of \( dE_{ab} \) errors of all the MBCCC patches. A distinction is made between the full set of images and the 'normal' images which were acquired with proper camera settings: correct manual or automatic white balance and no exposure bias. Indeed the full set contains several images which were strongly over- and underexposed or had a mismatched white balance on purpose. These extreme images demonstrate the effectiveness of the calibration method well, but are not representative for day-to-day photography. Moreover, the term 'proper patch' was used to indicate patches which were not saturated during acquisition (i.e. which had pixel values which were too close to either 255 or 0). Indeed, it is impossible to recover the proper pixel value of these saturated patches, and thus their calibration is infeasible. In other words, the accuracy and reproducibility results for the set of proper patches of normal images are representative for colours in properly photographed images which are different from the colours of the patches which were disregarded during calibration due to saturation (marked by 'x' on the calibrated image). Note that this does not happen very often with normal images, but if it does, it usually manifests itself as an overexposure of the white, deep red, yellow and orange MBCCC patches. Moreover, these strong or very saturated colours rarely occur in skin imaging. Also, if this problem is frequent with a certain camera it can be remedied by slightly underexposing images, say half an f-stop (exposure bias).

Tukey's five-number summary of the \( dE_{ab} \) colour differences of each proper patch of the normal images is calculated and visualized using a box plot (the minimum, the lower quartile, the median, the upper quartile and the maximum). Outliers are marked with a red 'x'. In the case of accuracy, the chart patches are split in 2 groups of 12 patches and only the second group is used for calibration, resulting in a lower quality calibration than if 24 patches were used. The first group of 12 patches is then used to check the accuracy. Colour differences between the measurements and real spectrophotometric measurements reveal median \( dE_{ab} \) values respectively 6.40 for the proper patches of calibrated normal images and 17.75 for uncalibrated images demonstrating an important improvement in accuracy after calibration. (Figure 8 and 9).

The result for the patches used in the calibration is also included with a median of 1.59 \( dE_{ab} \). Figure 10 shows the accuracy box plot for the proper patches of the normal images. As we explained before, we can only use the patches that were not used in computing the calibration to check accuracy, which is why only 12 patches are shown. As shown in Figure 12, the
reproducibility, visualized by the probability distribution of the \( dE_{ab} \) errors between 2 measurements of the patches of the images has a median of 3.43 \( dE^* \) for all calibrated images, 23.26 \( dE_{ab} \) for all uncalibrated images, a median of 2.83 \( dE_{ab} \) for all 'normal' calibrated images and 14.25 \( dE_{ab} \) for all 'normal' uncalibrated images. If we restrict ourselves to the proper patches of normal calibrated images the median is only 2.58 \( dE_{ab} \). Wilcoxon sum-rank testing (\( p<0.05 \)) between uncalibrated normal images and calibrated normal images with proper squares were equal to 0 demonstrating a highly significant improvement of reproducibility. (Figure 11.) A closer look at the \( dE_{ab} \) errors of each MBCCC patch individually reveals the largest errors are for the red, orange yellow, orange and yellow patches. Note the absence of the cyan patch as this patch cannot be properly be represented in the sRGB colour space (out of gamut).(Figure 13.)

3.2 Experiment 2
The reproducibility of the chart detection during the automatic calibration is presented using a probability distribution of \( dE_{ab} \) errors between 2 measurements of the same ROI. Clearly this distribution should be as narrow and as close to zero as possible as ideally the measurements of the same ROI in Figure 1. Reproducibility: box plot for the proper patches of the normal images. the rotated versions of an image should all be equal. Any deviation of this indicates a variability in the chart detection, which leads to a slightly different calibration and thus different measurements. (Figure 14.)

4. Conclusion
The experiments presented in this paper provide evidence that images taken with commercially available digital cameras can be calibrated independently of any camera settings and independently of illumination features (within reason of course), provided that illumination in the field of view is uniform and a calibration chart is used. This is especially useful in chronic wound assessment, as this is often performed in different locations and conditions. Practically, the proposed calibration transforms the acquired images, in an unknown colour space (usually RGB), to a standard, well defined colour space called sRGB which allows proper display of images and has a known relationship to the CIE colorimetric colour spaces. In a first experiment the calibration procedure was challenged with a large collection of images containing both 'normal' images with proper camera settings and images that were purposely over- or underexposed and/or had white balance mismatches. The reproducibility and accuracy of the calibration procedure can be seen in Figures. 7 and 9, and show marked improvements. Also note that the calibration procedure also works very well on the images with improper camera settings, as evidenced by the very small differences between the error distributions of the complete set of images and the set with only the 'normal' images. An innovative feature demonstrated in this paper is the automatic detection of the MacBeth Colour Checker Chart Mini [MBCCC] in the digital image.
In a second experiment we tested the effect of this MBCCC chart detection on subsequent colour measurements. Figure 11 shows the probability distribution of errors between 2 colour measurements of the same region of interest that can be attributed to variations in the chart detection process. The vast majority of these errors are well below 1 \( dE_{ab} \), demonstrating that the chart detection is quite robust. To our knowledge, the proposed technology demonstrates for the first time a fundamental and in our opinion essential tool enabling intra-individual (in different phases of wound healing) and inter-individual (for features and properties) comparisons of digital images in human wound healing. By implementing this extra step in the assessment, we hope to be a step closer to the scientific standards for research in this domain.[37]

5. References
Figure 11

Box plots showing variation in $dE^*$ across different patches.
Figure 14

$dE^*$ probability distribution between 2 measurements
Additional files provided with this submission:

Additional file 1: automatic colorimetric calibration of human wound imaging title., 11K
http://www.biomedcentral.com/immedia/8691204842772250/supp1.docx