KTH Technology
and Health

EVALUATION OF ERRORS AND LIMITATIONS IN ULTRASOUND IMAGING SYSTEMS

Mattias Mårtensson

Doctoral Thesis

Division of Medical Engineering,

School of Technology and Health,

Royal Institute of Technology

Trita-STH Report 2011:3
ISSN: 1653-3836
ISRN: ISRN KTH/STH/2011:3--SE
ISBN: 978-91-7501-026-7

© Mattias Mårtensson, Stockholm, 2011

PREFACE

This thesis is the result of five years research performed at the Royal Institute of Technology (KTH), School of Technology and Health (STH) in Huddinge, Sweden. The thesis is the research part of the requirements for the Doctoral degree in Technology. The work has been supervised by Professor Lars-Åke Brodin and co-supervised by Matilda Larsson, PhD, and Anna Bjällmark, PhD. The research projects were supported by grants from the Swedish Heart-Lung Foundation (Hjärt-Lungfonden) and the Swedish Research Council (Vetenskapsrådet). The thesis is publically defended at 13:00, 10 June 2011 in the lecture hall 3-221, Alfred Nobels Allé 10, Huddinge, Sweden.

ABSTRACT

There are binding regulations requiring safety and efficacy aspects of medical devices. The requirements ask for documentation that the devices are safe and effective for their intended use, i.e. if a device has a measuring function it must be correct. In addition to this there are demands for quality systems describing development, manufacturing, labelling, and manufacturing of a device. The requirements are established to guarantee that non-defective medical devices are used in the routine clinical practice. The fast rates in which the imaging modalities have evolved during the last decades have resulted in numerous new diagnostic tools, such as velocity and deformation imaging in ultrasound imaging. However, it seems as if the development of evaluation methods and test routines has not been able to keep up the same pace. Two of the studies in this thesis, Study I and IV, showed that computed tomography-based and ultrasound based volume measurements can yield very disparate measurements, and that tissue Doppler imaging-based ultrasound measurements can be unreliable.

Furthermore, the new ultrasound modalities impose higher demands on the ultrasound transducers. Transducers are known to be fragile, but defective transducers were less of a problem earlier when the ultrasound systems to a lesser extent were used for measurements. The two other studies, Study II and III, showed that serious transducer errors are very common, and that annual testing of the transducers is not sufficient to guarantee an error free function.

The studies in the thesis indicate that the system with Notified Bodies, in accordance with the EU's Medical Device Directive, checking the function and manufacturing of medical devices does not work entirely satisfactory. They also show that the evaluation of new methods have led to the undesirable situation, where new measuring tools, such as volume rendering from imaging systems, and tissue Doppler-based velocity and deformation imaging in echocardiography are available for clinicians without proven knowledge about their accuracy.

Keywords: ultrasound, transducer function, computed tomography, LV volume, tissue Doppler imaging, quality assurance

ACKNOWLEDGEMENT

First and foremost I would like to thank *Matilda Larsson* and *Anna Bjällmark*, whom I first met when I was working with my master thesis. I would like to thank you not just for being the best friends and colleagues one could ever wish for, but also for convincing me to continue as a PhD student after my graduation. This thesis had simply not existed without the support you have given me. As I write this I realize that words cannot express the depth of my appreciation for you.

I would like to thank my supervisor, *Lars-Åke Brodin*, for believing in me and giving me the opportunity to continue the hunt for errors in medical imaging devices. Furthermore, I would like to thank all my co-authors*: Reidar Winter, Kerstin Cederlund, Jonaz Ripsweden, Habib Mir-Akbari, Mats Olsson, Björn Segall, and Alan G. Fraser*, with a special gratitude to *Jacek Nowak* for all the invaluable help both with the articles and in dark streets of Prague.

I would also like to express my gratefulness to *Britta Lind* who have helped me and learned me so much during the last five years.

During the work with this thesis I have had the privilege to have two roommates. Despite being abandoned by her during very suspicious circumstances I would like to thank my first roommate *Nina Elmstedt* for being such a good friend. I would also like to thank her successor *Malin Larsson,* or *Lillhelganbofilifjonkan* as she is called daily, for being so kind to me every day.

Even though being the mastermind behind the suspicious circumstances in stealing my roommate to her own room, and giving me a lot of electronically executed punishments and using me as a guinea pig, I must thank *Frida Lindberg* for being a great friend and for all our fruitful coffee breaks.

For introducing me to the true Chinese kitchen and for always being so happy and sweet, I would like to thank *Xiaogai Li*.

I would like to thank *Madelen Fahlstedt,* the early bird of STH, for the many early morning chats at work. Your presence illuminates the darkest and dullest of mornings for me.

To *Rickard Axelsson*: I am writing this on the day of your wedding. I congratulate and wish you and Sandra much happiness together.

The "MÖP" says that he misses *Sofia Hedenstierna*. He thinks that it was better when Sofia worked at STH.

For the help with the phantom manufacturing I would like to thank *Peter Arfert*, and for the help with the proofreading I would like to thank *Peta Sjölander*, *Karen Gustafsson* and *Björn-Erik Erlandsson*.

Finally, I would like to thank all my other colleagues making STH such a nice workplace, especially *Gunilla Langvall* who has changed chaos to order.

Mattias Mårtensson, Järfälla, 2011

ABBREVIATIONS

2-D	Two-dimensional
3-D	Three-dimensional
CT	Computed tomography
CV	Coefficient of variation
LV	Left ventricle
MSCT	Multi slice computed tomography
SD	Standard deviation
TDI	Tissue Doppler imaging

TABLE OF CONTENTS

1 INTRODUCTION

Ultrasound is one of the most widely used medical imaging systems because of its many advantages. Compared to x-ray and magnetic resonance imaging it is a relatively inexpensive low risk imaging modality providing real-time information bedside. The pace with which medical imaging systems develop is today very fast, and ultrasound scanners are no exception. During the last two decades, ultrasound scanners and its appurtenant methods have become much more sophisticated**.** The rapid development is a result of and goes hand in hand with the development of computer technology. The development of faster processors and high capacity data memories has allowed the use of much more computationally demanding methods than previously possible. At first, ultrasound scanners were only an imaging system. The fast acceleration of the development rate of ultrasound techniques started in the late eighties and early nineties when the concepts of blood flow imaging[1] and Tissue Doppler Imaging (TDI) was introduced[2], and the ultrasound scanners started to include a measuring function or modality, than just a pure representation of an image. The two concepts have triggered the progress of a various number of methods for the evaluation of functional variables, such as strain and strain rate. Furthermore, the recent introduction of ultrasound transducers with three-dimensional (3-D) capabilities have enabled seemingly more accurate rendering of volumes, such as the volume of the left ventricle (LV). The functional variables provide opportunities to use the ultrasound scanners not just for diagnostic purposes but also for treatment follow-up and patient monitoring purposes. If the ultrasound scanners are intended to be used for treatment follow-up or in monitoring situations it is essential that the measurements are correct and robust over time.

From a patient safety perspective, this development has led to a situation where new testing and evaluation methods are needed to supplement earlier methods, such as gray-scale resolution phantoms. However, this is a highly neglected area which may induce a risk of incorrect clinical decision making. The testing of ultrasound scanners has for a long time been focusing on resolution, image quality, and flow velocities[3-21]. Noticeably, there is

despite a massive effort by several groups no international consensus about a complete quality assurance protocol for ultrasound devices. When new medical devices and methods are developed and introduced on the market they are often favourably evaluated by phantom setups or against a reference method. These evaluation studies, which are almost invariably based on only one particular device, often conclude that the novel method is accurate and reliable[22-23]. This is a far too wide conclusion, which can lead to the use of inferior devices. The fact that one specific device has been evaluated favourably does not mean that all devices of that kind works well. It is very important that the credibility of a method is not confused with the function of a particular device.

An example showing that resolution and Doppler phantoms is not enough to evaluate the condition of an ultrasound scanner is the clinical case that initiated the second study in this thesis. A patient with suspicion of a congenital heart disease was examined following a standard echocardiography protocol, but the examination did not confirm the suspicion; there was no sign of a pathological blood flow associated to the disease (Figure 1, left panel). Slightly more than a year later, the patient was re-examined with a different ultrasound scanner (Figure 1, right panel). This time the pathological flow of a patent ductus arteriosus was clearly visible in the echocardiogram. The green and red area in the image represents a turbulent jet passing back into the pulmonary artery from the aortic arch via the patent ductus arteriosus. The reason for this mistake was found to be a defective ultrasound transducer. This finding prompted a large study to investigate the magnitude of the problem.

All four studies in the thesis have been performed in areas where testing of medical devices is lacking. The general aim of the thesis has been to show the urgent need for more extensive testing and continuous evaluation of medical devices.

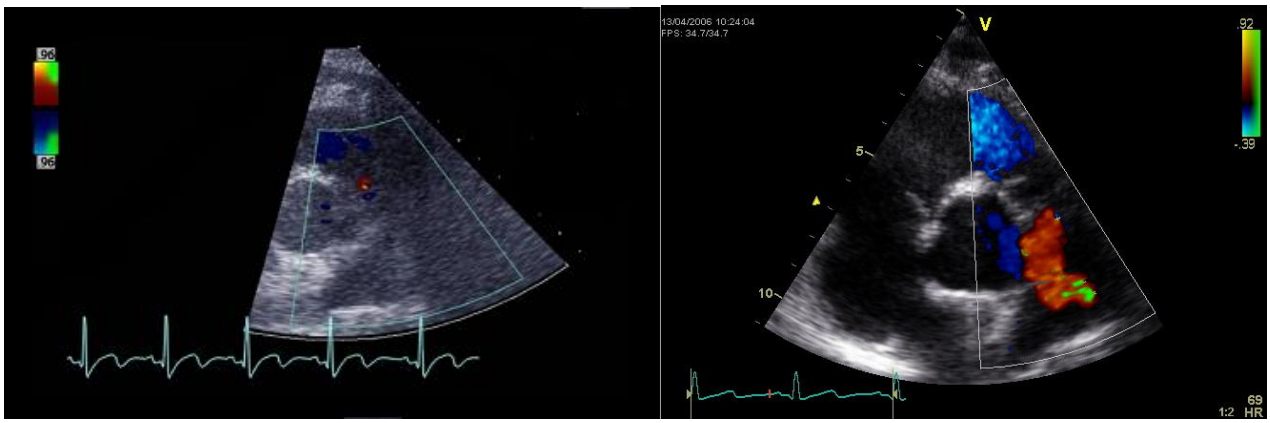

Figure 1: The clinical case showing a missed patent ductus arteriosus. Left panel: First examination without information about the pathological blood flow. Right panel: Second examination revealing the patent ductus arteriosus.

2 AIMS

The aim of the studies in this thesis was to evaluate the performance of medical devices with focus on ultrasound-based medical devices. The studies have been performed in areas where problems have been identified earlier or where evaluation of new untested methodologies was needed. The aims of the studies are listed below:

Study 1: To test the accuracy of simplified 3-D echocardiography for the quantification of LV volumes *in vivo* and *in vitro* phantoms using multi-slice computed tomography (MSCT) as reference method.

Study 2: To evaluate the function of ultrasound transducers in use in routine clinical practice and to estimate the incidence of defective transducers.

Study 3: To investigate whether annual testing is sufficient to decrease the number of defective ultrasound transducers in use in routine clinical practice that was reported in Study 2.

Study 4: To test the accuracy and to assess the diagnostic interchangeability of tissue Doppler-based displacement, velocity, strain and strain rate measurements in commercially available ultrasound systems using a dynamic phantom.

3 LIST OF INCLUDED PAPERS

The thesis is based on the four papers listed below. Full versions of the papers are attached as appendices at the end of the thesis.

I. Assessment of left ventricular volumes using simplified 3-D echocardiography and computed tomography - a phantom and clinical study. **Mattias Mårtensson**, Reidar Winter, Kerstin Cederlund, Jonaz Ripsweden, Habib Mir-Akbari, Jacek Nowak, Lars-Åke Brodin. *Cardiovascular Ultrasound,* vol 6:26, 2008

II. High incidence of defective ultrasound transducers in use in routine clinical practice. **Mattias Mårtensson**, Mats Olsson, Björn Segall, Alan G. Fraser, Reidar Winter, Lars-Åke Brodin. *European Journal of Echocardiography*, vol 10, pp. 389-94, 2009

III. Ultrasound transducer function: annual testing is not sufficient. **Mattias Mårtensson**, Mats Olsson, Lars-Åke Brodin. *European Journal of Echocardiography*, vol 9, pp. 801-5, 2010

IV. Evaluation of tissue Doppler-based velocity and deformation imaging: a phantom study of ultrasound systems. **Mattias Mårtensson**, Anna Bjällmark, Lars-Åke Brodin. *European Journal of Echocardiography*. doi:10.1093/ejechocard/jer056

4 BACKGROUND

This chapter provides a brief background regarding the techniques and methods used in the studies.

4.1 Computed tomography

Computed tomography, often abbreviated to just CT but sometimes also called CAT-scan from computed axial tomography, is an imaging method giving cross section images of the body. CT is an x-ray modality where the x-ray tube, which emits the x-rays, and the x-ray detector are rotated around the examined object. During this rotation, attenuation data are collected from a large number of angles. One angle gives information only from one image line, but all angles together create a cross section image of the examined object. Figure 2 shows a cross section image of a section of the heart where the LV is delineated. With modern CT scanners the x-ray tube and the detector are rotated while the body is translated through the CT scanner giving the possibility to generate 3-D data through many two-dimensional (2-D) cross section images. Furthermore, there are CT-scanners equipped with multiple x-ray detectors allowing faster image generation. When multiple detectors are used the method is called multi slice computed tomography (MSCT). By using cross section images with known spacing distance, the volume of a structure can be calculated. First, the cross section area of the region of interest is delineated in all available images. Thereafter, the unknown parts between the cross section images are mathematically stitched together by an interpolation algorithm.

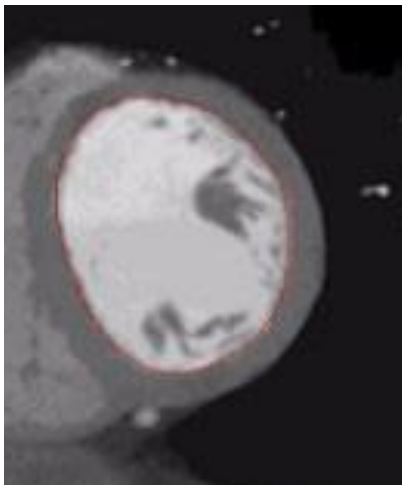

Figure 2: Cross section MSCT image of the LV in diastole. Image from Study I where the myocardial border has been delineated in the volume rendering process. MSCT; multi-slice computed tomography, LV; left ventricle.

4.2 Ultrasound imaging

Sound propagates through and interacts with tissue in such a way that it can be used for image rendering of tissue structures. Sound is a mechanical wave that varies in pressure while propagating through a medium, for instance audible sound in air. Audible sound has pressure variations (frequencies) in the range up to 20 KHz, and is sometimes called infrasound. Sound with frequencies higher than the audible range is called ultrasound. The most commonly used frequencies in medical diagnostic ultrasound are in the range of 1 to 15 MHz. The distance from one pressure maxima to the next is called wavelength. The wavelength is therefore directly related to the frequency; the higher the frequency the shorter the wavelength.

Echoes are produced when a propagating ultrasound wave interacts with anatomical tissue structures. In diagnostic medical ultrasound imaging (sonography) the echoes are used to create gray-scale images of the internal anatomy of the human body. In order to get an ultrasound image with the desired information, the choice of frequency is important. A high frequency results in a short wavelength which gives a high resolution image, but waves of high frequencies are dampened much faster than low frequency waves and have thus a shorter penetration depth. If the region of interest is situated deep in the body, high resolution has to be sacrificed for better penetration. Consequently, there is always a trade-off between these two factors.

The pulse-echo technique

Ultrasound imaging produces gray-scale images that consist of bright points of different intensities on a dark background. Darkness in the image means absence of echoes from that

area, or that the emitted ultrasound has failed to reach that specific area. Consequently, brightness in the image means that a part of the emitted ultrasound wave has echoed back.

The reasons for the generation of echoes/bright areas in the image are mainly twofold. Firstly, when a sound wave is propagating through a homogenous medium no energy of the sound wave is reflected back and consequently the resulting image is dark. However, if the examined volume constitutes of several media with different acoustic impedance, some of the energy is reflected back in the transition from one medium to another medium. Acoustic impedance is an important media characteristic in ultrasound imaging; it is the product of the density and the speed of sound of the media. The border between two different medium with the same acoustic impedance cannot generate a returning echo; the ultrasound wave passes through the interface without loss of energy in the form of a returning echo. It is the returning echoes from the interface between different medium that constitute the information about the examined structure in the ultrasound image. The proportion of the energy that is reflected back depends mainly on the difference in acoustic impedance and to a lesser extent by the angle of incidence.

Secondly, echoes can arise within inhomogeneous medium because of the scattering effect by very small objects in the size of the wavelength or smaller. These small objects are too small to be individually resolved in the ultrasound image. The echoes generated in this way are very weak compared to echoes from interface transitions and are undetectable. A very large number of echoes from many scatterers can however be added together by constructive interference and become detectable. This phenomenon is seen in almost every tissue in the body except blood and gives rise to so-called speckles in the ultrasound image. A speckle pattern is the combined result of both constructive and destructive interference when the sound propagates through the tissue back to the transducer. Speckles are thus not echo information of true tissue structures what the interface transition echoes are.

The ultrasound image is built up from scan lines. Sound waves propagate in straight lines. If the sound waves interact with the tissue media during the propagation along the scan lines and are reflected back, an echo can be registered by the ultrasound scanner. The placement of the echo signal along a scan line depends on two factors; the time for the ultrasound to return back to the transducer and the propagation speed of the sound in the material. The travel time can always be measured accurately by the ultrasound scanner. It is however more problematic to know the propagation speed of the sound since the human body consists of many tissue types with different properties. The speed of sound in soft tissues of the human body is ranging from about 1440 m/s in adipose tissue to 1580 m/s in muscle tissue. Therefore the manufacturers of ultrasound scanners use a fixed mean value regardless of tissue type. The mean value differs between the manufacturers, but is in the range of 1540 – 1560 m/s.

4.3 Ultrasound transducer design

The word *transducer* is a collective name for devices that converts one form of energy to another. The ultrasound transducer converts electricity to mechanical waves in the ultrasound frequency range, or converts the energy in mechanical waves to electricity. In medical imaging, transducers operating in the range from 1 to 15 MHz are typically used. The transducer design does not differ much from different manufacturers when it comes to the respective components. There are components that are common in all ultrasound transducers for medical imaging. The most important components are described below and showed in Figure 3. Sometimes the word probe or ultrasound probe is used instead of ultrasound transducer.

Piezoelectric element

Piezoelectricity is a phenomenon which means that certain materials can produce an electric current when a pressure is applied to them, and that an applied electric current deforms these materials. There are materials that hold this property naturally, for example quartz. The piezoelectric materials used in transducers for medical imaging consist of ceramics that are given this property by a manufacturing process where the ceramics are heated in a strong electric field.

The piezoelectric property gives the material the ability to operate both as transmitter and receiver of mechanical waves. If an alternating electric current is applied to a piezoelectric material it starts to vibrate and sends out a mechanical wave. The mechanical wave that is sent has the same frequency as the frequency of which the piezoelectric material vibrates, which in turn is the same frequency as the electric current applied to the piezoelectric material. When an incoming mechanical wave hits the surface of a piezoelectric material it produces an electric current with the same frequency as the frequency of the mechanical wave. In transducers for medical applications there is an array of many small piezoelectric elements (Figure 3).

The term piezoelectric element is often abbreviated simply to element, or sometimes crystal. The number of elements in a transducer can vary greatly but is often a multiple of 64; common numbers are 64, 128 and 192.

Matching Layer

If an alternating electric current is applied to a piezoelectric material it starts to vibrate. The vibration can be transferred into another material in the form of a mechanical wave, the tissue of a human body for instance. It would however be inefficient to transfer an ultrasound pulse through an interface of the high impedance piezoelectric elements directly in to the low impedance tissue. The difference in acoustic impedance between the piezoelectric elements and the tissue would be too large and result in the reflection of a

major portion of the total energy of the pulse when the pulse is transmitted. The same loss of energy would also occur when an echo returns and must pass through the same interface once more. Only a small fraction of the energy of the original pulse would return back to the piezoelectric elements. That would result in degradation of penetration depth and resolution. The problem is counteracted with a *matching layer* between the piezoelectric elements and the tissue. The matching layer is attached to the piezoelectric elements (Figure 3). The matching layer material has an acoustic impedance value between the values of the piezoelectric material and the tissue. There is usually more than one layer, where each layer has an acoustic impedance value closer to the acoustic impedance of the tissue. The outermost layer may also have a focusing function and be working as an ultrasound lens to minimize the thickness of the emitted ultrasound beam (Figure 3).

Backing material

A heavy metal based material, called *backing material* or damping material, is placed in contact with the piezoelectric elements on the opposite side of the matching layer (Figure 3). The vibration of the piezoelectric elements caused by the electric stimulation is an absolute necessity to generate the ultrasound pulse. But the same vibration can cause problems. If the vibration is not dampened out as fast as possible the generated ultrasound pulse would be useless; it would last too long. The axial resolution is directly related to the length of the pulse emitted by the transducer. A typical pulse length is 2-3 cycles. Another function of the backing material is to absorb residual ultrasound energy that is reverberating in the piezoelectric element between the matching layer and the backing material.

The electric cable

The electric cable which connects the transducer with the connector in the other end consists of several smaller wires. The number of wires within the cable can be as many as the number of elements. This is common in transducers with up to 128 elements. In wide array transducers with a larger number of elements there are however often fewer wires than elements. In these transducers, multiplexers are used to control the elements. A multiplexer is an electrical component that reduces the number of wires by having a set of wires controlling a larger number of elements.

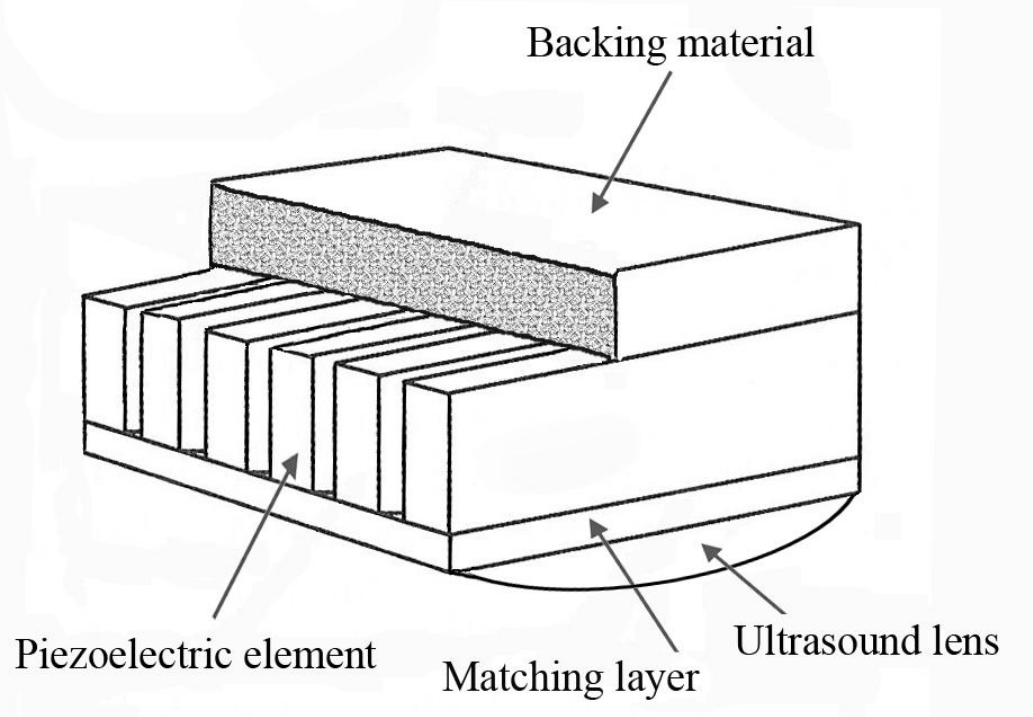

Figure 3: The internal parts of an ultrasound transducer.

Transducer types

There are a number of different transducer types. The most common types are linear transducers, curved linear (or curvilinear) transducers and the phased array transducers (Figure 4). The linear transducer offers a rectangular image with the same width as the transducer. It is therefore the preferred transducer design for imaging at short distances, e.g. vascular imaging of the carotid artery. The curved linear transducer type is of the same principle as the linear transducer but the piezoelectric elements are mounted on a curve instead of in a straight line. This type of assembly yields an increasingly wider field when the search depth increases, which makes it suitable for abdominal applications where larger tissue structures often are imaged.

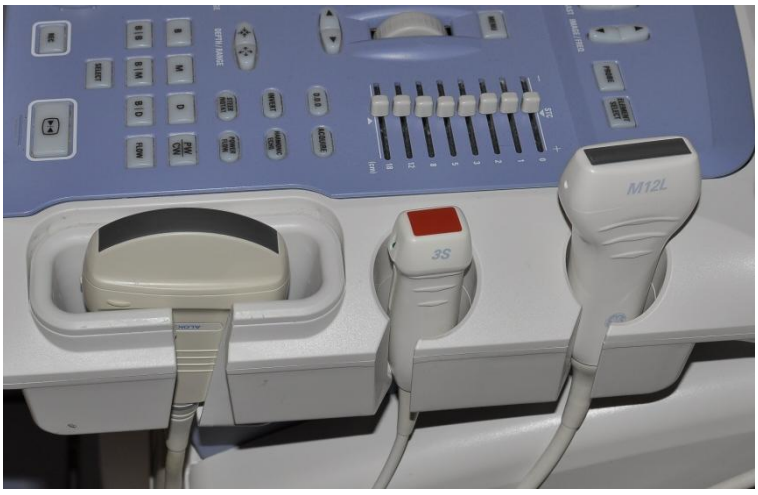

Figure 4: Three transducers of different types. The leftmost is a curved linear transducer, the middle is a phased array transducer, and the rightmost is a linear transducer.

The phased array transducer has an array of piezoelectric elements similarly mounted as the linear array. The phased array transducer does however not, in contrast to the linear array, transmit the ultrasound pulses in straight lines perpendicular to the transducer surface. The ultrasound pulses are steered from side to side generating a triangular shaped image. The phased array transducer is especially designed for cardiac applications, where the linear transducer would offer a too narrow sector width and where a curved linear transducer would not make it possible to get skin contact with the whole transducer due to the costal bones. One of the latest design features is to build transducers consisting of not just one, or a few arrays of piezoelectric elements, but a large number of rows, which result in a more rectangular shaped transducer. The design enables image rendering in three dimensions. The drawback is that the procedure is more time consuming compared to the 2-D alternatives.

Simplified 3-D ultrasound imaging — the tri-plane method

A method to compensate for the long computational time of 3-D ultrasound imaging is the simplified 3-D ultrasound imaging technique. The method is not a true 3-D image technique. Instead the examined object is scanned in three arbitrary 2-D planes (Figure 5). If a structure of interest is delineated manually, the shape and volume of the structure can be calculated by an interpolation algorithm. Figure 5 shows when the volume of the 334 ml phantom used in Study I was measured by use of the tri-plane method.

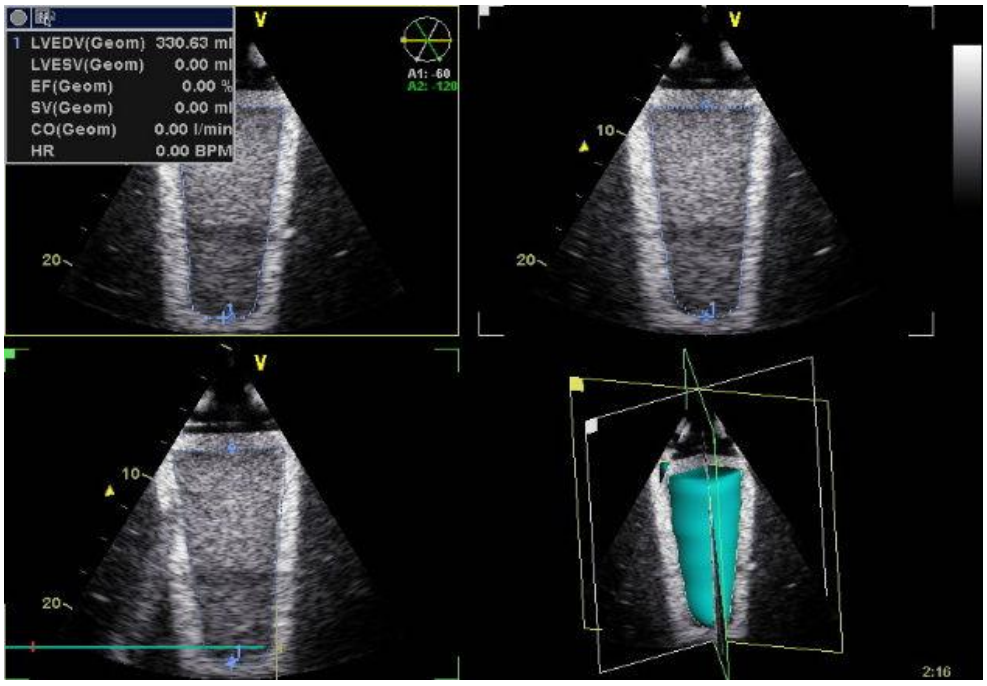

Figure 5: One of the volume phantoms in Study I imaged by the tri-plane method, a simplified 3-D ultrasound imaging technique. 3-D; three-dimensional.

4.4 The Sonora FirstCall Test System

The function of the transducers was tested with the Sonora FirstCall Test System (Sonora Medical Systems Inc. Longmont, CO, USA). When performing this test, the transducer is connected to the testing system instead of the ultrasound scanner. The test is performed in water where every individual element in the transducer is activated one by one. The emitted ultrasound pulse is reflected back to the transducer using a metal plate. The returning pulse is analyzed by means of the peak-to-peak amplitude, centre frequency, pulse width, bandwidth and the pulse waveform. Furthermore, the accumulated capacitance of every element and its wires is measured to test if cable brakes and short circuits are present. The peak-to-peak (p-p) amplitude (sensitivity), measures the ability of a transducer to emit and receive ultrasound pulses.

The two most important test results, the sensitivity and capacitance, are presented in histograms with the same number of bars in the histogram as piezoelectric elements within the tested transducer (Figure 6 and 7).

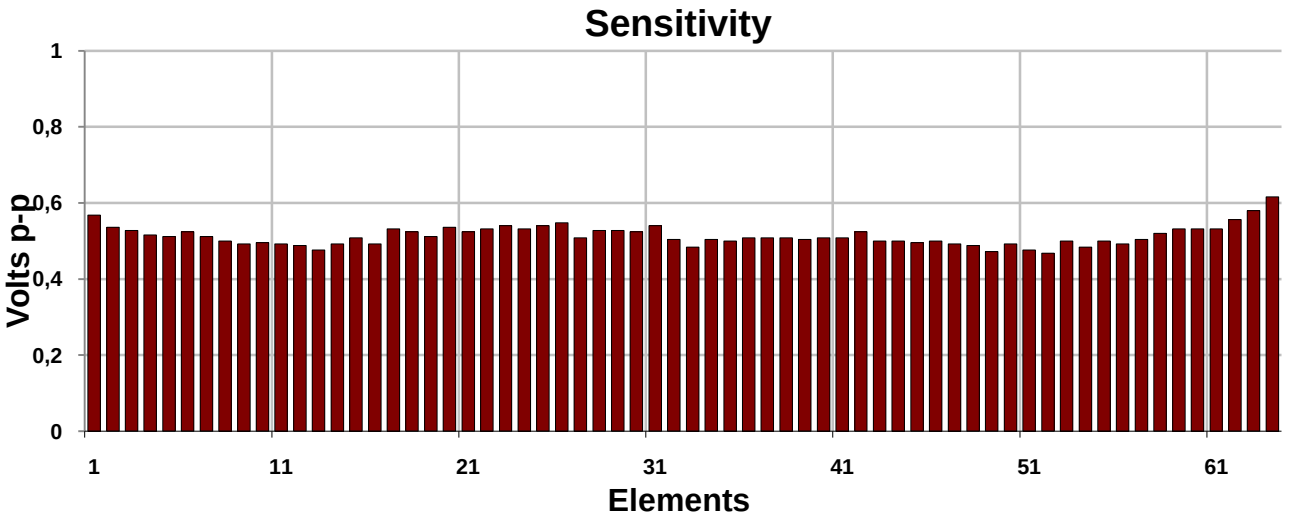

Figure 6: The element sensitivity histogram of a 64 element phased array transducer from the Sonora FirstCall Test System. The peak-to-peak (p-p) amplitude (sensitivity), measures the ability of a transducer to emit and receive ultrasound pulses.

The height of the sensitivity bars in the histogram is a result of the amplifier setting. The absolute value is therefore not significant. The most important aspect is that the sensitivity value is at a constant level for all transducer elements. There are however transducers with different kinds of piezoelectric elements and thus multiple sensitivity levels in the same transducer (Figure 21). In this case, the sensitivity should be constant within each sensitivity level. The bars in the capacitance histogram show the combined capacitance of each element and its signal wires. Since the elements are almost identical and the signal wires are of the same length, the capacitance values should also be constant for all transducer elements.

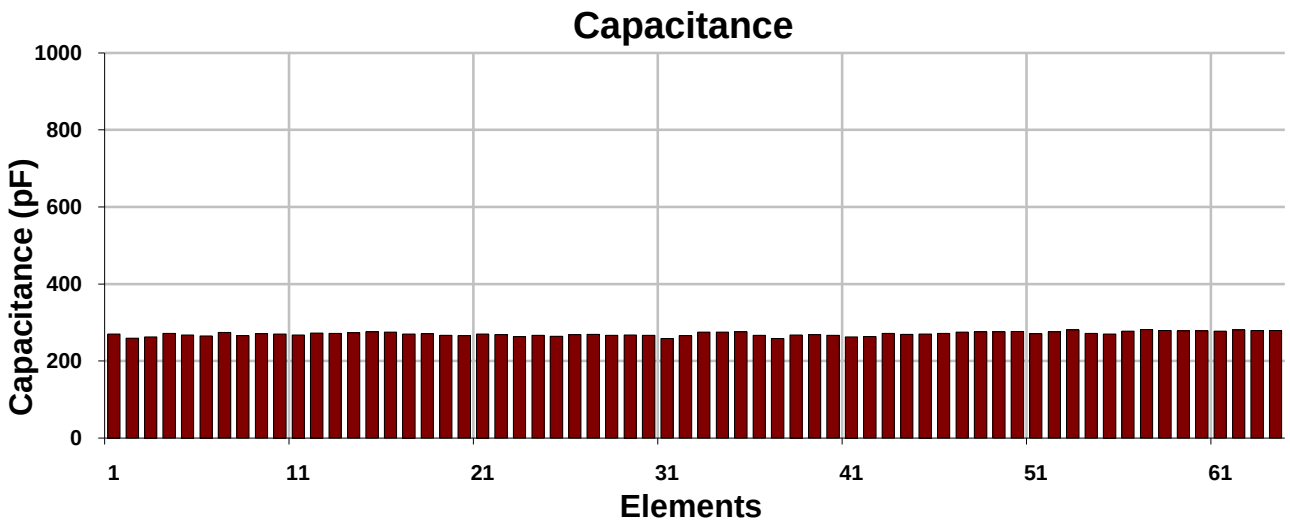

Figure 7: The histogram of a 64 element phased array transducer showing the total capacitance (element and wires in both directions) from the Sonora FirstCall Test System.

Figure 8 shows the sensitivity histogram for the defective transducer used in the clinical case with the missed persistent ductus arteriosus mentioned in the introduction. The capacitance histogram was normal for that transducer. Figure 9 shows the capacitance histogram for a transducer with two kinds of electrical errors. If the capacitance level is lower than normal there is a broken signal wire to the element. A capacitance value higher than normal means that there is a short circuit.

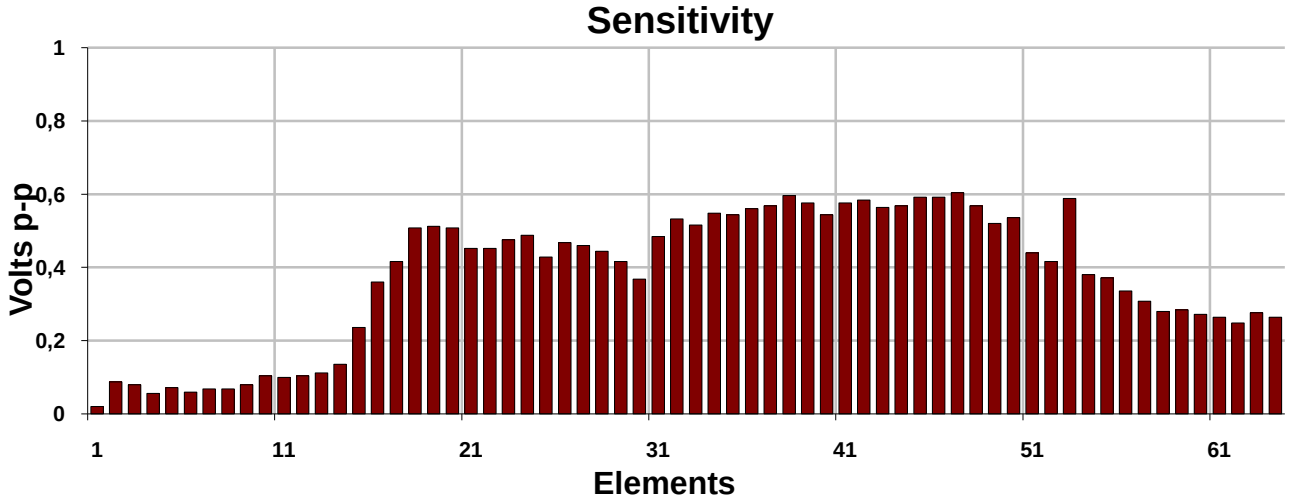

Figure 8: The sensitivity histogram of a delaminated 64 element phased array transducer. Note that 0.6 is the expected value and how differently the elements are affected. The peak-to-peak (p-p) amplitude (sensitivity), measures the ability of a transducer to emit and receive ultrasound pulses.

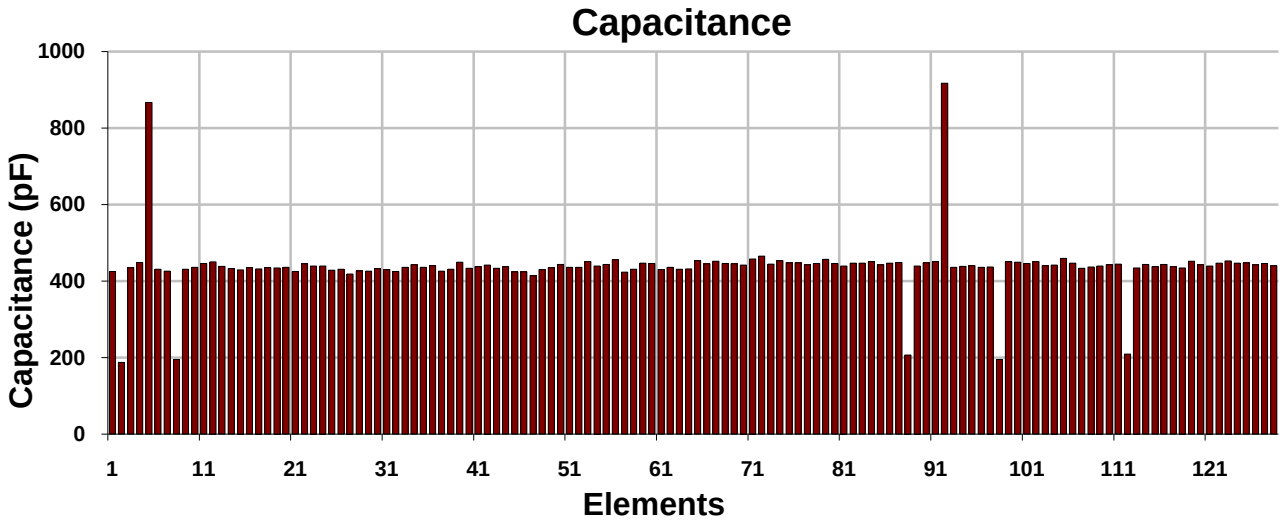

Figure 9: The capacitance histogram of a 128 element phased array transducer both with breaks in the cable and short circuits. Break in the cable give capacitance values lover than normal, whilst short circuits give higher. In this transducer there are 5 broken wires and two short circuits. Note that the abnormal values are arbitrary and can assume any value. When it looks like this the wires are broken at the same place.

4.5 Tissue Doppler imaging

Tissue Doppler Imaging (TDI)[2] is a further development of color Doppler imaging[1] used for blood flow velocity measurements. TDI is of special interest as a tool in cardiology to assess myocardial function. When a propagating ultrasound wave hits moving tissue, the frequency of the ultrasound wave is shifted. The size of the shift is directly related to the velocity of the moving object and the phenomenon is known as the Doppler shift. If the propagating wave interacts both with flowing blood (high velocities) and moving tissue (low velocities) there will be a combination of Doppler shifts from the two sources. These signals can be separated by applying a filter to keep the signal of interest. The 2-D image of the moving tissue are then coloured in a red to yellow spectrum when the movement is directed towards the transducer and in a blue to dark blue spectrum when the movement is directed away from the transducer.

A common feature of many modern ultrasound devices is that a certain point of interest can be tracked and displayed in velocity-time graphs. Velocities (cm/s) are not the only parameters of interest. Displacement (mm), the total performed motion calculated as the temporal integration of the velocity, can provide additional information to the velocity measurement. Another type of measurement is to assess the deformation parameters strain (%) and strain rate (1/s). Strain is the percentage change between two points in time and the strain rate is the speed with which the deformation occurs[24]. Strain and strain rate can respectively be calculated by spatial derivation the displacement and velocity data. It is in fact theoretically possible to calculate any of these parameters from each other. The calculation steps are shown in Figure 10.

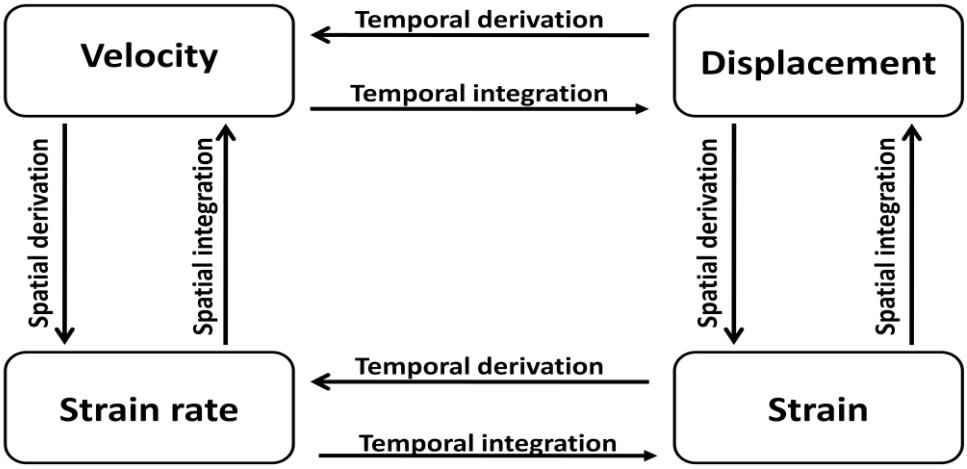

Figure 10: Diagram showing how the parameters velocity, displacement, strain, and strain rate can be calculated from each other.

4.6 The Bland-Altman plot

When methods for measuring are to be evaluated, the results from a method can either be compared to known true values or the results from another method measuring the same parameter. Two properties are often of interest; the correlation and the agreement of the methods. The correlation is preferably tested by standard linear regression. However, the correlation does not provide information about agreement or accuracy of the methods. When two different methods or devices are used to measure the same parameter it is often of interest to visualize the agreement of the methods. This can be done by the Bland-Altman plotting method [25]. In the Bland-Altman plot, the difference of the measurements is presented as a function of the measured mean. Lines indicating the mean difference and the limits of agreement (± 2 standard deviations (SD)) are commonly plotted in the Bland-Altman plot. Sometimes the 95% confidence intervals for the limits of agreement are also plotted, see Figure 17. If available, true values can be used instead of one method to show the accuracy of the other method.

4.7 Error types

When analyzing measurements, different kinds of error types may occur. The result of measurements can be described as how accurate they are and the spread of them (Figure 11). This description makes the discussion of measured results easier. The result in *a* is the perfect result and gives reliable data. The error type in *b* is not correct, but with this low spread it can easily be compensated for. The result in *c* is difficult to handle. The spread is in all directions and cannot easily be compensated for. For this error type, the requirements in the application where the measurements are intended to be used must

decide whether the accuracy and spread is acceptable. The low accuracy and large spread in *d* makes this error type uncorrectable and most likely completely useless in every application. The spread in measurements is described by the SD. The SD is low in *a* and *b*, whereas it is high in *c* and *d*.

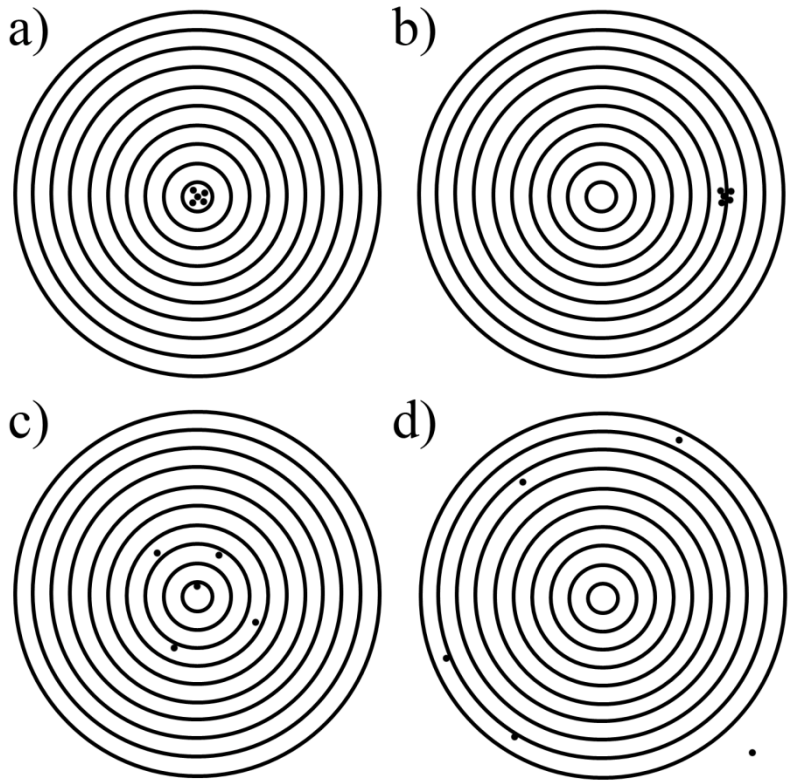

Figure 11: Error types illustrated by shooting targets. a) Accurate with low spread. b) Inaccurate with low spread. c) Accurate with large spread. d) Inaccurate with large spread.

4.8 Rules and regulations from the Medical Product Agency

In Sweden, the regulation and surveillance of development and manufacturing of medical products are under the jurisdiction of the Medical Product Agency, MPA, (Läkemedelsverket). Their task is to ensure that safe and effective medical devices are placed on the market. The regulation from the Medical Product Agency that controls the development and manufacturing of medical devices is the document LVFS 2003:11[26], which is based on the European medical device directive (Council Directive 93/42/EEC of 14 June 1993 concerning medical devices, with the latest update 2007/47/EC). In the first chapter of supplement 1 in this document, the following essential requirements are stated in the third note: "*the products must have the performance claimed by the manufacturer*". This requirement

applies to all medical devices, for instance imaging devices in general. There are however special requirements for medical devices used for measuring purposes. In the first note of the tenth chapter of supplement 1, it is stated that "*products with measuring functions must be designed and manufactured in such a way that the measurements are accurate and within tolerances sufficiently for the intended purpose*". The compliance is assessed by an accredited organisation called a *Notified Body*. The Notified Body must be accredited by a member state in the European Union. It is further stated in LVFS 2003:11 that a manufacturer shall give, the Notified Body that checks if the requirements are complied, complete access to any information necessary and the possibility to perform necessary checks.

5 METHODOLOGY

The studies in this thesis included evaluation of ultrasound systems both in experimental *in vitro* setups and in a clinical *in vivo* study. Moreover, ultrasound transducers were evaluated using a commercially available test system.

5.1 Phantom construction

In two of the studies (Study I and Study IV), phantoms for evaluation of LV volumes and TDI-based motion and deformation measurements were constructed.

Volume phantoms

Ten cone-shaped phantoms for LV volume measurements were manufactured in Study I. Eight of the phantoms with symmetric geometric structure and two of them with asymmetric geometry. The asymmetric phantoms were considered to mimic LV aneurysms. The phantoms were manufactured using water-based agar. The phantoms consisted of three agar-based compartments. The innermost compartment should emulate the LV and the second compartment the myocardium (Figure 5).

In order to increase the x-ray attenuation and thereby make the innermost compartment detectable for the MSCT scanner an x-ray contrast agent (Iodixanol, Visipaque, GE Healthcare, Little Chalfont, UK; 320 mg I/ml) was added to the agar. Glass powder (GL1, KMC, Järfälla, Sweden) with very fine granularity was added to the second compartment to mimic the blood-myocardium interface. The outermost compartment consisted only of agar and was the bulk material enclosing the phantom. The concentrations of the contrast agent and the glass powder were chosen to yield the same Hounsfield values and speckle pattern as in normal clinical scans of the tissue type they were supposed to mimic.

The density of the mixture constituting the inner compartment was measured before the manufacturing of the phantoms. During the manufacturing process the specific weight of

each phantom was measured before and after adding the mixture. From this information the volumes of the inner compartments of the phantoms were calculated. The speed of sound within the phantoms was found to be 1540 m/s in the innermost compartment and the bulk material and 1470 m/s in the second compartment. The volume of the innermost compartments ranged from 39 to 334 millilitres. All volumes are given in Table 1.

Table 1: Phantom volumes and their geometric structure.

Phantom (nr.)	Volume (ml)	Geometric structure
1	39	Symmetric
2	98	Symmetric
3	334	Symmetric
4	293	Symmetric
5	127	Symmetric
6	52	Symmetric
7	202	Symmetric
8	246	Symmetric
9	68	Asymmetric
10	157	Asymmetric

Tissue Doppler phantom

The dynamic phantom used in Study IV was based on the fatigue testing machine ElectroPuls E3000 (Instron, Norwood, Massachusetts, USA) (Figure 12). This machine, normally used for dynamic testing of material properties, can be programmed to perform motions with very complex wave forms. The programmed motion is a request to the machine, not necessarily the actually performed motion. The actually performed motion is accurately monitored by a sensor on the motor shaft, which enables comparisons of "true" phantom values and the values measured by the ultrasound device.

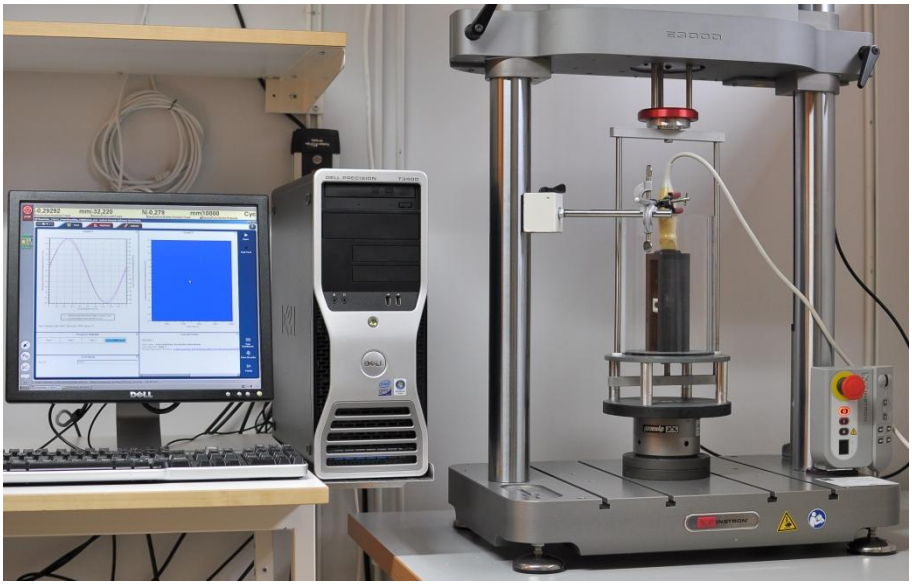

Figure 12: The tissue Doppler phantom. Phantom mounted with the strain/strain rate setup.

In order to evaluate both displacement, velocity, strain, and strain rate measurements two phantom setups were created; one setup to measure displacement and velocity and another setup to measure strain and strain rate (Figure 13). A static tissue mimicking object was moved up and down in the first setup and in the second setup a tissue mimicking object was stretched in an oscillating motion.

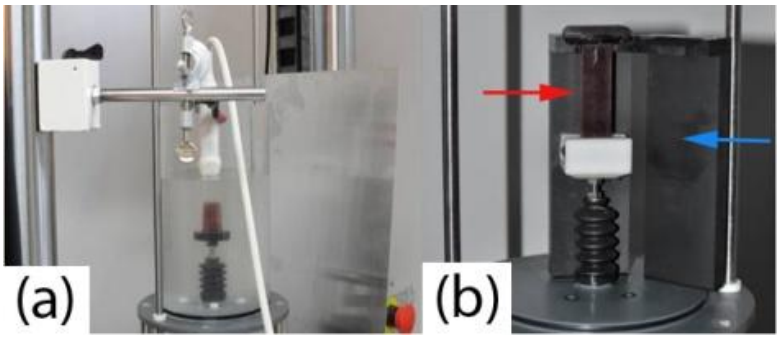

Figure 13: (a) Displacement/velocity setup of the phantom. (b) Strain/strain rate setup of the phantom. The water container has been removed and the front part of the suspension device (blue arrow) has been moved and rotated to the right to display the inside of the setup. The tissue mimicking object is shown by the red arrow.

To achieve a motion as true as possible, a displacement curve from the basal septum of a healthy individual, examined with a GE Vivid*i* ultrasound scanner (GE Healthcare, Horten, Norway) and post-processed on a GE Echopac BT10 (GE Healthcare, Horten, Norway) workstation, was used as input data to the phantom in both setups. The tissue mimicking object was *moved* towards the transducer in the velocity/displacement setup and *stretched* in the strain/strain rate setup in accordance with the imported *in vivo* displacement curve. A sequence of three heart beats was used (Figure 14).

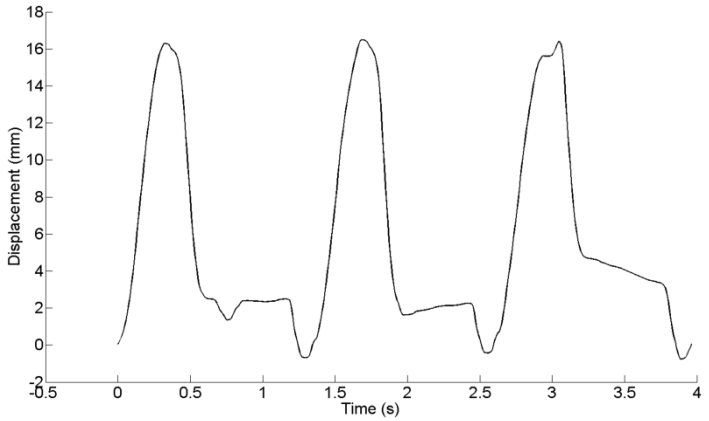

Figure 14: The displacement curve that was imported to the tissue Doppler phantom.

Polyurethane (RenCast 5073, Huntsman Advanced Materials, The woodlands, Texas, USA) was used as tissue mimicking material. In order to generate speckle patterns in the ultrasound images similar to the ones of myocardial tissue, a small amount of glass with very fine granularity was added (GL1, KMC, Järfälla, Sweden). The speed of sound of the tissue mimicking material was measured to 1490 m/s. The phantom setups were immersed in a mixture containing 11% glycerol (Sigma-Aldrich, St. Louis, Missouri, USA) in deionized water. The concentration of glycerol was chosen to yield a speed of sound of 1540 m/s.

The ability of the phantom to accurately repeat the imported displacement curve was monitored during the tests in Study IV. The three heart beats in the imported displacement curve were not identical, and consequently the peak amplitude values differed slightly. For that reason, the SD and coefficient of variation (CV) were calculated separately for the three peaks. The calculations were based in the combined data from all tests in Study IV. The SD and the CV for the phantom displacement were calculated to be 0.01 mm and 0.00, respectively. When the same statistics were calculated for the phantom velocity, the SD and CV were calculated to be in the ranges 0.12 – 0.16 cm/s and 0.01 – 0.02, respectively.

5.2 Study population

Nine patients (4 female, age range 51-82 years) were included in Study I. They were consecutively selected from another study evaluating MSCT for the detection of coronary artery disease in patients referred routinely for coronary angiography due to known or suspected coronary artery or valvular disease. In connection with this examination, a 3-D echocardiography was performed in all patients.

5.3 Echocardiography and ultrasound scanners

The echocardiographic part of Study I was performed using a GE Vingmed Vivid 7 ultrasound scanner equipped with a V3 matrix transducer. The matrix transducer allows simultaneous data acquisition in three different scan planes positioned at 60° angle to each other. In the phantoms, five repeated echocardiographic acquisitions were performed from the basal end with the phantoms submerged in water. In the patients, three consecutive cardiac cycles were acquired at end-expiration from a thoracic apical window. In order to obtain the LV volumes, the data sets were post-processed offline using a GE Echopac workstation version BT06. In the three simultaneously generated images, the interface of the two inner compartments in the phantoms and the myocardial-lumen border in the patients were manually outlined. The phantom volumes, the LV end-systolic and end-diastolic volumes were estimated using the Echopac software.

Four different scanner models were included in Study IV. Two different scanners of the same model were tested and one scanner acquisition was tested twice with two generations of the same workstation giving six test results in total. The scanners were in active clinical use and under regular maintenance routines. The transducers used during the acquisitions were tested in a transducer tester prior to the study, and showed normal function. All measurements were performed with the default scanner settings for a tissue Doppler-based echocardiographic examination, yielding frame rates from 92 to 168 frames per second, pulse repetition frequencies from 1000 to 1250 Hz, and transducer transmission frequencies around 1.7/3.4 MHz for the harmonic gray-scale image and 2.4 to 3.4 MHz for the TDI. Ten acquisitions were acquired with both phantom setups. The acquisitions were then post-processed on dedicated workstations to generate text files with displacement, velocity, strain and strain rate data. These text files, together with the text files from the phantom, were then imported to Matlab for further processing.

5.4 Multi-slice computed tomography

The MSCT examinations in Study I were performed using a GE LightSpeed VCT 64-slice spiral CT scanner. The MSCT settings, both for the phantoms and the patients yielded acquisitions with slice thickness and spacing of 5 and 4 millimetres respectively. The data sets were afterwards transferred to a GE Advantage Workstation for post-processing. The MSCT software that was used for the post-processing was the GE CardIQ Function version 1.0.3. The images from the patient data were reconstructed with retrospective electrocardiogram-gating with 10% increments of the cardiac cycle. The diastolic and systolic volumes were identified respectively at 0% and 40% of the cardiac cycle. Short-axis images were reconstructed from these positions and the inner phantom volume and the myocardial border were outlined with an automatic contour detection algorithm in the MSCT-CardIQ software. Images which did not include any parts of the LV or the inner phantom compartment were manually excluded before the volume calculation. Short-axis images, at the level of the mitral valve, usually show both parts of the LV and an area delineated by the mitral valve leaflets. The images with the ventricle representing more than half of the image area in this region were included in the volume calculation. Additionally, the volumes of the papillary muscles were included. The volume measurements were repeated five times in each phantom.

5.5 Ultrasound transducer testing

The function of the transducers was evaluated with the commercially available Sonora FirstCall Test System (Sonora Medical Systems Inc. Longmont, CO, USA), that was described in Chapter 4.4. When using this test system, the transducers were connected to the test system, which activated every individual element in the transducer one by one. The

transmitted ultrasound waves propagated through water and were reflected back to the transducer by a metal plate. The received signal was then analyzed to estimate the function of the transducer.

Transducers and clinics

The transducers tested in Study II were used on a daily basis in clinical departments in 32 hospitals in the south of Sweden. Regular transducer testing was not performed at the clinics prior to the study. A total of 676 transducers were tested in a one-time test without follow-up. The transducers were of seven different brands, called A through G in the results section. Brand names and model names are not given because the errors found are not necessarily related to the manufacturing process and the durability of the tested transducers. Furthermore, the number of each transducer type varies greatly and may therefore give a misleading picture. All transducers were tested except for rectal, vaginal and stand-alone continuous wave Doppler transducers.

In Study III, a total of 299 transducers were tested. The transducers were tested at 13 clinics in 5 hospitals in the Stockholm area. These clinics were also included in Study II and since that study the clinics had introduced transducer testing to their annular maintenance routine. The specialties of these clinics were radiology, cardiology, clinical physiology, obstetrics, and gynaecology. During 2008, each of these clinics performed between 6900 and 15000 ultrasound examinations. Single element and 3-D transducers were not included. The tested transducers had one year before passed the test during the annual test routine or had been put in to service as a replacement of a defective one. Since there were no significant differences in the error rates between the transducer brands or models their respective names were given. The transducers originated from three different manufacturers: Siemens, GE, and Philips. The number of transducer models from Siemens, GE, and Philips were 22, 15, and 4, respectively, and the number of transducers was 227 from Siemens, 59 from GE, and 13 from Philips.

Acceptance criteria for piezoelectric elements

The first step to decide whether a transducer is in working order or not, is to judge if the elements within the transducer work properly. The transducer elements were in the test either classified as *Functionally Acceptable Element*, *Weak Element,* or *Dead Element*. The acceptance criteria used in order to interpret the test results were defined as follows:

I. *Functionally Acceptable Element* – element with a sensitivity value of over 75% of the mean value for <u>all</u> elements within a transducer.
II. *Weak Element* – element with a sensitivity value of between 10% and 75% of the mean value for <u>all</u> elements within a transducer.
III. *Dead Element* – element with a sensitivity value of below 10% of the highest value within a transducer.

The sensitivity of an element is the amplitude of the returning pulse using a perfect reflector at a specific distance and gain setting. The sensitivity value will thus increase with decreased distance to the reflector or increased gain setting, and vice versa. There is consequently no "true" reference value for the sensitivity amplitude. The mean value of all elements is therefore considered when the function of an individual element is judged.

The second step is to decide if the numbers of weak and dead elements are acceptable. In Study II and III, a transducer was considered as defective if:

I. the transducer contained more than four contiguous weak elements,
II. or more than two dead elements,
III. or two contiguous dead elements.

The criteria are chosen to be sure that no transducer induced errors can occur. It has been shown that two contiguous dead elements can have a substantial negative impact on the transducer function when using convex or linear transducers[27].

Transducer errors

Five definitions of transducer errors were used; *delamination*, *break in the cable*, *short circuit*, *weak elements* and *dead elements*. Delamination is when the matching layer detaches from one or more elements or the different matching layers separate from each other. When delamination occurs, the affected elements have sensitivity level lower than normal, where the reduction may be minor to a complete cessation of the pulse (Figure 8). Break in the cable and short circuit prevents activation of the elements and will result in a zero value of the sensitivity amplitude, and they will respectively result in a reduction and in an increase of the capacitance value for the affected element-wire unit.

5.6 Data presentation and statistical analysis

In Study I, the relation between LV volumes measured by the simplified 3-D echocardiography and the MSCT-CardIQ software was tested with standard regression analysis. The agreement of the measurements was illustrated by using Bland and Altman plots [25]. Paired data were presented as mean ± SD and compared using Student's *t*-test. The results were expressed as percentage, absolute numbers, and with a 95% confidence interval in Study II and III.

Comparisons between transducer error types, the manufacturers, and transducer types were performed in Study II using chi-squared tests. For the comparison between the transducer types, the transducers were grouped into nine categories: transesophageal transducers (phased array), adult cardiac transducers (phased array), paediatric cardiac transducers (phased array), radiology transducers (linear & curved linear array), linear transducers below 8 MHz, linear transducers between 8-10 MHz, linear transducers above

10 MHz, curved linear transducers below 6 MHz, curved linear transducers between 6-8 MHz.

Comparisons between transducers from the three manufacturers, transducer models, clinics, and clinic types were performed in Study III using Fishers' exact tests. For the comparison between the clinic types, the clinics were grouped into three categories: radiology, cardiology/clinical physiology, and obstetrics/gynaecology. Furthermore, the clinics were grouped into two categories, depending on how they handled their transducers. The clinics in handling category 1 had their transducers constantly plugged into the ultrasound scanners and the transducers were hanging on a suspension device on the scanners when not in use. The clinics in handling category 2 had their transducers disconnected from the scanners when not in use. Between the examinations the transducers were transferred to, and stored in, a special transducer storage place in the clinics.

In Study IV, the measured parameter values are the combined result of an ultrasound scanner and a workstation. Therefore, an ultrasound scanner and a workstation combination is hereafter called an ultrasound system. There were four ultrasound scanners in the study. The Ultrasound systems were called A through D indicating that the ultrasound systems were based on different ultrasound scanner types. Since there were two ultrasound scanners of the same model and two generations of the same workstation, two ultrasound systems were called A (A1 and A2) and D (D1 and D2) to highlight this relationship.

The comparison of the performance of the six ultrasound systems and the comparison between the ultrasound systems and the phantom were done by emphasizing the three peak values in the curves of each parameter. The differences in amplitude and timing of the peak values between the measurements of the ultrasound system and the phantom values were presented as mean ± SD. The mean amplitude differences were also given in percent of the phantom values. The phantom displacement values were the values registered by the motor shaft sensor mentioned in the phantom description, and the phantom velocity values were derived from the displacement values by temporal derivation.

The initial length of the tissue mimicking object was known and with the elongation data from the motor shaft sensor the strain was calculated. When the initial length is used as reference in a strain calculation the so called Lagrangian strain is calculated. The values for strain and strain rate can be calculated either as normal (Eulerian) or as Lagrangian. The normal strain has been suggested as most appropriate for myocardial strain[24], but since the manufacturers of the ultrasound systems used in this study claim that their systems measure Lagrangian strain, both definitions are presented in the results. In the result figures showing strain and strain rate, the Lagrangian alternative will yield slightly higher values. The difference increases with higher values, which leads to larger differences in the peaks than for the lower values. The phantom strain rate values were calculated from the strain values by temporal derivation identical to how the velocity values were calculated

from the displacement data. The natural strain and strain rate results were calculated as the natural logarithm of one plus the Lagrangian strain value. All the calculation steps are shown in Figure 15.

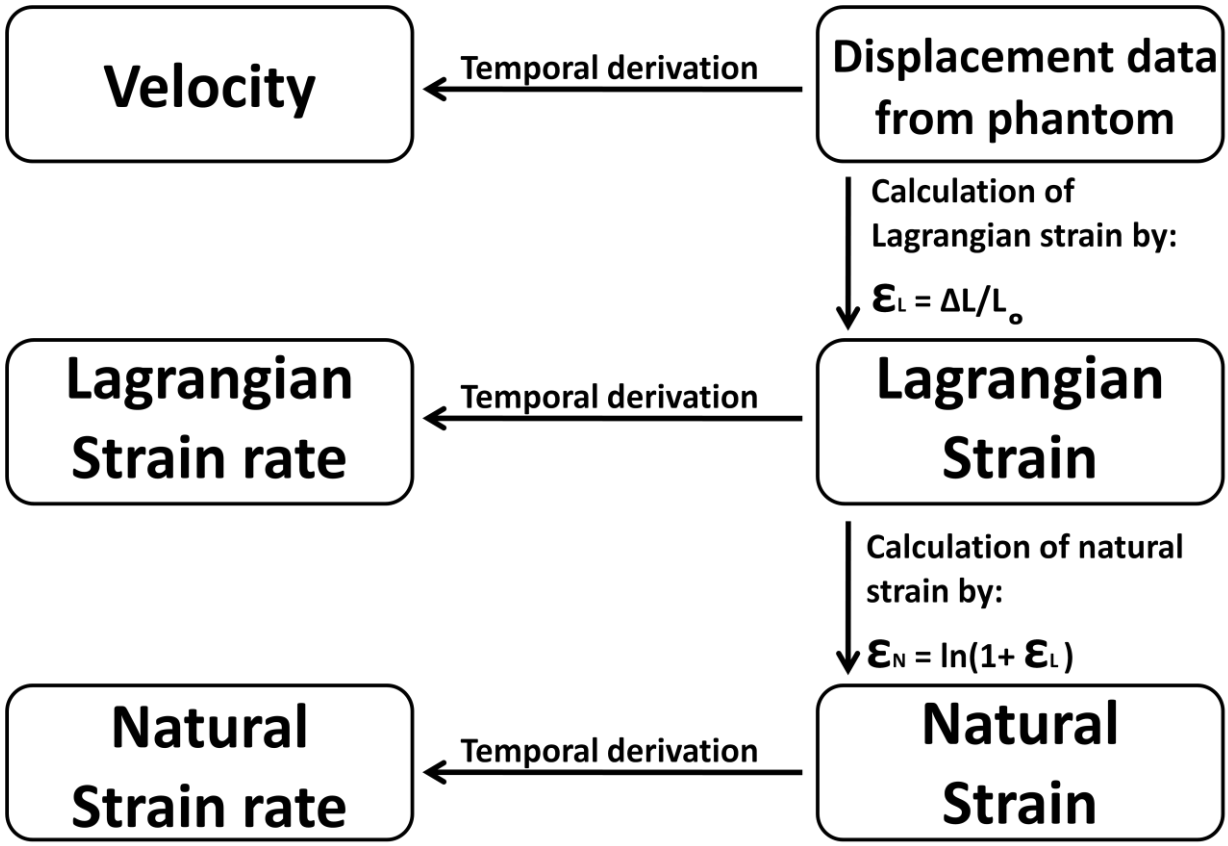

Figure 15: Schematic illustration of the different calculation steps used in Study IV. The velocity was calculated by temporal derivation of the displacement generated by the motor shaft sensor in the phantom. The strain was calculated by the Lagrangian strain definition and from this data, the Lagrangian strain rate was calculated by temporal derivation. The natural strain was calculated as the natural logarithm of one plus the Lagrangian strain value, and the natural strain rate was calculated by temporal derivation of the natural strain data.

6 RESULTS

The main results from the four studies will be presented in separate sub-chapters. Additionally, in Study II, three important observations are presented that were not included in the original article.

6.1 Study I: Assessment of left ventricular volumes using simplified 3-D echocardiography and computed tomography - a phantom and clinical study

In this study, the agreement between simplified 3-D echocardiography and MSCT was evaluated. The accuracy in the assessment of LV volumes formed the basis of the evaluation of the two modalities. The study consisted of two parts. The first part was *in vitro* calibration of the two modalities by volume assessments using phantoms with known volumes. The second part comprised *in vivo* LV volume assessments in clinical subjects.

***In vitro* measurements**

The automatic contour detection algorithm in the MSCT-CardIQ software provided fully adequate delineation of the inner phantom compartment, but failed to detect the irregularities of the asymmetric phantoms and manual correction was therefore necessary. Both tested methods correlated strongly with the phantom volumes. The regression analysis yielded correlation factors of one for both methods when compared with the phantom values, thus indicating a perfect correlation between the methods. However, the measured values from the two methods differed significantly in the individually performed phantom tests. The differences between the phantom volumes and the volumes measured by the two methods are shown in Figure 16. MSCT-CardIQ invariably overestimated the phantom volumes and the results were scattered. The mean difference for MSCT-CardIQ was about six millilitres (indicated by the dotted line in Figure 16) and about zero millilitres

for the 3-D echocardiography method (indicated by the solid line in Figure 16). When comparing these results with the shooting target illustrations in Chapter 4.7, the 3-D echocardiography results are of type *a*, since the measurements are accurate and with low spread, whereas the MSCT-CardIQ results are more like error type *c* (accurate, high spread), since the spread is higher.

The Bland-Altman plot in Figure 17 further highlights the mismatch between 3-D echocardiography and the MSCT-CardIQ software. Because of the constant overestimation by the MSCT-CardIQ software, the values on the x-axis are all negative; the values on the x-axis are the subtraction of the simplified 3-D echocardiography measurements and the MSCT-CardIQ software measurements. The mean difference is indicated by the solid line in Figure 17 (-6 ± 5 ml). The upper and lower limits of agreement, defined as mean difference ± 2SD, were +4 and -16 respectively. The 95% confidence intervals for the upper and lower limits were +9 to -3 ml and -10 to -21 ml. It means that, based on these phantom measurements, the difference between two arbitrary measurements on the same volume can be expected to be somewhere between +9 to -21 ml.

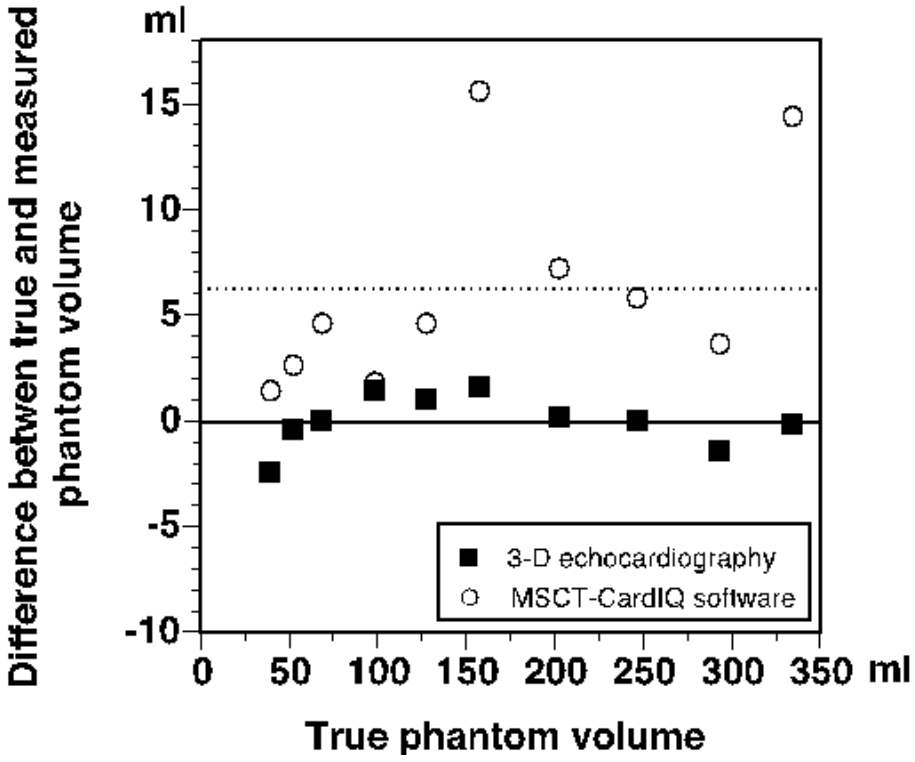

Figure 16: Differences between the phantom volumes and the measured volumes. The mean differences are indicated by the dotted (MSCT-CardIQ software) and solid (3-D echocardiography) lines. 3-D; three-dimensional, MSCT; multi slice computed tomography.

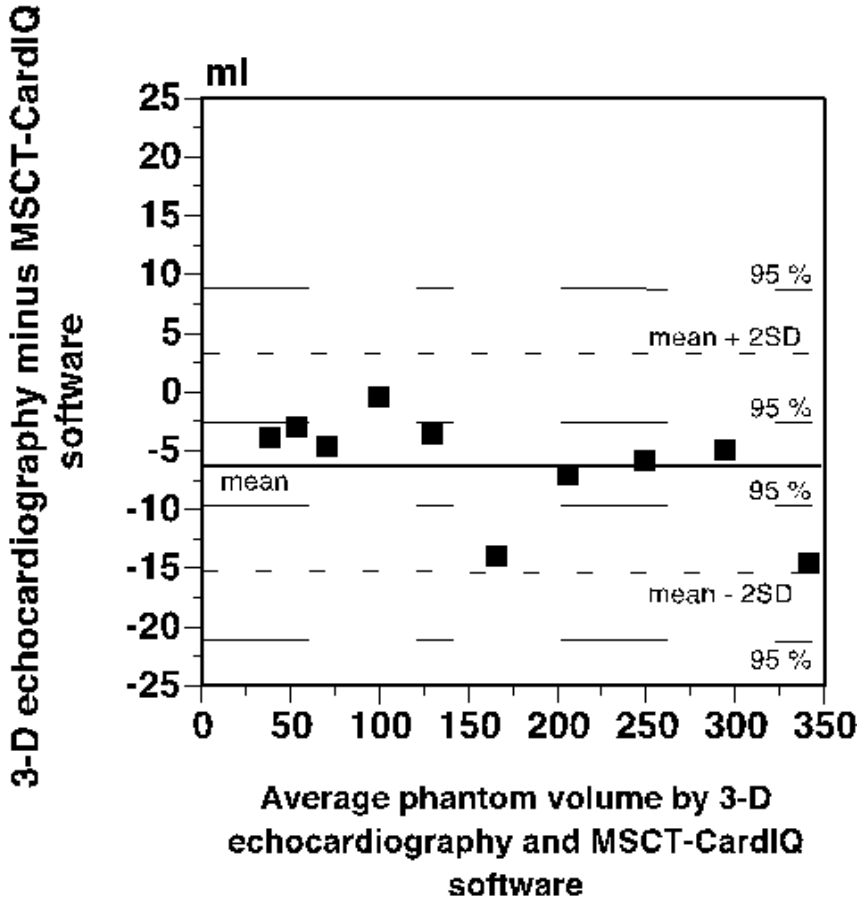

Figure 17: Bland-Altman plot of differences between phantom volumes measured by 3-D echocardiography and the MSCT-CardIQ software. The mean difference is indicated by the solid line. The limits of agreements (mean ± 2SD) are indicated by the dashed (narrow) lines. The 95% confidence intervals for the limits of agreement are indicated by the dashed (wide) line. 3-D; three-dimensional, MSCT; multi slice computed tomography, SD; standard deviation.

In vivo measurements

The automatic contour detection algorithm in the MSCT-CardIQ software was less accurate in the delineation of the LV in patients than in the phantom experiments and required in many cases manual correction. Manual correction was necessary in some of the end-diastolic images, but in all end-systolic images because of the sharper irregularities of the myocardial borders in those images. The tested methods correlated moderately. The regression analysis resulted in R-squared values of about 0.6 for both end-systolic and end-diastolic volumes.

Just as in the case of the phantom experiments, the LV volume measurements differed significantly also *in vivo*. When compared to the phantom results, evidenced by the Bland-Altman analysis presented in Figure 18, the agreement between the measured volumes with simplified 3-D echocardiography and the MSCT-CardIQ software in patients was undeniably poorer. As shown in Figure 18, the mean difference between the two methods

was -23 ± 40 ml (-10 ± 42 ml for LV end-systolic and -35 ± 37 ml for LV end-diastolic volume measurements). The limits of agreement (mean ± 2SD) were 57 and -103 ml, respectively. The 95% confidence interval for the upper limit of agreement was ranging from 23 to 93 ml and for the lower limit the corresponding interval was ranging from -138 to -69 ml. Thus, according to the results of the present study, the difference between simplified 3-D echocardiography and the MSCT-CardIQ software may vary between -138 and 93 ml when LV volumes are measured.

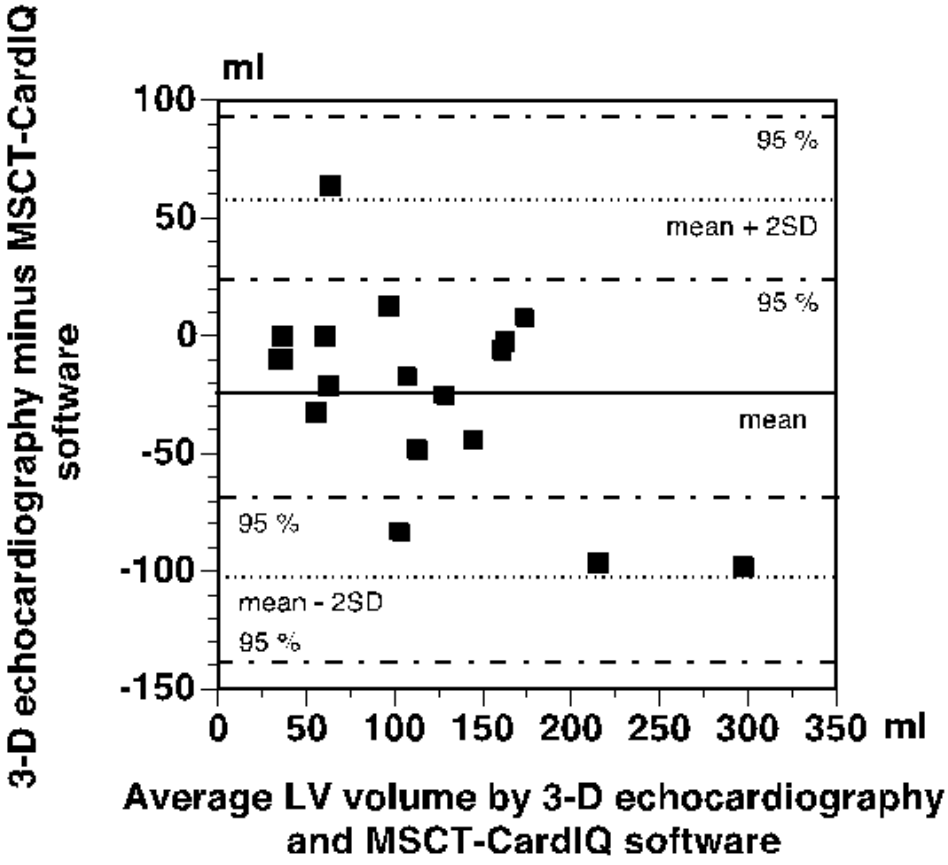

Figure 18: Bland-Altman plot of differences between LV volumes measured by simplified 3-D echocardiography and the MSCT-CardIQ software. The mean difference is indicated by the solid line. The limits of agreements (mean ± 2SD) are indicated by the dotted lines. The 95% confidence intervals for the limits of agreement are indicated by dash-dotted lines. LV; left ventricular, 3-D; three-dimensional, MSCT; multi slice computed tomography, SD; standard deviation.

Study II: High incidence of defective ultrasound transducers in use in routine clinical practice

In this study, the function of transducers in use in routine clinical practice was evaluated by the Sonora FirstCall test system and thereby the incidence of defective transducers was estimated.

Transducer errors

The main result from this study is presented in Table 2, where the tested transducers have been divided in to 2 groups: having normal function or being defective. Of the 676 tested transducers 269 (39.8%) were defective according to the criteria presented in the methodology chapter (Chapter 5.5).

Table 2: Ultrasound transducer classification in absolute numbers, percentage, and 95% confidence interval.

Transducer classification	Number	Frequency (%)	95% confidence Interval (%)
Normal function	407	60.2	56.5 – 63.9
Defective	269	39.8	36.1 – 43.5
Total	676	100	-

The most common occurring transducer errors were delamination and break in the cable (Table 3). The delaminated transducers (26.5%) constituted 66.5% of all transducer errors found, and together with the second and third most common errors, the electrical errors break in the cable (8.4%) and short circuit (3.4%), delamination constituted 96.3% of the transducer errors. Transducer errors related to the piezoelectric elements were thus uncommon. There were only six transducers with weak elements and four transducers with dead elements among the 269 defective transducers.

Table 3: Ultrasound transducer errors in absolute numbers, percentage, and 95% confidence interval.

Transducer error	Number	Frequency (%)	95% confidence interval (%)
Delamintaion	179	26.5	23.5 – 29.8
Break in the cable	57	8.4	6.3 – 10.5
Short circuit	23	3.4	2.0 – 4.8
Weak elements	6	0.9	0.2 – 1.6
Dead elements	4	0.6	0 – 1.2
Total	269	39.8	-

Comparison of the manufacturers

The error distribution among the seven manufacturers varied from 22.2 to 67.7%, with a mean value of 42.8%. The errors delamination and short circuit occurred in transducers from every manufacturer. However, the prevalence of break in the cable, weak elements, and dead elements were over-represented in transducers from certain manufacturers, and there were significant differences between the manufacturers regarding these errors when the variations were analyzed by a chi-squared test. Delamination, which was common in transducers from all manufacturers, and short circuit, that was found to a lesser extent in

transducers from four of the seven manufacturers, did not show a significant difference between the manufacturers.

Comparison of transducer types

The highest prevalence of errors was found in radiology transducers, linear transducers with frequencies between 8-10 MHz, curved linear transducers under 6 MHZ, and curved linear transducers between 6-8 MHz. In these groups, about half of the transducers were defective. The lowest prevalence of transducer errors was found in the groups of linear transducers above 10 MHz, paediatric cardiac transducers, and transesophageal transducers, where about one-fifth to one-fourth were defective.

Additional results

The observations presented below were not included in the original article. These results are more of observational nature and based on a few cases, but never the less of important value.

Observation 1: Quickly deteriorated transducer

The transducer function can deteriorate quickly when the transducer is delaminated. Figure 19 shows the same transducer tested with a three month interval.

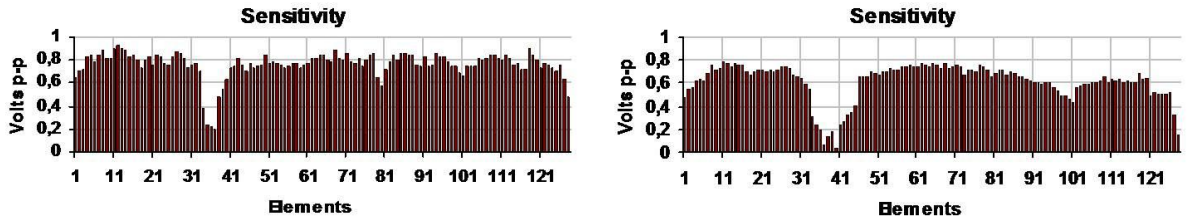

Figure 19: Quickly deteriorated transducer. Left panel: The sensitivity histogram of a tested phased array transducer with 128 elements. Right panel: Same ultrasound transducer tested three month later. The peak-to-peak (p-p) amplitude (sensitivity), measures the ability of a transducer to emit and receive ultrasound pulses.

Observation 2: Defective transducer not realized by the user

It can be very difficult for the user of a defective transducer to realize that the transducer actually is defective. The transducer in Figure 20 was used daily at the Department of Radiology at the Karolinska University Hospital without any suspicion from the user that it would be defective. However, the transducer had a very poor function. The left side was affected by delamination and the right side was affected by a combination of both delamination and breaks in signal wires.

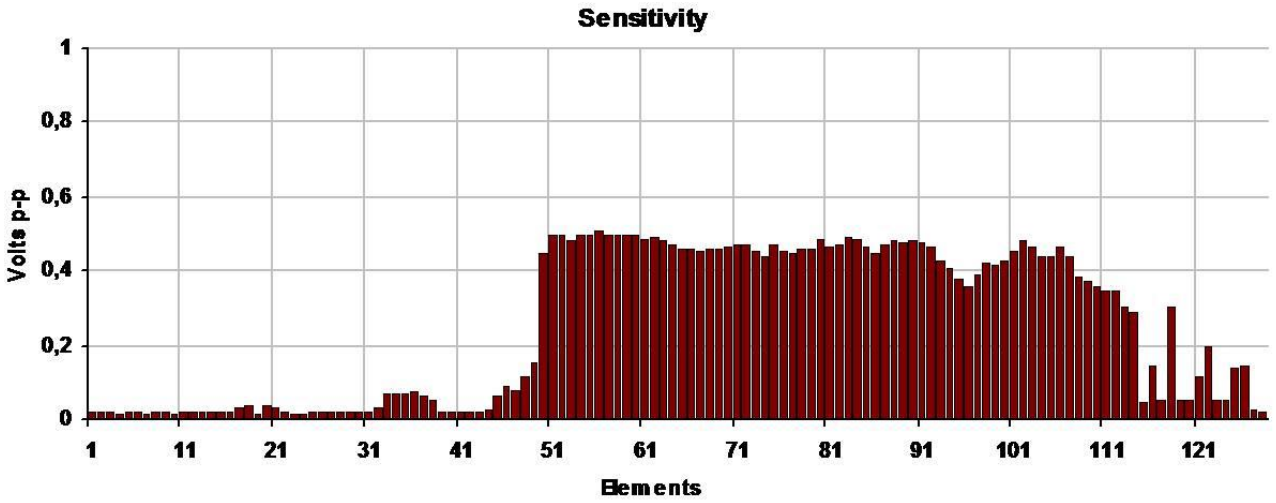

Figure 20: Defective transducer not realized by the user. Defective transducer found at the department of radiology at Karolinska University Hospital. The peak-to-peak (p-p) amplitude (sensitivity), measures the ability of a transducer to emit and receive ultrasound pulses.

Observation 3: New defective transducer

The third observation shows that brand new transducers can be defective. The transducer in Figure 21 was tested when delivered to the Karolinska University Hospital. It had two delaminated areas on the left side and two broken signal wires on the right side. The general lower sensitivity on the right side is not a defect in the transducer, but a design feature for increased focusing properties.

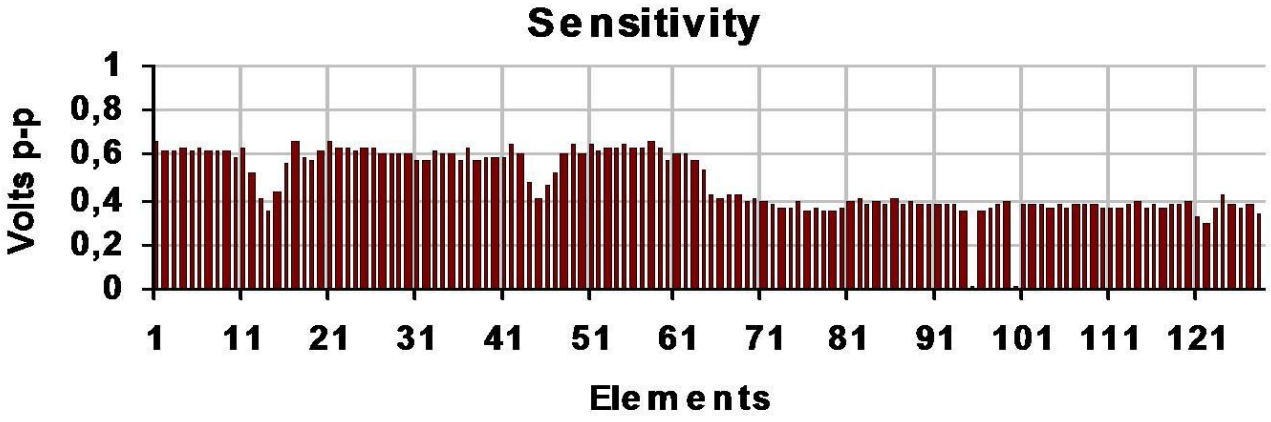

Figure 21: New defective transducer. Brand new transducer with two delaminated areas around elements 11 to 17 and 42 to 47, and two broken signal wires affecting element 95 and 100. The peak-to-peak (p-p) amplitude (sensitivity), measures the ability of a transducer to emit and receive ultrasound pulses. This transducer is designed with elements having two sensitivity levels for increased focusing properties.

6.2 Study III: Ultrasound transducer function: annual testing is not sufficient

In this study, it was investigated if annual transducer testing is sufficient to reduce the high number of defective transducers described in Study II.

Transducer errors

There were 81 defective transducers among the 299 tested; giving an error rate of 27.1% with a 95% confidence interval ranging from 22.1 to 32.1%. The error rate is significantly lower than in the earlier *High Incidence of Defective Ultrasound Transducers in the Clinical Routine study* where the 95% confidence interval for the error rate was ranging from 36.1 to 43.5%. The difference in error rate between the five error types was significant. The most common transducer errors were delamination (40), break in the cable (27), and weak elements (12). There was only one transducer with short circuit and dead elements, respectively.

Comparison of the manufacturers

There was no significant difference in error rates between the three manufacturers. There were 67 (29.5%) defective ones among the 227 tested Siemens transducers. The error distribution of the Siemens transducers was as follows: 31 delaminated, 23 with broken signal wires, 12 with weak elements, and 1 with a short circuit. The 95% confidence interval of the error rate for the Siemens transducers ranged from 23.6 to 35.4%. There were 11 (18.6%) defective ones among the 59 tested GE transducers. The error distribution of the GE transducers was as follows: 8 delaminated, 2 with broken signal wires, and 1 with dead elements. The 95% confidence interval of the error rate for the GE transducers ranged from 8.7 to 28.6%. There were 3 (23.1%) defective ones among the 13 tested Philips transducers. The error distribution of the Philips transducers was as follows: 1 delaminated and 2 with broken signal wires. The 95% confidence interval of the error rate for the Philips transducers ranged from 0.0 to 46.0%.

Comparison of the transducer models

There was no significant difference in the error rates between the 41 transducer models included in the study, and nor was the difference significant when calculated for the three manufacturers. The number of transducers of a specific model varied from 1 to 35 and the error rates from 0 to 100%.

Comparison of the clinics and clinic types

The number of transducers from each clinic ranged from 7 to 38. There were three radiological clinics having handling category 2, all the others having category 1. The error rates in the clinics were ranging from 0 to 51.5%, but the difference in error rates was not significant. The difference in error rates was however significant when the 13 clinics were divided into the three clinic types. The clinics of radiology showed the highest error rates with a mean error rate of 36.0%, whereas the mean error rate at the clinics of cardiology/physiology and obstetrics/gynaecology was 20.5 and 31.8%, respectively. The evidence for statistical difference was even stronger when calculated for the two handling categories. The clinics having handling category 2 had a mean error rate of 37.9% whereas category 1 had 21.4%. The 95% confidence interval for handling category 1 and 2 was 14.3 to 25.5% and 31.2 to 50.3%, respectively.

6.3 Study IV: Evaluation of tissue Doppler-based velocity and deformation imaging: a phantom study of ultrasound systems

In this study, the accuracy and diagnostic interchangeability of TDI-based displacement, velocity, strain and strain rate were assessed using six ultrasound systems.

Displacement

The displacement results are shown in Figure 22. When comparing the results with the error types in Chapter 4.7, two ultrasound systems (B and C) displayed results of type *a* (accurate, low spread), i.e. almost perfect results for both displacement measurements and timing errors. The measurements from A1 showed a constant overestimation with low spread and low timing errors. This is an error of type *b* (inaccurate, low spread) which easily could be compensated for. The other three ultrasound systems showed problematic tendencies to drift and have a high spread in the measurements for D1 and D2. Ultrasound system A1, B and C could thus be considered to be of interchangeable clinical use.

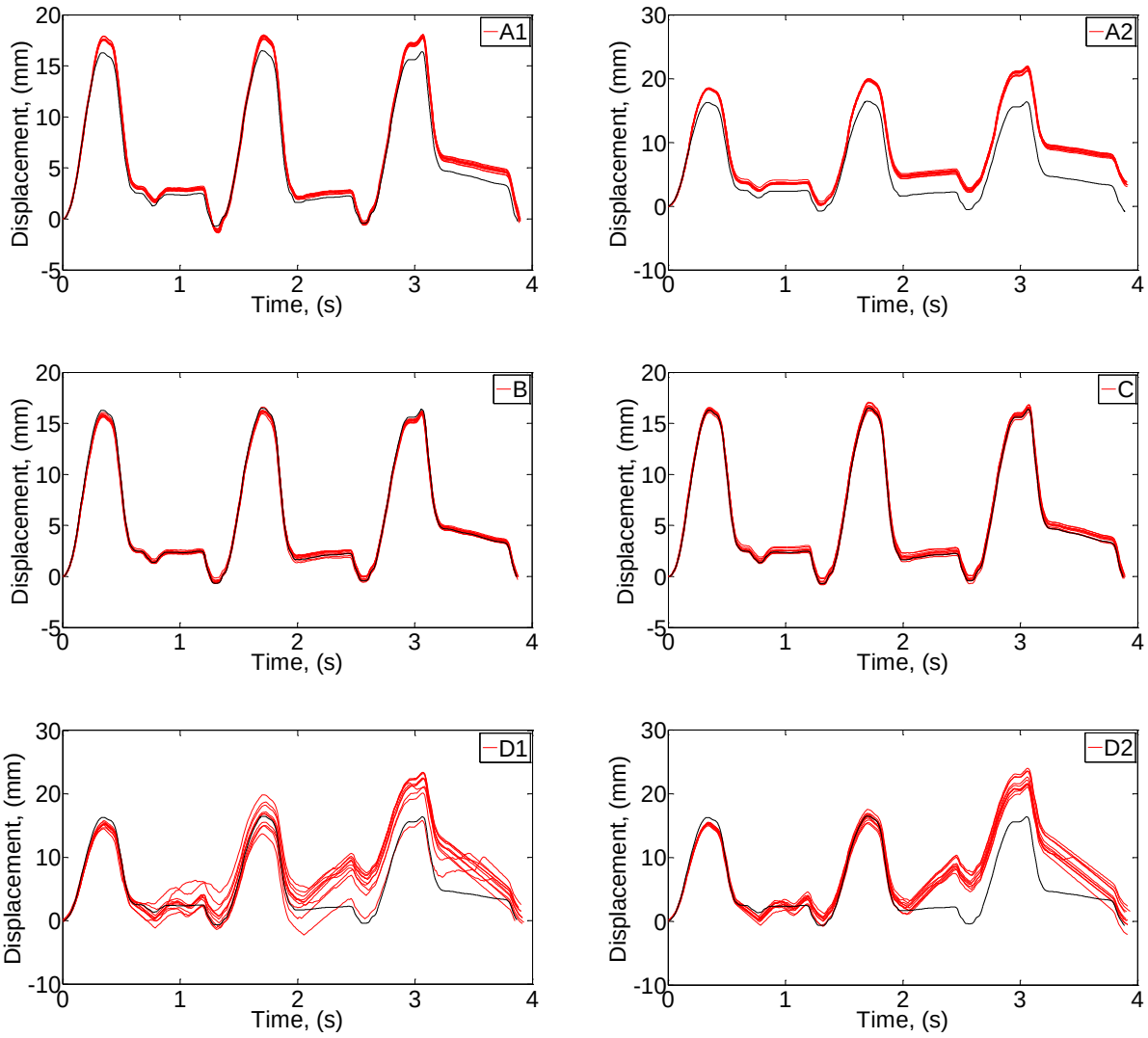

Figure 22: The displacement results of ten repeated measurements by the six ultrasound systems

Velocity

The velocity results are shown in Figure 23. As in the case of displacement, ultrasound system B and C produced reliable velocity measurements of type *a* (accurate, low spread). In contrast to the displacement measurements, ultrasound system A1 and D2 performed measurements that were of type *a* (accurate, low spread). Notably, the timing errors were extremely low for all these four ultrasound systems. All yielded mean timing errors and SDs of a few milliseconds. Ultrasound system A2 overestimated the peak values significantly; however with a repeatable error type *b* pattern (inaccurate, low spread). There were similar tendencies in the velocity curves from A1, i.e. there were incorrect overestimations of the descending part of the velocity curves after the peaks.

Ultrasound systems D1 and D2, which measurements were based on the same acquisition but post processed on different generations of software, showed different shapes of their velocity curves. D1 could not display the fast oscillations of the peaks, while D2 showed similar pattern as the phantom curve. Furthermore they showed both a larger spread of the whole curve than the other four ultrasound systems and evident problems tracing the low velocity parts between the heart beats. Ultrasound system A1, B and C could be considered to be of interchangeable clinical use regarding velocity measurements, most likely also together with ultrasound system D2 in applications with high velocities.

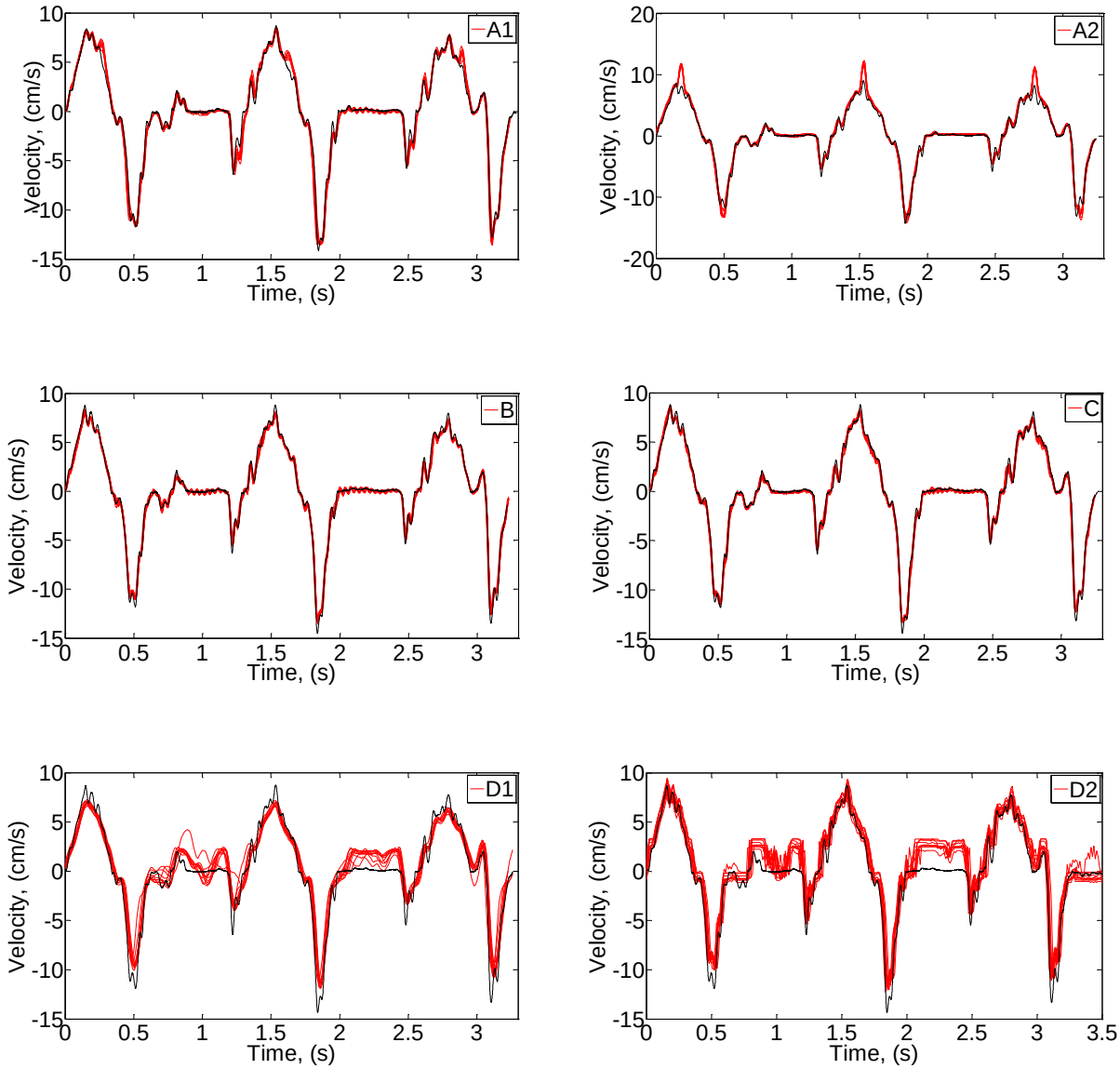

Figure 23: The velocity results of ten repeated measurements by the six ultrasound systems.

Strain

The strain results are shown in Figure 24. The measurements of strain turned out to be more difficult to perform than displacement and velocity for the ultrasound systems. There was in general much larger spread in these measurements, thus there were no results of error type *a* (accurate, low spread). Notably for the strain result is that all ultrasound systems showed better concordance with the natural strain than the Lagrangian strain.

Ultrasound system A1, which showed very low spread in the displacement and velocity measurements, showed for the strain measurements errors of type *d* (inaccurate, high spread) with a considerable spread sometimes resulting in large underestimations. Ultrasound system D1 and D2 were also showing spread and underestimation in their measurements. But there was not just an underestimation of the strain values. Ultrasound systems D1 and D2, especially D1, could not correctly differentiate positive strain from negative strain. The timing errors were generally larger than for displacement and velocity measurements. The degree of interchangeability was generally lower in the strain measurements than in the displacement and velocity measurements since the spread and timing errors were larger.

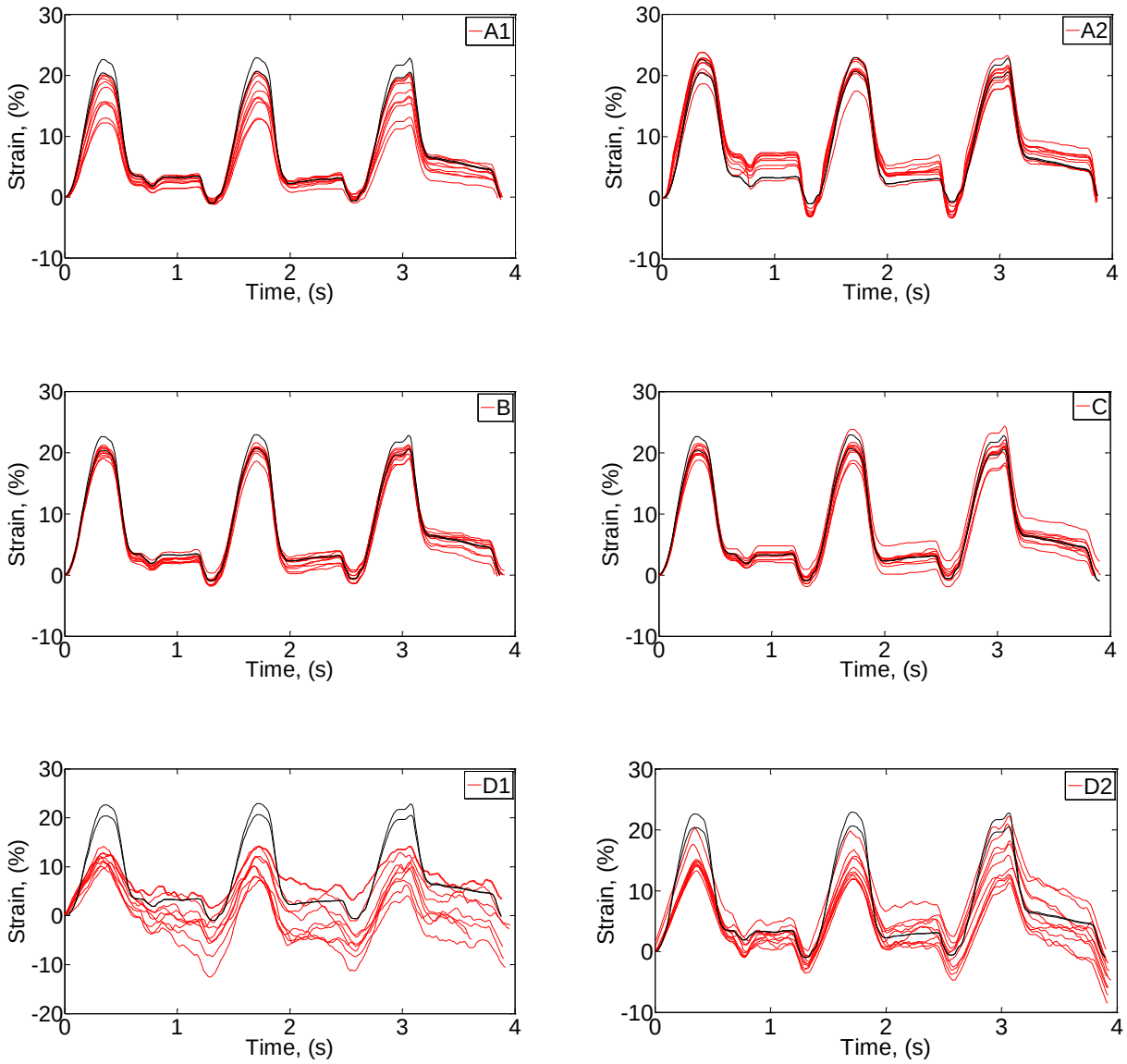

Figure 24: The strain results of ten repeated measurements by the six ultrasound systems.

Strain rate

The strain rate results are show in Figure 25. The results of strain rate measurements were invariably poorer than all other three parameters, ultrasound system A1, B and C displayed the least spread in the measurements. But all three underestimated the highest values and lacked the oscillating pattern of peaks in the strain rate curves. The strain rate curves produced by ultrasound system A2 appeared to contain strong artifactual peak values. That problem is not seen in the measurements from A1 which were post-processed on the same workstation. The strain rate curves from the D-systems were extremely noisy and showed considerable spread and underestimation. Since all ultrasound systems showed either error type *c* (accurate, high spread) or *d* (inaccuracy, high spread) errors in the strain rate measurements or for the timing errors, there were no ultrasound systems showing signs of clinical interchangeability.

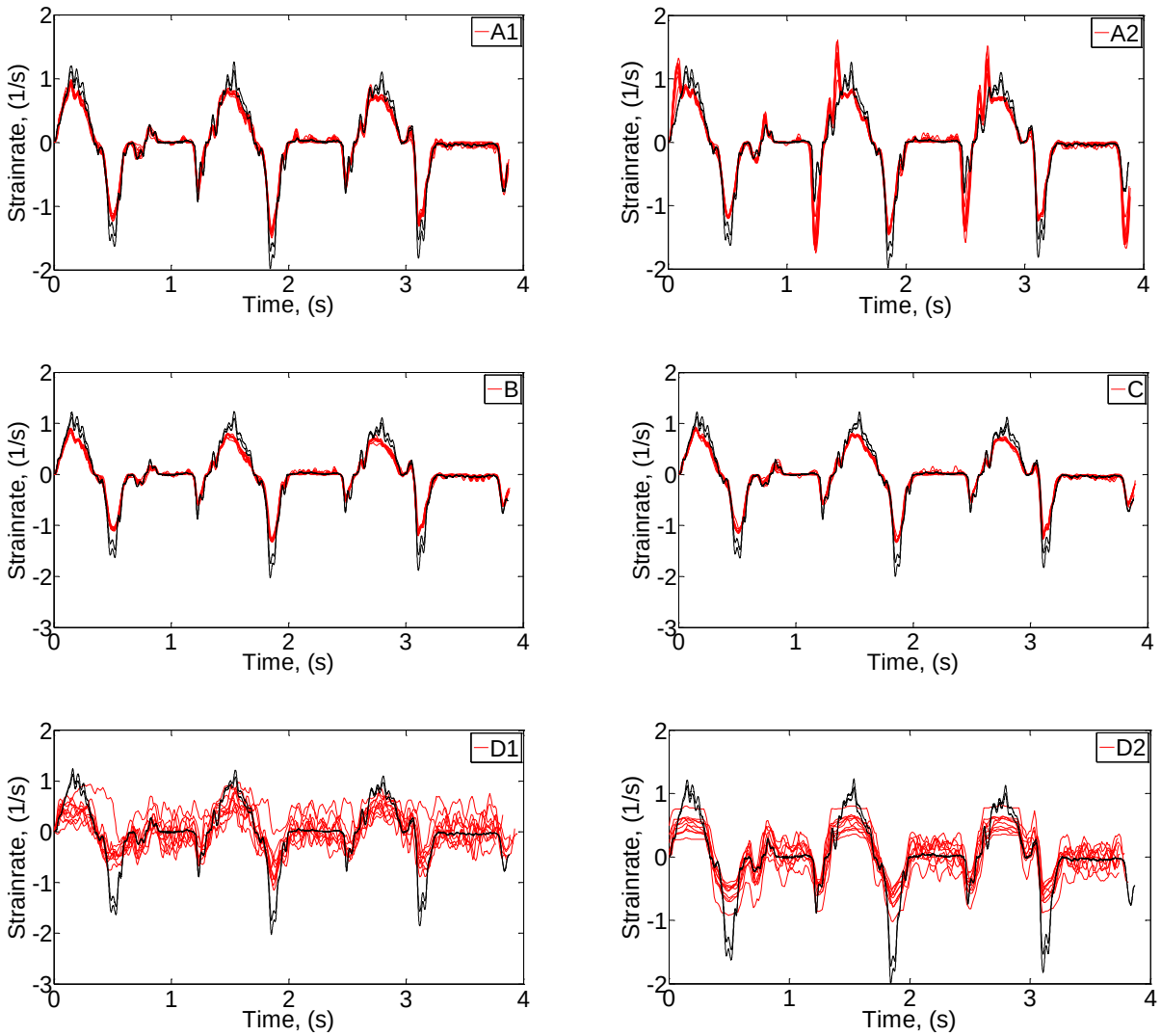

Figure 25: The strain rate results of ten repeated measurements by the six ultrasound systems.

7 DISCUSSION

The four studies in the thesis aimed to evaluate certain medical imaging devices, but the main focus was not on image quality. Study I and IV focused on accuracy when imaging systems are used for quantification, and Study II and III on investigating the function of ultrasound transducers. Results and conclusions about the specific focus areas are discussed in the articles. However, there are more profound insights to be gained from the studies, which are likely general for the whole industry of medical imaging systems.

7.1 Quantification by medical imaging systems

In Study I, the volume measurements by simplified 3-D echocardiography and 64-slice spiral CT were compared in two steps. In step one; the basic performance in volume measurements of the two methods was tested *in vitro* by comparison of the measured volumes with known phantom volumes. In step two, the measurements of the two methods were compared when measuring *in vivo* LV volumes.

The *in vitro* phantom volume measurements by simplified 3-D echocardiography yielded a result of type *a* (accurate, low spread), with fully acceptable agreement for use in the routine clinical practice. The mean difference was close to zero and the phantom comparison indicated a possible error in volume measurements by only a few millilitres.

The *in vitro* phantom measurements produced by the MSCT-CardIQ software were of error type *c* (accurate, high spread), and showed a significantly larger mean error with larger spread than in the measurements by simplified 3-D echocardiography. Despite the fact that there was a perfect correlation (r=1.00) between both methods and the phantom volumes, considerable discrepancies might occur between the phantom volumes and the volumes measured with the MSCT-CardIQ software (Figure 16), and consequently, also between the volumes measurements obtained by the two methods, thus limiting their interchangeable clinical use.

The limited interchangeable use of the two methods was even more prominent in the *in vivo* LV volume measurements where the variation between the methods increased drastically to a totally unacceptable level (Figure 18). This is a result somewhat in disagreement with earlier published data. Several studies where LV volume measurements have been evaluated have indicated a good agreement between MSCT-based and 2-D echocardiography-based methods[23, 28-31]. Good correlations have also been found between measurements by MSCT, 3-D echocardiography, and magnetic resonance imaging[29, 32-38]. Even though the methodological details may differ in these studies from the present study, similar good agreement with acceptable volume measurements could be expected to exist between MSCT and 3-D echocardiography.

On the other hand, there is at least one other study that is in line with the present results. The results presented by Sugeng et al.[38] showed significant overestimation of LV volumes by MSCT when compared to magnetic resonance imaging whereas the 3-D echocardiography showed more favourably results. Nevertheless, the majority of the studies in the literature would suggest better agreement between the simplified 3-D echocardiography and the MSCT-CardIQ software.

There are known methodological errors in the two methods used. These errors are likely greater in the *in vivo* situation than in the *in vitro* one. The LV volumes measured by simplified 3-D echocardiography are calculated by interpolating a 3-D LV volume from manual traces of the endocardial border in three different 2-D images. Consequently, any local changes in the LV geometry between the traced planes result in calculation errors. The same situation is on the other hand pertinent for the MSCT software. The delineation of the LV area contour by the edge detection algorithm may underestimate or overestimate the contribution of the contrast filled LV myocardial trabecular network. Furthermore, due to the relatively low temporal resolution of the MSCT method and the MSCT software procedure to calculate LV volumes by integrating LV volume data from short-axis slices the determination of volumes near the mitral annulus is uncertain.

However, these possible error sources can hardly explain the considerable discrepancy between the simplified 3-D echocardiography and the MSCT-CardIQ software in LV volume measurements. Since the MSCT software showed the worst performance already in the phantom experiments it appears reasonable to believe that the considerable discrepancies between the two tested methods were due to flaws or unknown error sources in the used MSCT algorithm. Be that as it may, the results in Study I clearly show that two different devices which originate from two different methodologies, which are considered to have been evaluated and be providers of reliable and clinically useful measurements, actually, can give very disparate results. The background to the problem is most probably multi-faceted, a part of the problem is that studies evaluating the performance of a medical device often conclude that a *method* (e.g. MSCT) is favourably evaluated, not *the specific device (model & manufacturer) that was actually used in the study*. Medical devices are most

often only tested by independent researchers when they are launched with a novel function. Therefore the user of later versions or unevaluated devices will most likely have an overreliance on their device.

Another aspect in the evaluation of the quantification performance of imaging systems is how the results are presented. Usually, the evaluation is based on a comparison with a reference method, usually the golden standard if possible. A rather common unsatisfactory way to present the results is the use of regression analysis and claim that the tested method is accurate and clinical valuable if the methods correlate well. However, correlation has nothing to do with accuracy. In the *in vitro* part of Study I, the methods correlated extremely well in the phantom experiments but the volumes measured by simplified 3-D echocardiography were more accurate (Figure 16). This was also seen in Study IV where the displacement, velocity, strain and strain rate curves of the tested parameters could correlate extremely well with the phantom values despite being very inaccurate. In order not to give misconception about the performance of the ultrasound scanners, the correlation analysis was omitted from the results in Study IV.

What also might contribute to the problem is how the requirements in the council directive concerning medical devices and the requirements from the Medical Product Agency are interpreted. It would result in more stringent interpretation of the requirements whether medical devices, such as ultrasound and CT-scanners, are considered as general medical devices or as devices with measuring functions. If they were considered as devices for measuring purposes, which probably would be the most proper interpretation when measurements actually are performed, the requirements would be more stringent. The Medical Product Agency require that "*products with measuring functions must be designed and manufactured in such a way that the measurements are accurate and within tolerances sufficiently for the intended purpose*" whilst the requirement for medical devices in general is that "*the products must have the performance claimed by the manufacturer*". Either way, the responsibility lies with the Notified Body. If the devices are considered as general medical devices they are checking the compliance with the wrong requirements, and if not, the control system used to evaluate the methods has been insufficient since the requirements are not met.

The problem with medical imaging systems giving unreliable measurements is further evidenced by Study IV, where the accuracy and diagnostic interchangeability of TDI-based displacement, velocity, strain and strain rate measurements were tested using an in-house phantom. The errors of the measurements in the phantom tests were of all types (*a* to *d*), and revealed that ultrasound systems cannot be assumed to measure correctly. Nor can it be assumed that measurements from different ultrasound systems are of clinical interchangeable use or that an ultrasound system yields the same value when the measurements are repeated. Furthermore, it cannot be assumed that displacement, velocity, strain and strain rate measurements are performed equally well by the same ultrasound

system; the performance of one parameter can be fully satisfactory while measurements of another parameter are so inaccurate that they are of no clinical value at all.

The default settings for a cardiac TDI acquisition and for the post processing on the workstations were used during the tests. As in Study I, there are known methodological sources of error. Important sources of error in TDI-based quantification are low frame rates and angle of incidence. These error sources are eradicated in the phantom setup since the frame rate is sufficiently high (92-168 Hz) and the phantom does not have any fast moving structures as valve leaflets, and the transducer position is fixed by a clamp at the phantom centre line. Other possible scanner factors that might influence the measurements are sector depth, sector width, filter setting and pulse repetition frequency. Workstation factors such as size, shape and placement of the region of interest, drift compensation and filter settings. However, all these factors can most likely not explain the found inaccuracy and certainly not the spread seen in the measurements from some of the tested ultrasound systems.

The results in Study IV are not what one could expect when reading the conclusion of earlier studies where TDI-based measurements have been evaluated. Several previous studies have concluded good agreement or good correlation with a phantom or a reference method[22, 39-41]. The reason for this discrepancy is difficult to state but it seems to be a combination of expectation of good results resulting in a focus on the best results and the statistical methods giving the result that are most in concordance with the expectations, as mentioned earlier. Discrepancies of the same magnitude as found in Study IV are actually present in some of these earlier studies. Kjaergaard et al. reports a mean bias of 32% in a phantom study with clinical relevant velocities and deformation values[40]. In another similar phantom study by Belohlavek et al. where strain rate was evaluated large spread and significant differences from the reference method were reported[39]. Moreover, in a phantom study by Matre et al. where lower myocardial velocities than normal were used, several measurements showed the same degree of spread in the strain results as seen by some of the ultrasound scanners in Study IV[41].

However, it is not only Study IV that indicates difficulties in TDI-based velocity and deformation measurements. There are other convincing evidences that these measurements are unreliable. As such, there is a large spread of measurements in studies trying to establish reference values[22, 42-43], the guidelines for echocardiography state that TDI-based strain and strain rate are research tools that should not be used clinically[44], and in a recent statement from an expert group, additional evaluation of TDI is suggested[45].

The regulations should make it impossible to implement a measuring method in a medical imaging system that turns out to produce inaccurate measurements. Since inaccurate measurements evidently occur in Study I and IV, something in the process has failed. There are three possible alternatives here: either the ultrasound devices are inaccurate from the beginning, or else they have aged or changed in some way that compromises their performance, or possibly a combination of the two. The results in Study IV could be a

combination of the alternatives. However, it is clear, evidenced by the D systems, where D2 performed better than D1 in the velocity measurements with the same software settings, that this is an error that has been present from day one. This is because the workstation is nothing but a software code which not could be expected to age or alter with time and the measurements are based on the same acquisition. The differences seen between ultrasound system A1 and A2 are however differences originating from the ultrasound scanner and not the workstation, which means that it in theory could be an error that have arisen over time. Still, it inevitably boils down to the testing and certification process by the Notified Bodies, which also include the lack of knowledge in general but also specifically to the given measurement and technological background in clinical practice.

7.2 Defective ultrasound transducers

The clinical case (Figure 1, missed patent ductus arteriosus), mentioned in the introduction of the thesis, where a congenital heart disease was misdiagnosed due to a defective transducer, was an eye-opener that raised the question for the first transducer evaluation study in this thesis (Study II). The importance of the clinical case is that it involves a patient with a congenital heart disease, which means that the pathology must have been present during the first examination. In most cases it can be difficult to know for certain whether if important findings have been missed.

Since the first examination of the clinical case was performed with an ultrasound scanner under maintenance protocol performed by the manufacturer and the clinic, it shows that the maintenance that had been carried out was insufficient in finding this kind of transducer generated problem. As part of the regular maintenance protocol, the performance of the scanners can be tested in many different ways, most often with some kind of resolution and/or flow phantom[5-7, 9-13, 15-21, 46]. That kind of performance testing had been performed on the ultrasound scanner used in the clinical case without finding the transducer defect. The situation in the clinical case where a defective transducer was missed in the regular maintenance routine turned out not to be a one-off incident, but a common occurrence in the clinics involved in the Study II. Even though maintenance measures and phantom testing were performed at the clinics in the study, about 40% of their transducers were marred with errors.

The clinical case, which was further evidenced by *Observation 2* (Figure 20, Defective transducer not realized by the user) in the result part, also suggests that the sonographer using the ultrasound scanner is without any chance to realize that a severe transducer error is present. The reason why such defective transducers as the one in *Observation 2* could be used without major complications is most likely that this was a curved linear transducer with a large contiguous segment working properly. The effect of this delamination would be a narrower sector and the sonographer has probably without thinking of it compensated

for the narrower sector by tilting the transducer. It is a very common misunderstanding that it is easy for an experienced sonographer to realize the presence of transducer errors. Again, the transducer type influences the situation. The difference in activation pattern makes phased array transducers more robust against a small number of dead elements since all elements are activated at the same time. This will also make it far more complicated to realize from the ultrasound image if a transducer is defective. A dead element in a linear transducer, even just one and in contrast to phased array transducers, will give rise to prominent black lines in the image. What happened in the clinical case (Figure 1), where a phased array transducer was used, was that a much smaller number of elements were activated during the blood flow Doppler measurement than during normal image acquisition. When a majority of the activated elements were dead or weak, no Doppler shift was registered and hence no colour flow information in the image.

That the functioning of the ultrasound transducer is essential for the overall performance of the ultrasound system has been shown previously[27]. The manufacturers themselves indicate it in some of their patents. GE writes in the United States Patent 6120449 that "*the effect of dead elements on the image can be significant, particularly in the near field of the image where a fewer number of elements are used to form the beam*", and Siemens writes in the United States Patent 5676149 that "*Given the number of piezoelectric crystals found in an ultrasound transducer, it is inevitable that one or more of the transducer elements will inevitably malfunction and become inoperative*", and ATL writes in the United States Patent 5517994 that "*But some problems, such as the failure of a single element of a multielement probe or a single channel of a multichannel system, are more subtle and not immediately observable by a user. Such undetected failures can lead to a degradation in diagnostic performance which is difficult to detect or remedy*"[47-49]. The statement from ATL is further evidenced by *Observation 2* in Study II (Figure 20, Defective transducer not realized by the user).

The patent statements from the manufacturers together with the findings in Study II and III show that transducer testing is a necessary part of the maintenance protocol in ultrasound labs. Transducer testing should be a maintenance step prior to the phantom testing, since there is no point in testing an ultrasound scanner with a defective transducer. The second step in transducer testing, after the testing of the transducers, is to decide whether the test result is acceptable or if the transducer should be considered as defective and be replaced. To be able to make such decision, the number of weak and/or dead elements that could be accepted must be established.

Today there is no consensus in how many weak and/or dead elements that could be accepted. It must be the work of future studies. Different acceptance criteria should probably be used for different transducer types, e.g. linear arrays compared to phased arrays. Due to differences in the activation pattern phased array transducers, where all elements are activated instead of a smaller number, are probably more robust against weak and dead elements than linear type transducers. The intended use of the transducer would also affect the decision. However unlikely, in applications were no measurements are

performed but only image acquisition, probably a larger number of weak and/or dead elements could be accepted.

However, the number of weak or dead elements in a transducer should not be the only focus point in the judgement of defective transducers. The underlying cause to the transducer errors is probably equally important. Delamination and break in the cable were the most common transducer errors in Study II and III, constituting about 88% and 83% of all transducer errors. These two errors might adversely affect an arbitrary number of piezoelectric elements. If this number is small, the whole transducer function is necessarily not adversely affected.

Compared to other errors, delamination and break in the cable are not stable errors and will most certainly increase with time. If the cable for instance has sustained damage and some of the signal wires are broken, it is unlikely that all other signal wires are undamaged; this probably results in a transducer that is more fragile than normal. Furthermore, as shown in *Observation 1* (Figure 19, Quickly deteriorated transducer) in the result chapter, delamination is an error also likely to grow with time and affect an increasing portion of the transducer.

The suggestion for these two errors is that the transducers can be used if not more elements than described in the methodology (Chapter 5.5) are affected, but the function must be monitored regularly. Judging from the result in the two transducer studies, where an unacceptable large number of the tested transducers were defective despite the introduction of annual testing to the quality procedures, there must be a serious underlying problem. The intriguing question here is whether the transducer errors are due to normal fatigue or quality problems associated to the manufacturing or handling processes.

The statistical tests in Study III showed that break in the cable was significantly more common in transducers from some manufacturers and that linear transducers over 10 MHz had the lowest prevalence of errors, thus indicating differences in quality among different transducer types and manufacturers. Together with *Observation 3* (Figure 21, New defective transducer) which showed that brand new transducers might be defective reinforces the suspicion about problems outside the walls of the clinics. The difficulty in understanding where those problems might be is that the transducers' earlier workload and any mishaps with the transducers are always unknown. This makes it almost impossible to state if certain transducer types or transducers from a certain manufacturer are more durable. The fact that the most common transducer errors are delamination and break in the cable indicates that these transducer parts in general are of too poor quality. However, it could equally well be interpreted as if these transducer parts are the most likely to be damaged by careless handling. The proper interpretation of the results is probably just to conclude that the problem with defective transducers is common and therefore that regular transducer testing is equally necessary, and that the handling and storage routines of the transducers at the clinics must be designed to minimize the risk of human error mishaps.

The handling and storage system of the transducers proved in Study III to be vital for the transducer function. In handling category 2, where the transducers were frequently connected and disconnected to the ultrasound scanners and transported a lot showed significantly higher error rates. These handling systems were only observed in the radiology clinics, where a mean error rate was about 77% higher than in the other clinics. Interestingly, one radiology department was not using this system and had the second lowest error rate, which indicates that the high error rates not only depend on inadequate transducer testing, but also on of how the transducers are handled.

The type of transducer test system used in Study II and III is a convenient measure to counteract the problem with defective transducers. The test is easy to learn and can be performed within minutes. However, the problem with defective transducers is not just a technical issue about assessing transducer function, but also largely an economical issue and a patient safety one. The clinic will only replace a transducer if it is absolutely necessary and if the patient safety is likely to be compromised. The transducers are expensive and likewise are the transducer test systems. The test systems are far more expensive than other performance testing devices such as grey scale and flow phantoms. When the present study was performed the price for the test system was about €30 000.

A possible remedy for minor transducer errors affecting only a small number of the piezoelectric elements is mentioned in the United States Patent 5676149, where a compensation method for inoperative elements in an ultrasound transducer is described[48]. The compensation method is also mentioned in the United States Patent 6120449, where the method is described as a yield increasing measure for the transducer production: "*While this technique is not the optimal solution from the image standpoint, it enables the loosening of manufacturing tolerances, thereby providing more cost-efficient manufacturing*"[47]. The compensation method shorts the circuit of a weak or dead element to a fully functional adjacent element and increases the drive signals and the gain of the received signals of the adjacent element. The data regarding which elements that are weak or dead are stored in a flash memory in the transducer, enabling the access of the data to the ultrasound scanner the transducer is connected to. It would thus be possible, at least theoretically, for the manufacturers to create a feature where the user can feed the flash memory in the transducer with new information every time the transducer is tested. Or even better, have a transducer tester built into the ultrasound scanner so that the test result is automatically transferred to the flash memory in the transducer. That would simplify the test procedure and the compensation information would follow the transducer if moved to another ultrasound scanner. It would prolong the life span of the transducer and the health care provider would thereby benefit economically.

The origin of the problem with defective transducers differs from the problem with inaccurate measurements from imaging systems. A transducer is more of a consumable item than a CT-scanner or the ultrasound scanner itself. It is known that a transducer is

fragile and must inevitably someday be replaced. Here, there is a large responsibility on the clinics using the transducers that they evaluate the function of their devices. Nevertheless, significant issues could be raised about the certification process of the quality assurance protocol for the production of the transducers. Since the number of defective transducers is large, it is unlikely to believe that the transducer errors seen in Study II and III not at all are due to quality problems in the manufacturing process. *Observation 3* (Figure 21, New defective transducer) in Study II is definitely due to a bad quality assurance protocol. That kind of problem would not be possible if the general requirements from the council directive for medical devices were fully met.

7.3 What to do?

The problem with defective transducers should firstly be addressed by the clinics by introducing transducer testing and routines that minimize the risk of damaging the transducers. Secondly, the Notified Bodies must secure that the transducers are fully functional and of sufficient quality for its intended use, which first and foremost means that the transducers should be manufactured with quality assurance protocols preventing that defective transducers leave the plant, and secondly that fragile transducer parts, such as the matching layers and the cable, withstand the expected lifetime claimed by the manufacturer.

In order to remedy the problem with inaccurate imaging system-based measurements, and the problem with defective transducers, the gravity of the situation must be appreciated by several actors. It is important that authorities like the Medical Product Agency realize the problem and clearly states that *every* implemented measuring function must provide robust and accurate measurements. Furthermore, the Notified Bodies must take their responsibility and check that the products which are placed on the market fulfil basal requirements for their intended use. However, the studies in this thesis show that this is not always the case. It is important that the Notified Bodies use testing methods, which evaluate all functions. Furthermore, it would be better if the Notified Bodies continuously evaluated the function of the produced medical devices in order to keep the certification valid.

The gravity of the situation with defective medical devices must also be appreciated by the Swedish National Board of Health and Welfare (Socialstyrelsen). One of their responsibilities is to monitor the usage of medical devices in clinical practice and also to remove, or prohibit usage, of malfunctioning medical devices when necessary. This is a responsibility where the Swedish National Board of Health and Welfare often fails and therefore seriously compromises the patient safety. The clinical case described in the introduction proves that it is possible to misdiagnose due to defect ultrasound devices. In Sweden, there are more than 20 cases reported to the Swedish National Board of Health

and Welfare, where pregnant women were incorrectly told after an ultrasound examination in early pregnancy that their living foetuses were dead. In several of these cases, the practitioners have been penalized by the Swedish National Board of Health and Welfare. It should be noted that, in these cases, the function of the devices has not been scrutinized in the investigation. If so, the results have not been included in the published reports about the cases. It would be wise, as soon as possible after an incident have been noticed, to quarantine the medical device used in the examination until its function has been examined and evaluated. It would not just make it possible to learn from technical problems but also free the practitioner from suspicion of causing the incident if a defective medical device is the underlying reason for the incorrect decision.

When purchasing medical devices, such as ultrasound systems or CT-scanners, the clinicians as a group have the power to demand better evaluated functions in the medical devices. However, Study I and IV, indicate that the clinicians in general lack the knowledge or courage to demand evidence that the purchased medical device is fully suitable for its intended use. The clinicians should demand that all functions of a medical device are methodologically evaluated and tested as part of the incoming inspection, with a known reliable test method. If such a test is not performed with successful results, the tested medical device should simply not be purchased. Furthermore, the test should be performed with standardized scientific methods and phantoms that are approved by authorities, Notified Bodies, and relevant scientific groups.

8 CONCLUSIONS

The studies in this thesis have revealed an immense need of more extensive evaluation of medical devices, both new devices and devices that are in clinical practice. The present fulfilment of regulations does not seem to be adequate in order to keep inferior medical devices off the market.

Study I: Assessment of left ventricular volumes using simplified 3-D echocardiography and computed tomography - a phantom and clinical study

Simplified 3-D echocardiography provided reliable and significantly more accurate assessment of *in vitro* phantom volumes than the MSCT-CardIQ software. The discrepancy between the results of both methods increased considerably when *in vivo* LV volumes were measured and the limits of agreement were not acceptable for interchangeable diagnostic use of the two methods.

Study II: High incidence of defective ultrasound transducers in use in routine clinical practice

Defective transducers are common in the routine clinical practice and there is from a patient safety perspective an urgent need for increased transducer testing in the clinical departments.

Study III: Ultrasound transducer function: annual testing is not sufficient

Annual testing is not sufficient to reduce the incidence of defective ultrasound transducers in routine clinical practice. Still, more than one-fourth of the transducers were defective. Furthermore, it is important to have routines that minimize the handling of the transducers.

Study IV: Evaluation of tissue Doppler-based velocity and deformation imaging: a phantom study of ultrasound systems

Ultrasound systems cannot be assumed to measure tissue Doppler-based displacement, velocity, strain, and strain rate with enough accuracy. The performance of tissue Doppler measurements by ultrasound systems must be further evaluated before it is used clinically.

9 FUTURE WORK

Different problems regarding quality assurance of medical devices have been highlighted during this thesis work. Future work is needed in order to investigate the reasons to these problems and to improve the evaluation methods and procedures. Work is needed both from authorities, clinicians, and independent researchers. Some of the most urgent needs are listed below:

- Investigate the reasons why there are medical devices with unreliable measuring functions on the market.
- Create standardized scientifically sound and proven methods for evaluation of medical devices.
- Demand all Notified Bodies to use the same evaluation methods.
- Investigate if different ultrasound-based applications are affected differently by transducer errors.
- Investigate thoroughly how different transducer designs are affected by different transducer errors.
- Every clinic should investigate the necessary maintenance routines for their ultrasound systems more exactly.
- Evaluate the accuracy of speckle tracking-based velocity and deformation measurements.

10 OTHER SCIENTIFIC CONTRIBUTIONS

Articles

No evidence of cardiac fatigue in tissue velocity curves at rest after 6 days of ultra-undurance exercise. Mattsson CM, Lind B, Enqvist JK, Mårtensson M, Ekblom B, Brodin L-Å. *Submitted manuscript.*

Conference abstracts

Reproducibility of echocardiographic strain estimates in a high fidelity phantom. Mattias Mårtensson, Anna Bjällmark, Lars-Åke Brodin, *Myocardial Velocity and Deformation Imaging,* Leuven, Belgium, 2011.

Ultrasound transducer function: annual testing is not sufficient. Mattias Mårtensson, Mats Olsson, Lars-Åke Brodin. *EuroEcho*, Copenhagen, Denmark, 2010.

Ultrasound transducer function: annual testing is not sufficient. Mattias Mårtensson, Mats Olsson, Lars-Åke Brodin. *Medicinteknikdagarna*, Umeå, Sweden, 2010.

No evidence of cardiac fatigue in tissue velocity curves at rest after 6 days of ultra-endurance exercise. C. Mikael Mattsson, Britta Lind, Jonas K. Enqvist, Mattias Mårtensson, Björn Ekblom and Lars-Åke Brodin. *ESC 2010*, Stockolm, Sweden, 2010.

High Incidence of Defective Ultrasound Transducers in the Clinical Routine. Mattias Mårtensson, Mats Olsson, Björn Segall, Reidar Winter, Lars-Åke Brodin, *Medicinteknikdagarna*, Örebro, Sweden, 2007.

Quality Control of Ultrasound Transducers in the clinical routine. Mattias Mårtensson, Mats Olsson, Björn Segall, Reidar Winter, Lars-Åke Brodin, *EuroEcho*, Prague, Czech republic, 2006.

Comparison of 64 channel multi slice CT and simplified three dimensional echocardiography for volume measurements for patients and cardiac phantoms. Jonaz

Ripsweden, Habib Mir-Akbari, Kerstin Cederlund, Mattias Mårtensson, Lars-Åke Brodin, Reidar Winter. *Kardiovaskulära vårmötet*, Linköping, Sweden, 2006.

11 REFERENCES

1. Kasai C, Namekawa K, Koyano A, Omoto R. Real-Time Two-Dimensional Blood Flow Imaging Using an Autocorrelation Technique. IEEE Trans Sonic Ultrasonics. 1985;SU-32(3).
2. McDicken WN, Sutherland GR, Moran CM, Gordon LN. Colour doppler velocity imaging of the myocardium. Ultrasound in Medicine & Biology. 1992;18(6-7):651-4.
3. AIUM. The AIUM 100 mm test object and recommended procedures for its use. AIUM. 1974.
4. AIUM. Technical Standard Committee. Methods for Measuring Performance of Pulse-Echo Ultrasound Imaging Equipment. AIUM. 1990.
5. AIUM. Technical Standard Committee. Quality Assurance Manual for Gray-Scale Ultrasound Scanners, Stage 2. AIUM. 1995.
6. AIUM. Technical Standard Committee. Methods for Measuring Performance of Pulse-Echo Ultrasound Imaging Equipment, Part 2: Digital Methods stage 1. AIUM. 1995.
7. Brendel K, Filipczynski LS, Gerstner R, Hill CR, Kossoff G, Quentin G, Reid M, Saneyoshi J, Somer JC, Tchevnenko AA, Wells PN. Methods of measuring the performance of ultrasonic pulse-echo diagnostic equipment. Ultrasound in Medicine & Biology. 1977;2(4):343-50.
8. Carson PL. Rapid evaluation of many pulse echo system characteristics by use of a triggered pulse burst generator with exponential decay. Journal of Clinical Ultrasound. 1976;4(4):259-63.
9. Dudley NJ, Griffith K, Houldsworth G, Holloway M, Dunn MA. A review of two alternative ultrasound quality assurance programmes. European Journal of Ultrasound. 2001;12(3):233-45.
10. Gibson NM, Dudley NJ, Griffith K. A computerised quality control testing system for B-mode ultrasound. Ultrasound in Medicine & Biology. 2001;27(12):1697-711.
11. Goodsitt MM, Carson PL, Witt S, Hykes DL, Kofler JM, Jr. Real-time B-mode ultrasound quality control test procedures. Report of AAPM Ultrasound Task Group No. 1. Journal of Medical Physics. 1998;25(8):1385-406.
12. Hall AJ, Fleming JEE. A method of calibrating contact B-scanners. Ultrasonics. 1978;16(6):277-81.
13. MacGillivray TJ, Ellis W, Pye SD. The resolution integral: visual and computational approaches to characterizing ultrasound images. Physics in Medicine and Biology. 2010;55(17):5067.
14. Madsen E. Quality assurance for gray-scale imaging. Ultrasound in Medicine & Biology. 2000;26(Supplement 1):48-50.

15. Price R. IPSM Report No. 71 Routine quality assurance of ultrasound imaging systems. The Institute of Physical Sciences in Medicin. 1995.
16. Pye SD, et al. Medical ultrasound: a new metric of performance for grey-scale imaging. Journal of Physics: Conference Series. 2004;1(1):187.
17. Rownd JJ, Madsen EL, Zagzebski JA, Frank GR, Dong F. Phantoms and automated system for testing the resolution of ultrasound scanners. Ultrasound in Medicine & Biology. 1997;23(2):245-60.
18. Satrapa J, Schultz HJ, Doblhoff G. Automated quality control of ultrasonic B-mode scanners by applying an TMM 3D cyst phantom. Ultraschall in der Medizin. 2006;27(03):262,72.
19. Thijssen JM, van Wijk MC, Cuypers MHM. Performance testing of medical echo/Doppler equipment. European Journal of Ultrasound. 2002;15(3):151-64.
20. Thijssen JM, Weijers G, de Korte CL. Objective performance testing and quality assurance of medical ultrasound equipment. Ultrasound in Medicine & Biology. 2007;33(3):460-71.
21. Tierlink C, Bezemer R, Kollman C, Lubbers J, Hoskins P, Fish P, Fredfeldt K, Schaarschmidt U. Development of an example flow test object and comparison of five of these test objects, constructed in various laboratories. Ultrasonics. 1999;36:653-60.
22. Edvardsen T, Gerber BL, Garot J, Bluemke DA, Lima JAC, Smiseth OA. Quantitative assessment of intrinsic regional myocardial deformation by Doppler strain rate echocardiography in humans - Validation against three-dimensional tagged magnetic resonance imaging. Circulation. 2002;106(1):50-6.
23. Salm LP, Schuijf JD, de Roos A, Lamb HJ, Vliegen HW, Jukema JW, Joemai R, van der Wall EE, Bax JJ. Global and regional left ventricular function assessment with 16-detector row CT: Comparison with echocardiography and cardiovascular magnetic resonance. European Journal of Echocardiography. 2006;7(4):308-14.
24. D'hooge J, Heimdal A, Jamal F, Kukulski T, Bijnens, Rademakers F, Hatle L, Suetens P, Sutherland GR. Regional Strain and Strain Rate Measurements by Cardiac Ultrasound: Principles, Implementation and Limitations. European Journal of Echocardiography. 2000;1(3):154-70.
25. Altman D, Bland J. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;32:307-17.
26. Läkemedelsverket. Läkemedelsverkets författningssamling, LVFS 2003:112003. Report No.: 1101-5225.
27. Weigang B, Moore GW, Gessert J, Phillips WH, Schafer M. The Methods and Effects of Transducer Degradation on Image Quality and the Clinical Efficacy of Diagnostic Sonography. Journal of Diagnostic Medical Sonography. 2002;19(1):3-13.
28. Dirksen MS, Jukema JW, Bax JJ, Lamb HJ, Boersma E, Tuinenburg JC, Geleijns J, van der Wall EE, de Roos A. Cardiac multidetector-row computed tomography in patients with unstable angina. The American Journal of Cardiology. 2005;95(4):457-61.
29. Yamamuro M, Tadamura E, Kubo S, Toyoda H, Nishina T, Ohba M, Hosokawa R, Kimura T, Tamaki N, Komeda M, Kita T, Konishi J. Cardiac Functional Analysis with Multi–Detector Row CT and Segmental Reconstruction Algorithm: Comparison with Echocardiography, SPECT, and MR Imaging1. Radiology. 2005;234(2):381-90.
30. Henneman M, Bax J, Schuijf J, Jukema J, Holman E, Stokkel M, Lamb H, de Roos A, van der Wall E. Global and regional left ventricular function: a comparison between gated SPECT, 2D echocardiography and multi-slice computed tomography. European Journal of Nuclear Medicine and Molecular Imaging. 2006;33(12):1452-60.
31. Schuijf JD, Bax JJ, Jukema JW, Lamb HJ, Salm LP, de Roos A, van der Wall EE. Assessment of left ventricular volumes and ejection fraction with 16-slice multi-slice computed

tomography; comparison with 2D-echocardiography. International Journal of Cardiology. 2007;116(2):201-5.
32. Buck T, Hunold P, Wentz KU, Tkalec W, Nesser HJ, Erbel R. Tomographic Three-dimensional Echocardiographic Determination of Chamber Size and Systolic Function in Patients With Left Ventricular Aneurysm : Comparison to Magnetic Resonance Imaging, Cineventriculography, and Two-dimensional Echocardiography. Circulation. 1997;96(12):4286-97.
33. Caiani EG, Corsi C, Zamorano J, Sugeng L, MacEneaney P, Weinert L, Battani R, Gutierrez JL, Koch R, Perez de Isla L, Mor-Avi V, Lang RM. Improved Semiautomated Quantification of Left Ventricular Volumes and Ejection Fraction Using 3-Dimensional Echocardiography with a Full Matrix-array Transducer: Comparison with Magnetic Resonance Imaging. J Am Soc Echocardiogr. 2005;18(8):779-88.
34. Gutiérrez-Chico JL, Zamorano JL, Pérez de Isla L, Orejas M, Almería C, Rodrigo JL, Ferreirós J, Serra V, Macaya C. Comparison of left ventricular volumes and ejection fractions measured by three-dimensional echocardiography versus by two-dimensional echocardiography and cardiac magnetic resonance in patients with various cardiomyopathies. The American Journal of Cardiology. 2005;95(6):809-13.
35. Chan J, Jenkins C, Khafagi F, Du L, Marwick TH. What Is the Optimal Clinical Technique for Measurement of Left Ventricular Volume After Myocardial Infarction? A Comparative Study of 3-Dimensional Echocardiography, Single Photon Emission Computed Tomography, and Cardiac Magnetic Resonance Imaging. J Am Soc Echocardiogr. 2006;19(2):192-201.
36. Heuschmid M, Rothfuss J, Schroeder S, Fenchel M, Stauder N, Burgstahler C, Franow A, Kuzo R, Kuettner A, Miller S, Claussen C, Kopp A. Assessment of left ventricular myocardial function using 16-slice multidetector-row computed tomography: comparison with magnetic resonance imaging and echocardiography. European Radiology. 2006;16(3):551-9.
37. Raman SV, Shah M, McCarthy B, Garcia A, Ferketich AK. Multi-detector row cardiac computed tomography accurately quantifies right and left ventricular size and function compared with cardiac magnetic resonance. American Heart Journal. 2006;151(3):736-44.
38. Sugeng L, Mor-Avi V, Weinert L, Niel J, Ebner C, Steringer-Mascherbauer R, Schmidt F, Galuschky C, Schummers G, Lang RM, Nesser H-J. Quantitative Assessment of Left Ventricular Size and Function: Side-by-Side Comparison of Real-Time Three-Dimensional Echocardiography and Computed Tomography With Magnetic Resonance Reference. Circulation. 2006;114(7):654-61.
39. Belohlavek M, Bartleson VB, Zobitz ME. Real-Time Strain Rate Imaging: Validation of Peak Compression and Expansion Rates by a Tissue-Mimicking Phantom. Echocardiography. 2001;18(7):565-71.
40. Kjaergaard J, Korinek J, Belohlavek M, Oh JK, Sogaard P, Hassager C. Accuracy, Reproducibility, and Comparability of Doppler Tissue Imaging by Two High-end Ultrasound Systems. J Am Soc Echocardiogr. 2006;19(3):322-8.
41. Matre K, Ahmed AB, Gregersen H, Heimdal A, Hausken T, Ødegaard S, Gilja OH. In vitro evaluation of ultrasound Doppler strain rate imaging: modification for measurement in a slowly moving tissue phantom. Ultrasound in Medicine & Biology. 2003;29(12):1725-3
42. Andersen NH, Poulsen SH. Evaluation of the longitudinal contraction of the left ventricle in normal subjects by doppler tissue tracking and strain rate. J Am Soc Echocardiogr. 2003;16(7):716-23.
43. Sun JP, Popovic ZB, Greenberg NL, Xu X-F, Asher CR, Stewart WJ, Thomas JD. Noninvasive quantification of regional myocardial function using Doppler-derived velocity,

displacement, strain rate, and strain in healthy volunteers: effects of aging. J Am Soc Echocardiogr. 2004;17(2):132-8.
44. Rudski LG, Lai WW, Afilalo J, Hua L, Handschumacher MD, Chandrasekaran K, Solomon SD, Louie EK, Schiller NB. Guidelines for the Echocardiographic Assessment of the Right Heart in Adults: A Report from the American Society of Echocardiography: Endorsed by the European Association of Echocardiography, a registered branch of the European Society of Cardiology, and the Canadian Society of Echocardiography. J Am Soc Echocardiogr. 2010;23(7):685-713.
45. Mor-Avi V, Lang RM, Badano LP, Belohlavek M, Cardim NM, Derumeaux G, Galderisi M, Marwick T, Nagueh SF, Sengupta PP, Sicari R, Smiseth OA, Smulevitz B, Takeuchi M, Thomas JD, Vannan M, Voigt JU, Zamorano JL. Current and evolving echocardiographic techniques for the quantitative evaluation of cardiac mechanics: ASE/EAE consensus statement on methodology and indications endorsed by the Japanese society of echocardiography. European Journal of Echocardiography. 2011;12(3):167-205.
46. Walker A, Olsson E, Wranne B, Ringqvist I, Ask P. Accuracy of spectral Doppler flow and tissue velocity measurements in ultrasound systems. Ultrasound in Medicine & Biology. 2004;30(1):127-32.
47. Snyder J; Method and apparatus for compensating for inoperative elements in ultrasounic transducer array. USA patent 6,120,449. 1998 Sep. 19, 2000.
48. Lin-Xin Y; Method of compensating for inoperative elements in an ultrasound transducer. USA1996 Oct. 14, 1997.
49. Burke T, Mehi J; Self diagnostic ultrasonic imaging systems. USA patent 5,517,994. 1996 May 21, 1996.