For publication in *Biomedical Physics & Engineering Express*

Please note: this is the final draft of the accepted article:

Mechelli F, Arendt-Nielsen L, Stokes M, Agyapong-Badu S.

Inter-rater and intra-rater reliability of ultrasound imaging for measuring quadriceps muscle and non-contractile tissue thickness of the anterior.

***Biomed Phys Eng Express* 5 (2019) 037002**

Accepted: 15th March 2019

Please use the following link for the final, fully proofed and peer-reviewed journal article online: https://doi.org/10.1088/2057-1976/ab102f

Inter-rater and intra-rater reliability of ultrasound imaging for measuring quadriceps muscle and non-contractile tissue thickness of the anterior thigh

Mechelli Filippo, PT, MSc[1,2]
Arendt-Nielsen Lars, Dr. Med, PhD[1]
Stokes Maria, FCSP, PhD[3,4]
Agyapong-Badu Sandra, PT, PhD[5].

[1]Centre of Sensory Motor Interaction, Department of Health Science and Technology, School of Medicine, University of Aalborg, Aalborg, Denmark.
[2]PT, MSc, private practice, Urbino, Italy.
[3]School of Health Sciences, University of Southampton, Southampton, UK.
[4]Arthritis Research UK Centre for Sport, Exercise and Osteoarthritis, UK.
[5]School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham, Edgbaston, Birmingham, UK.

Address correspondence to:

Filippo Mechelli,
Centre of Sensory Motor Interaction, Department of Health Science and Technology, School of Medicine, University of Aalborg, Aalborg, Denmark.
PT, MSc, private practice, Via dei Morti, 4, 61029 Urbino, Italy.
E-mail: fmechelli@gmail.com
Tel: +39 0722 327188
Fax: +39 0722 540491

Keywords: muscle thickness, rectus femoris, rehabilitative ultrasound imaging, subcutaneous fat thickness, vastus intermedius

Title

Inter-rater and intra-rater reliability of ultrasound imaging for measuring quadriceps muscle and non-contractile tissue thickness of the anterior thigh

Abstract

Objective: To determine intra-rater and inter-rater reliability of ultrasound imaging for measuring muscle and non-contractile (subcutaneous fat and perimuscular fascia) tissue thickness of the anterior thigh.

Approach: Ultrasound imaging has been used for clinical research to assess the morphology and cross-sectional area of muscles and other musculoskeletal structures. Repeatability of measurements with the ultrasound imaging technique between operators and test re-test reliability are important and need to be established for specific muscles. Twenty-four healthy moderately active adults (aged 36-64 years), underwent B-mode ultrasound imaging by two investigators. The anterior thighs were scanned at a site two-thirds of the distance between the antero-superior iliac spine and the superior pole of the patella. Intraclass Correlation Coefficients (ICCs, model 3,1 for inter-rater and 3,2 for intra-rater reliability between-days) with 95% confidence intervals (CI) was used to assess reliability.

Main results: Inter-rater reliability of ultrasound imaging measurements were ICC3,1 of 0.98 (95% CI: 0.95–0.99) for muscle thickness, 0.81 (95% CI: 0.60–0.91) for subcutaneous fat, 0.78 (95% CI: 0.56–0.90) for non-contractile tissue (subcutaneous fat combined with perimuscular fascia), and 0.70 (95% CI: 0.42–0.86) for perimuscular fascia. Intra-rater reliability values were ICC3,2 0.96 (95% CI: 0.90–0.98) for muscle thickness, 0.99 (95% CI: 0.97–0.99) for subcutaneous fat, 0.98 (95% CI: 0.96–0.99) for non-contractile tissue, and -0.02 (95% CI: -0.41–0.38) for perimuscular fascia.

Significance: The present findings indicate very high inter-rater and intra-rater reliability of ultrasound imaging thickness measurements of the quadriceps muscles and non-contractile tissue of the anterior thigh, while perimuscular fascia measurements alone were not reliable between days.

Introduction

Ultrasound (US) imaging has been successfully applied in clinical practice and clinical research to evaluate the architecture, thickness, and cross-sectional area of muscles, and in studying morphology and/or pathology of other musculoskeletal structures, such as tendons and ligaments (Whittaker *et al* 2007). Quadriceps muscle weakness and atrophy are commonly reported in patients with knee osteoarthritis (Petterson *et al* 2008) or any painful condition that affect the knee (Henriksen *et al* 2011, Rice *et al* 2014), and in critically ill patients in intensive care units, in whom muscle weakness occurs rapidly (Hadda *et al* 2017). The US imaging technique is non-invasive, sufficiently accurate, relatively low cost, widely available and a safe modality to study musculoskeletal tissues, but the accuracy of the procedure is operator-dependent (Wakefield *et al* 2005). Accurate, reliable measurement of muscle thickness is a powerful tool for use in research and in the clinical setting if the technique is shown to be reliable for a specific muscle, in this case, the quadriceps and its associated non-contractile tissues (i.e. subcutaneous fat and perimuscular fascia). The relative proportions of muscle and subcutaneous fat would be useful to know, for example, when losing weight to ensure fat and not muscle was being lost, and when gaining weight to ensure muscle and not fat was increasing, such as in critically ill patients receiving nutritional support. Obtaining a clear image and taking accurate measurements both require the ability to interpret the image, so this could vary between operators and also between images of different complexity, so reliability needs to be examined between different operators for specific tissues (Whittaker *et al* 2007). Therefore, ensuring the repeatability of measurements between investigators and measurements made on different occasions need to be established when imaging a particular muscle.

It was necessary to conduct the present study of healthy participants to establish the variability when measuring healthy tissues before assessing reliability in pathological cases, as pathology could contribute to variability in measurements due to poorer tissue quality.

Test-retest (intra-rater) reliability of US imaging has already been determined for various muscles: e.g. O' Sullivan *et al* (2007) reported excellent reliability for lower trapezius thickness measurements, with intraclass correlation coefficient (ICC) values greater than 0.90; Herbert *et al* (2009) performed a systematic review on the reliability of the US imaging technique for measurements of abdominal and lumbar trunk muscle thickness, and reported moderate to excellent reliability found from high quality studies, with ICC values ranging from 0.62 to 0.97. Similar reliability (ICC values) was reported by Costa *et al* (2009) in a systematic review for studies that measured abdominal muscle thickness. Excellent reliability of lumbar multifidus muscle thickness was established by Wallwork *et al* (2007), for both experienced raters ($ICC_{3,1}$ of 0.94) and for a novice rater ($ICC_{3,1}$ of 0.89). Very high inter-rater and intra-rater reliability, with ICC values above 0.90, were reported for supraspinatus (Temes *et al* 2011), gluteus medius, gluteus minimus, and vastus medialis muscles (Whittaker and Emery 2014), and intra-rater reliability of anterior thigh tissues, including muscle and non-contractile tissues (Agyapong-Badu *et al* 2014).

The contribution of fascia thickness relative to measurements of muscle thickness has received little attention but a study of the abdominal muscles that measured fascia separately found that thickness of fascia was greater in participants with back pain than those without, whereas muscle was thinner in those with back pain, suggesting that measuring these tissues separately may help us understand mechanisms of muscle contractile abnormalities (Whittaker *et al* 2013). This was an important finding, as fascia is often included in muscle thickness measurements, so the implication is that assessment of muscle atrophy may be underestimated if fascia is included in US measurements of muscle thickness. Regarding quadriceps, a novelty of the present study is that inter-rater reliability of the non-contractile tissue measurements has not been studied previously, i.e. subcutaneous fat and fascia (superficial perimuscular fascia of the rectus femoris muscle and deeper intermuscular fascia between the rectus femoris and vastus intermedius muscles. Another novel aspect is that the relative contributions of muscle and non-contractile tissues have only been studied in young and older age groups (Agyapong-Badu *et al* 2014), and not the present middle-aged group that would typically be the age of osteoarthritis onset. The purpose of the present study was to determine inter-rater and intra-rater between-day reliability of US imaging in measuring rectus femoris and vastus intermedius muscles of the quadriceps, and non-contractile tissue (subcutaneous fat and perimuscular fascia) thickness of the anterior thigh, in healthy middle-aged individuals.

Methods

Participants

A group of 24 (12 females, 12 males) healthy moderately active adults (Physical Activity Guidelines Advisory Committee, 2018), aged (years) 48.91± 9.78 (36-64), height (m) 1.71 ± 0.06 (1.59–1.82), body mass (kg) 72.87± 12.66 (47.8–100.2) participated in the study. Exclusion criteria included diseases and conditions that affect muscle function, lower limb musculoskeletal injuries and pathologies including fracture, surgery, neoplasm, and neurological conditions. Participants were asked not to undertake vigorous exercise within the 24 hours prior to testing. The study was approved by the local Ethics Committee on human experimentation (CESU 1/2015). Written informed consent was obtained from all participants after full explanation of the aims and procedures, and the study was conducted in accordance with the Helsinki Declaration of 1975, as revised in 2008. Participants' rights were protected.

Procedure

Transverse B-mode images of the anterior thigh were acquired by two physical therapists, using an ultrasound scanner (MicrUs EXT-1H; Telemed, Vilnius, Lithuania) with a 5 MHz linear transducer (39 mm length). Ultrasound scans were obtained with the participant resting in supine lying (Figure 1), with the hip in neutral and the knee fully extended. Sandbags were placed at the ankle to avoid lateral rotation of the hip, with the ankle relaxed in slight plantar flexion. The distance between the antero-superior iliac spine and the superior border of the patella was measured, then all ultrasound measurements were performed at two thirds of the measured distance from the antero-superior iliac spine (Delaney *et al* 2010). For inter-rater reliability, measurement of this distance was performed independently by each investigator. Ultrasound scans were taken by two investigators on the same session but they did not observe one another scanning, which was performed independently. For image acquisition, a thick layer of ultrasound gel was applied between the transducer and the skin, and minimal contact pressure was applied when placing the transducer on the skin to obtain the image, to avoid compression of the underlying tissues, which would influence the measurement of tissue thickness. For intra-rater reliability, participants returned one week later for scans to be repeated by only one investigator, to examine test-retest reliability.

Ultrasound Imaging Data Processing

Images were analyzed off-line using ImageJ software (available from https://imagej.nih.gov/ij/). Subcutaneous fat thickness was measured from the skin to the outside edge of the superficial fascial layer (Figure 2), muscle thickness of rectus femoris (RF) and vastus intermedius (VI) were measured between the inside edges of muscles borders to exclude perimuscular fascia. The superficial perimuscular fascial layer was considered between the outside edges of the connective tissue layers superior to RF, while deep fascial layer between RF and VI. Each anonymized ultrasound image was measured twice and the mean used in the analysis. The same investigator (FM) performed the measurements on all images.

Data Analysis

Data analysis was conducted using SPSS 22 (SPSS Inc, Chicago, IL). The data were examined for normality using the Shapiro-Wilk test and found to be normally distributed and therefore parametric. Descriptive statistics were used to summarize the data as means and standard deviations. Intraclass coefficients (ICC) provide a measure of how consistent measurements from multiple observations are, so ICCs were used to assess agreement between measurements made on the two occasions (intra-rater, test-retest).

For intra-rater reliability, the association between measurements made by the same investigator from the two sessions was analyzed using a two-way mixed repeated measure ANOVA ($ICC_{3,2}$).

Agreement between the two raters was analyzed using the ICC calculated by a two-way mixed single measure ANOVA (ICC3,1) to test inter-rater reliability. Precision of measurements was assessed using the Standard Error of Measurement (SEM), which provides values in meaningful units for measuring thickness (mm) and was used to calculate Minimum Detectable Change (MDC), which is the minimal change that falls outside the measurement error. The SEM and MDC were calculated from the ICC as follows: $SEM=SD(\sqrt{1-ICC})$; $MDC=1.96\times\sqrt{2}\times SEM$

Interpretation of ICCs was based on criteria described by Munro (2005), in which ICC > 0.90 indicate very high reliability, 0.70-0.89 high reliability, 0.50-0.69 moderate reliability, and 0.26-0.49 low reliability.

Results

Inter-rater reliability

The measurements obtained by rater 1 and rater 2 demonstrated very high reliability for muscle thickness (ICC3,1 of 0.98; 95% CI: 0.95–0.99), and high reliability for subcutaneous fat (ICC3,1 of 0.81; 95% CI: 0.60–0.91) and perimuscular fascia (ICC3,1 of 0.78; 95% CI: 0.56–0.90), as seen in (Table 1). Inter-rater reliability of perimuscular fascia thickness was high, $ICC_{3,1}$ of 0.70 (95% CI: 0.42–0.86). Standard error of measurement (SEM) was 0.99mm for muscle thickness, 1.44mm for subcutaneous fat, 1.55mm for non-contractile tissue and 0.16mm for fascia thickness. Minimal detectable change (MDC) was 2.74mm for muscle thickness, 3.99mm for subcutaneous fat, 4.29mm for non-contractile tissue and 0.44mm for perimuscular fascia.

Intra-rater reliability

Intra-rater reliability of the measurements obtained by rater 1 on day 1 compared with the measurements taken by the same rater one week apart, are shown in Table 2. The US imaging measurements showed a very high test-retest reliability between days with $ICC_{3,2}$ of 0.96 (95% CI: 0.90–0.98) for muscle thickness measurements, $ICC_{3,2}$ of 0.99 (95% CI: 0.97–0.99) for subcutaneous fat, $ICC_{3,2}$ of 0.98 (95% CI: 0.96–0.99) for non-contractile tissue. Inter-rater reliability of perimuscular fascia thickness was negative with $ICC_{3,2}$ of -0.02 (95% CI: -0.41–0.38). Standard error of measurement (SEM) was 1.3mm for muscle thickness, 0.47mm for subcutaneous fat, 0.66mm for non-contractile tissue and 0.3mm for fascia. Minimal detectable change (MDC) was 3.6mm for muscle thickness, 1.3mm for subcutaneous fat, 1.83mm for non-contractile tissue and 0.83mm for perimuscular fascia.

Discussion

US imaging is an operator-dependent technique (Wakefield *et al* 2005), so demonstrating the level of repeatability of measurements between investigators and test re-test reliability is of high importance and needs to be established for individual muscles. Several factors can influence variability of tissue thickness measurement, many of which can be minimized by standardizing the data collection protocol (Whittaker *at al* 2007). For example, participant positioning in lying was standardized, in terms of posture, position of the limb affecting muscle length and joint angles. The scanner settings were kept constant for each individual, such as type and frequency of the transducer, as well as minimal contact pressure, so as not to distort the image through compression of the tissues. The status of the tissues can also affect their thickness, hence participants were asked not to undertake vigorous exercise within the 24 hours prior to testing and were studied at the same time of day when they returned for repeated testing, to keep conditions as constant as possible.

The present study demonstrated that US imaging is a highly reliable methodology for measuring anterior thigh muscle thickness, though reliability was poor for measuring perimuscular fascia on its own.

Results of the present study are in keeping with the very high intra-rater reliability reported by Agyapong-Badu *et al* (2014) for rectus femoris and vastus intermedius, as well subcutaneous fat thickness of anterior thigh in young and older adults. The present study provides further evidence producing data for both intra- and inter-rater reliability of anterior thigh measurements in middle aged adults.

The MDC values recorded in the present study for intra-rater reliability were relatively lower (muscle thickness=3.6mm, subcutaneous fat=1.3mm) compared to those reported (young: muscle thickness=5.51mm, subcutaneous fat=2.28mm; older: muscle thickness=5.86mm, subcutaneous fat=2.49mm) by Agyapong-Badu *et al* (2014) in healthy people. The authors reported differences with sex and ageing, highlighting the sensitivity to differences between groups but not changes in individuals over time or minimal differences with the technique. Comparisons with data in patient groups may be more informative to identify clinically meaningful differences for specific patient groups (Sabatino *et al* 2017).

The present results for perimuscular fascia thickness measurements alone were not reliable between-days, with ICC values of -0.2, which is below the level of 0.9 recommended for clinical measurements by Portney and Watkins (2000). Similarly, low intra-rater reliability has been reported for abdominal wall perimuscular connective tissue by Whittaker *et al* (2014). Inter-rater reliability for fascia was more acceptable (ICC 0.7), possibly because the two raters performed imaging on the participants within the same session, when participants remained in the same position and the same scanning site marked on the skin. It may be that relocating the scanning site for test-retest reliability was not adequately robust in the present study protocol. From a clinical point of view, it may be more valuable to evaluate the integrity and the continuity of the fascia, as it transmits mechanical tension generated by muscular activity (Maas and Sandercock 2010, Wilke *et al* 2018, Yucesoy 2010), rather than measuring its thickness. The contrary applies to muscle, where reliable measurement of thickness is of high clinical importance. For example, in situations where quadriceps muscle wasting is known to occur, such as critically ill patients in intensive care units (Hadda et al 2017), in patients with knee osteoarthritis (Petterson et al 2008) or any painful condition affecting the knee (Henriksen et al 2011, Rice et al 2014). Monitoring of recovery in these patients could be aided using US imaging to ensure that gain in weight with nutritional loading is due to increases in muscle size and not subcutaneous fat. Other clinical areas where reliable measurement of quadriceps muscle thickness is potentially valuable include assessing quadriceps atrophy due to disuse, sarcopenia, knee pain, and to evaluate quadriceps hypertrophy during specific training programs.

Conclusion

The present findings provide evidence for excellent inter-rater and intra-rater reliability of US imaging for measuring quadriceps muscle and non-contractile tissue (subcutaneous fat combined with perimuscular fascia) thickness of the anterior thigh in healthy middle-aged adults. These results add to the body of knowledge about musculoskeletal soft tissues that can be measured reliably using ultrasound imaging, supporting its use as an objective research and clinical tool.

FIGURE 1. The experimental set-up, with a subject positioned in supine lying, with the ultrasound transducer placed to obtain a transverse section of the anterior mid-thigh, and sandbags around the legs to keep the hips in a neutral position.

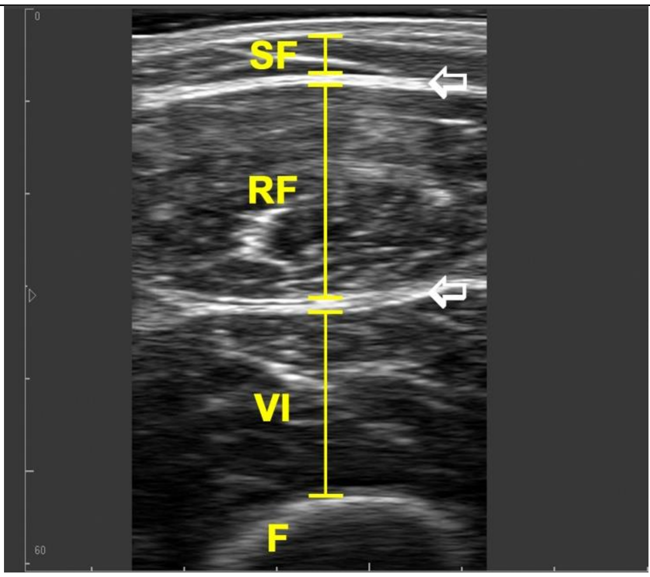

FIGURE 2. Subcutaneous fat (SF); rectus femoris (RF); vastus intermedius(VI); femur (F); arrows indicate superficial and deep fascial layers.

TABLE 1. Inter-rater reliability between two raters

Participants n=24	Subcutaneous fat	Fascia	Non-contractile tissue	Muscle thickness
$ICC_{3,1}$	0.81	0.70	0.78	0.98
95% CI	0.6-0.91	0.42-0.86	0.56-0.9	0.95-0.99
SEM (mm)	1.44	0.16	1.55	0.99
MDC (mm)	3.99	0.44	4.29	2.74

CI=Confidence interval; SEM=standard error of measurement; MDC=minimum detectable change

TABLE 2. Intra-rater reliability on two days

Participants n=24	Subcutaneous fat	Fascia	Non-contractile tissue	Muscle thickness
$ICC_{3,2}$	0.99	-0.02	0.98	0.96
95% CI	0.97-0.99	-0.41-0.38	0.96-0.99	0.90-0.98
SEM (mm)	0.47	0.3	0.66	1.3
MDC (mm)	1.3	0.83	1.83	3.6

CI=Confidence interval; SEM=standard error of measurement; MDC=minimum detectable change

References

Agyapong-Badu S, Warner M, Samuel D, Narici M, Cooper C and Stokes M 2014 Anterior thigh composition measured using ultrasound imaging to quantify relative thickness of muscle and non-contractile tissue: a potential biomarker for musculoskeletal health *Physiological Measurement* **35** 2165–76

Costa L O P, Maher C G, Latimer J and Smeets R J E M 2009 Reproducibility of rehabilitative ultrasound imaging for the measurement of abdominal muscle activity: a systematic review *Physical Therapy* **89** 756–69

Delaney S, Worsley P, Warner M, Taylor M and Stokes M 2010 Assessing contractile ability of the quadriceps muscle using ultrasound imaging *Muscle & Nerve* **42** 530–8

Hadda V, Khilnani G C, Kumar R, Dhunguna A, Mittal S, Khan M A, Madan K, Mohan A and Guleria R 2017 Intra- and Inter-observer Reliability of Quadriceps Muscle Thickness Measured with Bedside Ultrasonography by Critical Care Physicians *Indian Journal of Critical Care Medicine: peer-reviewed, official publication of Indian Society of Critical Care Medicine* **21** 448–52

Hebert J J, Koppenhaver S L, Parent E C and Fritz J M 2009 A systematic review of the reliability of rehabilitative ultrasound imaging for the quantitative assessment of the abdominal and lumbar trunk muscles *Spine* **34** E848-56

Henriksen M, Rosager S, Aaboe J, Graven-Nielsen T and Bliddal H 2011 Experimental knee pain reduces muscle strength *The Journal of Pain : official journal of the American Pain Society* **12** 460–7

Maas H and Sandercock T G 2010 Force transmission between synergistic skeletal muscles through connective tissue linkages *Journal of Biomedicine & Biotechnology* **2010** 575672

Munro B H 2005 *Statistical methods for health care research* 5th edn (Philadelphia: Lippincott Williams and Wilkins)

O'Sullivan C, Bentman S, Bennett K and Stokes M 2007 Rehabilitative ultrasound imaging of the lower trapezius muscle: technical description and reliability *The Journal of Orthopaedic and Sports Physical Therapy* **37** 620–6

Petterson S C, Barrance P, Buchanan T, Binder-Macleod S and Snyder-Mackler L 2008 Mechanisms underlying quadriceps weakness in knee osteoarthritis *Medicine and Science in Sports and Exercise* **40** 422–7

Physical Activity Guidelines Advisory Committee. 2018 Physical activity guidelines advisory committee scientific report. Part C-7. Washington, DC: U.S. Department of Health and Human Services. 2018

Portney L G and Watkins M P 2000 *Foundations of clinical research: Applications to practice* 2nd edn (Upper Saddle River: Prentice Hall)

Rice D A, McNair P J, Lewis G N and Dalbeth N 2014 Quadriceps arthrogenic muscle inhibition: the effects of experimental knee joint effusion on motor cortex excitability *Arthritis Research & Therapy* **16** 502

Temes W C, Temes Clifton A, Hilton V, Girard L, Strait N and Karduna A 2014 Reliability and validity of thickness measurements of the supraspinatus muscle of the shoulder: an ultrasonography study *Journal of Sport Rehabilitation* **23**

Wakefield R J *et al* 2005 Musculoskeletal ultrasound including definitions for ultrasonographic pathology *The Journal of Rheumatology* **32** 2485–7

Wallwork T L, Hides J A and Stanton W R 2007 Intrarater and interrater reliability of assessment of lumbar multifidus muscle thickness using rehabilitative ultrasound imaging *The Journal of Orthopaedic and Sports Physical Therapy* **37** 608–12

Whittaker J L and Emery C A 2014 Sonographic measures of the gluteus medius, gluteus minimus, and vastus medialis muscles *The Journal of Orthopaedic and Sports Physical Therapy* **44** 627–32

Whittaker J L, Teyhen D S, Elliott J M, Cook K, Langevin H M, Dahl H H and Stokes M 2007 Rehabilitative ultrasound imaging: understanding the technology and its applications *The Journal of Orthopaedic and Sports Physical Therapy* **37** 434–49

Whittaker J L, Warner M B and Stokes M 2013 Comparison of the sonographic features of the abdominal wall muscles and connective tissues in individuals with and without lumbopelvic pain *The Journal of Orthopaedic and Sports Physical Therapy* **43** 11–9

Wilke J, Schleip R, Yucesoy C A and Banzer W 2018 Not merely a protective packing organ? A review of fascia and its force transmission capacity *Journal of Applied Physiology (Bethesda, Md. : 1985)* **124** 234–44

Yucesoy C A 2010 Epimuscular myofascial force transmission implies novel principles for muscular mechanics *Exercise and Sport Sciences Reviews* **38** 128–34