

# Community Challenge towards Consensus on Characterization of Biological Tissue: C<sup>4</sup>Bio's First Findings

Nele Famaey<sup>a,b,1,\*</sup>, Heleen Fehervary<sup>a,1</sup>, Yoann Lafon<sup>c,1</sup>, Ali Akyildiz<sup>i,j,1</sup>,  
Silke Dreesen<sup>i,j,1</sup>, Karine Bruyère-Garnier<sup>c,1</sup>, Jean-Marc Allain<sup>t,u</sup>, Marta  
Alloisio<sup>d</sup>, Alejandro Aparici-Gil<sup>ab</sup>, Chiara Catalano<sup>l</sup>, Fanette Chassagne<sup>k</sup>,  
Snehal Chokhandre<sup>n</sup>, Kimberly Crevits<sup>a</sup>, Hanneke Crielaard<sup>i</sup>, Eoghan  
Cunnane<sup>r</sup>, Connor Cunnane<sup>r</sup>, Karen De Leener<sup>ac</sup>, Amisha Desai<sup>v</sup>, Rob  
Driessen<sup>s</sup>, Ahmet Erdemir<sup>n</sup>, Mona Eskandari<sup>o</sup>, Sam Evans<sup>e</sup>, Christian  
Gasser<sup>d</sup>, Marc Gebhardt<sup>y</sup>, Birgit Glasmacher<sup>l</sup>, Gerhard A. Holzapfel<sup>g</sup>,  
Mikel Isasi<sup>ad</sup>, Louise Jennings<sup>v</sup>, Sascha Kurz<sup>z</sup>, Sara Leal-Marin<sup>l</sup>, Pauline  
Lecomte<sup>x</sup>, Annie Morch<sup>x</sup>, John Mulvihill<sup>r</sup>, Fulufhelo Nemavhola<sup>p</sup>, Thanyani  
Pandelani<sup>q</sup>, Salvatore Pasta<sup>l</sup>, Estefania Peña<sup>ab</sup>, Baptiste Pierrat<sup>k</sup>,  
Heidi-Lynn Ploeg<sup>w</sup>, Stanislav Polzer<sup>f</sup>, Manuel Rausch<sup>h</sup>, David Schwarz<sup>f</sup>,  
Hazel Screen<sup>m</sup>, Selda Sherifova<sup>g</sup>, Gerhard Sommer<sup>g</sup>, Shengzhang Wang<sup>aa</sup>,  
Darragh Walsh<sup>r</sup>, Deepesh Yadav<sup>a,b</sup>, Thierry Marchal<sup>ae,2</sup>, Liesbet Geris<sup>af,2</sup>

<sup>a</sup>*FIBEr Core Facility for Biomechanical Experimentation, KU Leuven, Celestijnenlaan  
300, Heverlee, 3001, Belgium*

<sup>b</sup>*Soft Tissue Biomechanics Group, Department of Mechanical Engineering, KU  
Leuven, Celestijnenlaan 300, Heverlee, 3001, Belgium*

<sup>c</sup>*Univ Lyon Gustave Eiffel Univ Claude Bernard Lyon 1 LBMC  
UMR.Ω.-T9406, Lyon, 69622, France*

<sup>d</sup>*Material and Structural Mechanics, Department of Engineering Mechanics, KTH Royal  
Institute of Technology, Teknikringen 8D, Stockholm, 11428, Sweden*

<sup>e</sup>*School of Engineering, University of Cardiff, The Parade, Cardiff, CF24  
3AA, Wales, United Kingdom*

<sup>f</sup>*VSB - Technical University Ostrava, 17.listopadu 2172/15, Ostrava, 708 00, Czech  
Republic*

<sup>g</sup>*Graz University of Technology, Institute of Biomechanics, Stremayrgasse  
16/2, Graz, 8010, Styria, Austria*

<sup>h</sup>*The University of Texas at Austin, Department of Aerospace Engineering and  
Engineering Mechanics, 2617 Wichita Street, Austin, 78712, TX, USA*

---

\*Corresponding author

Email address: [nele.famaey@kuleuven.be](mailto:nele.famaey@kuleuven.be) (Nele Famaey)

<sup>1</sup>joint first author

<sup>2</sup>joint last author

- <sup>i</sup>Department of Cardiology, Biomedical Engineering Group, Erasmus Medical Center, Dr. Molewaterplein 40, Rotterdam, 3015 GD, the Netherlands
- <sup>j</sup>Department of Biomechanical Engineering, Delft University of Technology, Mekelweg 2, Delft, 2628 CD, the Netherlands
- <sup>k</sup>Mines Saint-Etienne, Univ Jean Monnet, Etablissement Francais du Sang, INSERM, U 1059 Sainbiose, Centre CIS, 158 cours Fauriel, Saint-Etienne, F-42023, France
- <sup>l</sup>Department of Engineering, University of Palermo, viale delle scienze, Ed.8, Palermo, 90151, Italy
- <sup>m</sup>School of Engineering and Materials Science, Queen Mary University of London, Mile End Road, London, E1 4NS, UK
- <sup>n</sup>Department of Biomedical Engineering, Lerner Research Institute, Cleveland Clinic, 9500 Euclid Avenue, Cleveland, 44195, Ohio, USA
- <sup>o</sup>Department of Mechanical Engineering, University of California Riverside, 3401 Watkins Drive, Riverside, 92521, CA, USA
- <sup>p</sup>Durban University of Technology, Faculty of Engineering and the Built Environment, Department of Mechanical Engineering, 1 Steve Biko Road, Durban, 4000, Kwazulu-Natal Province, South Africa
- <sup>q</sup>Johannesburg, 1709, Gauteng, South Africa
- <sup>r</sup>School of Engineering and Bernal Institute, AD Building, University of Limerick, Castletroy, Limerick, V94 T9PX, Ireland
- <sup>s</sup>Eindhoven University of Technology, PO Box 513, Eindhoven, 5600 MB, The Netherlands
- <sup>t</sup>LMS, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, Route de Saclay, Palaiseau, 91120, France
- <sup>u</sup>Inria, 1 Rue Honoré d'Estienne d'Orves, Palaiseau, 91120, France
- <sup>v</sup>Institute of Medical and Biological Engineering, School of Mechanical Engineering, University of Leeds, Woodhouse Lane Leeds, LS2 9JT, West Yorkshire, United Kingdom
- <sup>w</sup>Queen's University, Smith Engineering 130 Stuart St., McLaughlin Hall 303A Kingston, K7L 2V9, Ontario, Canada
- <sup>x</sup>Univ. Lille, CNRS, Centrale Lille, UMR 9013 - LaMcube - Laboratoire de Mécanique, Multiphysique, Multiéchelle, Cité scientifique, Villeneuve d'Ascq, 59651, France
- <sup>y</sup>Institute of Experimental Mechanics, Faculty of Civil Engineering, Leipzig University of Applied Sciences, Karl-Liebknecht-Str. 132, Leipzig, 04277, Saxony, Germany
- <sup>z</sup>ZESBO - Center for Research on Musculoskeletal Systems, Department of Orthopedic Surgery, Traumatology and Plastic Surgery, Faculty of Medicine, Leipzig University, Semmelweisstraße 14, Leipzig, 04103, Saxony, Germany
- <sup>aa</sup>Institute of Biomechanics, Fudan University, 220 Handan Road, Shanghai, 200433, China
- <sup>ab</sup>Aragón Institute of Engineering Research (I3A), University of Zaragoza, María de Luna, s/n, Zaragoza, 50015, Spain
- <sup>ac</sup>Materialise NV., Technologielaan 15, Leuven, 3001, Belgium
- <sup>ad</sup>Leartiker Elikagaien Teknologia, Edificio Tecnología de Alimentos — Xemein Etorbidea 19, Markina-Xemein, Bizkaia, 48270, Spain
- <sup>ae</sup>Avicenna Alliance, Rue Guimard 10, Brussels, 1040, Belgium
- <sup>af</sup>Virtual Physiological Human Institute, Celestijnenlaan 300c, Heverlee, 3001, Belgium

---

## Abstract

This study investigates methodological variability across various expert laboratories worldwide, with regards to characterizing the mechanical properties of biological tissues. Two testing rounds were conducted on the specific use case of uniaxial tensile testing of porcine aorta. In the first round, 24 labs were invited to apply their established methods to assess inter-laboratory variability. This revealed significant methodological diversity and associated variability in the stress-stretch results, underscoring the necessity for a standardized approach.

In the second round, a consensus protocol was collaboratively developed and adopted by 19 labs in an attempt to minimize variability. This involved standardized sample preparation and uniformity in testing protocol, including the use of a common cutting and thickness measurement tool. Despite protocol harmonization, significant variability persisted across labs, which could not be solely attributed to inherent biological differences in tissue samples.

These results illustrate the challenges in unifying testing methods across different research settings, underlining the necessity for further refinement of testing practices. Enhancing consistency in biomechanical experiments is pivotal when comparing results across studies, as well as when using the resulting material properties for in silico simulations in medical research.

*Keywords:* Biomechanical characterization, Standardization, Methodological variability, Uncertainty

---

## 1. Introduction

Characterization of the mechanical properties of biological tissue serves many purposes. Firstly, it can help to improve our fundamental understanding of physiological function, due to a strong structure-function relationship for many organs and tissue systems. Secondly and as a consequence of the former, ageing as well as various diseases are known to alter the mechanical properties of the involved tissues, such that quantification of these alterations can aid diagnosis and monitoring. Thirdly, in the field of tissue engineering and regenerative medicine, it is imperative that the engineered tissues replicate the mechanical behaviour of native tissues to ensure proper functioning. Likewise mechanical data informs the design of implants, prosthetics, and surgical tools to ensure compatibility with tissue mechanics and minimize adverse reactions. Indirectly, mechanical properties can be used to inform constitutive models which can in turn be used in computational models, for various applications in the field of *in silico* medicine (Motiwale & Sacks, 2025). However, although the scientific literature therefore abounds with articles experimentally characterizing the mechanical properties of biological tissues, there are still no widely recognized testing standards for these experiments. This shortcoming has consequences for the interpretability of the results, especially when comparing across studies. Moreover, when this experimental data is used to derive material properties used as input parameters to *in silico* models, the associated error and uncertainty propagates into these simulations.

Simulation-driven medical device development and *in silico* trials are gaining importance (Pappalardo et al., 2022; Viceconti & Emili, 2024). Clearly, the quality of these simulations is of utmost importance if *in silico* medicine is to take its rightful place in medical research and development. Consequently and analogously to the well-known concepts of ‘good medical practice’ and ‘good laboratory practice’ (Robertson & Williams, 2009; Stevens, 2003), there is a growing emphasis on the development of guidelines for ‘good simulation practice’. In 2018, the American Society of Mechanical Engineers introduced a standard known as ‘ASME V&V 40 - Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices’ (ASME, 2018), and in 2023, the FDA issued a guidance document entitled ‘Assessing the Credibility of Computational Modeling and Simulation in Medical Device Submissions’ (FDA, 2023). For example, a reporting checklist for verification and validation of finite element analysis was

made more specifically for orthopaedic and trauma biomechanics (Oefner et al., 2021).

Such guidelines span various aspects related to model development, verification, calibration and validation, and focus their attention on the quantification of the uncertainty that accumulates in each of these aspects. Apart from modeling assumptions and discretization errors, it is intuitive that model input data, whether used to calibrate model parameters or validate the model outcomes, are an important source of uncertainty. Indeed, credible numerical simulations require input parameters and underlying acquisition methods that are both traceable and reliable (ASME, 2018). However, where numerical analysts are usually well-equipped to test the validity of their modeling assumptions or to minimize discretization errors, not every numerical analyst has the expertise and/or facilities to assess the quality and uncertainty of their model input data, especially when it involves data related to biological tissue properties.

Ideally, the required material properties and their uncertainty should come from shared, trusted databases. However, various challenges are faced when attempting to create such a database. Firstly, depending on the context of use of a simulation (Viceconti et al., 2021), different types of material properties might be required (mechanical, thermal, electrical, etc.). Even when focusing on mechanical properties alone, a simulation might require properties related to a specific material constitutive law and its characteristics, such as the tensile elasticity, viscoelasticity, anisotropy, compressibility, nonlinearity, ultimate tensile strength, pre-stress state etc. Each of these properties can, in turn, be acquired by various test setups and methods, such as uniaxial or biaxial tensile tests, compression tests, shear tests, indentation tests, dynamic mechanical analysis, and creep tests. The selection of the appropriate testing method depends on the tissue type, its anatomical location, and the desired application/objective to which the simulation responds. Moreover, apart from the actual testing method, the sample collection, preservation and preparation method will also influence the resulting properties, even if such effects are still under debate (Blaker et al., 2024; Oswald et al., 2017; Chow & Zhang, 2011; Stemper et al., 2007).

The second challenge relates to the population- and subject-related variability. Indeed, material properties of biological tissues exhibit significant

110 variability due to physiological and demographic factors. Sex-related vari-  
 111 ations and ethnic diversity contribute to changes in tissue composition and  
 112 structure, leading to differences in material properties for a given tissue type  
 113 (Smoljkić et al., 2023; Åstrand et al., 2011). Age-related changes, such as  
 114 decreased elasticity and increased brittleness in tissues like skin and bone, or  
 115 other pathological conditions, induce further variability (Kirilova-Doneva &  
 116 Pashkouleva, 2022; Mirzaali et al., 2016). Even within a subject, properties  
 117 will vary depending on the location within the body (Krueger et al., 2011).  
 118 Ideally, a material properties database would encompass this broad repre-  
 119 sentation, which would require the collection of an extremely large range  
 120 of experimental data. Moreover, in certain research communities, there ap-  
 121 pears to be an unsubstantiated trust in the ‘known values’ taken from the  
 122 literature. As Hammer & Klima (2019) also noted in their review, the imple-  
 123 mentation of further region-specific studies is partially inhibited, as reliance  
 124 is placed on a supposedly existing broad database, instead of carrying out  
 125 new studies adapted to the research question. Using numerical studies of  
 126 the biomechanics of the sacro-iliac joint as an example, they point out the  
 127 frequent nested literature references and the often accompanying decrease in  
 128 relevance of the basic source for the intended loading scenario.

129  
 130 With that in mind, perhaps the most important challenge related to build-  
 131 ing up such a database is related to methodological variability. Indeed, even  
 132 when comparing literature results of the same tissue type of a similar an-  
 133 imal strain or patient group, tested according to - reportedly - the same  
 134 method, results vary widely. As mentioned, there are currently no estab-  
 135 lished standards for measuring the material properties of biological tissue,  
 136 nor is there a unified approach for sample preparation and storage, or for  
 137 reporting and assessing the obtained results. Some groups have proposed  
 138 guidelines for specific aspects or applications, e.g. Kurz et al. (2023) have re-  
 139 cently suggested a standardized approach for characterization of the human  
 140 lumbopelvic system, Wale et al. (2021) have investigated sample preparation  
 141 and optimal sample shape to induce reproducible failure for tensile testing of  
 142 musculoskeletal soft tissues, Scholze et al. (2020) have proposed a clamping  
 143 system for simple and reproducible sample clamping. Though valuable, these  
 144 studies are limited to a single research group, and it remains to be seen if and  
 145 how they will be adopted by the scientific community. Lin et al. (2024) have  
 146 recently performed a systematic review of uniaxial tensile testing of human  
 147 soft tissues across research groups. They found large variations in sample

148 shape and especially underreporting of various aspects of the protocol, in-  
 149 cluding the clamping mechanism. Similarly, Fehervary (2018) investigated  
 150 variations in planar biaxial testing of arterial tissue, both in experimental  
 151 setup and data processing, and formulated guidelines for testing and report-  
 152 ing thereof. Although this lack of standardization and high variability is  
 153 commonly known, to the best of our knowledge, a quantitative analysis of  
 154 the current degree of methodological variability has not yet been attempted.  
 155 It is therefore an essential first step towards standardization, ultimately lead-  
 156 ing to reliable material properties, including their uncertainty, which can be  
 157 used in computational analyses for *in silico* medicine.

158  
 159 In 2019, the Virtual Physiological Human institute together with the Avi-  
 160 cenna Alliance launched a task force on tissue characterization, out of which  
 161 the C<sup>4</sup>Bio (C<sup>4</sup>Bio, 2025) initiative was born, an acronym for ‘Community  
 162 Challenge towards Consensus on Characterization of Biological Tissue’. Con-  
 163 sidering the wide range of properties, tissue types and testing methods, the  
 164 decision was made to direct the initial focus to ‘simple’ uniaxial tensile testing  
 165 (i.e. the method most commonly found in literature) of porcine aorta. This  
 166 paper first describes the methods employed to compare the results obtained  
 167 by 24 expert laboratories from around the globe who characterized biological  
 168 tissues and synthetic samples during two test rounds. In a first testing round,  
 169 laboratories were instructed to use their own established method, to allow  
 170 an unbiased evaluation of existing methodologies. In a second testing round,  
 171 participants worked together to create a ‘consensus methodology’ and were  
 172 then instructed to perform their tests accordingly.

## 173 2. Materials & methods

### 174 2.1. Participant recruitment & overall methodology

175 The initiative was announced in the fall of 2020 through various com-  
 176 munication channels. All research groups with experience in material char-  
 177 acterization of biological tissue were invited to join, provided the following  
 178 prerequisites:

- 179 • the availability of infrastructure suitable for uniaxial tensile testing of  
 180 samples within a length range of 10-60 mm and a linearized Young’s  
 181 modulus range of 0-5 MPa,
- 182 • the clearance to test biological tissue in the laboratory,

- 183 • a demonstrated experience in mechanical experiments on biological tis-  
184 sue through scientific references,
- 185 • the willingness to publicly share the experimental results.

186  
187 In round 1, 24 laboratories from all over the world participated (see Fig-  
188 ure S1 in the supplementary material, section Appendix A, for a map), and  
189 19 of them further participated in round 2. The participants that were no  
190 longer able to participate indicated that this was due to a lack of budget  
191 and/or person power.

192  
193 For both testing rounds, all samples were procured and prepared cen-  
194 trally and shipped as described in Section 2.2. In round 1, 24 laborato-  
195 ries performed mechanical testing and analysis according to their preferred  
196 methodology, with only a limited number of instructions given (see Section  
197 2.3). After testing, raw and processed data were collected and analysed cen-  
198 trally according to the methods described in Section 2.4. When providing the  
199 raw and the processed data, participants were also asked to fill in a survey,  
200 querying on various aspects of the testing methodology they used. Following  
201 the central analysis, the observations were discussed with the participants  
202 over multiple online meetings, and a consensus protocol was established and  
203 used in round 2. The full protocol as shared with the participants can be  
204 found in the supplementary material (section Appendix A).

## 205 *2.2. Sample preparation & shipping*

### 206 *2.2.1. Biological tissue*

207 In both testing rounds, 2 sets of a proximal and a distal segment of one  
208 porcine descending aorta was prepared per participant at FIBEr, the KU  
209 Leuven core facility for Biomechanical Experimentation. Each round, aortas  
210 were collected from 70 animals from a local slaughterhouse, and brought to  
211 FIBEr in physiological medium (saline solution) at 4°C. Care was taken to  
212 minimize biological variability, by harvesting material as much as possible  
213 from animals of the same strain, average weight and age, and distributing  
214 these randomly over the participants. From each aorta a distal and proxi-  
215 mal segment of approximately 2 cm in diameter and 5 cm in length was cut,  
216 with a small proximal anterior incision, to allow tracking of the orientation of  
217 the sample (see Figure S2 in the supplementary material, section Appendix

218 A). Samples were placed in individual containers with physiological medium,  
219 while keeping track of the animal and the location (proximal vs distal). All  
220 samples were transferred to a  $-80^{\circ}\text{C}$  freezer inside the containers within 8  
221 hours of excision from the animal, and stored in FIBEr until shipment. The  
222 samples were shipped to each of the participants on dry ice in styrofoam  
223 boxes. Participants were instructed to keep the aortic tissue in frozen condi-  
224 tions (i.e. at  $-80^{\circ}\text{C}$ ) until they were ready for testing, or to start the thawing  
225 and testing process immediately if no freezing capability was available in  
226 the laboratory. Upon receipt of the package, the participants verified that a  
227 sufficient amount of dry ice was still present in the box to ensure proper sam-  
228 ple preservation. For all samples, the duration of frozen storage was recorded.  
229

### 230 2.2.2. *Synthetic samples*

231 In round 1, two sheets of synthetic elastomer material of approximately  
232  $4.5 \times 9 \text{ cm}$  were shipped to the participating research group along with the  
233 biological tissue, in casu ‘Leartiker-K73’ and ‘Leartiker-K51’, provided by  
234 Leartiker s.a. (Spain).

235 In round 2, to use a material with more biologically relevant properties  
236 as compared to round 1, samples were 3D printed by Materialise NV (Bel-  
237 gium) using the company’s proprietary HeartPrintFlexPlus (HPF+) material  
238 (Schickel et al., 2019) in the dogbone shape as agreed upon in the consensus  
239 protocol (see section 2.3.2). In both rounds, the synthetic material was kept  
240 and shipped in dry conditions at ambient temperature.  
241

## 242 2.3. *Mechanical testing*

### 243 2.3.1. *Round 1*

244 In the first testing round, only minimal instructions were given to par-  
245 ticipants. This was done on purpose to allow for an objective comparison  
246 of the participating laboratories preferred methodologies. It was requested  
247 that all tests were executed by the same person, to avoid any inter-user vari-  
248 ability. Also, the same equipment and methodology had to be used for both  
249 the aortic and the synthetic samples. The only exception in the process was  
250 that the synthetic material should always be tested in dry conditions.  
251

252 The evening before the testing day, the participating groups placed the  
253 aortic specimen in a fridge ( $4^{\circ}\text{C}$ ) to thaw overnight. This allowed a thawing

254 time of around 16 h, whereby the exact time per specimen was recorded.  
255 Next, the specimens were divided into test samples. From each tubular seg-  
256 ment, at least 3 uniaxial test samples were prepared, with geometry and  
257 cutting method varying per participant. Proximal and distal segments were  
258 used to prepare circumferentially and longitudinally oriented samples, re-  
259 spectively. Participants were instructed to take a picture or make a sketch to  
260 indicate the location of each of sample with respect to the original specimen.

261  
262 For the synthetic specimens, at least 6 samples were cut out of each pro-  
263 vided sheet, in a direction of choice since the material is considered to be  
264 isotropic. Figure S3 in the supplementary material (section Appendix A)  
265 gives an overview of all samples (biological and synthetic) tested per partic-  
266 ipant.

267  
268 For each (biological and synthetic) sample, the undeformed sample thick-  
269 ness, undeformed sample width in the neck region, and undeformed length  
270 of the neck region were measured according to each participant’s method-  
271 ology. Finally, participants also selected individually the appropriate loading  
272 protocol and data processing methodology to determine the elastic stiffness  
273 moduli at various strain levels and the ultimate tensile strength.

### 274 275 2.3.2. Round 2

276 In round 2, the same thawing instructions were given as in round 1. After  
277 thawing, test samples were excised from the aortic segments and at least three  
278 circumferentially oriented samples were prepared from the proximal segments  
279 and three longitudinally oriented samples from the distal segments. This  
280 time, the samples were cut into a dogbone shape using a cutting tool that  
281 was shipped to the participating research group along with the tissue. The  
282 dogbone shape and cutting tool are shown in Figure 1a and b and more details  
283 are available in 2.2. The time required for each phase in the preparation and  
284 testing protocol was logged for each sample.

285 Participants were required to measure the undeformed width in the neck  
286 region and overall length of each sample using calibrated images taken with a  
287 high-resolution camera on millimeter paper. For the thickness measurement,  
288 the sample was inserted into a tool that was provided to the participants,  
289 shown in Figure 1c. This tool gently squeezes the sample with the help of  
290 identical elastic bands between two plates with known dimensions, such that

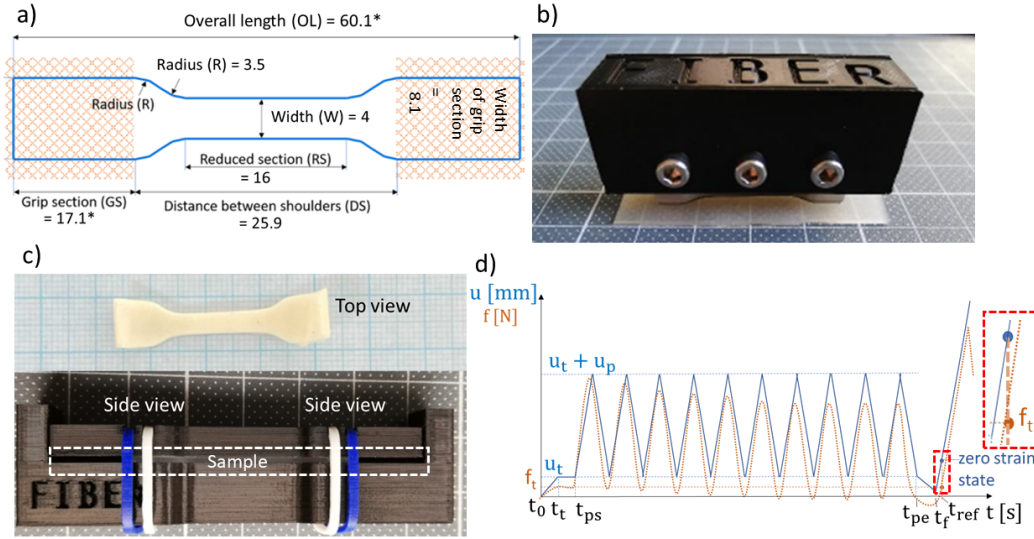


Figure 1: In round 2, a) samples were cut into a dogbone shape as agreed upon by all participants. \*These dimensions can be increased if more gripping surface is required. All dimensions are in millimeter. b) A cutting tool was provided to ensure consistency in the shape. c) Calibrated pictures were taken during sample preparation for dimension measurements. Top: aortic sample on millimeter paper. Bottom: synthetic sample in a custom-made thickness measurement tool, tightened with elastic bands. d) Loading protocol for round 2, showing actuator displacement  $u$  and expected reaction force  $f$  as a function of time  $t$ . Time  $t_t$  corresponds to the time at which the threshold force of  $f_t$  is reached for the first time. Times  $t_{ps}$  and  $t_{pe}$  correspond to the start and end of the preconditioning cycles, during which a preload amplitude  $u_p$  is applied for 10 cycles. Time  $t_f$  corresponds to the start of the final upward ramp and  $t_{ref}$  corresponds to the assumed zero strain state.

291 the sample thickness can be derived from a side view image focussed on the  
 292 tissue surface.

293

294 Participants were advised to use digital image correlation (DIC) to track  
 295 sample deformation during the test, which required application of a speckle  
 296 pattern on the sample's intimal surface, by sprinkling graphite powder on a  
 297 piece of paper and subsequently placing the intimal side of the sample onto  
 298 the paper. Alternatively, if no DIC system was available, participants used  
 299 a marker tracking technique, placing four markers at  $\sim 6$  mm intervals in the  
 300 central area of the sample.

301 Similar to round 1, the participants were asked that all samples were

302 tested using the same equipment by a single operator. After sample accli-  
 303 matisation in saline solution at 37°C for approximately 10 minutes, the load  
 304 cell was zeroed. For a horizontal set-up, this was done before mounting the  
 305 sample, whereas for a vertical set-up, this was done after the sample was  
 306 fixed in the upper clamp. During testing, the biological samples were im-  
 307 mersed in a saline bath at 37°C, whereas the synthetic samples were kept in  
 308 dry condition.

309 The specific loading protocol as depicted in Figure 1d was applied. After  
 310 reaching a threshold force  $f_t = 0.1$  N, 10 preconditioning cycles were applied  
 311 in the physiological loading regime with an amplitude  $u_p = 8$ mm. This  
 312 was followed by a displacement-controlled ramp loading until failure at a  
 313 displacement rate of 1.3 mm/s. During the test, resulting displacement,  
 314 force, applied strain, local samples strains, and where applicable, images,  
 315 were acquired with a minimum sampling rate of 10 Hz. The full description  
 316 of the loading protocol is provided in the supplementary material (section  
 317 Appendix A, see `Instructions_2nd_testround.pdf`).

## 318 2.4. Data analysis

### 319 2.4.1. Method variation

320 In addition to the testing data submission, in round 1 the participants  
 321 were also requested to fill in a questionnaire to report their choices regarding  
 322 pre-, intra-, and post-testing conditions. These included pre-testing stor-  
 323 age conditions, test sample preparation and shape characterization, testing  
 324 device characteristics, and testing protocol.

325 While the consensus protocol was used in round 2, the research groups  
 326 were asked if they had any intentional or unintentional deviations from the  
 327 protocol to identify potential sources of variability.

### 328 2.4.2. Data submission

329 For each test, a table as shown in supplementary material was filled in,  
 330 which differed slightly depending on the testing round. Apart from the afore-  
 331 mentioned sample geometry information, actuator displacement, load cell  
 332 values and local stretch estimation in the testing direction were obtained as  
 333 a function of time. The participants were also asked to indicate whether the  
 334 sample ruptured in the central area or at the clamp site.

335

### 336 2.4.3. Data processing

337 **Dimensions of interest.** In round 2, the method and software used to  
 338 measure dimensions from the calibrated images (see section 2.3.2) were left  
 339 to the teams' discretion.

340 **Engineering stress & stretch.** The raw force data was divided by  
 341 the cross-section area of the reduced section to compute the raw engineering  
 342 stress. The stretch data was retrieved by the participants using their own  
 343 DIC analysis or marker tracking software. Subsequently, the raw stress and  
 344 stretch data of all groups were processed by two data analysts to avoid any  
 345 processing errors, before being synthesized by just one of them using Mat-  
 346 lab 2019a and 2023a (Mathworks.com): only the test data between the end  
 347 of preconditioning and the highest stress value was kept, filtered (forward-  
 348 backward moving average with a window of 5%) and resampled to 1000  
 349 points. A qualitative double-check of the cleaned data against the raw data  
 350 was performed. Per step of 0.1 stretch ratio, a tangent modulus (TM) was  
 351 computed using a least-squares line-fitting.

352 **Comparison of stress/stretch curves.** Various parameters for each  
 353 team and per type of sample (i.e. a set) were computed to compare the  
 354 results. The mean and standard deviation of the stress was computed at  
 355 every 0.1 interval of the stretch, and the mean and standard deviation of the  
 356 stretch was then computed at every 0.1 interval of the stress. Next, the mean  
 357 stress/stretch curve a set of curves was computed point per point. I.e, for  
 358 any given stress/stretch point of the 1000 points discretizing each curve, the  
 359 mean of all the values at that point within the set was used.

360 **Analytic uncertainty on stress.** A theoretical uncertainty on the  
 361 obtained stress values  $U(\sigma)$  was determined for each team and sample type,  
 362 incorporating measurement uncertainties as follows:

$$U(\sigma) = \sqrt{\sigma^2 \left( \left( \frac{U(N)}{\bar{N}} \right)^2 + \left( \frac{U(w)}{\bar{w}} \right)^2 + \left( \frac{U(t)}{\bar{t}} \right)^2 \right)} \quad (1)$$

363 In the above equation,  $\sigma$  represents the engineering stress (in MPa),  $N$  is  
 364 the normal force and  $w$  and  $t$  are the undeformed width and thickness of  
 365 the sample, respectively. The measurement uncertainty on the force,  $U(N)$ ,  
 366 is considered the same for all groups and on average assumed to be 0.1 N  
 367 (based on a ISO 376 class 1 load cell measuring around 20 N), while the  
 368 uncertainties on the width,  $U(w)$ , and thickness,  $U(t)$ , are both set to 0.02

369 mm (based on the maximum permissible error for external jaws on calipers  
 370 with measurement length 0-5 cm, according to ISO 13385-1). The notation  
 371  $\bar{\cdot}$  represents the mean value of each parameter for a given team and sample  
 372 type.

373 **Zero-strain state** In the consensus protocol, the zero-strain state of the  
 374 final upward ramp was considered to correspond to the time point  $t_{ref}$  (see  
 375 Figure 1d) at which the threshold force  $t_f = 0.1$  N was reached. However,  
 376 during post-processing to increase consistency, this was modified to the strain  
 377 state corresponding to the time point at which the measured stress was equal  
 378 to  $2*U(\sigma)$  (see equation 1). This was equal to 0.03 MPa for biological samples  
 379 and 0.06 MPa for synthetic ones.

### 380 3. Results

#### 381 3.1. Submissions

382 In round 1, 20 participants successfully submitted their results for both  
 383 the biological and synthetic data. In round 2, we received 17 successful  
 384 submissions for the synthetic samples and 13 for the biological samples.

#### 385 3.2. Method variation

386 The questionnaire results, regarding the choices of the research groups  
 387 in round 1 for pre-, intra-, and post-testing conditions, are given in Figure  
 388 2 together with how many groups (in %) employed a certain technique or  
 389 device.

390 Although the research groups were asked to use the consensus protocol  
 391 in round 2, intentional or unintentional deviations from the protocol were  
 392 observed by some groups and summarized in supplementary material (sec-  
 393 tion Appendix A). For example, as the surface of the synthetic samples  
 394 started showing cracks during testing, the stretch measurements could not  
 395 be obtained with the intended DIC approach. Hence, many research groups  
 396 reported the clamp-to-clamp distance instead. Another common deviation  
 397 related to the removal of supposedly surrounding connective tissue in the  
 398 biological samples. Some groups removed this tissue, possibly removing the  
 399 adventitia along with it, others did not. Besides these, there were some  
 400 occasional, unintentional deviations from the loading protocol, such as dif-  
 401 ferent preconditioning displacement or time point at which load-taring was  
 402 performed.

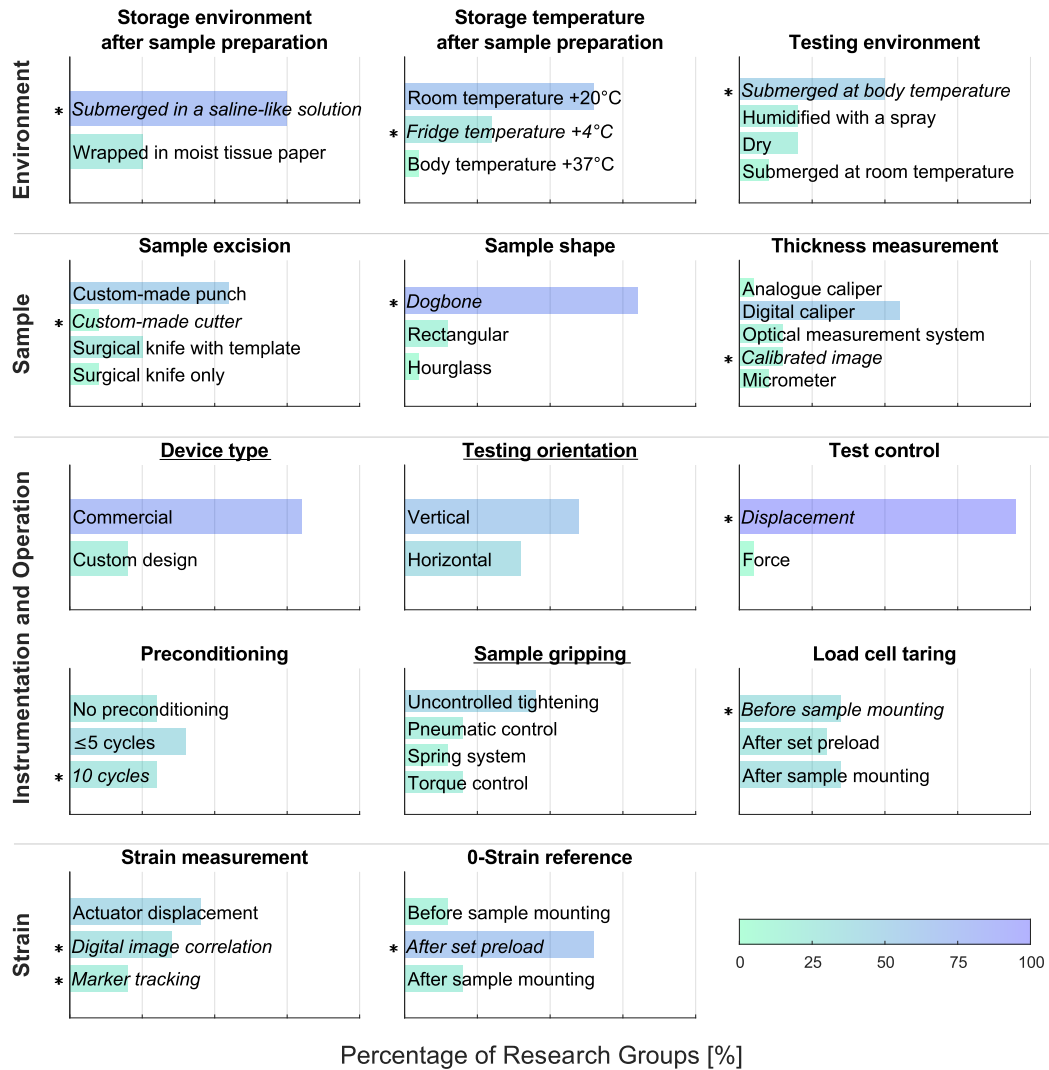


Figure 2: Method variation in round 1 and the percentage of groups using a certain approach. Parameters fixed in the consensus protocol of round 2 are marked with '\*'. Test criteria not constrained by the consensus protocol in round 2 are underlined.

### 3.3. Main results

An overview of all the engineering stress vs. stretch curves obtained by the participants for both testing rounds is given in Figure 3. These curves allow to obtain an impression of the overall variability in stress-stretch response and ultimate strain values of the tested samples.

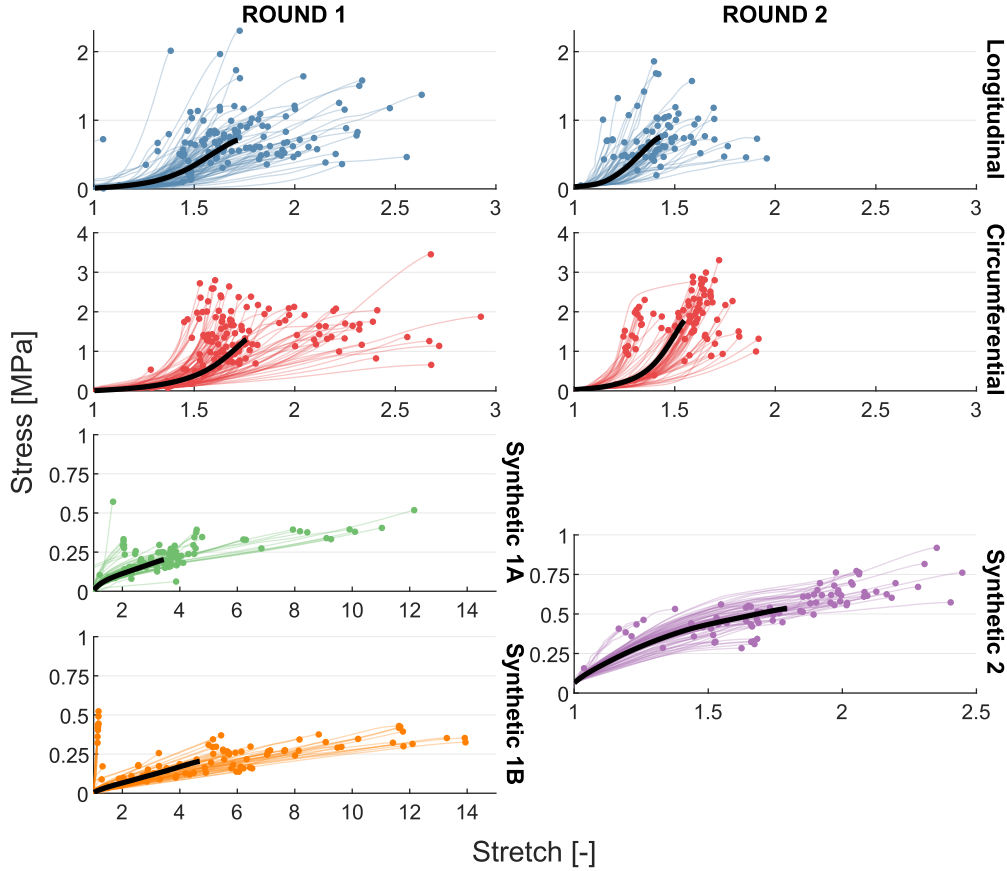


Figure 3: Engineering stress (in MPa) as a function of stretch obtained from uniaxial tension test until rupture, categorised by specimen type and separated per testing round. The average of all curves corresponding to each specimen type and testing round is represented in black.

408        Figures 4a and 4b show the stress vs. stretch curves obtained per research  
409        group on aortic tissue in the longitudinal direction during round 1 and 2 re-  
410        spectively. The corresponding curves for the circumferential direction and for  
411        the synthetic samples can be found in the supplementary material (Figure S4  
412        and S5). The graphs also show the standard deviations of stress and stretch  
413        values at certain intervals. Figure 5 zooms in on this standard deviation and  
414        how it varies between research groups and between rounds. Now once again  
415        grouped for all participants, the graphs show the coefficient of variation of  
416        stress and stretch at two specific stretch levels. The graph also shows the

percentage of research groups that fall under the arbitrary threshold of 0.33 in terms of this coefficient of variation.

The distributions of the tangent moduli measured at 1.2 and 1.4 stretch are illustrated on Figure 6 and 7, respectively. To consider both the stress and strain measurement variabilities in the assessment of the tangent moduli, Figure 8a illustrates the coefficient of variation of stress and stretch at 1.2 and 1.4 stretch, for each group and for the various sample types. This figure discriminates the groups that fall under the arbitrary threshold of coefficient of variation of 0.33 for both stretch and stress.

The distributions of the measurements of the thickness and the width of the samples among the groups and for both rounds are illustrated on Figure 9 and Figure 10, respectively. The coefficient of variation values of these measurements are illustrated on Figure 8b.

## 4. Discussion

Even for a relatively simple method like uniaxial tensile testing, the results of round 1 demonstrate staggering variability between the participating laboratories. This is true for the obtained stress-stretch curves (Figure 3), the corresponding coefficient of variation values (Figure 5), as well as the resulting tangent moduli (Figure 8a). Looking at Figure 2, one can also observe a large variation in the choices made regarding the test protocol in this first round. However, statistical analysis (not shown here) could not identify any correlation between these protocol variations and the results. This implies that the observed variability cannot be directly attributed to a single queried aspect of the protocol. Indeed, the observed variability is a superposition of biological variability, methodological (or aleatory) uncertainty *and* methodological (or epistemic) discrepancy.

### 4.1. Variability versus uncertainty in round 1

To minimize biological variability, aorta samples were harvested from animals of the same age, weight and strain, as detailed in section 2.2, and randomly distributed over the participants. However, even in such a population, biological variability is expected. Therefore, in an attempt to isolate the effect of biological variability, we also integrated synthetic samples into our analysis, as they suffer far less from intrinsic variability. Figure 8a shows how the overall coefficient of variation is indeed significantly reduced for the synthetic materials compared to the biological tissue, but still reaches levels

452 close to 50%.

453

454 A rough estimate of the degree of methodological uncertainty on the stress  
455 measurement was made using equation 1, based on estimates of the accuracy  
456 of the used load cells and geometry measurement methods. Figure 4a shows  
457 how this estimated measurement uncertainty relates to the resulting stan-  
458 dard deviation in stress. For most groups, this uncertainty represents only a  
459 fraction of the total observed variability, even for the synthetic material (see  
460 the supplementary material).

461

462 A significant portion of the observed variability must therefore be at-  
463 tributed to methodological variability, where we can in turn distinguish be-  
464 tween intra-research group variability (also known as repeatability) and inter-  
465 research group variability (also known as reproducibility). Although one  
466 might expect the former to be minimal, given the instructions to test consis-  
467 tently, i.e. on the same day and by the same operator, the results indicate  
468 a significant and varying amount of intra-research group variability. Conse-  
469 quently, one might expect this repeatability to correlate with the reported  
470 protocol variations (see Figure 2). However, as mentioned, further statistical  
471 analysis did not reveal any correlation between these protocol variations and  
472 the actual results.

473 To disentangle sources of variability, one could attempt to obtain infor-  
474 mation on biological variability from other studies reported in literature.  
475 Indeed, there are several studies that report mechanical properties of aorta.  
476 For example, Shahbad et al. (2025) report on regional variations in stiff-  
477 ness of the human aorta along its length, whereas Ryu et al. (2022) report  
478 region-dependent stiffness parameters of porcine aorta. Both studies show a  
479 response and degree of variability (per region) in the same order of magnitude  
480 as our results. Of course, and this is the main message of our publication,  
481 it is not possible to distinguish methodological variability from biological  
482 variability in any of the studies. Indeed, there is no reason to assume that  
483 other laboratories would not suffer from the same degree of methodological  
484 variability as our reported intra-research group variabilities.

#### 485 4.2. Consensus protocol and its trade-offs

486 Therefore, in round 2, we aimed to minimize methodological variability  
487 of as many aspects of the protocol as possible, leading to a lengthy consen-  
488 sus protocol. All participants also used the same cutting tool and thickness

489 measurement tool in this second round. Still, in this quest for a harmonized  
490 protocol, we needed to consider two trade-offs. The first was a trade-off  
491 between methods that are most likely to yield biologically relevant results  
492 and methods with a lower complexity. The latter are more accessible for a  
493 broad group of researchers, and might also induce a lower amount of vari-  
494 ability. As an example, it is easier to test samples in dried conditions than in  
495 physiological, submersed conditions. The latter requires mounting of a fluid  
496 bath, which makes it harder to mount the sample, and, depending on the  
497 orientation of the set-up, puts restrictions on the type of load cell. Although  
498 the dry state of the sample might not alter the intrinsic variability between  
499 samples compared to a submerged state, the extra requirements and manip-  
500 ulations could induce further methodological variability. Still, we opted for  
501 testing in immersed conditions. A counterexample in this respect is the fact  
502 that all tissue was frozen until the test day, rather than testing the tissue in  
503 fresh conditions. Whereas fresh tissue testing is, of course, more biologically  
504 relevant (Chow & Zhang, 2011; Stemper et al., 2007), freezing the tissue was  
505 necessary to allow for uniformity in the tissue preservation step. The second  
506 trade-off was to be made between methods that are likely to yield the most  
507 unbiased results and methods that were available to most participants (in  
508 terms of infrastructure and personnel requirements). As an example, the  
509 thickness measurement method should ideally not compress the tissue (Kim  
510 et al., 2011; O’Leary et al., 2013) or correct for any compression during the  
511 measurement (Schwarz et al., 2023), but requires specific hardware and soft-  
512 ware that was not available to all participants, which is why we opted for  
513 the supplied thickness measurement tool.

514

515     ASTM D412 is a common standard to characterize mechanical properties  
516 of rubbers and elastomers using a tensile test. Our resulting consensus pro-  
517 tocol deviates from this standard in various ways. Firstly, we used a lower  
518 displacement rate (1.3 mm/s vs. 500 mm/min) to approximate ‘physiological’  
519 loading conditions. Secondly, where the standard suggests a fixed temper-  
520 ature, we allowed variations between laboratories in ambient temperature  
521 when testing the synthetic samples, which might contribute to the variabil-  
522 ity of these measurements. Due to the limited size of an aorta, our chosen  
523 dogbone shape was smaller than the standard, especially regarding the width  
524 of the grip section (8.1mm vs. 25mm) and the length of the reduced section  
525 (16mm vs. 33mm). This reduced size and aspect ratio might lead to a non  
526 uniform strain field within the area of measurement (Lin et al., 2024).

#### 527 4.3. *Variability on geometrical measurements*

528 Despite efforts to reach a consensus on the testing methodology, the sec-  
529 ond round did not display significant improvement in the quantities of inter-  
530 est. One metric that was however drastically improved was the variation on  
531 the sample width measurement (see Figures 8b and 10). This makes sense  
532 given the fact that this dimension was not specified in the first round, whereas  
533 all participants used the same cutting tool in the second round. Thickness  
534 measurement variability was also improved (as can be observed from the re-  
535 duced coefficient of variation of the thickness measurement in Figure 8b),  
536 but not to the same extent as the width measurement. This may be ex-  
537 plained by a greater variation in the thickness compared to the width of the  
538 samples, as the samples were prepared using the cutting tool. Of course,  
539 release of residual stresses may still have led to some width variation in the  
540 samples. Secondly, the samples are approximately twice as wide as they are  
541 thick, reducing the signal to noise ratio for the latter measurement. The  
542 variability on the thickness measurement is indeed significantly lower for the  
543 synthetic samples in both rounds, although one can notice an outlier group  
544 for the measurements on synthetic 1A samples. Apparently, this group had  
545 used an optical laser measurement method, which suffered from laser light  
546 penetration for this sample type, yielding a systematic error in their results  
547 (O’Leary et al., 2013).

#### 548 4.4. *Variability on stress-stretch curves and tangent modulus*

549 Figures 5 and 8a illustrate how the improvements on the dimension mea-  
550 surements did not propagate to the resulting stress-stretch curves or tangent  
551 moduli. For low stretch values, coefficient of variation levels are even slightly  
552 increased, and there is a decrease in research groups reaching the threshold  
553 value of coefficient of variation for the longitudinal stress-strain behaviour of  
554 the aorta. For the higher strain level, we do observe a reduction in coefficient  
555 of variation levels and a higher number of participants reaching the thresh-  
556 old, but the improvement is far lower than initially anticipated.

557  
558 An explanation for this lack of improvement in terms of tangent modulus  
559 variability could lie in the following: due to the standardization of the zero  
560 strain state in the consensus protocol, the origin of the curves is more ho-  
561 mogeneous and the curves have therefore on average shifted towards the left,  
562 leading to higher stresses for a certain stretch level (see Figures 6 and 7). As

563 a consequence, the values of the tangent moduli obtained in round 2 are sig-  
564 nificantly higher than in round 1, increasing the amplitude of their variability.

565  
566 In all cases, we can observe less variability for synthetic samples than  
567 for aortic samples. In round 2, all participants are below the coefficient  
568 of variation threshold value for the stress-stretch behaviour and the overall  
569 coefficient of variation in tangent modulus is lower for both stretch levels.  
570 The resulting tangent modulus is also significantly higher for the synthetic  
571 material used in round 2 compared to those of round 1. Indeed, a different  
572 material was used, which at least at low stretch levels, matched the average  
573 tangent modulus of the aortic material more closely.

#### 574 4.5. *Remaining sources of variability*

575 The estimated methodological uncertainty on the stress measurement  
576 (equation 1) is significantly smaller than the observed one, even more so  
577 for round 2 (see Figure 4b). This means that much of the variability is still  
578 due to other factors, a number of which are discussed below.

- 579 • **Stretch measurement uncertainty** - Either marker tracking or DIC  
580 was performed in round 2. The reliability of the former can be affected  
581 by marker placement accuracy and tracking resolution, which was not  
582 harmonized between the participants. Also for DIC, its reliability de-  
583 pends on the quality of the speckle pattern, the resolution of the imag-  
584 ing system, as well as on software analysis settings, as was investigated  
585 at length in a series of ‘DIC challenges’ (Reu et al., 2018, 2022). Nev-  
586 ertheless, these DIC challenges were not specifically aimed at biological  
587 tissue applications, which pose particular challenges. For example, im-  
588 mersion of the sample in a water bath possibly causes optical distortion,  
589 speckle pattern application should ensure proper adherence to the tis-  
590 sue, not cause or require dehydration of the tissue, and last through  
591 large deformations (Lionello et al., 2014). The sample shape, shorter  
592 than the ASTM recommendations for synthetic materials, can induce  
593 an inhomogeneous deformation pattern, the degree of which should be  
594 investigated (Lin et al., 2024). These inhomogeneities should be dealt  
595 with, either through a sample-size specific correction factor, or by ap-  
596 plying DIC rather than marker tracking where possible.
- 597 • **Zero-strain state** - The samples were considered to be in their zero  
598 strain state at the start of the test, specifically at 0.03 MPa for biolog-

ical samples and 0.06 MPa for synthetic ones. This stress-based definition clearly induces error as well as uncertainty proportional to the force and surface area measurement uncertainty. Moreover, it is generally known that biological tissues exhibit residual strains (Vaishnav & Vossoughi, 1983), further confounding the definition of the zero-strain state.

- **Adherence to the testing protocol in round 2** - Despite the provided instructions, we observed variations in how different groups adhered to the testing protocol. For instance, some groups removed the connective tissue of the adventitial layer while others did not. The adventitial layer is known to be collagen-rich and therefore contribute significantly to the nonlinearity of the stress-stretch curve. Additionally, there were discrepancies in the applied preconditioning cycles and in the definition of the state to be reached before starting the final ramp loading. This might be attributed to misunderstanding of the instructions, but also due to the incompatibility of certain testing machine controllers w.r.t. the desired protocol. These deviations are documented in detail in the supplementary material (see ProtocolDeviationsRound2.pdf). Note also that the hypothesis was made that each participant used calibrated sensors and testing machines, for example according to ASTM E4 standards, but this was not explicitly verified.
- **Intra-group variability** - Some teams exhibited greater intra-group variability in round 2 compared to round 1. This could be attributed to the fact that they were following a testing protocol different from their usual practices. For example, managing samples in an immersed test environment can be more challenging and may introduce additional variability, especially when the user is still going through a learning curve.
- **Sample misalignment** - For highly anisotropic samples, ensuring that samples are cut precisely along the circumferential or longitudinal direction is crucial for consistent results. Misalignment can lead to variations in the mechanical properties measured, as the orientation of the fibers in the tissue can significantly affect its behavior. Careful attention to cutting and aligning the samples is necessary to minimize this source of variability, but can never be fully excluded.

- 634 • **Variability of the synthetic materials** - The synthetic samples were  
635 considered as isotropic and homogeneous, although it is known that  
636 the manufacturing process may introduce a flow-related anisotropy and  
637 other imperfections. Hence, even for these materials, it is important  
638 to differentiate methodological variability from production-related vari-  
639 ability.
- 640 • **Participant preferences** - Certain aspects of the testing protocol  
641 were left to the discretion of the participants, such as the orientation  
642 of the testing device (vertical vs horizontal), the type of clamps used  
643 (pneumatic or manual). Narrowing this down would have led to a much  
644 smaller number of participants, which is why this trade-off was made.  
645 Scholze et al. (2020) have proposed an easy-to-fabricate clamping sys-  
646 tem for uniaxial tensile testing, that could be investigated for use in  
647 future testing rounds.
- 648 • **Number of test samples** - 6 samples were tested per condition per  
649 research group, each time coming from two different animals. This  
650 number is based on previous experiments combined with the practical  
651 limitations on the number of aortas that could be collected from the  
652 slaughterhouse. Ideally, a dedicated power analysis of the current re-  
653 sults should be performed to reveal the required number of tests for  
654 future testing rounds for this specific test and sample type.

#### 655 4.6. *Future work*

656 This pilot campaign serves as an exploration of the current landscape,  
657 highlighting the existing variability in the mechanical testing to characterize  
658 biological tissues. As such, it marks only the start of a much needed effort  
659 to further quantify and minimize this variability, and the following areas of  
660 future work have been identified.

- 661 • Further statistical analysis of the data should be performed to help  
662 identify specific sources of variability. Also, even more detailed analysis  
663 of each of the participants' submissions might uncover further devia-  
664 tions from the protocol.
- 665 • Each of the above listed remaining sources of variability accounts for  
666 just a portion of the overall variability. For example, when observing  
667 the degree of anisotropy of the biological tissues tested here, one could

668 argue that the difference between the circumferential and longitudinal  
669 behavior is much smaller than the overall variability, such that sample  
670 misalignment isn't such a big issue. Indeed, the challenge is not only to  
671 identify sources of uncertainty, but also to rank each of these sources in  
672 terms of their relative importance. Once a clear overview and ranking  
673 of these sources of variability is obtained, further fine-tuning of the  
674 consensus protocol can be performed after which a new test round can  
675 be launched.

- 676 • To assess the impact of operator variability, it would be beneficial to  
677 have the same operator perform the test on different set-ups. In other  
678 words, rather than sending samples around, one could exchange re-  
679 searchers. Organising training sessions or workshops for all participants  
680 could help improve consistency in following the testing protocols.
- 681 • The shape and amplitude of the stress-stretch curves of all the syn-  
682 thetic materials deviate strongly from those of the curves for biological  
683 samples. This severely limits the relevance of these synthetic samples  
684 in terms of protocol verification capacity. Indeed, these synthetic sam-  
685 ples do not exhibit the characteristic nonlinear stiffening behaviour,  
686 and this absence of a so-called 'toe-region' significantly reduces the  
687 sensitivity to slight differences in strain. Therefore, for future testing  
688 rounds, we need synthetic materials that do exhibit this nonlinear stiff-  
689 ening behaviour. Possible candidates are PVA hydrogels (Millon et al.,  
690 2006), fabric-reinforced polymers (Zhalmuratova et al., 2019) or melt  
691 electrowritten scaffolds (Mirani et al., 2024). A thorough examination  
692 of these materials is required to ensure maximal resemblance to aortic  
693 tissue with minimal variability due to the production process, storage,  
694 etc.
- 695 • To generalize our findings it is necessary to expand the study to include  
696 other types of tests and tissues. This can help to identify whether the  
697 observed variability is specific to certain tests or tissues, or if it is a  
698 more general issue.

699 This first C<sup>4</sup>Bio campaign serves as a wake-up call that current method-  
700 ological variability is too large to enable reliable comparison of results be-  
701 tween research groups. This undermines the credibility of resulting material

702 properties reported in literature, and their consequent use in *in silico* simu-  
703 lations. This study has made it clear that proper uncertainty quantification  
704 is essential, and that through community effort we should aim to further  
705 increase the quality of our mechanical testing methods and reduce the un-  
706 certainty to a workable level.

## 707 5. Acknowledgements

708 The authors would like to thank the Avicenna Alliance and VPH Institute  
709 for their financial support, as well as KU Leuven (OPSTAMET C2M/24/031  
710 and KERNFAC FIBER) and the Research Foundation - Flanders (I011822N  
711 and G0H1116N). Furthermore, we wish to acknowledge the efforts of the  
712 staff of FIBER, KU Leuven core facility, specifically Adriana Wilmots Helsen,  
713 Markos Kapeliotis and Wouter Willekens, for procuring, preparing and ship-  
714 ping all samples. Next, we would like to express our gratitude to Eddy  
715 Smets and Thomas Pilkington from the Fablab of Mechanical Engineering,  
716 KU Leuven for their assistance in the development of the cutting tool used  
717 in test round 2. Finally, we would also like to thank the following researchers  
718 for their contribution to the conception of C<sup>4</sup>Bio, to the mechanical testing  
719 and/or to the data-analysis:

- 720 • Dr.Ir. Klaas Vander Linden, KU Leuven
- 721 • Dr. Kris Bogaerts, KU Leuven
- 722 • Ir. Thibault Vervenne, KU Leuven
- 723 • Ir. Matijs Loeckx, KU Leuven
- 724 • Dr. Alex Caulk, Medtronic
- 725 • Dr. Markus Reiterer, Medtronic
- 726 • Dr. Venkateswaran Perumal, Stryker
- 727 • Dr. Raja Jayendiran, Stryker
- 728 • Dr. Naveen Kumar Subramani, Stryker
- 729 • Dr. Martin Tanaka, University of Western Carolina
- 730 • Dr. Connor Gains, Queen Mary University of London

## 731 Appendix A. Supplementary material

732 All the supplementary material can be found in a repository on Zenodo  
733 through the following link.

734 <https://zenodo.org/uploads/14179432?token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjRiMzM5MGlyLTNlYTItNGY4NS04ZTliLWZhMjA2ZTMzN2QyNiIsImRhdGEiOnt9LCJyYW5kb20iOiJjMGZmZDUyZWVlZDIzYzc3MTYwNjAyMmRlNjZhYzAwNyJ9.PG4xkyAfg2JTQuDuZfFh341z2F2nWUedUd9D33ey0aP1AD4s4XTtyyJpmtbftX14oKjP4kSFhu8GZC4ZiJhRfA>  
735  
736  
737  
738

739 *Note to reviewers: This link will be shortened in the final submission,*  
740 *once the repository is published.*

## 741 References

742 ASME (2018). *Assessing credibility of computational modeling through ver-*  
743 *ification and validation: application to medical devices: ASME V&V 40 -*  
744 *2018*. New York: The American Society of Mechanical Engineers.

745 Åstrand, H., Stalhand, J., Karlsson, J., Karlsson, M., Sonesson, B., & Länne,  
746 T. (2011). In vivo estimation of the contribution of elastin and collagen  
747 to the mechanical properties in the human abdominal aorta: effect of age  
748 and sex. *Journal of applied physiology*, 110, 176–187. doi:10.1152/japp  
749 lphysiol.00579.2010.

750 Blaker, C. L., Ashton, D. M., Hartnell, N., Little, C. B., & Clarke, E. C.  
751 (2024). Tendon biomechanical properties are altered by storage duration  
752 but not freeze-thaw temperatures or cycles. *Journal of Orthopaedic Re-*  
753 *search®*, 42, 1180–1189. doi:10.1002/jor.25783.

754 C<sup>4</sup>Bio (2025). C<sup>4</sup>bio — community challenge towards consensus on charac-  
755 terization of biological tissue. URL: <https://c4bio.eu/>.

756 Chow, M.-J., & Zhang, Y. (2011). Changes in the mechanical and biochemical  
757 properties of aortic tissue due to cold storage. *Journal of Surgical Research*,  
758 171, 434–442. doi:10.1016/j.jss.2010.04.007.

759 FDA (2023). Assessing the Credibility of Computational Modeling and Sim-  
760 ulation in Medical Device Submissions. FDA guidance.

761 Fehervary, H. (2018). Planar biaxial testing of soft biological tissues, .

- 762 Hammer, N., & Klima, S. (2019). In-silico pelvis and sacroiliac joint mo-  
 763 tion—a review on published research using numerical analyses. *Clinical*  
 764 *biomechanics*, *61*, 95–104. doi:10.1016/j.clinbiomech.2018.12.005.
- 765 Kim, D.-H., Song, C.-G., Ilev, I. K., & Kang, J. U. (2011). Axial-scanning  
 766 low-coherence interferometer method for noncontact thickness measure-  
 767 ment of biological samples. *Applied Optics*, *50*, 970–974. doi:10.1364/AO  
 768 .50.000970.
- 769 Kirilova-Doneva, M., & Pashkouleva, D. (2022). The effects of age and sex  
 770 on the elastic mechanical properties of human abdominal fascia. *Clinical*  
 771 *Biomechanics*, *92*, 105591. doi:10.1016/j.clinbiomech.2022.105591.
- 772 Krueger, N., Luebberding, S., Oltmer, M., Streker, M., & Kerscher, M.  
 773 (2011). Age-related changes in skin mechanical properties: a quantita-  
 774 tive evaluation of 120 female subjects. *Skin research and technology*, *17*,  
 775 141–148. doi:10.1111/j.1600-0846.2010.00486.x.
- 776 Kurz, S., Gebhardt, M., Grundmann, F., Heyde, C.-E., & Steinke, H. (2023).  
 777 Approach to standardized material characterization of the human lum-  
 778 bopelvic system—specification, preparation and storage. *PLOS ONE*, *18*,  
 779 e0289482. doi:10.1371/journal.pone.0289482.
- 780 Lin, A. C., Pirrung, F., Niestrawska, J. A., Ondruschka, B., Pinter, G.,  
 781 Henyš, P., & Hammer, N. (2024). Shape or size matters? towards stan-  
 782 dard reporting of tensile testing parameters for human soft tissues: sys-  
 783 tematic review and finite element analysis. *Frontiers in bioengineering and*  
 784 *biotechnology*, *12*, 1368383. doi:10.3389/fbioe.2024.1368383.
- 785 Lionello, G., Sirieix, C., & Baleani, M. (2014). An effective procedure to cre-  
 786 ate a speckle pattern on biological soft tissue for digital image correlation  
 787 measurements. *Journal of the mechanical behavior of biomedical materials*,  
 788 *39*, 1–8. doi:10.1016/j.jmbbm.2014.07.007.
- 789 Millon, L. E., Mohammadi, H., & Wan, W. K. (2006). Anisotropic polyvinyl  
 790 alcohol hydrogel for cardiovascular applications. *Journal of Biomedical*  
 791 *Materials Research - Part B Applied Biomaterials*, *79*, 305–311. doi:10.1  
 792 002/jbm.b.30543.
- 793 Mirani, B., Mathew, S. O., Latifi, N., Labrosse, M. R., Amsden, B. G., &  
 794 Simmons, C. A. (2024). Programmable melt electrowriting to engineer soft

connective tissues with prescribed, biomimetic, biaxial mechanical properties. *Advanced Functional Materials*, *34*. doi:10.1002/adfm.202302786.

Mirzaali, M. J., Schwiedrzik, J. J., Thaiwichai, S., Best, J. P., Michler, J., Zysset, P. K., & Wolfram, U. (2016). Mechanical properties of cortical bone and their relationships with age, gender, composition and microindentation properties in the elderly. *Bone*, *93*, 196–211. doi:10.1016/j.bone.2015.11.018.

Motiwale, S., & Sacks, M. S. (2025). Structural constitutive models for soft biological tissues and biomaterials: the role of mechanical interactions. *Mechanics of Soft Materials*, *7*, 1.

Oefner, C., Herrmann, S., Kebbach, M., Lange, H.-E., Kluess, D., & Woiczinski, M. (2021). Reporting checklist for verification and validation of finite element analysis in orthopedic and trauma biomechanics. *Medical Engineering & Physics*, *92*, 25–32. doi:10.1016/j.medengphy.2021.03.011.

Oswald, I., Rickert, M., Brüggemann, G.-P., Niehoff, A., Ulloa, C. A. F., & Jahnke, A. (2017). The influence of cryopreservation and quick-freezing on the mechanical properties of tendons. *Journal of biomechanics*, *64*, 226–230. doi:10.1016/j.jbiomech.2017.08.018.

O’Leary, S. A., Doyle, B. J., & McGloughlin, T. M. (2013). Comparison of methods used to measure the thickness of soft tissues and their influence on the evaluation of tensile stress. *Journal of biomechanics*, *46*, 1955–1960. doi:10.1016/j.jbiomech.2013.05.003.

Pappalardo, F., Wilkinson, J., Busquet, F., Bril, A., Palmer, M., Walker, B., Curreli, C., Russo, G., Marchal, T., Toschi, E., Alessandrello, R., Costignola, V., Klingmann, I., Contin, M., Staumont, B., Woiczinski, M., Kaddick, C., Salvatore, V. D., Aldieri, A., Geris, L., & Viceconti, M. (2022). Toward A Regulatory Pathway for the Use of in Silico Trials in the CE Marking of Medical Devices. *IEEE Journal of Biomedical and Health Informatics*, *26*, 5282–5286. doi:10.1109/JBHI.2022.3198145.

Reu, P. L., Blaysat, B., Andó, E., Bhattacharya, K., Couture, C., Couty, V., Deb, D., Fayad, S. S., Iadicola, M. A., Jaminion, S., Klein, M., Landauer, A. K., Lava, P., Liu, M., Luan, L. K., Olufsen, S. N., Réthoré, J., Roubin,

- 827 E., Seidl, D. T., Siebert, T., Stamati, O., Toussaint, E., Turner, D., Vem-  
828 ulapati, C. S. R., Weikert, T., Witz, J. F., Witzel, O., & Yang, J. (2022).  
829 Dic challenge 2.0: Developing images and guidelines for evaluating accu-  
830 racy and resolution of 2d analyses. *Experimental Mechanics*, *62*, 639–654.  
831 doi:10.1007/s11340-021-00806-6.
- 832 Reu, P. L., Toussaint, E., Jones, E., Bruck, H. A., Iadicola, M., Balcaen, R.,  
833 Turner, D. Z., Siebert, T., Lava, P., & Simonsen, M. (2018). Dic challenge:  
834 Developing images and guidelines for evaluating accuracy and resolution  
835 of 2d analyses. *Experimental Mechanics*, *58*, 1067–1099. doi:10.1007/s1  
836 1340-017-0349-0.
- 837 Robertson, D., & Williams, G. H. (2009). *Clinical and translational science:  
838 principles of human research*. Academic Press.
- 839 Ryu, D., Baek, S., & Kim, J. (2022). Region-dependent mechanical char-  
840 acterization of porcine thoracic aorta with a one-to-many correspondence  
841 method to create virtual datasets using uniaxial tensile tests. *Frontiers in  
842 Bioengineering and Biotechnology*, *10*, 937326.
- 843 Schickel, M., Townsend, K., & Farotto, D. (2019). Material Characterization  
844 of 3D Printed HeartPrint Flex Plus Models and Comparison with Arterial  
845 Tissue Properties. URL: [https://www.materialise.com/en/inspiration/](https://www.materialise.com/en/inspiration/whitepapers/material-characterization-heartprint-models)  
846 [whitepapers/material-characterization-heartprint-models](https://www.materialise.com/en/inspiration/whitepapers/material-characterization-heartprint-models).
- 847 Scholze, M., Safavi, S., Li, K. C., Ondruschka, B., Werner, M., Zwirner, J.,  
848 & Hammer, N. (2020). Standardized tensile testing of soft tissue using a  
849 3d printed clamping system. *HardwareX*, *8*, e00159. doi:10.1016/j.ohx.  
850 2020.e00159.
- 851 Schwarz, D., Fleisman, J., Vitasek, R., & Polzer, S. (2023). Methodology for  
852 estimation of undeformed thickness of arterial tissues. *Scientific Reports*,  
853 *13*, 2816. doi:10.1038/s41598-023-28871-y.
- 854 Shahbad, R., Kazim, M., Razian, S. A., Desyatova, A., & Jadidi, M. (2025).  
855 Variations in stiffness and structure of the human aorta along its length.  
856 *Scientific Reports*, *15*, 11120.
- 857 Smoljkić, M., Vander Sloten, J., Segers, P., & Famaey, N. (2023). In vivo  
858 material properties of human common carotid arteries: Trends and sex

859 differences. *Cardiovascular Engineering and Technology*, 14, 840–852.  
860 doi:10.1007/s13239-023-00691-1.

861 Stemper, B. D., Yoganandan, N., Stineman, M. R., Gennarelli, T. A., Bais-  
862 den, J. L., & Pintar, F. A. (2007). Mechanics of fresh, refrigerated,  
863 and frozen arterial tissue. *Journal of Surgical Research*, 139, 236–242.  
864 doi:10.1016/j.jss.2006.09.001.

865 Stevens, W. (2003). Good clinical laboratory practice (gclp): the need for  
866 a hybrid of good laboratory practice and good clinical practice guide-  
867 lines/standards for medical testing laboratories conducting clinical trials  
868 in developing countries. *Quality Assurance*, 10, 83–89.

869 Vaishnav, R. N., & Vossoughi, J. (1983). Estimation of residual strains in  
870 aortic segments. In *Biomedical Engineering II* (pp. 330–333). Elsevier.  
871 doi:10.1016/B978-0-08-030145-7.50078-7.

872 Viceconti, M., & Emili, L. (2024). *toward Good Simulation Practice: Best*  
873 *Practices for the Use of Computational Modelling and Simulation in the*  
874 *Regulatory Process of Biomedical Products*. Springer Nature.

875 Viceconti, M., Emili, L., Afshari, P., Courcelles, E., Curreli, C., Famaey, N.,  
876 Geris, L., Horner, M., Jori, M. C., Kulesza, A., Loewe, A., Neidlin, M.,  
877 Reiterer, M., Rousseau, C. F., Russo, G., Sonntag, S. J., Voisin, E. M.,  
878 & Pappalardo, F. (2021). Possible Contexts of Use for *In Silico* Trials  
879 Methodologies: A Consensus-Based Review. *IEEE Journal of Biomedical*  
880 *and Health Informatics*, 25, 3977–3982. doi:10.1109/JBHI.2021.3090469.

881 Wale, M. E., Nesbitt, D. Q., Henderson, B. S., Fitzpatrick, C. K., Creechley,  
882 J. J., & Lujan, T. J. (2021). Applying astm standards to tensile tests of  
883 musculoskeletal soft tissue: Methods to reduce grip failures and promote  
884 reproducibility. *Journal of Biomechanical Engineering*, 143. doi:10.1111  
885 5/1.4048646.

886 Zhalmuratova, D., La, T. G., Yu, K. T. T., Szojka, A. R., Andrews, S. H.,  
887 Adesida, A. B., Kim, C. I., Nobes, D. S., Freed, D. H., & Chung, H. J.  
888 (2019). Mimicking "j-shaped" and anisotropic stress-strain behavior of  
889 human and porcine aorta by fabric-reinforced elastomer composites. *ACS*  
890 *Applied Materials and Interfaces*, 11, 33323–33335. doi:10.1021/acsami  
891 .9b10524.

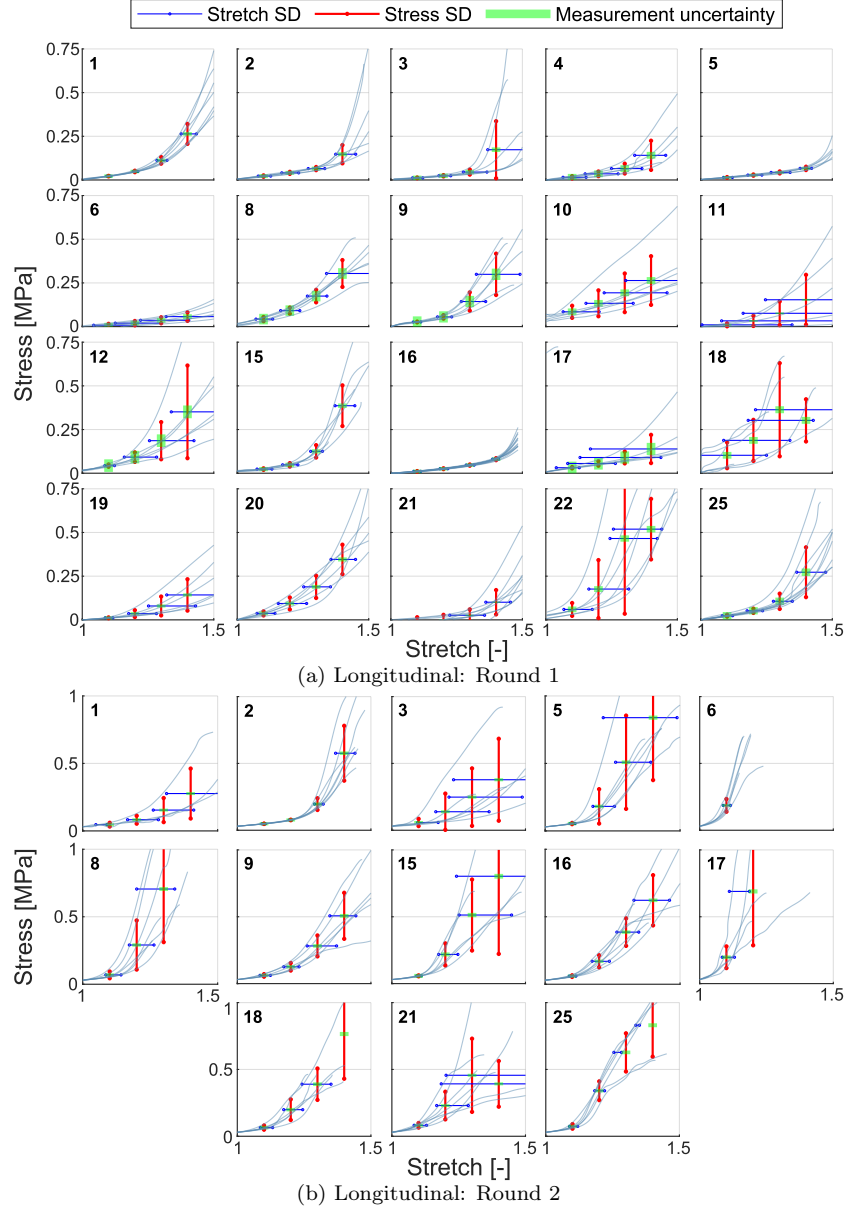


Figure 4: Engineering stress (in MPa) - stretch curves of longitudinal specimen types obtained in (a) round 1 and (b) round 2, grouped per participating research group indicated in the top left corner. Vertical lines represent the standard deviation (SD) of stress values at stretch levels of 1.1, 1.2, 1.3 and 1.4. Horizontal lines indicate the SD of stretch values at the mean stress value corresponding to these stretch levels. Measurement uncertainty on the mean stress value is shown at each stretch level, accounting for a 0.1 N load cell uncertainty and 0.02 mm uncertainty in specimen thickness and width measurements.

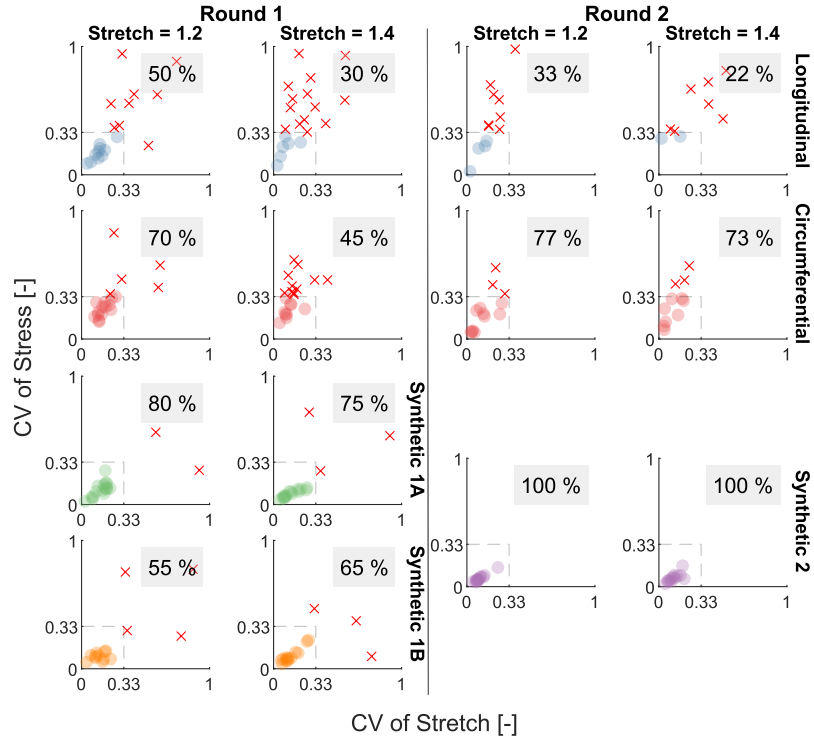


Figure 5: Coefficient of variation (CV) of engineering stress versus CV of stretch at two specific stretch levels, 1.2 and 1.4. CV values are calculated per research group. Research groups for which both CV values fall within the threshold of 0.33 are represented by dots, while those with either CV value exceeding this threshold are marked with ‘x’ markers. The percentage of research groups whose CV values remain within the threshold is indicated on each plot. The data are organised by specimen type and separated by testing round.

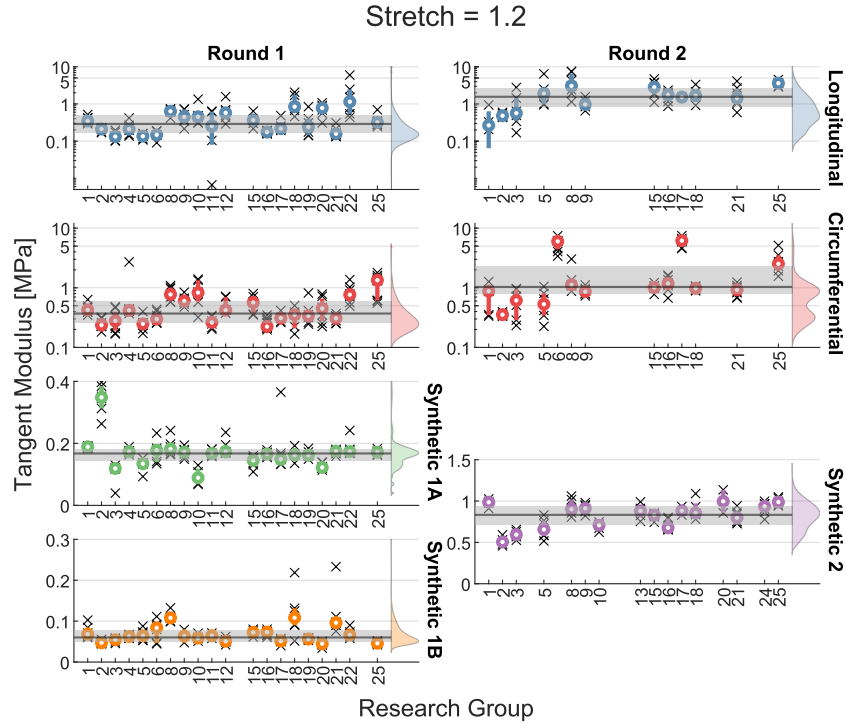


Figure 6: Tangent modulus of the stress-stretch curve at 1.2 stretch per research group, organised by testing round and specimen type. Individual moduli are represented by ‘x’ markers, while the mean for each research group is indicated by a dot and first- to third quantile (Q1-Q3) values are shown as bars. The overall mean and Q1-Q3 values, aggregated across all research groups, are depicted by a line and shaded band, respectively. A density plot is included to represent the distribution of all moduli for each specimen type and testing round, providing insights into the data distribution and variability.

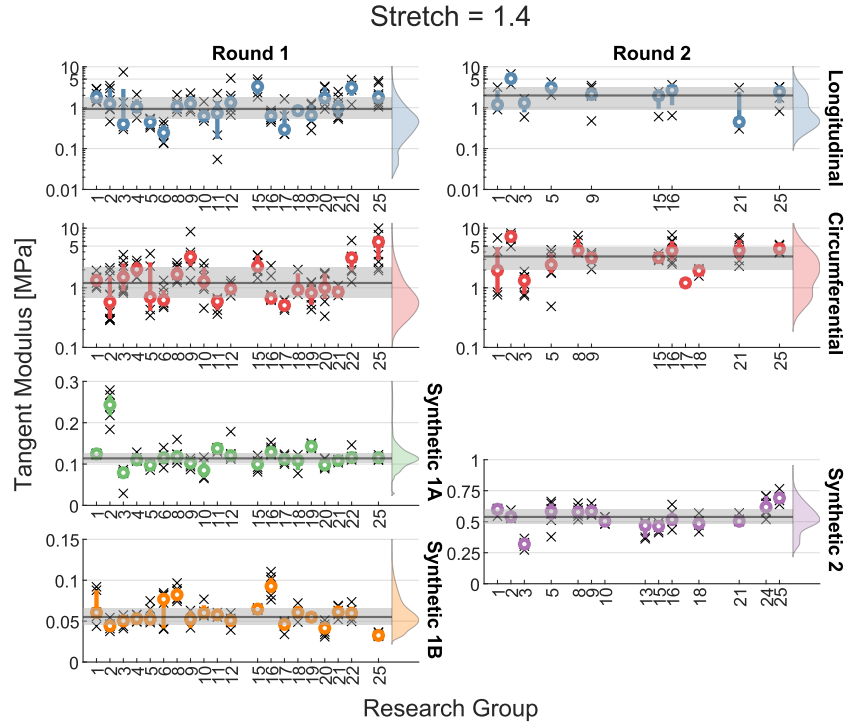


Figure 7: Tangent modulus of the stress-stretch curve at 1.4 stretch per research group, organised by testing round and specimen-type. Individual moduli are represented by ‘x’ markers, while the mean for each research group is indicated by a dot and first-to third quantile (Q1-Q3) values are shown as bars. The overall mean and Q1-Q3 values, aggregated across all research groups, are depicted by a line and shaded band, respectively. A density plot is included to represent the distribution of all moduli for each specimen type and testing round, providing insights into the data distribution and variability.

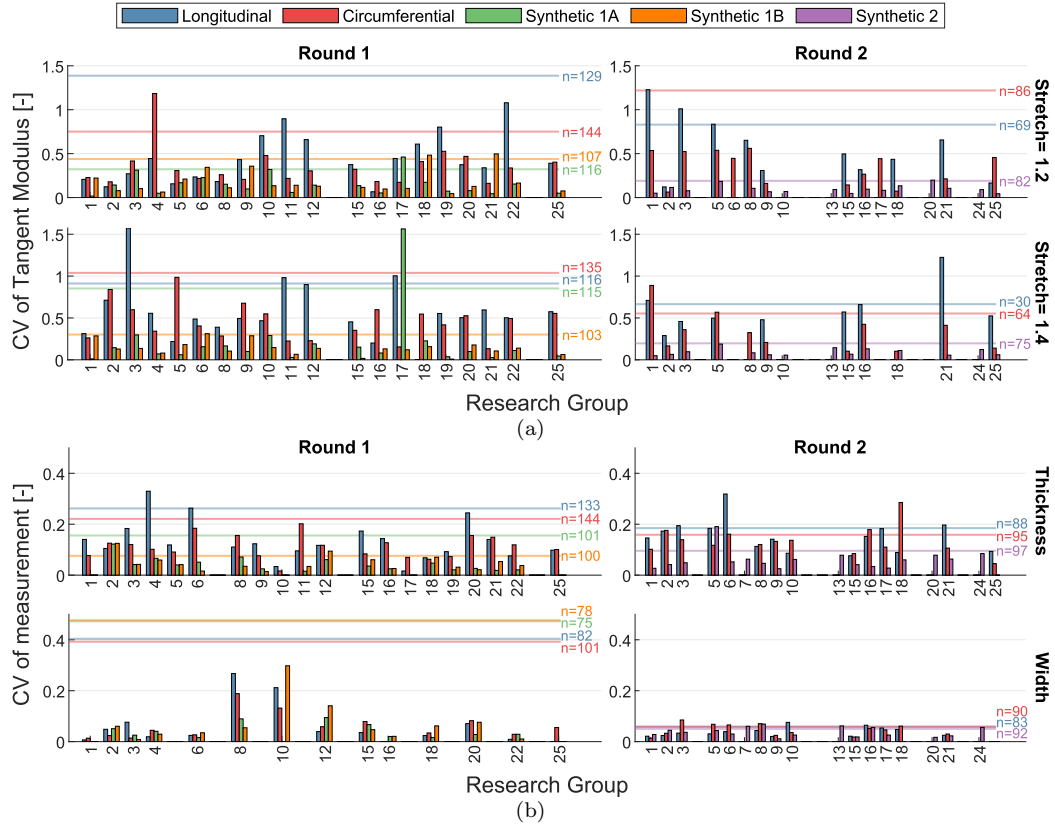


Figure 8: Bar plots showing the coefficient of variation (CV) of (a) the tangent modulus of the stress-stretch curve at 1.2 and 1.4 stretch values, and (b) thickness and width measurements. CV values are represented for each research group across different specimen types, separated by testing round. The overall CV value, calculated by aggregating measurements across all research groups for each specimen type, is represented by a line. The corresponding sample size is annotated adjacent to the line.

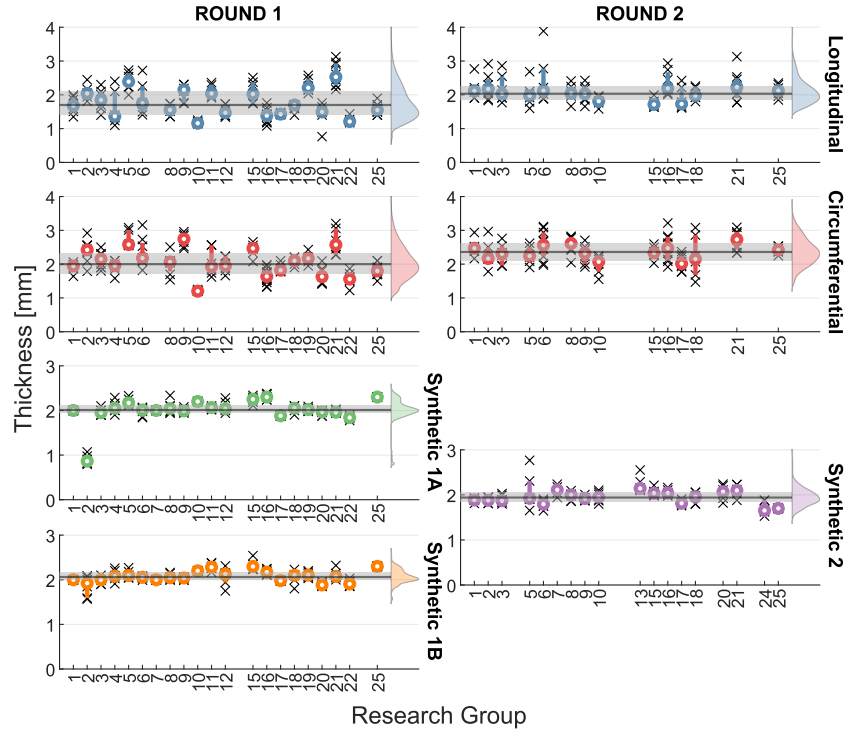


Figure 9: Specimen thickness measurements per research group, organised by testing round and specimen type. Individual measurements are represented by ‘x’ markers, while the mean for each research group is indicated by a dot and first- to third quantile (Q1-Q3) values are shown as bars. The overall mean and Q1–Q3 values, aggregated across all research groups, are depicted by a line and shaded band, respectively. A density plot is included to represent the distribution of all measurements for each specimen type and testing round, providing insights into the data distribution and variability.

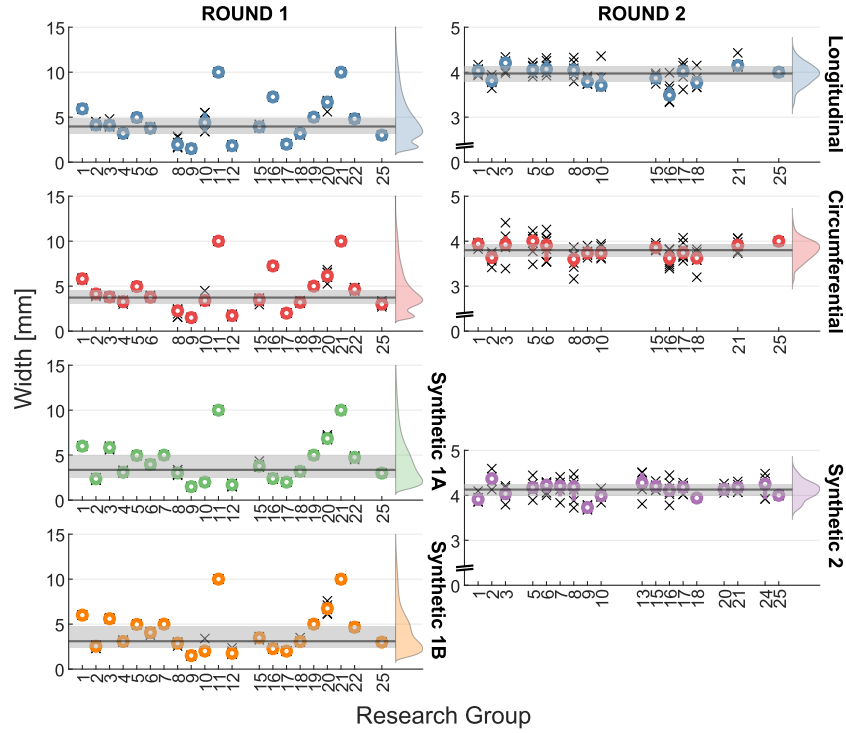


Figure 10: Specimen width measurements per research group, organised by testing round and specimen type. Individual measurements are represented by ‘x’ markers, while the mean for each research group is indicated by a dot and first- to third quantile (Q1-Q3) values are shown as bars. The overall mean and Q1-Q3 values, aggregated across all research groups, are depicted by a line and shaded band, respectively. A density plot is included to represent the distribution of all measurements for each specimen type and testing round, providing insights into the data distribution and variability.