Credibility Assessment Case Studies Panel Discussion

Panel Members

Paul Briant
Exponent

Mark Driscoll
McGill University

Finn Donaldson
FDA

Prasanna Hariharan
FDA

Brandon Lurie
WL Gore

Semih Oktay
CardioMed Device Consultants

#1: MITRAL VALVE DEVICE

#1: Mitral Valve Device – Introduction

- Mitral valve repair device analyzed for fatigue assessment for PMA submission.

- **QOI:** Is the device resistant to fatigue fracture under conservative *in vivo* loads?

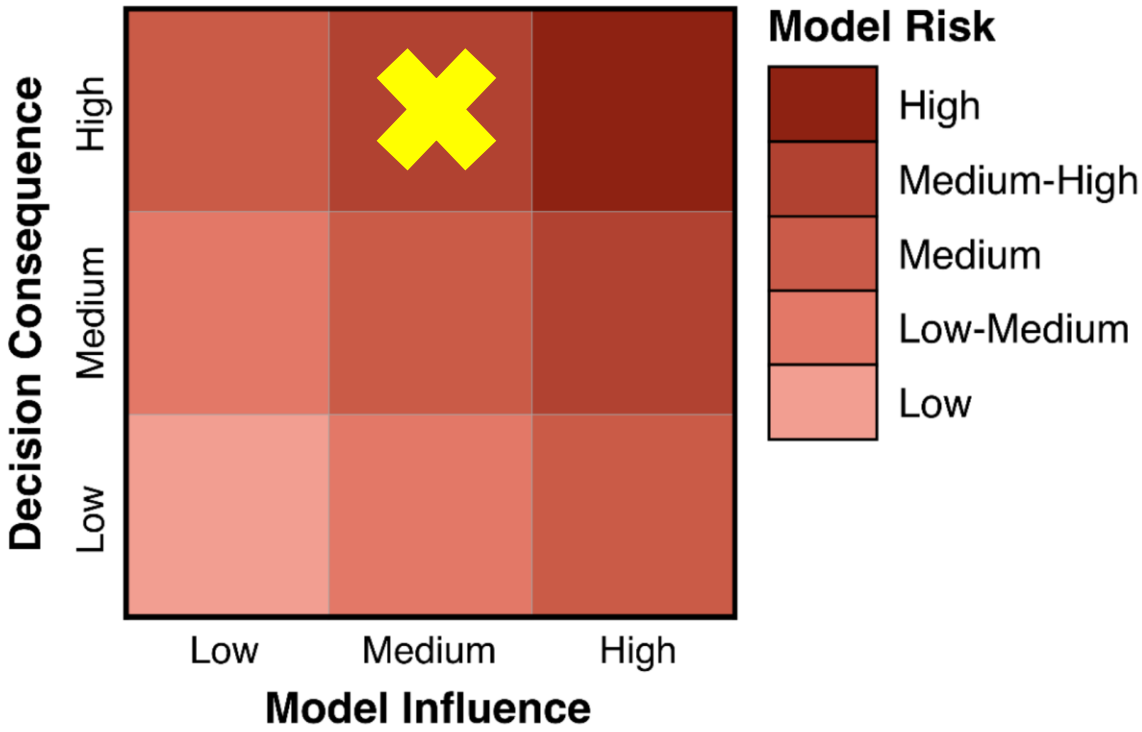
- **COU:** The model will be used to determine the fatigue safety factor of the frame under conservative *in vivo* loads based on literature. The results will be combined with experimental pulsatile fatigue testing under hypertensive conditions and annular loading based on literature.

Example Mitral Valve Replacement Device Shown

#1: Mitral Valve Device – Model Risk

- **Model Influence:** The model influence is determined to be **medium** since the FSF information will be combined with experimental results and the experimental conditions are independently justified.

- **Decision Consequence:** An incorrect decision regarding fatigue safety could result in one or more strut fractures. The clinical sequalae from a strut fracture can range from inconsequential to significant patient harm including death, depending on number of fractures and patient specifics. Decision consequence is therefore determined to be **high**.

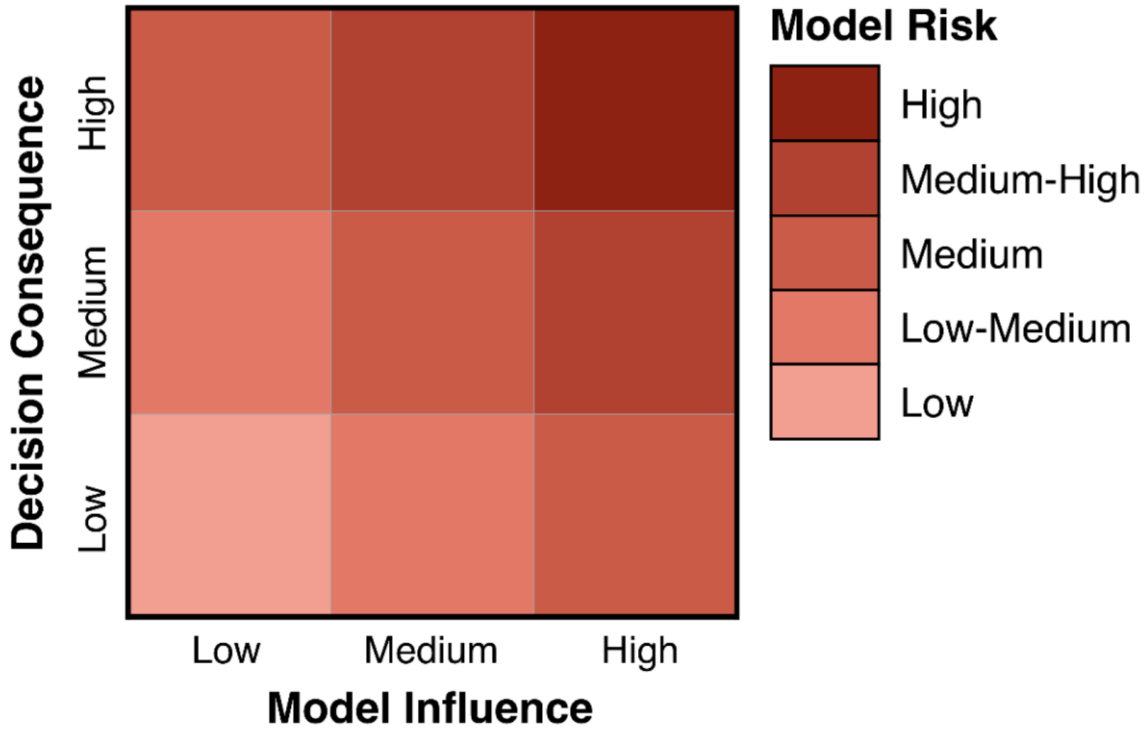


#1: Mitral Valve Device – Model Risk

- This case study used a 3x3 model risk table with 5 levels of model risk.

- **Are there best practices for gradations of decision consequence, model influence, and model risk?**

- **Is it best practice to keep model risk assessment agnostic to the type model?**
  - **Are there times to deviate from 3x3 with 5 levels of model risk?**



#1: Mitral Valve Device – Model Influence Scenarios

- How would you rank model influence?
- In all cases modeling and experimental device testing combined to evaluate fatigue performance.

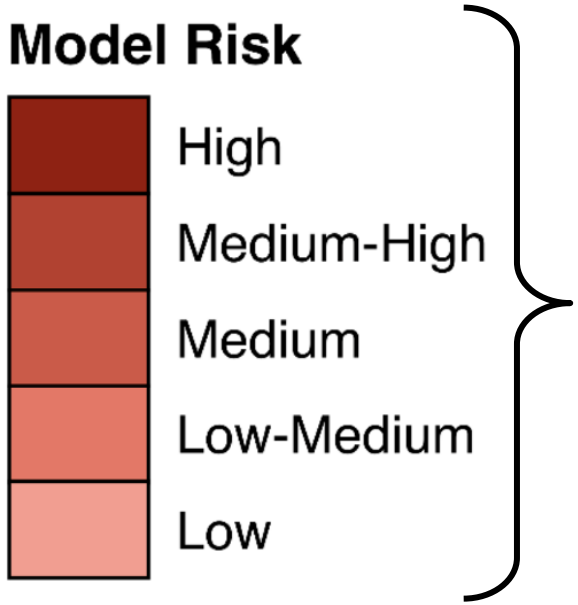
Determine FSF?	Determine Worst Case Size?	Determine Test Diameters?	Model Influence? Low/Med/High
Yes	No	No	?

#1: Mitral Valve Device – Credibility Goals

- In vitro comparator

- Force-displacement testing under four sizing scenarios performed
  - Device sizing adjusted through component positioning

- Goal: force-displacement stiffness within 10% of experimental average for all sizing scenarios tested



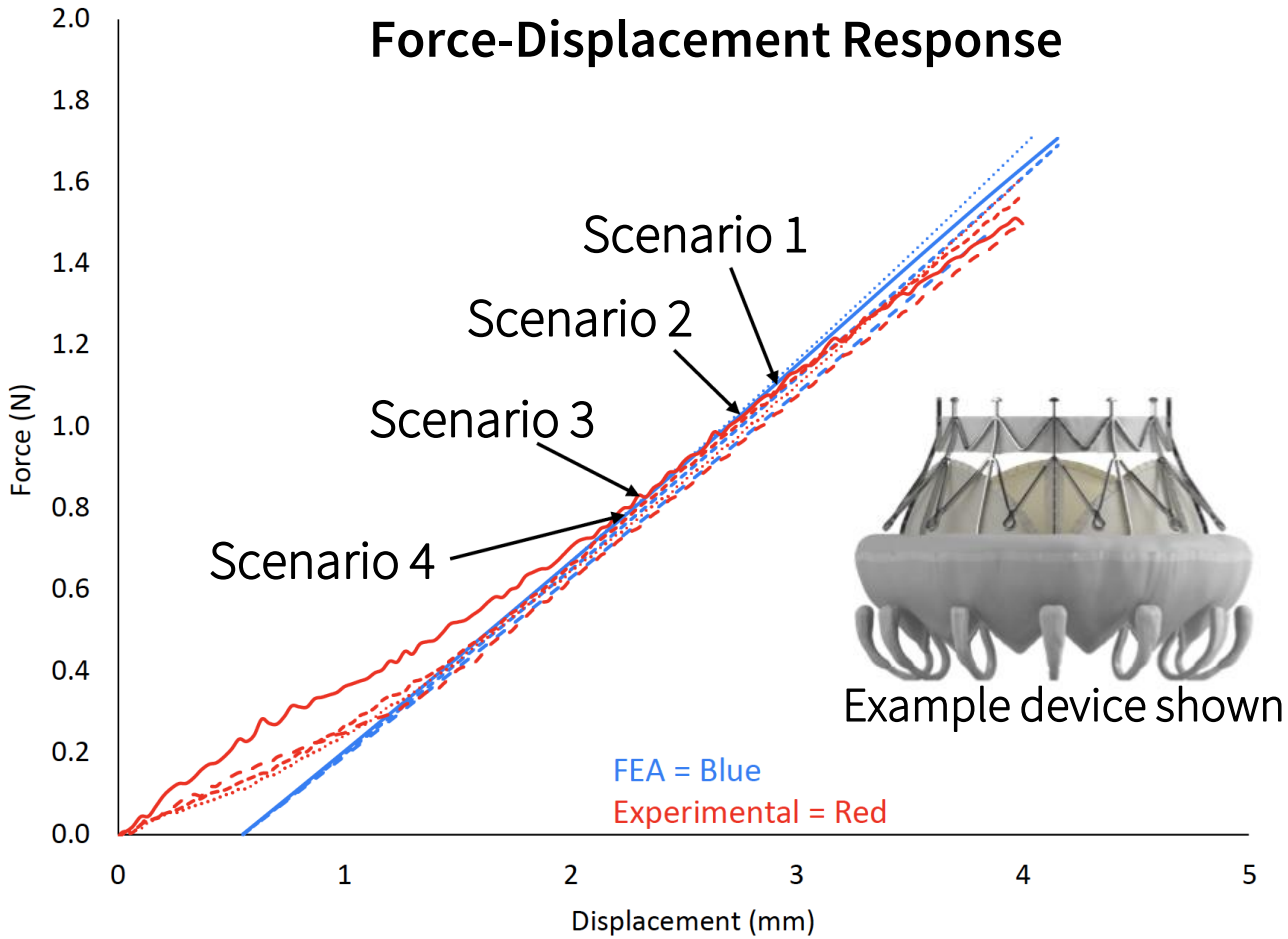
For quantifiable comparators, are there best practices or guidance on desired agreement for each risk level?

How are credibility goals established and what evidence do you cite to justify them?

Keep constant across model types?

#1: Mitral Valve Device – Credibility Goals

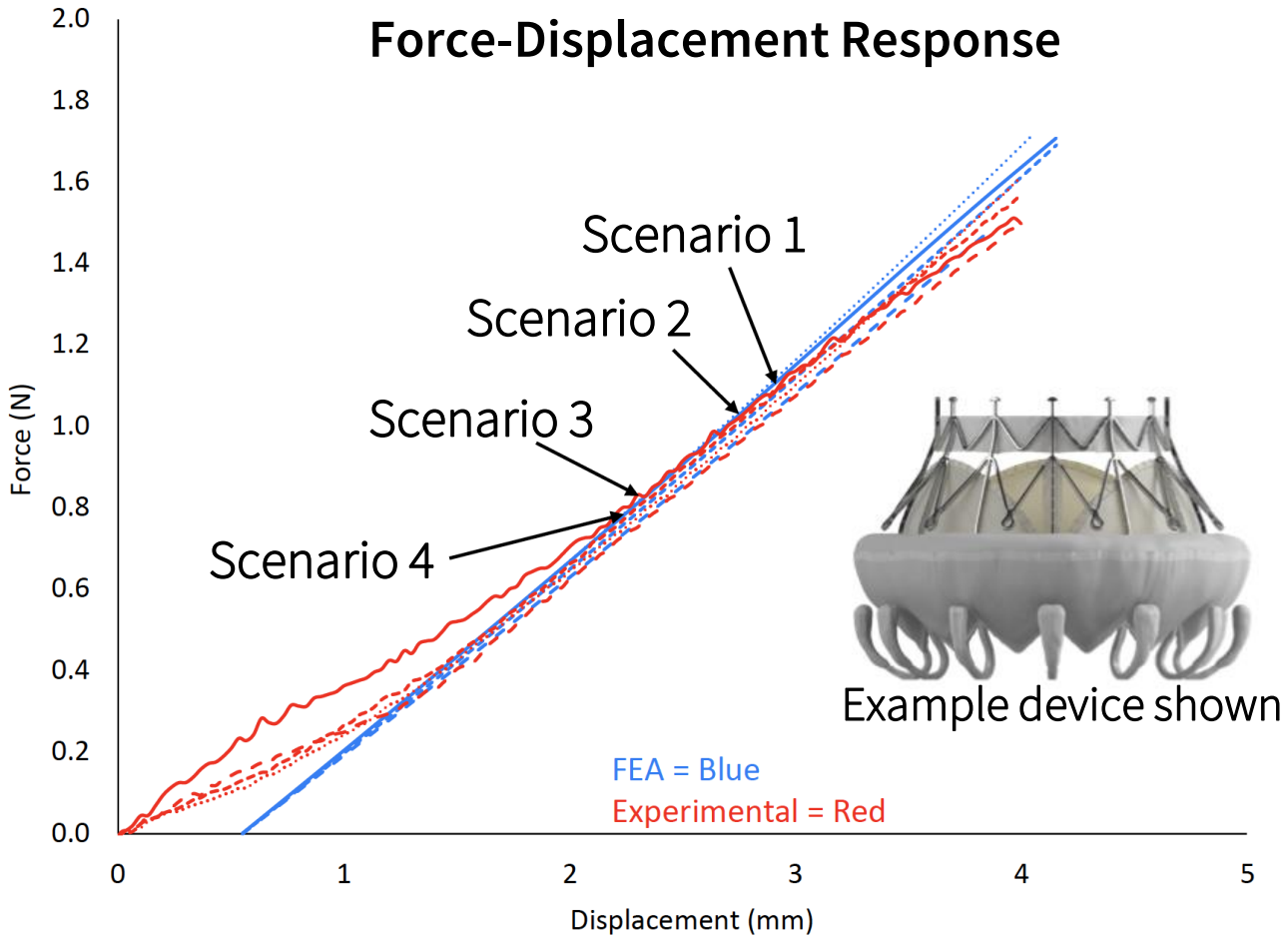
- Device coupon difficult to grab on to.
- Fixture grips originally modeled as rigid, which resulted in high forces in the model (started simple).
- Based on reviewing video of device motion during pull, determined that fixture stiffness not negligible.
- Desired agreement achieved, after tuning fixture constraints.
  - Same fixture modification worked for all four scenarios.



#1: Mitral Valve Device – Credibility Goals

- **Are we good?**
  - Somewhere between calibration and validation.
  - Device model was not modified.

- **Are there best practices for in vitro test simplicity and applicability?**
  - Role of friction – if friction adjusted within reasonable range improves match, do we need a new test?



#1: Mitral Valve Device – Results and Adequacy

- Peak strain amplitude under conservative conditions = 0.25%
- SN testing determined 10th percentile, 95% confidence fatigue strength = 0.35%
- FSF = 1.4

- FDA guidance discusses how the proximity of model predictions to thresholds can factor into assessing model adequacy.
- **Can we also incorporate conservativism in applied loads or threshold criteria into evaluating model adequacy?**

#2: BMD FOR HIP

#2: BMD of Hip - Introduction

- Objective: use patient specific model to assist clinician determine risk for femur fracture

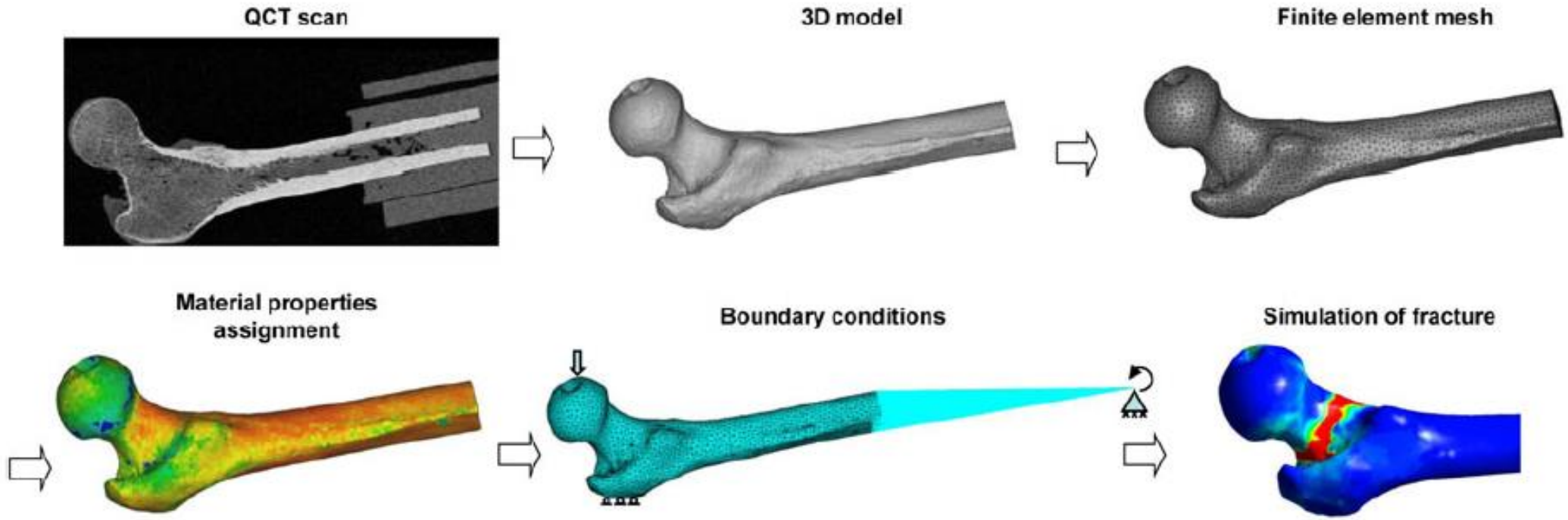
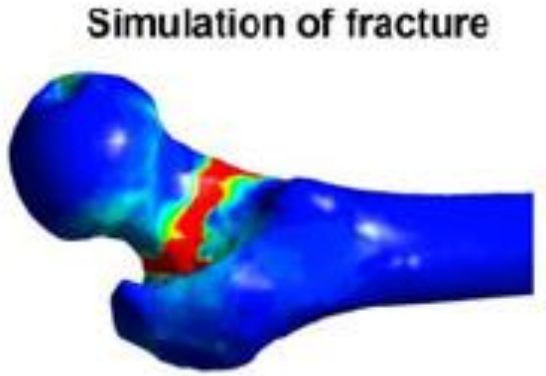


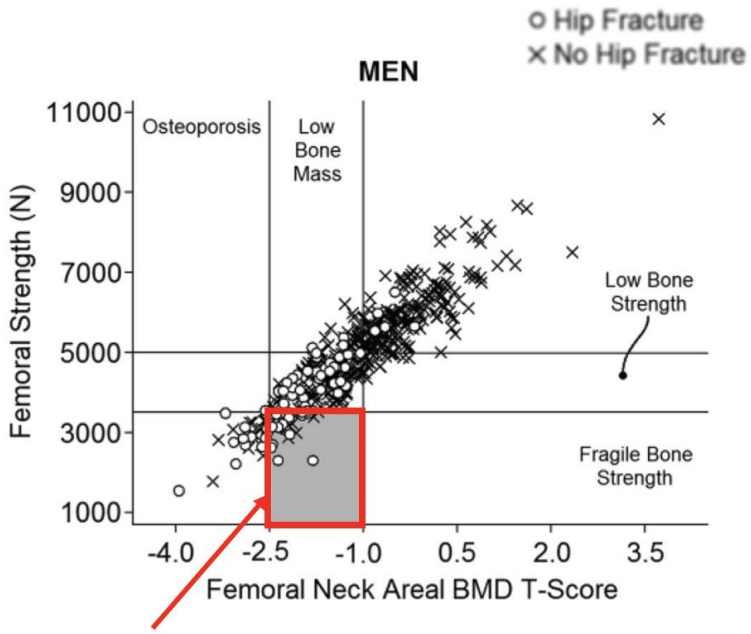
FIGURE 2. Overview of the QCT/FEA modeling. The QCT scan was segmented to obtain a 3D model. This model was meshed with finite elements. Material properties were assigned to the elements based on the HU values. Boundary conditions were then applied and the model was solved to simulate the fracture.

#2: BMD of Hip – Introduction

- **QOI:** Is the patient at an increased femoral fracture risk?

- **COU:**
  - The computational model will use patient-specific data from CT scans to predict bone strength under physiological loading conditions.
  - The predicted bone strength, in addition to other clinical parameters, will be used to categorize the patient's relative fracture risk.
  - This risk assessment can be used by a physician to monitor therapy and/or decide if the patient should be considered for pharmacologic or other treatments to maintain bone health.



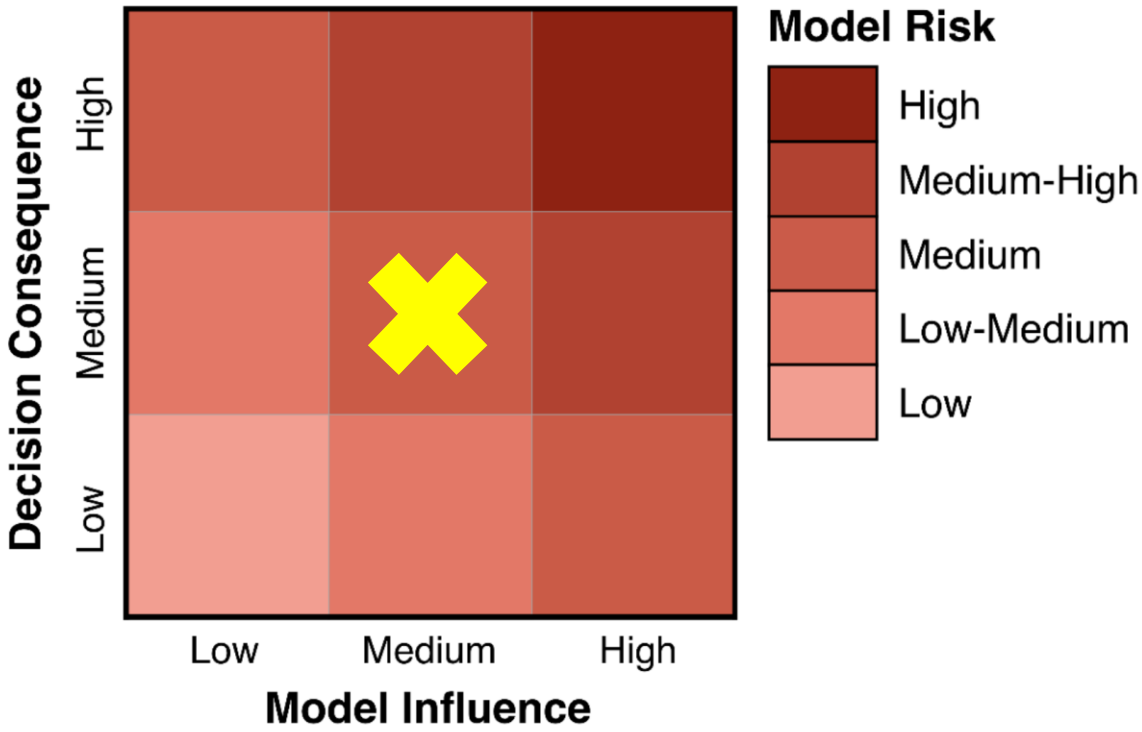


Patients at high risk not detected by aBMD alone but detected by BCT+aBMD.

#2: BMD of Hip – Model Risk

- **Model Influence:** The model influence is determined to be **medium** since the model fracture risk will be combined with other clinical data to make the patient decision.

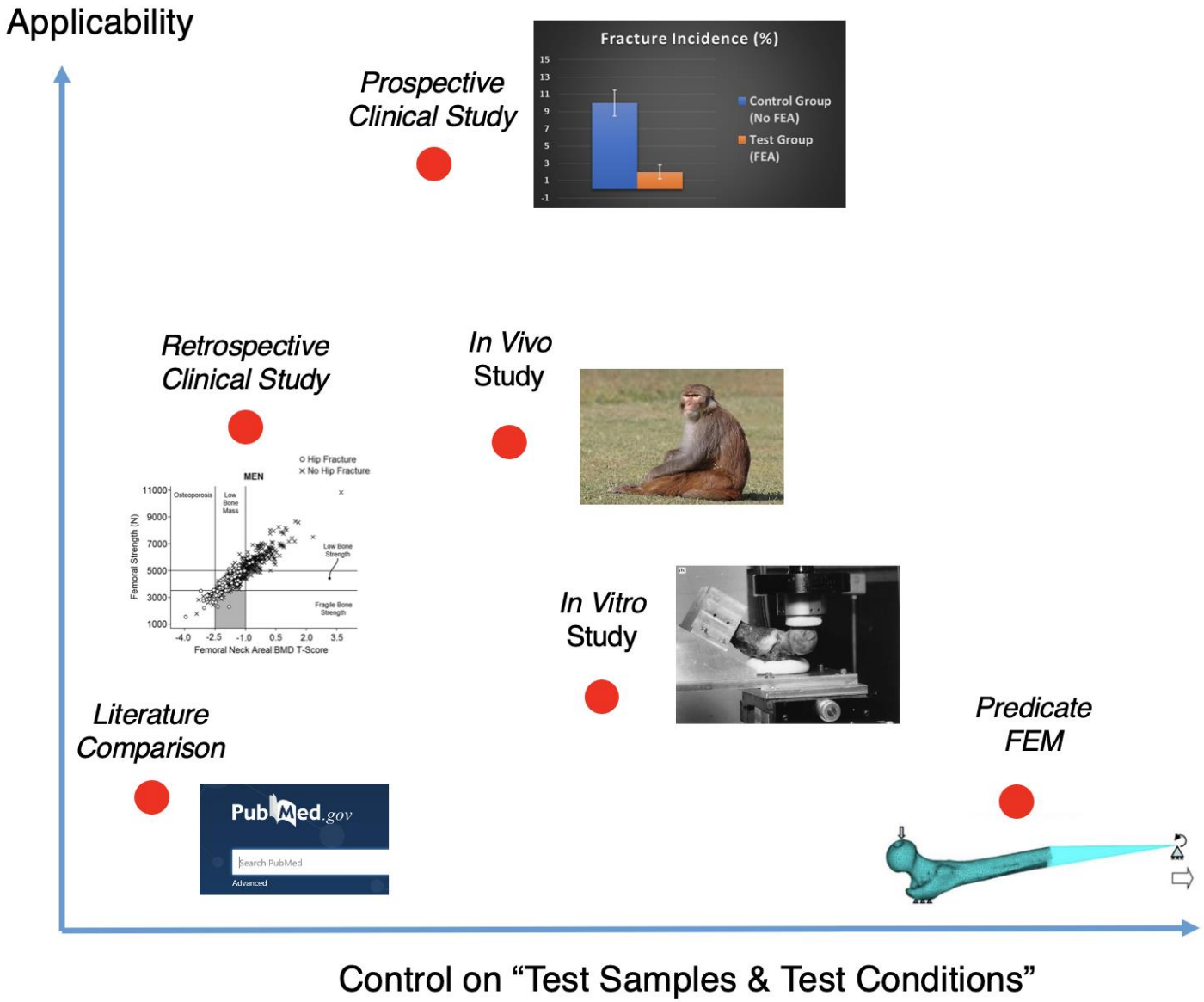
- **Decision Consequence:** An incorrect decision regarding femur fracture could result in multiple clinical sequalae. At a minimum, patient will suffer notable pain. Patient could also injury themselves falling from fracture. Decision consequence is therefore determined to be **medium**.



Do you agree with the risk level in this example?

#2: BMD of Hip – Credibility Goals

- Tiered validation approach with multiple comparators
- Are there best practices for ranking the comparators?
  - In vivo?
  - In vitro?
  - Literature?
  - Published FEM data?
- Can you reduce or increase credibility by adding more comparators?



#2: BMD of Hip – Credibility Goals

- Comparators: common practice in published literature.

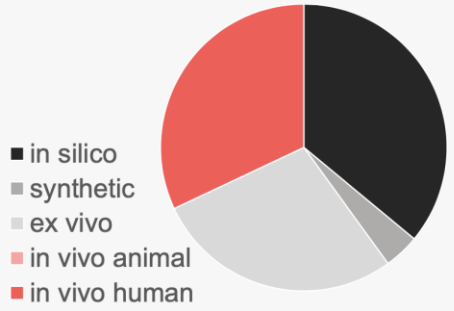


Figure 3: Quantity of each type of comparators used

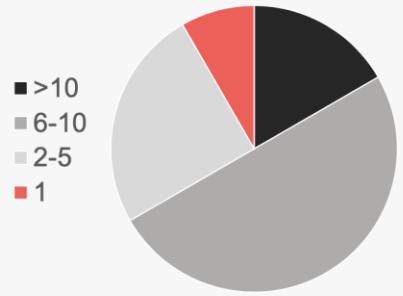


Figure 4: Number of comparators used per model

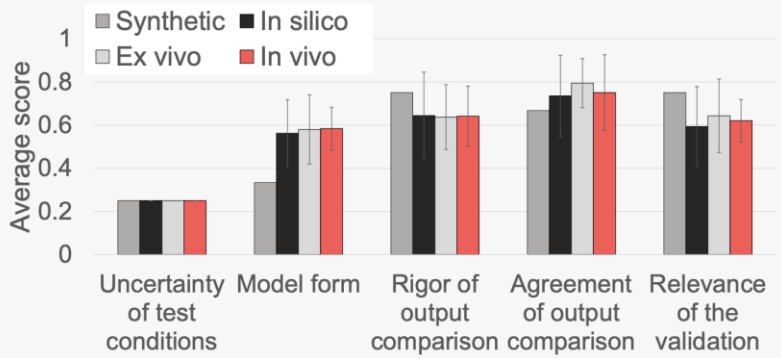


Figure 5: Average scores based on the V&V40 scoring metrics

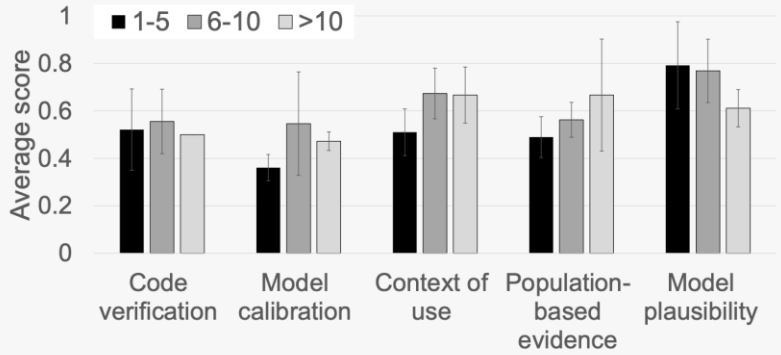


Figure 6: Average scores based on the FDA categories for credibility evidence

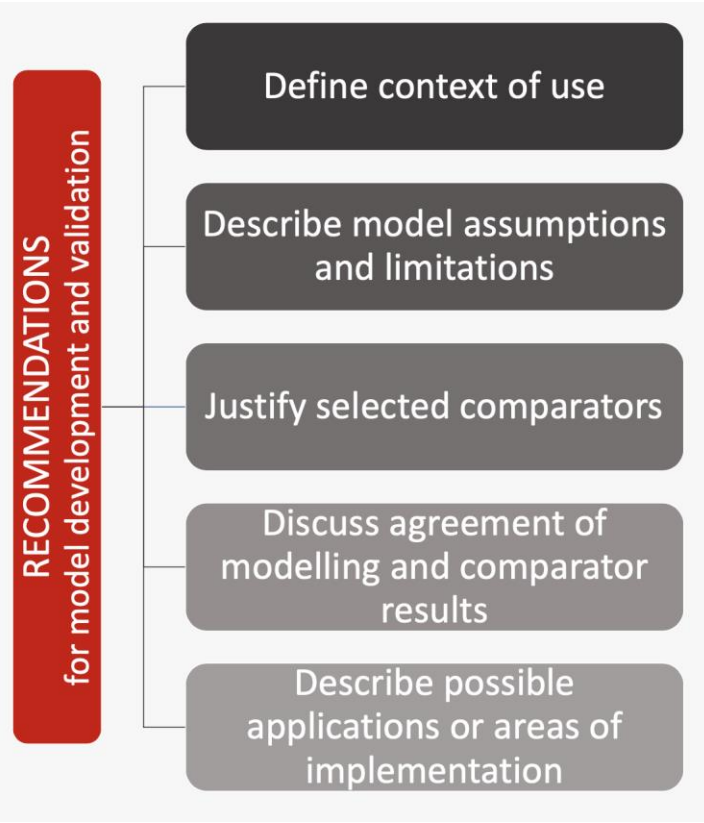
Annals of Biomedical Engineering, Vol. 51, No. 1, January 2023 (© 2022) pp. 150–162
https://doi.org/10.1007/s10439-022-03069-x

BMES BIOMEDICAL ENGINEERING SOCIETY

S.I. : Modeling for Advancing Regulatory Science

A Critical Comparison of Comparators Used to Demonstrate Credibility of Physics-Based Numerical Spine Models

Brittany Stott,[1,2] Payman Afshari,[3] Jeff Bischoff,[4] Julien Clin,[5] Alexandra Francois-Saint-Cyr,[6] Mark Goodin,[7] Sven Herrmann,[8] Xiangui Liu,[9] and Mark Driscoll[1,2]



#2: BMD of Hip

- Three groups:
  - **Sham group** (Sham-Vehicle): 17 monkeys, no ovariectomy, received only a vehicle.
  - **Control group** (OVX-Vehicle): 20 monkeys, ovariectomized, received only a vehicle.
  - **Test group** (OVX-DMAb): 17 monkeys, ovariectomized, received denosumab.

- A post hoc power analysis indicated that it would have required 1814 specimens for that difference to reach a level of statistical significance at the a 0.05% level of confidence.

What resources do you use to assess the probability of an event for a given test?

Finite Element Analysis of Denosumab Treatment Effects on Vertebral Strength in Ovariectomized Cynomolgus Monkeys

David C Lee,[1] Aurore Varela,[2] Paul J Kostenuik,[3] Michael S Ominsky,[4] and Tony M Keaveny[5]

[1]O.N. Diagnostics LLC, Berkeley, CA, USA
[2]Charles River Laboratories Inc., Montréal, QC, Canada
[3]Phylon Pharma Services, Newbury Park, CA, USA
[4]Amgen Inc., Thousand Oaks, CA, USA
[5]University of California, Berkeley, CA, USA

Credibility	Tests Samples: Quantity
C	A single sample was used
B	Multiple samples were used, but not enough to be relevant.
A	A relevant number of samples was used.

Credibility	Clinical Cohort Comparator: Number of Subjects
D	The comparator includes a limited number of subjects (preliminary validation study).
C	The comparator includes a broad range of subjects, but no power-analysis was done before or after the study to ensure this number allows performing statistically significant comparisons.
B	The number of subjects included in the study was fixed à priori using a power analysis and allows theoretically to perform statistically significant comparisons. However, it was not verified à posteriori that this number of subjects was really sufficient to reach statistical significance.
A	It was demonstrated that the number of subjects included in the study is sufficient to reach statistical significance (power analysis).

#2: BMD of Hip

- 9- to 18-year-old female cynomolgus (cynos) monkeys.

Finite Element Analysis of Denosumab Treatment Effects on Vertebral Strength in Ovariectomized Cynomolgus Monkeys

David C Lee,[1] Aurore Varela,[2] Paul J Kostenuik,[3] Michael S Ominsky,[4] and Tony M Keaveny[5]

[1]O.N. Diagnostics LLC, Berkeley, CA, USA
[2]Charles River Laboratories Inc., Montréal, QC, Canada
[3]Phylon Pharma Services, Newbury Park, CA, USA
[4]Amgen Inc., Thousand Oaks, CA, USA
[5]University of California, Berkeley, CA, USA

Credibility	Description
E	One or more samples with a single set of characteristics were included.
D	Samples representing a range of characteristics near nominal were included.
C	Samples representing a large range of characteristics, but not large enough to represent the expected extreme values of the parameters, were included.
B	Samples representing the expected extreme values of the parameters were included.
A	Samples representing the entire range of parameters were included

Range of Characteristics of Test Samples.

If seeking an *in vivo* comparator (pretend budget was infinite), what would you pick and why?

Credibility	Subjects Features Range
C	The subjects included in the comparator represent only a limited range of the key characteristics of the target population as defined in the COU.
B	The subjects included in the comparator represent a large range of the key characteristics of the target population as defined in the COU.
A	The subjects included in the comparator represent the whole range of the key characteristics of the target population as defined in the COU.

#2: BMD of Hip

- Characteristics measured:
  - Age
  - Body Weight
  - Whole body Mineral content (BMC)
  - Lumbar Spine BMD

- Note: The three groups did not present any sign difference for these measured characteristics.

Finite Element Analysis of Denosumab Treatment Effects on Vertebral Strength in Ovariectomized Cynomolgus Monkeys

David C Lee,[1] Aurore Varela,[2] Paul J Kostenuik,[3] Michael S Ominsky,[4] and Tony M Keaveny[5]

[1]O.N. Diagnostics LLC, Berkeley, CA, USA
[2]Charles River Laboratories Inc., Montréal, QC, Canada
[3]Phylon Pharma Services, Newbury Park, CA, USA
[4]Amgen Inc., Thousand Oaks, CA, USA
[5]University of California, Berkeley, CA, USA

Credibility	Description
C	Test samples were not measured and/or characterized.
B	One or more key characteristics of the test samples were measured.
A	All key characteristics of the test samples were measured.

Measurements of Test Samples

Credibility	subjects Characteristics Measurements
C	Characteristics of the subjects were not measured.
B	Some key characteristics of the subjects were measured but it was not possible to measure all key characteristics.
A	All key characteristics of the subjects were measured.

#2: BMD of Hip

- Test conditions :
  - Mechanical testing conditions:
    - Actuator initially displaced to generate a preload of approximately 10 N to ensure initial contact.
    - Crosshead displacement then zeroed.
    - Uniform displacement applied at a rate of 20 mm/mi up to a limit not exceeding 8 mm.

- Because the mechanical test is destructive, only one test condition can be performed on each specimen.

Finite Element Analysis of Denosumab Treatment Effects on Vertebral Strength in Ovariectomized Cynomolgus Monkeys

David C Lee,[1] Aurore Varela,[2] Paul J Kostenuik,[3] Michael S Ominsky,[4] and Tony M Keaveny[5]

[1]O.N. Diagnostics LLC, Berkeley, CA, USA
[2]Charles River Laboratories Inc., Montréal, QC, Canada
[3]Phylon Pharma Services, Newbury Park, CA, USA
[4]Amgen Inc., Thousand Oaks, CA, USA
[5]University of California, Berkeley, CA, USA

Credibility	Description
C	A single test condition was examined.
B	Multiple (two to four) test conditions were examined.
A	More than four test conditions were examined.

Quantity of Test Conditions.

#2: BMD of Hip

- In view of discussed test conditions for this study:

- **Ex vivo tests with tissue can often have more variability than tests with engineering materials. Do you have best practices for balancing test relevance vs. variability in the results?**

Finite Element Analysis of Denosumab Treatment Effects on Vertebral Strength in Ovariectomized Cynomolgus Monkeys

David C Lee,[1] Aurore Varela,[2] Paul J Kostenuik,[3] Michael S Ominsky,[4] and Tony M Keaveny[5]

[1]O.N. Diagnostics LLC, Berkeley, CA, USA
[2]Charles River Laboratories Inc., Montréal, QC, Canada
[3]Phylon Pharma Services, Newbury Park, CA, USA
[4]Amgen Inc., Thousand Oaks, CA, USA
[5]University of California, Berkeley, CA, USA

Measurements of Test Conditions.

Credibility	Description
D	Tests conditions were not measured and/or characterized.
C	One or more key characteristics of the test conditions were qualitatively measured and/or characterized.
B	One or more key characteristics of the test conditions were quantitatively measured.
A	All key characteristics of the test conditions were quantitatively measured.

#3: PERIPHERAL STENT

#3: Peripheral Stent – Introduction

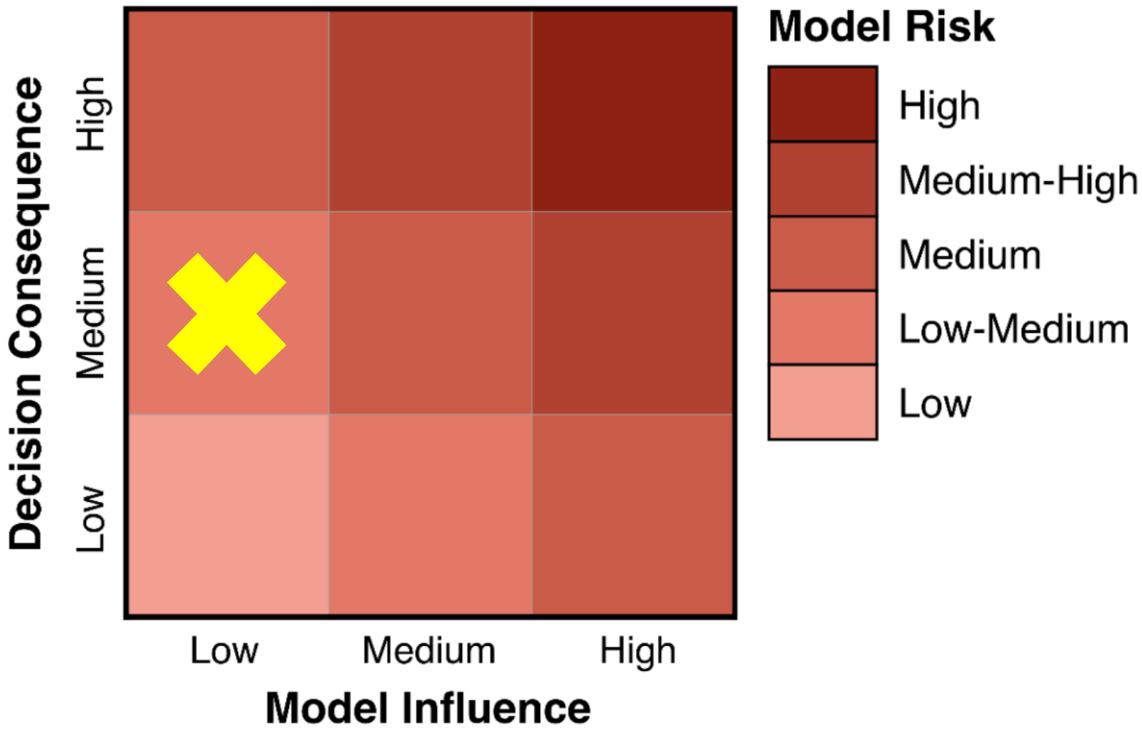
- Peripheral stent repair device analyzed for fatigue assessment for IDE submission.
- **QOI:** Is the device resistant to fatigue fracture under conservative radial and ambulatory *in vivo* loads?
- **COU:** The model will be used to determine the fatigue safety factor of the frame under radial and ambulatory *in vivo* loads based on literature. The model will also determine the worst case size and loading mode for fatigue testing.

Example Peripheral Stents Shown

#3: Peripheral – Model Risk

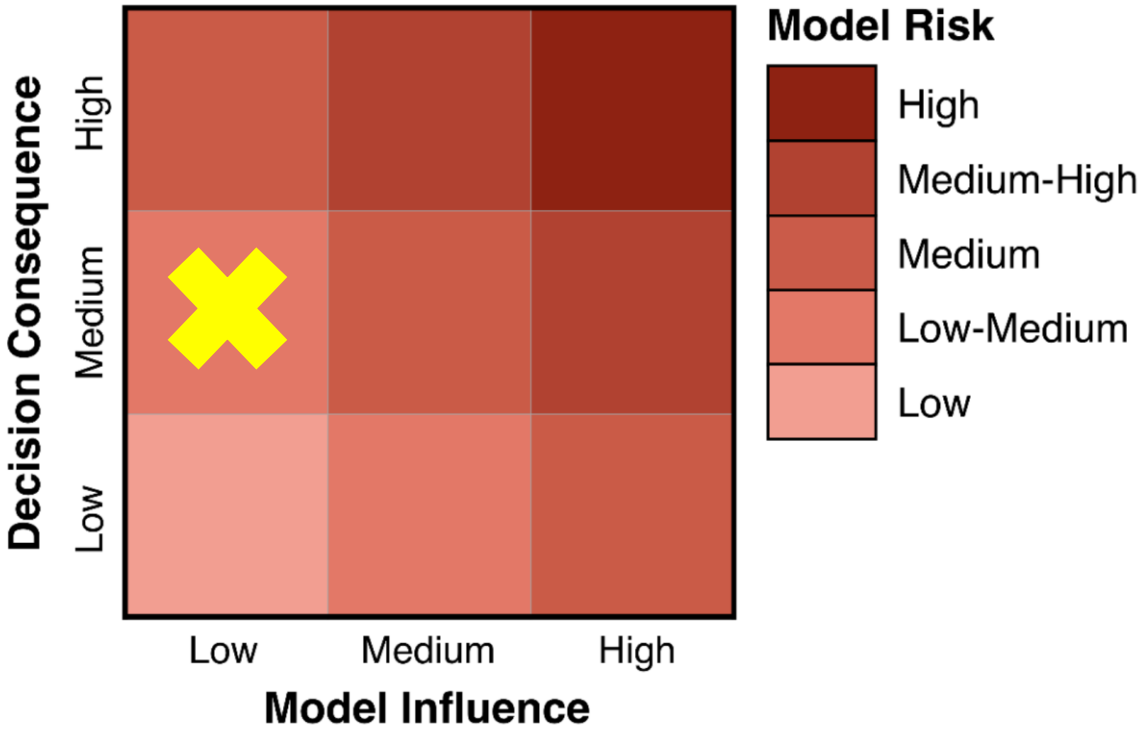
- **Model Influence:** The model influence is determined to be **high** since the model is determining the FSF, the worst case size, and testing conditions.

- **Decision Consequence:** An incorrect decision regarding fatigue safety could result in one or more strut fractures. The clinical sequalae from a strut fracture can range from inconsequential to significant patient harm, but death is unlikely. The sequalae depend on number of fractures and patient specifics. Decision consequence is therefore determined to be **medium**.



#3: Peripheral – Model Risk

- Current study is for IDE submission and indication is for peripheral vascular lesions, which are less common.
- Also, stent fracture by itself is unlikely to lead to significant patient harm, but additional clinical factors could induce consequences, if present (e.g. clotting).
- **Should the probability of complication occurrence be considered when evaluating model risk?**
- **Should the number of potential patients be considered?**



#3: Peripheral Stent – Credibility Goals

- In vitro comparator
- Radial force after deployment for two device sizes compared to experimental results
- Goal: force-displacement response within 10% of experimental average

Radial Force After Deploying

Device Size	FEA	Experimental mean value	% Difference
4 mm	0.62 N	0.63 N	-1.7%
8 mm	0.78 N	0.77 N	+0.5%

- **Is radial loading sufficient for multi-mode analyses?**
  - **Test is simple and focuses on device.**
- **If wanted to perform DIC and it demonstrated fundamental limitation of standard analysis methods, does it reduce credibility?**

#3: Peripheral Stent – Results and Adequacy

- Analysis results demonstrated that 4 mm device under cyclic bending had lowest FSF at 2.3.

- Two out of three primary results from model are comparative (worst case size and loading mode).
- **When establishing credibility goals do you consider whether model results will be relative or absolute?**