

의료기기 분야의 컴퓨터 모델링 및 시뮬레이션(CM&S)을 위한 ASME V&V 40 표준의 실무적 적용에 관한 연구

A Study on the Practical Application of ASME V&V 40 Standard for Computational Modeling and Simulation (CM&S) in Medical Devices

이주연¹, 이태희¹, 이주선¹, 김소희¹, 허희선¹, 고동현¹, 김현정¹, 박해대¹, 이수경^{1, #}
Ju-Yeon Lee¹, Tae-Hee Lee¹, Ju-Seon Lee¹, So Hee Kim¹, Hee Seon Heo¹, Dong Hyun Go¹,
Hyeon Jeong Kim¹, Hae Dae Park¹, and Su-Kyoung Lee^{1, #}

1 식품의약품안전평가원 의료제품연구부 의료기기연구과 (Medical Device Research Division, Department of Medical Product Research, National Institute of Food and Drug Safety Evaluation)

Corresponding Author / E-mail: sk1218@korea.kr, TEL: +82-43-719-4916
ORCID: 0009-0008-4054-3997

KEYWORDS: Computational modeling and simulation (컴퓨터 모델링 및 시뮬레이션), Credibility assessment (신뢰성 평가), Medical devices (의료기기), ASME V&V 40, Verification and validation (검증 및 유효성 확인)

The increasing use of computational modeling and simulation (CM&S) in the medical device sector has heightened the need for ensuring simulation credibility. The ASME V&V 40 standard offers a structured framework for assessing credibility, consisting of 23 factors divided into three main categories: Verification, Validation, and Applicability. However, practical guidance for implementing these factors is still scarce. This study systematically reviewed and analyzed ten CM&S-related publications in the medical device field that utilized the ASME V&V 40 framework. It examined how each publication addressed the credibility factors and compared their implementation methods, evaluation criteria, and credibility levels. From this comparative analysis, we developed implementation strategies focused on credibility factors, field-specific characteristics, and model risk levels in real-world regulatory and development contexts. Key considerations for the practical application of each factor were identified, and recommendations for effective implementation were proposed. These findings offer practical guidance for ensuring credibility in CM&S-based medical device development, performance evaluation, and regulatory processes. By clearly demonstrating the applicability of the ASME V&V 40 framework, this work provides valuable direction for related industries and research institutions, aiming to improve CM&S credibility and promote its broader adoption in healthcare.

Manuscript received: September 9, 2025 / Revised: January 6, 2026 / Accepted: January 15, 2026

1. 서론

컴퓨터 모델링 및 시뮬레이션(Computational Modeling and Simulation, CM&S)은 의료기기의 개발 및 평가에 있어 점차 필수적인 도구로 자리매김하고 있다[1]. CM&S는 In Silico 시험을 통해 물리적 시험의 의존도를 감소시켜 개발 기간 단축과 비용 절감을 가능케 하며, 환자 안전성 향상에 기여할 수 있다. 이러한 장점이 부각됨에 따라, 미국 식품의약국(FDA)[2]을 비롯한 여러 규제 당국에서는 CM&S를 규제 과학 및 의사결정 과정에

점진적으로 도입하고 있는 추세이다.

시뮬레이션 기반 증거의 신뢰성과 타당성을 확보하기 위해, 미국기계학회(ASME)는 ASME V&V 40 표준을 제정하였다[3]. 이 표준은 사용 맥락(Context of Use, COU)과 모델 위험도(Model Risk)를 기반으로, 검증(Verification), 유효성 확인(Validation), 적용 가능성(Applicability)의 세 과정을 통해 컴퓨터 모델의 신뢰성을 구조적으로 평가할 수 있는 프레임워크를 제공한다. 여기서 검증(Verification)이란 컴퓨터 모델이 개념적 모델의 수학적 표현을 정확하게 구현하고 있는지, 즉 “모델을

올바르게 만들었는가(Are we building the model right?)”를 확인하는 과정이다. 이는 소프트웨어 품질보증(Software Quality Assurance, SQA), 수치 코드 검증(Numerical Code Verification, NCV), 이산화 오차 분석, 수치 솔버 오차 검증 등을 포함한다. 유효성 확인(Validation)은 컴퓨터 모델이 실제 물리적 현상을 얼마나 정확하게 재현하는지, 즉 “올바른 모델을 만들었는가(Are we building the right model?)”를 평가하는 과정이다. 이는 시험 데이터 또는 임상 데이터와의 비교를 통해 모델의 예측 성능을 정량적으로 평가하며, 모델 형식(Model Form) 타당성, 입력 변수 민감도 분석, 불확실성 정량화, 비교자(Comparator) 기반 유효성 평가 등을 포함한다. 적용 가능성(Applicability)은 정량적 관심 지표(Quantities of Interest, QOIs)의 적합성과 유효성 확인 활동이 COU 조건을 얼마나 포괄하는지를 기준으로 평가하는 과정이다.

ASME V&V 40의 중요성과 활용 가능성은 지속적으로 확대되고 있으나[4], 실제 연구 및 산업 현장에서는 신뢰성 요소별 적용 방식에 대한 해석과 수행 전략이 다양하게 나타나고 있으며, 일부 사용자는 실무 적용에서의 구체적인 방향 설정에 어려움을 겪는 경우도 있다. 이는 ASME V&V 40이 각 신뢰성 요소를 명확히 정의하고 있으나, 프레임워크의 특성상 개별 요소에 대한 해석 및 실무 적용 방식에 대해서는 추가적인 구체화가 이루어질 수 있는 여지를 남기고 있기 때문으로 해석된다.

이에 본 연구는 의료기기의 성능 평가 등을 위해 CM&S를 활용한 기존 문헌 가운데 신뢰성 평가 활동이 수행된 사례들을 체계적으로 검토하고, ASME V&V 40에서 제시한 23개 신뢰성 요소별로 실무적으로 쉽게 적용할 수 있는 구체적 수행전략을 도출하였다. 이를 통해 해당 프레임워크의 실제 적용 가능성을 높이고, 연구자, 모델 개발자, 규제 당국이 보다 일관되고 명확한 신뢰성 평가를 수행할 수 있도록 실질적인 지침을 제공하고자 한다.

2. 방법

2.1 분석 대상 연구의 수집 및 선별

의료기기 분야의 CM&S에서 신뢰성 평가를 다룬 관련 문헌을 체계적으로 식별하기 위해, 구조화된 문헌 검색 전략을 수립하였다. 검색어는 ASME V&V 10[5], 20[6], 40[3] 표준에서 사용되는 주요 용어를 기반으로 도출하였으며, 예비 검색을 통해 검색어의 정확도(Precision)와 재현율(Recall)을 조정하였다.

핵심 검색어로는 “computational”과 “model”을 사용하여 수치 모델링 전반을 포괄하였고, V&V 40의 핵심 개념인 “credibility”를 포함시켰다. 또한, 신뢰성 평가의 다양한 활동을 반영하기 위해 “verification”, “validation”, “uncertainty quantification”, “sensitivity analysis”, “applicability” 등의 용어를 추가하였다. 최종 검색식은 다음과 같다:

computational AND model AND credibility AND (verification

OR validation OR “uncertainty quantification” OR “sensitivity analysis” OR applicability)

문헌 검색은 의료기기 규제 과학 분야의 대표성과 포괄성을 고려하여 PubMed, Embase, IEEE Xplore, ScienceDirect, Cochrane Library에서 수행하였다. 이 중 PubMed와 Embase는 생명과학 및 의학 분야를 포괄하고, IEEE Xplore와 ScienceDirect는 공학 및 수치 모델링 관련 문헌을 포함한다. 검색 대상은 ASME V&V 40 표준이 공식 발표된 2018년부터 2024년까지 출판된 영어 저널 논문으로 한정하였다. 이는 ASME V&V 40의 신뢰성 평가 요소들(예: 검증, 유효성 확인, 불확실성 정량화 등)에 대한 선행 연구를 확인하기 위함이다. 동료 평가(Peer-review)를 거치지 않은 출판물, 사설(Editorial), 학술대회 초록(Abstract), 서한(Letter), 학위논문(Thesis)은 제외하였다.

문헌 검색의 포괄성을 높이기 위해 수기 검색도 병행하였다. 먼저, 선정된 문헌의 참고문헌을 검토하는 Snowballing 기법[7]을 활용하였고, 검색 키워드 기반 자동 검색에서 누락되었을 가능성이 있는 중요 문헌을 보완하기 위해 전문가 자문을 병행하였다. 이 수기 검색은 사전에 정의된 검색 전략의 일환으로 계획되었다.

문헌 선별의 일관성을 확보하기 위해 다음과 같은 포함 및 배제 기준을 적용하였다.

포함 기준:

- (1) 의료제품 개발 또는 평가에 CM&S를 적용한 연구
- (2) 정식 학술지에 게재된 원저 논문
- (3) ASME V&V 40의 신뢰성 평가 활동(예: 검증, 유효성 확인, 불확실성 정량화 등)을 하나 이상 명시적으로 기술한 경우

배제 기준:

- (1) 리뷰 논문, 편집자 서신, 학위논문, 학술대회 발표자료 등
- (2) 이론적 논의에 그치고 실질적인 모델 평가가 없는 경우
- (3) “credibility” 등의 용어만 언급하고 구체적인 방법이 제시되지 않은 경우
- (4) 모델 유형, COU, 평가 결과 등의 핵심 정보가 부족한 경우

초기 검색에서는 총 325편의 논문이 수집되었으며, 중복 제거 후 제목 및 초록을 기준으로 1차 선별을 진행하였다. 이후 33편에 대해 본문 정밀 검토를 실시하였고, 최종적으로 포함 기준을 만족한 10편의 논문이 분석 대상으로 선정되었다. Fig. 1은 문헌 선정 및 분석 절차를 요약한 PRISMA [8] 도식이다.

2.2 연구 분석 방법

본 연구에서는 ASME V&V 40 표준을 실무적으로 적용하기 위한 전략을 도출하고자, 선별된 10편의 의료제품 분야 CM&S 관련 논문에 대해 체계적인 정성적 분석을 수행하였다. 분석은 다음의 네 가지 측면에서 단계적으로 진행되었다. 1) 연구의 일반적

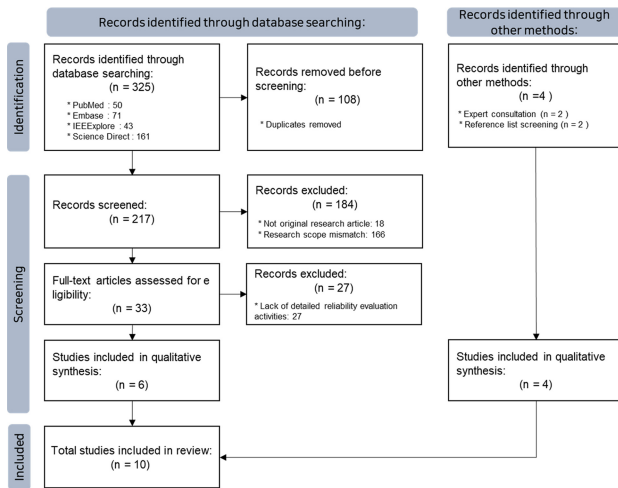


Fig. 1 Flowchart of literature selection depicted using the PRISMA diagram

특성, 2) 신뢰성 평가 요소별 수행 전략, 3) 해석 분야별 수행 전략, 4) 모델 위험도 수준별 수행 전략.

2.2.1 연구의 일반적 특성 분석

우선, 각 논문의 기본 정보를 바탕으로 ASME V&V 40의 적용정도, QOIs, COU, 모델 위험도, 적용된 의료기기 종류를 정리하였다. 이를 통해 연구 간 전반적인 경향성과 특성을 비교·분석하였다.

2.2.2 신뢰성 평가 요소별 수행 전략 분석

각 논문이 수행한 신뢰성 평가 활동을 ASME V&V 40에서 정의한 23개 신뢰성 평가 요소에 대응시켜 분석하였다. 분석은 세 가지 대분류(검증, 유효성 확인, 적용 가능성) 하의 하위 요소별로 구조화하여 수행되었으며, 각 요소에 대해 구체적으로 어떤 방법이 사용되었는지, 문헌 간 수행 수준 및 기술 방식에 어떤 차이가 있는지를 정성적으로 비교하였다. 이 과정에서 식별된 공통적 수행 방식 또는 실무 적용에서 활용 가능한 전략은 요소별 수행 전략으로 정리하였다.

2.2.3 분석 분야별 수행 전략 비교

선정된 10편의 논문에 대해 모델이 적용된 분석 분야(예: 혈류 역학, 생체역학, 약물전달 등)를 기준으로 분류한 후, 각 분야에서 활용된 신뢰성 평가 접근 방식의 차이를 분석하여 분야별 실무 적용을 위한 차별화된 전략을 제시하였다.

2.2.4 모델 위험도 수준별 수행 전략 분석

각 논문에서 명시하거나, 연구 내용을 바탕으로 추정한 모델 위험도 수준(Model Risk Level)에 따라 문헌을 분류하였다. 이후, 위험도 수준별로 수행된 신뢰성 평가 활동의 깊이, 정량화 수준, 평가 도구의 복잡도 등을 비교·분석하였다. 이를 통해 모델 위험도에 따른 평가 범위 조정 및 자원 배분 전략 등 실무에서의

적용 가이드라인을 도출하였다.

이러한 다층적 분석 접근을 통해, 본 연구는 ASME V&V 40 표준의 각 신뢰성 요소가 실제 CM&S 기반 의료기기 연구에서 어떻게 구현되고 있는지를 총체적으로 파악하고, 이를 실무에 직접 활용할 수 있는 구체적 전략으로 정리하고자 하였다.

3. 결과

3.1 연구의 일반적 특성 분석

분석에 포함된 10편의 연구는 모두 의료제품 개발 및 평가에 CM&S를 활용한 사례로, 다양한 제품 유형과 분석 목적을 포함하고 있었다. 선별된 문헌의 상세 정보는 Table 1과 같다. 주요 적용 대상은 원심 혈액 펌프, 인공관절, 약물 주입 시스템 등으로, 생리학적 반응 예측부터 구조적 안정성 평가에 이르기까지 광범위하였다. QOIs는 용혈 예측, 유속 분포, 강성, 최적 용량 등 다양하게 설정되었으며, 연구 목적에 따라 성능 검증 또는 안전성 확보에 중점을 두었다.

모델의 COU는 주로 시뮬레이션 기반 제품 성능 검증 혹은 환자 맞춤형 치료 시나리오 예측에 집중되었으며, 모델 위험도 수준은 Level 2-5까지 다양하게 분포되었다. 그러나 일부 연구에서는 위험도 수준이 명확히 기술되지 않거나 간접적으로만 판단 가능한 경우도 있었다[9,10]. 모델 위험도는 사용 목적, 인체 직접 영향 여부, 임상 의사결정에의 기여도 등을 기준으로 결정하였다.

Table 1에 제시된 바와 같이, 분석 대상 10편 문헌의 ASME V&V 40 적용 수준은 Full Application 4편[11-14], Partial application 4편[15-18], ASME V&V 40 Referenced 2편[9,10]으로 분포하였다. 즉, 8편(80%)의 연구가 ASME V&V 40의 요소를 부분 또는 전면적으로 실제 적용하여 신뢰성 활동을 수행한 반면, 2편(20%)은 ASME V&V 40을 참고(Reference) 수준으로 언급하였으며 표준의 전 요소를 체계적으로 적용·보고한 형태는 아니었다. 또한 Full Application으로 분류된 연구들에서도 적용 방식과 보고의 구체성에는 차이가 있었으며, 특히 요소별 목표(Credibility Goals) 및 달성 수준을 명시적으로 제시한 사례는 1편으로 제한적이었다[11]. 이러한 분포는 의료기기 CM&S 분야에서 ASME V&V 40의 실무 적용이 확대되고 있으나, 연구 목적, 모델 위험도 수준 등에 따라 적용의 깊이와 보고 수준이 달라질 수 있음을 시사한다.

3.2 신뢰성 평가 요소별 수행 전략

ASME V&V 40 표준에서 제시한 23개 신뢰성 요소에 따라, 분석 대상 논문들의 신뢰성 확보 활동을 검증, 유효성 확인, 적용 가능성의 세 영역으로 구분하여 분석하였다. 각 요소별 수행 전략은 개별 문헌에서 실제로 수행·보고된 신뢰성 관련 활동을 종합적으로 비교·정리하여 도출하였으며, 그 결과를 Table 2에 제시하였다.

Table 1 Summary of selected literature

No. [Ref.]	Assessment level	QOIs	COU	Model risk	Applied medical device
1 [15]	Partial application	Are the flow-induced hemolysis levels of the centrifugal pump acceptable for the intended use?	COU1: CPB (Class II); COU2: Short-term VAD (Class III)	L2 (COU1) / L5 (COU2) ^a	Centrifugal blood pump
2 [9]	ASME V&V 40 referenced	Prediction of metal–polyethylene contact area ^b	Design evaluation for TAA contact mechanics ^b	-	Total ankle arthroplasty
3 [10]	ASME V&V 40 referenced	Mechanical response of spinal rods under 3-point bending ^b	UQ and validation for mechanical response ^b	-	Spinal rod
4 [11]	Full application	For an apically implanted LVAD, does the selected pump speed produce: (a) complete aortic valve opening >0.3[L/min]; and (b) a Cardiac output compatible with life>(4.2[L/min]) for a range of HR and EF covering a HF patient population?	See note c	L3 ^a	LVAD
5 [16]	Partial application	Which is the optimal effective dose for a new anti-osteoporosis drug in adults and older adults (from 55 years) according to multi-dose Phase II studies?	See note d	L3 ^a	-
6 [17]	Partial application	QOI for COU1: “How do decisions regarding the material and design influence the functional parameters of the custom-made 3D printed WHO?”; QOI for COU2: “How do decisions regarding the material and design influence the occurring strains and stresses on the custom-made 3D printed WHO?”	COU1: Performance evaluation of the functional properties of the custom-made 3D printed WHO; COU2: Superiority evaluation of the strain distribution of the custom-made 3D printed WHO	L2~3 (COU1) / L2 (COU2) ^a	Custom 3D-printed WHO (wrist hand orthosis)
7 [18]	Partial application	“what is the most immunogenic dose of the new therapeutic vaccine to be used in patients affected by tuberculosis?”	See note e	L2 ^a	-
8 [12]	Full application	1. What are the crimp strains and fatigue strains (strain amplitude and mean strain) and peak locations in the TAV under simulated in vivo conditions? 2. What test conditions are required to replicate in vivo strain amplitudes (and mean strains) for structural component fatigue testing 3. Will the TAV survive 600M cycles under in vivo loading, without fracture?	To predict the fatigue strains in multiple device sizes under in vivo loading conditions. Results are used to identify the worst-case device size, location of peak fatigue strains and test conditions required to reproduce in vivo strain level in a benchtop structural component fatigue test.	Medium	TAV (transcatheter aortic valve) frame
9 [13]	Full application	Does the hypothetical new total knee arthroplasty design provide sufficient resistance to wear of the polyethylene (PE) inlay under ISO 14243-1 and activities of daily living test conditions in displacement control?	WearPy is used to determine the amount of PE volumetric wear of the new design and identify the worst case condition (test and size). Bench testing will be performed on the identified worst-case.	Low-medium	Knee implant
10 [14]	Full application	“does adding a 1.6 mm diameter cannulation to an existing 7.5 mm diameter pedicle screw design compromise mechanical performance of the rod-screw construct in static compression-bending?”	“model predictions of the original non-cannulated screw construct will be validated with benchtop testing per ASTM F1717 static compression bending conditions. The validated model framework will be used to evaluate the cannulated screw design undergoing identical static compression-bending conditions. No benchtop testing on the cannulated screw design will be performed to answer the ?OI.”	High-medium (L4)	Pedicle screw system

^aModel risk was categorized into five levels (L1–L5) in the selected literature.

^bQOI and COU were not explicitly stated in the original publication and were inferred by the authors based on the study objectives and methodology.

^cThe heart-LVAD computational model may be used by design engineers to assist in the preclinical development of LVAD, by characterising aortic root, LVAD and intra-LV flows for a given pump speed. The goal of the heart-LVAD computational model is to provide a computational replica of a benchtop experiment for a quantitative analyses in parametric explorations. The heart-LVAD computational model by no means is replacing animal experiments or clinical trials, but augmenting the totality of evidence.

^dBBCT-hip is a methodology where a stochastic biophysics model provides an estimate, for a given subject, of the Absolute Risk of proximal femur Fracture upon falling at time zero (ARF0), from their height, weight, and a Quantitative Computed Tomography (QCT) scan of the hip region. This ARF0 is to be used as a response variable in multi-dose Phase II studies in place of the measured DXA-based aBMD. The average change in ARF0 over the period of treatment for all subjects treated with a given dose (AveΔARF0) can be used as response variable, by assuming the optimal dose amongst those tested is the one for which AveΔARF0 is most positive (or least negative).

^eThe UISSTB-DR model will be used to support the decision about the most immunogenic dose of the new therapeutic vaccine against TB and inform phase II dose selection studies by predicting the human immune system response.

Table 2 Implementation strategies for credibility assessment by credibility factor

No.	Activity	Credibility factor	Implementation strategy	Ref.	
1	Code	SQA	① Verification of regular quality maintenance	[10,11,18]	
			② Conducting simple internal quality assurance activities, such as identifying COU-related bugs	[14,15]	
			③ Verification of compliance with international quality management standards (third-party certification)	[12,13,16,17]	
2		NCV	① Selection of benchmark solutions representing COU-related physical phenomena – Well-known solutions in relevant fields where analytical solutions exist – Benchmarks provided by the manufacturer – Numerical comparisons based on existing literature	[10-18]	
			② Alongside ①, conduct mesh or time step convergence studies and verify convergence	[11]	
			③ Verification of compliance with international quality management standards (third-party certification)	[12,13,16,17]	
3	Verification	Discretization error	① Perform mesh or time sensitivity analysis and evaluate convergence	[9-17]	
			② Define permissible discretisation error criteria (e.g., ≤5%) and evaluate results	[10,14]	
			③ Apply standard convergence analyses such as GCI based on Richardson Extrapolation	[15]	
4		Calculation	Numerical solver error	① Parameter setting via literature, expert consultation, etc.	[11]
				② Solver parameter sensitivity analysis affecting numerical stability	[12,14,16,18]
				③ Apply standard convergence analyses such as GCI based on Richardson Extrapolation	[15]
5	Use error	Use error	① Practitioner directly reviews validity of key inputs (boundary conditions, material properties, etc.) and outputs	[12,13,16]	
			② Internal peer independently reviews input files and modelling settings	[11,12,14,18]	
			③ Independent execution by two or more different organisations, with results compared to assess user and execution environment effects	[10]	
6		Computational model	Model form	① Clearly present all model assumptions and simplifications (e.g., geometric simplifications, symmetry assumptions, boundary conditions, loading conditions, material models, physiological characteristics, etc.)	[9,10,13,15]
				② Alongside ①, qualitatively/quantitatively analyse the impact of these assumptions on results	[11,12,14,16,18]
				③ Independent execution by two or more different organisations, with results compared to assess user and execution environment effects	[10]
7	Model input		Quantification of sensitivities	① Conduct sensitivity analysis for key or comprehensive inputs to identify those with significant impact on results	[9-14,16]
				② Solver parameter sensitivity analysis affecting numerical stability	[12,14,16,18]
				③ Apply standard convergence analyses such as GCI based on Richardson Extrapolation	[15]
8		Quantification of uncertainties	Quantification of uncertainties	① Quantify the uncertainty of key inputs identified through sensitivity analysis, etc.	[10,12,18]
				② Apply sampling-based uncertainty propagation techniques (LHS, Monte Carlo, etc.) to quantitatively assess the impact of input variable uncertainty on final outputs (QOLs)	[11,14-16]
				③ Independent execution by two or more different organisations, with results compared to assess user and execution environment effects	[10]
9	Quantity		Quantity	① Select two or more arbitrary sample quantities	[9,12]
				② Select sample quantities according to relevant standards or guidelines	[10,13,14]
				③ Select statistically significant sample quantities	[16]
10		Test samples	Range characteristics	① Select samples representing typical usage conditions, including those corresponding to the Nominal Value	[14,16]
				② Select samples to include extreme values, such as those from high-risk groups or under severe conditions	[16]
				③ Select diverse samples to represent the variety of the product/patient population	[12]
11	Measurements		Measurements	① Measure some or all key characteristics of the test samples required for the comparator study	[9-14,16]
				② Solver parameter sensitivity analysis affecting numerical stability	[12,14,16,18]
				③ Apply standard convergence analyses such as GCI based on Richardson Extrapolation	[15]
12		Uncertainty of test sample measurements	Uncertainty of test sample measurements	① Manage measurement uncertainty by performing the study using an already calibrated measurement system	[10,13]
				② Measure test condition values repeatedly to reflect measurement uncertainty	[14,16]
				③ Estimate the combined measurement uncertainty for the test samples	[12]
13	Comparator		Quantity	① Establish two or more diverse test conditions representing the clinical usage environment or mechanical operating conditions the model aims to evaluate	[11-13,15-17]
				② Select condition quantities according to relevant standards or guidelines	[9,10,13,14]
				③ Independent execution by two or more different organisations, with results compared to assess user and execution environment effects	[10]
14		Test conditions	Range characteristics	① Select a range including the most common or average usage conditions, or those specified by relevant standards or guidelines	[10,13,14]
				② Include boundary conditions or extreme conditions that could most significantly impact safety	[12,15]
				③ Select conditions to cover the entire range, including various points between normal and extreme conditions	[9,11]
15	Measurements		Measurements	① Measure some key or all conditions required for comparator study	[10-15,17]
				② Solver parameter sensitivity analysis affecting numerical stability	[12,14,16,18]
				③ Apply standard convergence analyses such as GCI based on Richardson Extrapolation	[15]
16		Uncertainty of test condition measurements	Uncertainty of test condition measurements	① Measure test condition values using an already calibrated measurement system	[13]
				② Measure test condition values repeatedly to reflect measurement uncertainty	[10,15,17]
				③ Estimate the combined measurement uncertainty for the test conditions	[12,14]
17	Assessment		Equivalency of input parameters	① If identical values or units (types) cannot be used, employ similar types	[12,18]
				② Use the inputs from the comparator as inputs for the CM&S model	[9,11,13-16]
				③ Independent execution by two or more different organisations, with results compared to assess user and execution environment effects	[10]
18		Output comparison	Quantity	① Select QOLs for multiple parameters when choosing them	[9-17]
				② Solver parameter sensitivity analysis affecting numerical stability	[12,14,16,18]
				③ Apply standard convergence analyses such as GCI based on Richardson Extrapolation	[15]
19	Equivalency of output parameters		Equivalency of output parameters	① If identical variable values cannot be measured in the computational model, select a similar physical quantity and document the correlation	[12,16]
				② Match the output parameters of the comparator to the outputs of the computational model	[9-11,13-15]
				③ Independent execution by two or more different organisations, with results compared to assess user and execution environment effects	[10]

Table 2 Continued

No.	Activity		Credibility factor	Implementation strategy	Ref.
20	Validation	Assessment	Output comparison	① Compare visual similarities such as the shape and trend of output curves	[13,17]
				② Compare arithmetic differences in key metrics such as maximum values, mean values, and mean root mean square error (NRMSE)	[9,10,12-16]
				③ Compare the uncertainty of computational model results or comparator results together (e.g., checking whether confidence intervals overlap)	[11,12,14,15,17]
21			Agreement of output comparison	① Where quantitative comparison is impossible, conduct a qualitative comparison of output consistency	[13]
				② Conduct a quantitative, statistical comparison of output consistency	[9-17]
				③ Confirm whether validation acceptance criteria exist in the same or similar field before analysis	[9,16]
22	Applicability		Relevance of the QOIs	① Even if QOIs are not directly linked to the COU, explicitly state their mathematical/logical association with the COU and demonstrate this relationship	[12,16,17]
				② Select parameters directly related to the core issues of the COU as QOIs for validation activities	[9-11,13-18]
23			Relevance of the Validation Activities to the COU	① If validation activities cover only part of the COU, clearly describe these limitations	[11,14,16,18]
				② Design validation activities to encompass as much of the COU as possible	[10,12,13,15,17]

Table 2의 수행 전략은 분석 대상 문헌들에서 관찰된 구현 사례를 기반으로 정리된 것으로, 상대적으로 접근이 용이한 전략(낮은 번호)부터 보다 높은 수준의 신뢰성 확보를 지향하는 전략(높은 번호)까지 단계적으로 배열하였다. 이를 통해 각 신뢰성 요소가 의로기기 CM&S 연구에서 어떠한 방식으로 구현되고 있는지를 구조적으로 비교할 수 있도록 하였다.

3.2.1 검증

검증 활동에서는 SQA, NCV, 이산화 오차 분석, 수치 솔버 오차 검증 등이 주요 항목으로 확인되었다. 상용 소프트웨어를 활용한 사례에서는 제조사의 국제 표준(ISO 9001 등) 준수 여부, 품질 문서(릴리즈 노트, 버그 목록 등) 참조, 제공되는 표준 벤치마크 문제의 활용을 통해 일정 수준의 소프트웨어 품질에 대한 신뢰성을 확보하였다. 반면, 상용 소프트웨어로 명시되지 않은 경우에는 버전 관리, 내부 QA 절차, COU 관련 버그 추적 등이 일부 수행되었으나, 독립적 검토나 체계적 문서화가 구체적으로 보고된 경우는 제한적이었다. 이는 실제 미이행이라기보다 보고 방식이나 지면 제약으로 인한 비기재 가능성을 함께 고려해야 한다.

NCV는 해석 해가 존재하는 문제, 널리 알려진 표준 벤치마크, 문헌의 수치 해를 활용하는 경우가 많았으나, 선정 기준을 체계적으로 제시한 경우는 제한적이었다. 이산화 오차 분석은 메시/시간 단계 수렴연구가 일부 보고되었으나, 허용 오차 기준 정의나 Richardson 외삽법, GCI 적용은 드물었다. 수치 솔버 오차 검증은 수렴 기준, 자동 감쇠(Damping) 등 안정성 관련 설정이 언급되었으나, 민감도 분석의 정량적 기법 적용은 제한적이었다.

따라서 상용 소프트웨어 사용이 가능한 경우에는 제조사가 제공하는 품질보증과 벤치마크를 적극적으로 활용하고, 비상용 소프트웨어의 경우에는 COU에 적합한 버전 관리, 버그 추적 등을 포함하는 품질보증 체계를 구축할 것을 권고한다. 이산화 오차 및 솔버 파라미터 오차와 관련해서는 가능하다면 정량적 수렴성 분석(GCI, 민감도 분석 등)을 적용하여 신뢰성을 강화해야 할 것이다.

3.2.2 유효성 확인

본 검토에서 확인된 유효성 확인 활동 중 컴퓨터 모델에 대한 신뢰성 활동의 범위는 주로 영향도가 높은 주요 변수에 집중되는 경향이 있었다. 주요 가정사항에 대한 모델링 과정을 설명하거나, 주요 입력변수에 대해서만 민감도 분석 및 불확실성 정량화를 수행하였는데, 이는 연구자들이 제한된 자원으로 인해 핵심 변수에 우선순위를 둔 결과로 해석된다. 높은 신뢰성 확보를 위해서는, 민감도 분석을 통해 주요 변수들을 선정하고, 선정된 변수들에 대한 불확실성을 정의하는 과정이 필수적으로 수행되어야 할 것이다.

비교자에 대한 활동을 살펴보면, 다수의 연구에서 국제 표준(ISO, ASTM 등)에 근거하여 샘플 수와 시험 조건을 설정하거나[10,13,14] 극단값(예: 고위험군, 가혹 조건)을 포함해 검증을 수행함으로써 다양한 임상적·물리적 상황을 반영하였으며[16], 일부는 무작위 샘플 선정을 통해 제품의 변동성을 반영하였다[12]. 시험 샘플의 주요 특성(형상, 재료, 표면 특성 등)을 교정된 기기를 통해 정량적으로 측정하거나[10,13], 반복 측정을 통해[12,14,16] 측정 불확실성을 관리하려는 노력이 다수 확인되었다. 컴퓨터 모델과의 엄격한 비교를 위해서는, 되도록 여러 시료를 대상으로 여러 번 시험을 수행하는 것이 권장된다.

평가 활동 중 입력 변수의 등가성은 대체로 컴퓨터 모델과 비교자 간 동일한 물리량을 적용하거나[9,11,13-15], 비교자와 모델 간 차이를 문서화하고 간접적으로 일치성을 확보[12,16]하는 방식으로 이루어졌다. 출력 비교에서는 컴퓨터 모델과 실제 시험의 출력 항목이 직접적으로 일치하는 경우가 많았다[9-15]. 비교의 엄격성(Rigor)은 단순 시각적 비교는 극히 일부였으며[13,17], 표준편차 등 통계적 비교를 통해 동등성을 평가한 경우가 가장 많았다[9,10,12-16]. 일치도(Agreement)는 대체로 정량적 비교를 통해 수용 가능한 수준임을 보고하였으나[9-17], 동일 또는 유사분야에 동등성 허용기준이 존재하는지 확인 후 이에 대한 분석을 수행한 연구는 제한적이었다[9,16]. 따라서, 모델-비교자 설계 시 주요 입력·출력 파라미터에 대해 가능한 한 유사하게 정의하여 객관적 비교가 가능하도록 설계해야 할 것이며, 또한

비교 케이스를 충분히 확보하여 다양한 조건에서 모델의 예측 성능을 평가하고, 사전에 수용 기준과 통계적 비교 지표를 유사 분야조사를 통해 명시함으로써 결과 해석의 신뢰성을 강화할 필요가 있다.

3.2.3 적용 가능성

수행 현황을 보면, 여러 연구에서 유효성 확인 시험이 COU에서 정의한 핵심 QOIs와 직접적으로 연계되었다고 평가하였다. 예를 들어, 혈류 예측 연구에서는 용혈량이 QOIs로 설정되어 COU의 핵심 문제와 일치하였고[15], 구조적 안정성 평가에서는 강성(Stiffness)이나 마모 부피, 표면 손상 등 COU에서 요구되는 지표를 그대로 반영하였다[13,17]. 또한 ASTM F1717, ISO 14243-1과 같은 국제 표준을 기반으로 Stiffness, Yield Force 등 구체적인 QOIs를 일치시킨 사례도 확인되었다[14]. 다만 일부 연구에서는 In Vivo 데이터의 단일 시점 지표를 사용하여 시간 축 평균 변화량 중심의 COU와는 불완전하게 대응한 경우도 있었다[16].

유효성 확인 활동의 COU 연계성 측면에서도 다양한 사례가 관찰되었다. 일부 연구에서는 COU 조건(예: 보조기에 적용되는 모멘트 범위, TAV 프레임의 하중 조건 등)을 물리 시험으로 그대로 구현하여 높은 일관성을 확보하였다[10,12,13,15,17]. 반면, 일부 연구에서는 검증 포인트가 제한되거나 후향적 데이터에 의존하여 COU 전체를 포괄하기 어려운 한계가 있었다[11,14,16,18]. 이는 유효성 확인이 COU의 대표적 조건을 충실히 반영하는 경향은 있으나, 데이터 가용성과 설계 범위에 따라 COU 전체 조건을 완전히 커버하지 못한 사례가 존재함을 의미한다. 따라서 유효성 확인 시험의 설계 단계에서 COU의 조건을 충분히 대표·포괄할 수 있도록 QOIs와 시험 조건을 설정하는 것이 중요하며, 모든 QOIs가 COU와 직접적·수직적으로 연계될 수 있도록 정의하고, 부분적 연관성에 그치는 경우에는 그 한계를 명확히 문서화할 필요가 있다.

3.3 해석 분야별 수행 전략

3.3.1 혈류역학 분야(Hemodynamics)

혈류역학 연구에서는 상용 소프트웨어보다는 개발되거나 수정된 소프트웨어가 주로 사용되어 SQA를 연구자가 직접 수행해야 하거나 널리 알려진 식을 활용한 벤치마크 테스트가 요구되는 어려움이 있었다[11,15]. 또한 분야 특성상 모델화 과정이 복잡하고 불확실성이 커서 비교자와의 일치율이 낮게 나타나는 경향이 있었다[15]. 이러한 특성은 신뢰성 확보에 한계로 작용하므로, 향후에는 기존 연구[11,15]를 적극 활용해 벤치마크 테스트를 수행하고, 축적된 연구 결과를 심층 분석하여 합리적인 일치율 기준을 마련하는 것이 필요하다.

3.3.2 고체역학 분야(Solid/Structural Mechanics)

고체역학 연구에서는 ASTM, ISO 등 국제 표준 시험을 기반으로 유효성 확인이 수행되었고, 반복 시험과 표준편차 분석을 통해

변동성이 정량화되었으며[10,14], 입력 변수(예: 직경, 재료 물성)는 모델과 비교자 간에 비교적 잘 일치하였다. 국제 표준 시험의 활용 덕분에 검증 지침은 비교적 명확하였으나, 실제 임상 환경에서 발생할 수 있는 피로 손상 등 복잡한 조건을 모두 반영하기에는 제한이 있었다. 따라서 향후 고체역학 검증에서는 표준 시험에 더해 실제 사용 환경을 최대한 반영한 시험 조건을 설계하고, 모델-시험 간 입력 파라미터의 등가성을 사전에 정의·검증하며, 일치율이 높은 분야인 만큼 반복 시험과 교차 검증을 통해 결과의 재현성을 강화하는 전략이 요구된다.

3.3.3 기타 분야

약물전달학 연구에서는 IFN- γ 농도, ARF0, 골밀도(aBMD) 등 임상 데이터 기반 QOIs가 활용되었으나[16,18], 시험적 비교자 확보가 어렵고 유효성 확인 포인트가 제한적이었으며 일부는 후향적 데이터를 기반으로 하여 COU와의 직접적 대응이 제한적이었다[11,16]. 이처럼 임상 데이터 기반 모델은 실제 적용 가능성이 크지만 데이터 수집 제약으로 유효성 확인 범위가 협소한 한계가 있다. 따라서 향후에는 기존 문헌 및 임상 데이터베이스를 적극적으로 활용해 유효성 확인 포인트를 다각화하고, 부분적 일치에 그치는 경우 그 한계를 명확히 기술하며, 환자군 다양성을 반영할 수 있는 QOIs를 확장함으로써 모델의 일반화 가능성을 높여야 한다.

3.4 모델 위험도별 수행 전략

분석 대상 연구들의 모델 위험도 수준은 Level 2에서 Level 5까지 분포하였으며, 일부 연구는 COU에 따라 두 개의 위험도를 동시에 보고하였다[15,17]. 또한 위험도를 정량적 등급 대신 중간(Medium)[12], 낮음-중간(Low-medium)[13]과 같이 정성적으로 기술한 사례도 확인되었다.

모델 위험도 수준에 따라 가장 뚜렷한 차이가 나타난 요소는 유효성 확인 단계에서의 비교 엄격성이었다. 낮은 위험도(Level 2) 모델의 경우, 시각적 비교 또는 불확실성이 포함되지 않은 제한적인 정량 비교에 기반한 평가가 수행된 반면[13,17], 높은 위험도(Level 4-5) 모델에서는 컴퓨터 모델과 비교자 시험 양측의 불확실성을 정량화하고, 해당 불확실성이 반영된 결과를 직접 비교하는 접근이 적용되었다[14,15].

이러한 고위험 모델에서는 입력 변수의 불확실성을 정량화한 후 이를 출력(QOIs)까지 전파하는 방식이 사용되었으며, 불확실성 전파를 위해 Monte Carlo 시뮬레이션[15], Latin Hypercube Sampling (LHS)[11]과 같은 샘플링 기반 기법이 활용되었다. 또한 일부 연구에서는 상용 불확실성 정량화 도구(예: Ansys OptisLang)를 이용하여 반복 시뮬레이션을 수행하고, 불확실성이 반영된 출력 분포를 비교·분석하였다[14]. 이러한 접근은 단일 값 비교가 아닌, 결과 분포를 기반으로 모델 예측의 신뢰성을 평가하는 방식으로 구현되었다.

반면, 코드 검증 활동(SQA, NCV, 이산화 오차)은 모델 위험도 수준과 관계없이 공통적으로 수행되는 양상을 보였다. 상용

소프트웨어를 사용한 연구에서는 제조사가 제공하는 품질보증 자료 및 벤치마크 문제를 활용하였고[12,13,15-17], 상용 소프트웨어를 사용했다고 언급하지 않은 경우에는 자체 버전 관리나 버그 추적과 같은 내부 품질보증 활동이 보고되었다[11,18]. NCV와 이산화 오차 분석 역시 가능한 범위 내에서 거의 모든 연구들이 수행하였다.

4. 고찰

CM&S의 “Credibility(신뢰성)” 개념은 1970년대 후반부터 시뮬레이션 학계에서 논의되기 시작하였으며[19,20], 의료기기 분야에서는 최근 들어 그 중요성이 빠르게 확대되고 있다. CM&S는 임상시험 부담을 완화하고 환자 맞춤형 평가를 가능하게 하는 동시에 개발 비용과 시간을 절감할 수 있어, 연구자·산업계·규제기관 모두의 관심을 받고 있다[21,22]. 이러한 흐름 속에서 ASME V&V 40 표준은 CM&S 신뢰성을 확보하기 위한 핵심적인 참조 틀로 자리매김하였으며, 의료기기 분야의 혁신을 뒷받침하는 중요한 기반이 되고 있다. 본 연구는 이러한 맥락에서 의료기기 CM&S 사례를 체계적으로 검토하고, ASME V&V 40의 실무적 적용 가능성과 동시에 남아 있는 과제를 확인함으로써 향후 발전 방향을 제시한다.

검증 활동은 대부분의 연구에서 기본적인 절차로 수행되었으나, 수행 주체와 구현 방식에서는 차이가 나타났다. 상용 소프트웨어를 활용한 연구에서는 제조사가 제공하는 품질보증 체계에 대한 신뢰를 기반으로 검증 활동이 수행되는 경향이 관찰되었으며[12,13], 연구자가 코드 수준에서 독립적인 품질보증 활동을 직접 수행하지 않는 경우도 다수 확인되었다. 반면, 자체 개발되었거나 상용 소프트웨어를 수정하여 사용한 연구에서는 버전 관리, 내부 QA, COU와 연관된 오류 관리와 같은 활동이 보고되었다. 이러한 차이는 검증 활동에서 연구자가 직접 수행해야 하는 요소와 제3자 품질보증에 의존할 수 있는 요소가 혼재되어 있음을 보여준다.

유효성 확인에서는 불확실성 정량화와 비교 평가의 기준 설정이 여전히 과제로 남아 있었다. 각 분야 특성에 따라 적용 지표가 달라 통일된 기준을 제시하기는 어렵지만, 기존 연구를 충분히 조사·분석하여 재현 가능한 평가 기준을 마련하는 후속 연구가 필요하다. 또한 유효성 확인 수행과 관련하여, 일부 연구에서는 시험 조건 구성의 제한과 비용 및 기간의 부담으로 인해 비교자(시험 시료)를 소수만 확보할 수 있었으며, 이에 따라 유효성 확인 활동이 제한된 범위에서 수행되었다[11,18]. 이러한 수행 범위는 모델 위험도가 낮거나, CM&S 결과가 의사결정의 보조적 근거로 활용되는 경우에는 현실적으로 허용 가능한 수준으로 사료된다. 그러나 향후 CM&S 결과가 실제 시험 결과를 대체하거나 규제 의사결정에 직접적으로 활용되는 등 모델 위험도가 증가하는 적용 단계에서는, 이러한 제약을 그대로 유지하기는 어려울 것으로 예상된다. 비교자 수의 확대가 어려운 경우

시험 조건의 범위를 확장하거나, 반복 시험을 통해 변동성을 정량화하는 방식[11,12] 또는 COU에서 영향도가 높은 조건을 중심으로 한 선택적 유효성 확인 등을 통해 평가의 깊이를 단계적으로 보완할 필요가 있다.

분야별로 살펴보면, 혈류역학 분야는 모델과 비교자간의 비교가 가능하다는 장점으로 정량적 평가가 비교적 활발히 수행되었으나, 모델화의 복잡성과 높은 불확실성이 지속적인 과제로 나타났다. 고체역학 분야는 ISO 및 ASTM 등 국제 표준 시험을 기반으로 비교적 명확한 유효성 확인 절차를 적용할 수 있었으나, 실제 임상 환경의 복잡한 조건을 모두 반영하기에는 한계가 있었다. 약물동태학 등 임상 데이터 기반 연구에서는 임상적 QOIs를 활용하였으나, 후향적 자료에 의존하는 경우가 많아 유효성 확인 범위가 제한적으로 설정되었다. 다만 본 연구에서 분석한 문헌 수가 제한적이며, 전자기장 해석과 같이 의료기기 CM&S 활용이 증가하고 있는 일부 분야가 포함되지 않았다는 점에서 분야별 특성에 대한 논의에는 한계가 존재한다.

모델 위험도 수준에서는 낮은 위험도 모델은 제한적 불확실성 분석에 머문 반면, 높은 위험도 모델은 컴퓨터 모델과 비교자 모두에서 불확실성을 정량적으로 평가하고 이를 반영해 결과 비교를 수행하였다[14,15]. 이는 위험도 수준에 따라 불확실성 정량화의 깊이가 달라져야 한다는 점을 보여주며, 동시에 검증은 위험도와 관계없이 기본적으로 수행되는 핵심 절차를 시사한다.

본 연구는 의료기기 CM&S 분야에서 ASME V&V 40 표준이 신뢰성 확보를 위한 핵심적인 프레임워크로 활용되고 있음을 확인하는 한편, 실무 적용 과정에서 나타나는 구현 양상의 다양성과 한계를 함께 보여준다. 향후에는 다양한 의료기기 분야를 포괄하는 추가 사례 분석을 통해 ASME V&V 40 요소별 적용 범위와 구현 수준에 대한 이해를 확장할 필요가 있다. 이러한 노력이 축적될 경우, CM&S 기반 의료기기 개발과 평가에서 보다 일관된 신뢰성 확보가 가능해질 것으로 기대된다.

5. 한계

본 연구는 2025년 이전 연구를 기준으로 문헌 조사를 수행하였으며, 해당 시점까지 공개된 문헌을 기반으로 분석을 진행하였다. 의료기기 분야에 한정할 경우 ASME V&V 40 적용 사례가 부족하여, 조사 범위를 의료제품 전반으로 확장하였다. 그럼에도 불구하고 최종적으로 포함된 문헌은 10편에 불과하여, 분석 결과의 일반화에는 한계가 있을 수 있다.

또한 본 연구는 공개된 학술 문헌에 기술된 내용을 기반으로 분석을 수행하였다. 이로 인해 실제 연구 과정에서 수행되었을 수 있는 신뢰성 확보 활동이 논문 본문에 충분히 보고되지 않았거나, 부록 자료에만 제한적으로 제시된 경우까지 모두 반영하지는 못하였다. 예를 들어, 일부 문헌[11]에서는 ASME V&V 40의 신뢰성 목표와 요소별 달성 수준을 명시적으로 보고하였으나, 이러한

수준의 상세한 보고는 다른 문헌에서 공통적으로 제공되지는 않았다.

마지막으로, 본 연구에서 분석한 문헌들은 의료제품 CM&S 활용의 일부 기술 분야에 국한되어 있다. 분석 대상은 주로 혈류역학 및 고체역학 분야에 집중되어 있었으며, 전자기장 해석과 같이 최근 의료기기 분야에서 CM&S 활용이 증가하고 있는 일부 영역은 본 연구의 검색 조건 하에서 확인되지 않아 포함되지 않았다. 따라서 본 연구의 분석 결과는 의료기기 CM&S 전 분야를 포괄하는 일반적 결론으로 해석되기에는 범위의 한계를 갖는다. 향후 연구에서는 보다 다양한 의료기기 기술 분야와 적용 사례를 포함한 분석을 통해, ASME V&V 40 표준의 적용 범위와 수행 전략을 확장·보완할 필요가 있다.

6. 결론

본 연구는 ASME V&V 40 표준을 기반으로 의료기기 분야의 CM&S 사례를 분석하고 신뢰성 확보를 위해 실제로 수행된 활동을 정리하여 요소별 수행 전략을 제시하였다. 향후 연구에서는 평가 기준의 구체화와 새로운 분야로의 확대 적용을 통해 ASME V&V 40의 실효성을 높이는 노력이 요구된다. 이를 통해 CM&S는 의료기기 개발에서 더욱 신뢰성 높은 도구로 자리매김하여 규제과학 발전과 환자 안전 강화에 기여할 것으로 기대된다.

ACKNOWLEDGEMENT

본 연구는 2025년도 식품의약품안전처의 연구개발비(24204MFDS197)로 수행되었으며 이에 감사드립니다.

REFERENCES

- Morrison, T. M., Pathmanathan, P., Adwan, M., Margerrison, E., (2018), Advancing regulatory science with computational modeling for medical devices at the FDA's office of science and engineering laboratories, *Frontiers in Medicine*, 5, 241.
- Food and Drug Administration, (2016), Reporting of computational modeling studies in medical device submissions: guidance for industry and food and drug administration staff, (Docket No. FDA-2013-D-1530). <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/reporting-computational-modeling-studies-medical-device-submissions>
- American Society of Mechanical Engineers, (2018), V V 40 - Assessing credibility of computational modeling through verification and validation: application to medical devices. <https://www.asme.org/codes-standards/find-codes-standards/assessing-credibility-of-computational-modeling-through-verification-and-validation-application-to-medical-devices>
- Food and Drug Administration, (2023), Assessing the credibility of computational modeling and simulation in medical device submissions: guidance for industry and food and drug administration staff, (Docket No. FDA-2021-D-0980). <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/assessing-credibility-computational-modeling-and-simulation-medical-device-submissions>
- American Society of Mechanical Engineers, (2019), V V 10 - Standard for verification and validation in computational solid mechanics. <https://www.asme.org/codes-standards/find-codes-standards/standard-for-verification-and-validation-in-computational-solid-mechanics>
- American Society of Mechanical Engineers, (2009), V V 20 - Standard for verification and validation in computational fluid dynamics and heat transfer. <https://www.asme.org/codes-standards/find-codes-standards/standard-for-verification-and-validation-in-computational-fluid-dynamics-and-heat-transfer>
- Wohlin, C., (2014), Guidelines for snowballing in systematic literature studies and a replication in software engineering, *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, 1-10.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., (2021), The PRISMA 2020 statement: An updated guideline for reporting systematic reviews, *BMJ*, 372.
- Dharia, M. A., Snyder, S., Bischoff, J. E., (2020), Computational model validation of contact mechanics in total ankle arthroplasty, *Journal of Orthopaedic Research*, 38(5), 1063-1069.
- Nagaraja, S., Loughran, G., Gandhi, A., Inzana, J., Baumann, A. P., Kartikeya, K., Homer, M., (2020), Verification, validation, and uncertainty quantification of spinal rod computational models under three-point bending, *Journal of Verification, Validation and Uncertainty Quantification*, 5(1), 011002.
- Santiago, A., Butakoff, C., Eguzkitza, B., Gray, R. A., May-Newman, K., Pathmanathan, P., Vu, V., Vázquez, M., (2022), Design and execution of a verification, validation, and uncertainty quantification plan for a numerical model of left ventricular flow after LVAD implantation, *PLoS Computational Biology*, 18(6), e1010141.
- McVeigh, C., Harewood, F., King, P., Driscoll, M., Kulkarni, S., Zhao, T., Goodin, M., Iles, T. L., Perspectives on heart valve modelling: Contexts of use, risk, validation, verification and uncertainty quantification and end-to-end example, in: *Heart valves: From design to clinical implantation*, Springer.
- Dreyer, M. J., Nasab, S. H. H., Favre, P., Amstad, F., Crockett, R., Taylor, W. R., Weisse, B., (2024), Experimental and computational evaluation of knee implant wear and creep under in vivo and ISO boundary conditions, *BioMedical Engineering OnLine*, 23(1), 130.
- Nagaraja, S., Loughran, G., Baumann, A. P., Kartikeya, K., Homer, M., (2024), Establishing finite element model credibility of a

pedicle screw system under compression-bending: An end-to-end example of the ASME V&V 40 standard, *Methods*, 225, 74-88.

15. Morrison, T. M., Hariharan, P., Funkhouser, C. M., Afshari, P., Goodin, M., Horner, M., (2019), Assessing computational model credibility using a risk-based framework: Application to hemolysis in centrifugal blood pumps, *Asaio Journal*, 65(4), 349-360.
16. Aldieri, A., Curreli, C., Szyszko, J. A., La Mattina, A. A., Viceconti, M., (2023), Credibility assessment of computational models according to ASME V&V40: Application to the bologna biomechanical computed tomography solution, *Computer Methods and Programs in Biomedicine*, 240, 107727.
17. Carl, A.-K., Kirillov, M., Hochmann, D., Quadrat, E., (2023), Towards credible computational models: Application of a risk based framework for establishing credibility, *Transactions on Additive Manufacturing Meets Medicine*, 5(1), 804-804.
18. Curreli, C., Di Salvatore, V., Russo, G., Pappalardo, F., Viceconti, M., (2023), A credibility assessment plan for an in silico model that predicts the dose-response relationship of new tuberculosis treatments, *Annals of Biomedical Engineering*, 51(1), 200-210.
19. Schlesinger, S., (1979), Terminology for model credibility, *Simulation*, 32(3), 103-104.
20. Balci, O., (1986), Credibility assessment of simulation results, *Proceedings of the 18th conference on Winter simulation*, 38-44.
21. Pathmanathan, P., Gray, R. A., Romero, V. J., Morrison, T. M., (2017), Applicability analysis of validation evidence for biomedical computational models, *Journal of Verification, Validation and Uncertainty Quantification*, 2(2), 021005.
22. Pappalardo, F., Wilkinson, J., Busquet, F., Bril, A., Palmer, M., Walker, B., Curreli, C., Russo, G., Marchal, T., Toschi, E., (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(11), 5282-5286.



Ju-Yeon Lee

Researcher, Medical Device Research Division, Department of Medical Product Research, National Institute of Food and Drug Safety Evaluation.

E-mail: jylee0118@korea.kr



Tae-Hee Lee

Researcher, Medical Device Research Division, Department of Medical Product Research, National Institute of Food and Drug Safety Evaluation.

E-mail: 2taehee@korea.kr



Ju-Seon Lee

Researcher, Medical Device Research Division, Department of Medical Product Research, National Institute of Food and Drug Safety Evaluation.

E-mail: idcg01@korea.kr



So Hee Kim

Researcher, Medical Device Research Division, Department of Medical Product Research, National Institute of Food and Drug Safety Evaluation.

E-mail: tg5010@korea.kr



Hee Seon Heo

Researcher, Medical Device Research Division, Department of Medical Product Research, National Institute of Food and Drug Safety Evaluation.

E-mail: heoheeseon@korea.kr



Dong Hyun Go

Assistant Director, Medical Device Research Division, Department of Medical Product Research, National Institute of Food and Drug Safety Evaluation.

E-mail: godh@korea.kr



Hyeon Jeong Kim

Research Officer, Medical Device Research Division, Department of Medical Product Research, National Institute of Food and Drug Safety Evaluation.

E-mail: khjyoi@korea.kr



Hae Dae Park

Director, Medical Device Research Division, Department of Medical Product Research, National Institute of Food and Drug Safety Evaluation.

E-mail: hdspace@korea.kr



Su-Kyoung Lee

Assistant Director, Medical Device Research Division, Department of Medical Product Research, National Institute of Food and Drug Safety Evaluation.

E-mail: sk1218@korea.kr