

From RECIST to Reality: Replacing 2D Judgment with Quantifiable Tumor Intelligence

Sean Khozin, MD, MPH

Chief Executive Officer

CEO Roundtable on Cancer

Project Data Sphere





RECIST: Response Evaluation Criteria in Solid Tumors

Standardized guidelines introduced to measure tumor response in oncology clinical trials.



Measurement Methodology

RECIST simplifies complex disease by measuring the **longest diameter** of preselected target lesions (up to five, maximum two per organ).



Response Categories

Patients are classified into **complete response**, **partial response**, **stable disease**, or **progressive disease** based on changes in measurements.



Regulatory Standard

RECIST is the **primary endpoint standard** used in the approval of nearly all modern cancer drugs by FDA and EMA.

Eisenhauer, E. A. *et al.* New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *European Journal of Cancer* (2009).

RECIST: A Regulatory Endpoint, Not a Clinical Reality



Regulatory Framework

While RECIST provides standardized measurement thresholds essential for clinical trials, most practicing oncologists—particularly in community settings—rarely incorporate the criteria into their treatment decision-making process.



Clinical Reality

In real-world practice, treatment decisions emerge from clinical impressions ("likely progression," "stable disease"), patient functional status, symptom trajectories, and clinical intuition rather than rigid measurement protocols.



Science-Practice Disconnect

Analogy: This is equivalent to approving statins based on LDL cholesterol levels, while physicians prescribe them based on completely different endpoint.





The Limits of Human Vision in Tumor Assessments

Significant variability (30%) in
Tumor Assessments Based on
Gold Standard (RECIST)

CR = Complete Response ✓
PR = Partial Response ✓
SD = Stable Disease 😐
PD = Progressive Disease ☹️
NE = Not Evaluable ?

PRIMARY
INVESTIGATOR



INDEPENDENT
RADIOLOGIST

	CR	PR	SD	PD	NE	Total
CR	305	250	116	14	25	710
PR	240	3366	908	126	33	4673
SD	64	849	3522	668	74	5177
PD	9	58	507	1890	134	2598
NE	1	13	31	49	425	519
Total	619	4536	5084	2747	691	13677

Khozin S, et al. 2018. FDA Analysis. Unpublished.



The Reality Is Worse than 30%

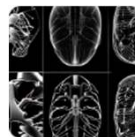


RECIST's 50% Buffer for Human Error

Progression: 20% growth from baseline

Response: 30% shrinkage from baseline

50% in between: zone of "stable disease"



Despite the buffer,..

There is a 30% interobserver variability in tumor assessments!

Higher variability when comparing two radiologists without the buffer



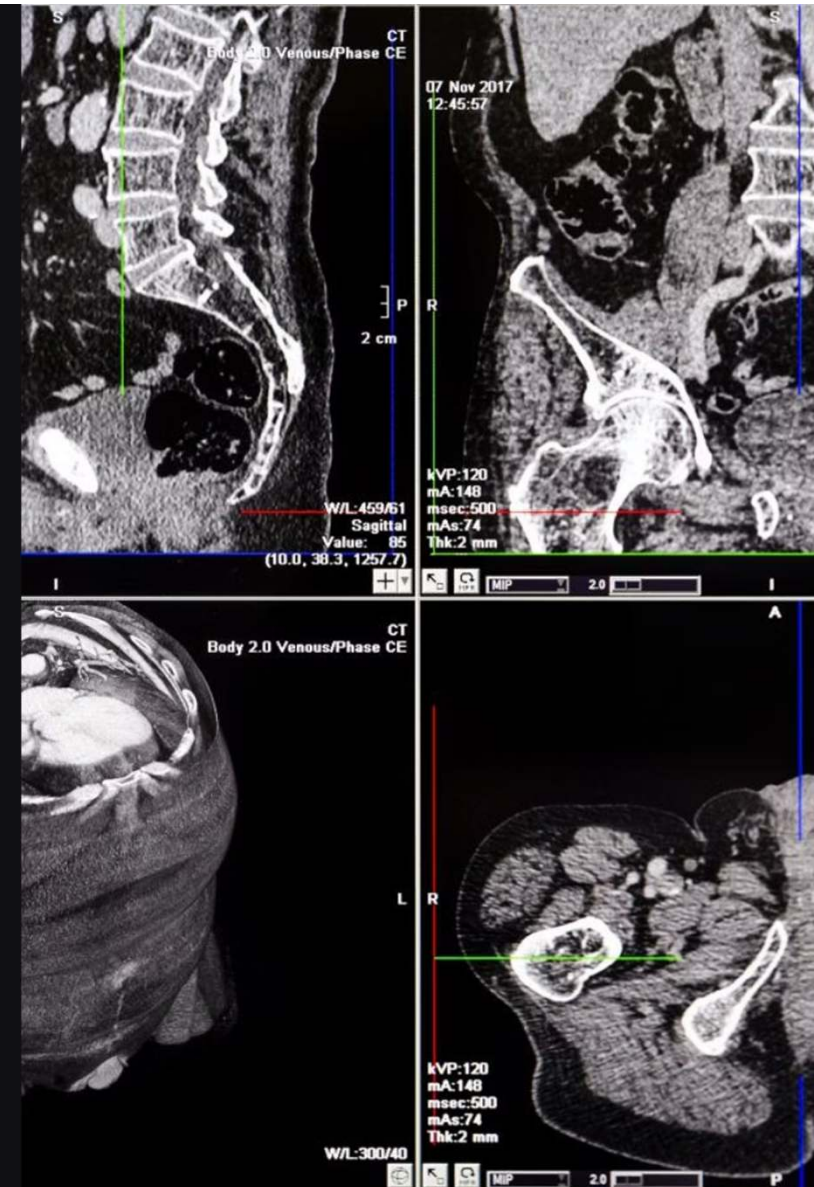
Other Shortcomings of "Gold Standard"

1 Inability to capture tumor heterogeneity

Tumors exhibit substantial morphological, functional, and molecular heterogeneity that cannot be adequately represented by unidimensional measurements.

2 Limited sensitivity to early changes

RECIST may not detect early signs of treatment response or resistance, delaying decision-making and potentially missing opportunities for optimizing treatment strategies.





The Billion-Dollar Paradox in Cancer Drug Development



Investment

We invest hundreds of millions of dollars to develop a single cancer drug, leveraging cutting-edge science and technology at every step.



Regulatory Approval

When it comes to the most critical stage for regulatory approval—assessing treatment efficacy in human trials—we rely on subjective visual tumor measurements prone to significant human error.



Reliability and Precision

This paradox introduces significant assessment volatility due to inherent measurement inconsistency, resulting in imprecise phase transition decision-making and delayed efficacy signal detection. Therapeutic effect quantification remains ambiguous until late-stage development, potentially leading to resource allocation decisions based on measurement artifact rather than true biological response.



FDA U.S. FOOD & DRUG ADMINISTRATION Project Data Sphere Stanford MEDICINE

FDA - PDS Symposium VIII

Artificial Intelligence in Tumor Imaging: Needs, Opportunities & Challenges

Monday, November 11, 2019
7:30 AM - 5:15 PM

Looking for agenda and more information about this event? Scan here.



Project Data Sphere

An initiative of the Life Sciences Consortium of the CEO Roundtable on Cancer

projectdatasphere.org





Developed and validated a robust AI framework for automated **Liver lesion** detection and segmentation, showcasing generalizability across diverse imaging datasets and modalities

Combined over **5,000 annotated patient cases**, representing one of the largest datasets used for lesion segmentation validation.

1 Improved Detection

Achieved significantly higher sensitivity than radiologists, demonstrating the ability to identify subtle and difficult-to-detect lesions.

2 Superior Segmentation

Delivered precision in lesion boundaries comparable to expert radiologists, while being faster and more reproducible.

3 Robust Consistency

Maintained consistent performance across different datasets, lesion sizes, and imaging modalities, proving generalizability and scalability.

4 Novel Architecture

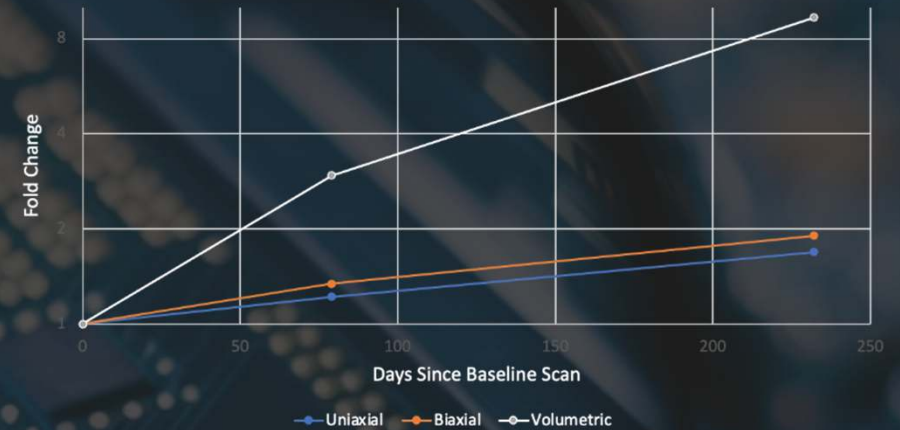
First-ever application of ScaleNAS-based architecture for lesion detection and segmentation, solidifying leadership in AI-driven oncology research.

Ma, J. *et al.* Generalizability of lesion detection and segmentation when ScaleNAS is trained on a large multi-organ dataset and validated in the liver. *Medical Physics*. 2025.



High Precision in Lung Lesion Detection and Segmentation

Successful application of AI in auto-detecting and volumetrically segmenting lung lesions with high precision.

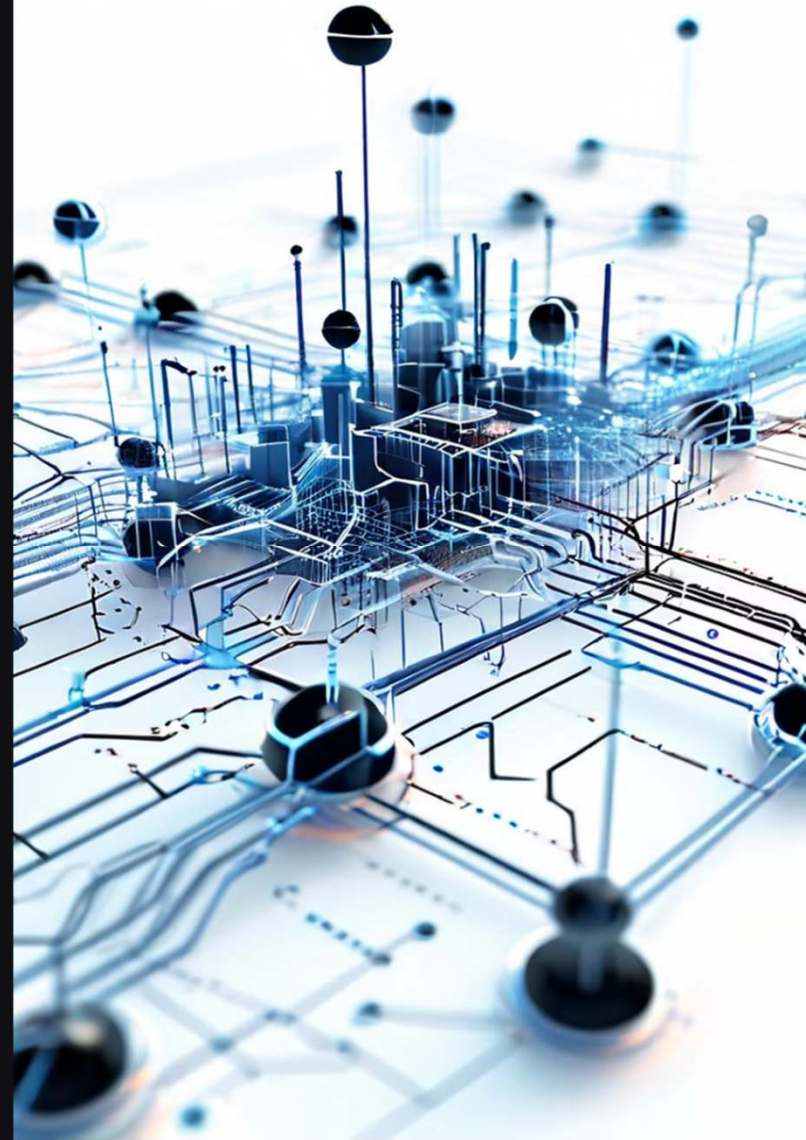


Automated segmentation results across three time points, comparing uniaxial, biaxial, and volumetric measurements.

Volumetric analysis provides the most detailed and informative assessment of tumor growth dynamics, showcasing its potential to surpass traditional methods in accuracy and insight.

Going Beyond RECIST: Harnessing the Power of Foundation Models

Total Tumor Burden index (TTBI)





TTBI: Automated Tumor Detection and Assessment

1 Automated Detection

AI-powered algorithms precisely locate and outline tumors

2 Volumetric Segmentation

Advanced 3D modeling captures tumor shape and size

3 Fusion and Scoring

Data integrated into a comprehensive, objective TTBI score

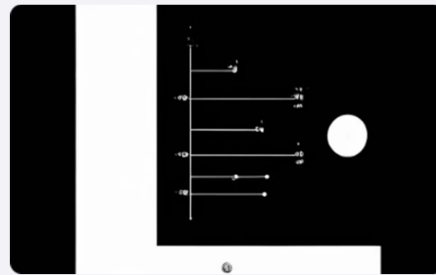


TTBI: Benefits



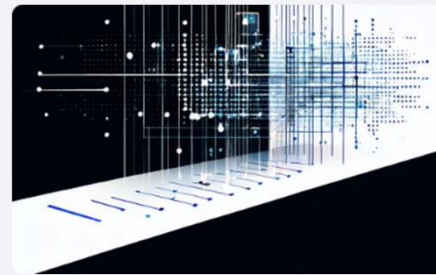
Enhanced Statistical Power

Reduces variability, captures tumor heterogeneity, enabling smaller trials and reproducible insights.



Accelerated Decisions

Early detection of response and resistance, faster go/no-go decisions, expediting promising therapies.



Precise Benefit Prediction

Reduces uncertainty in assessing treatment efficacy.



Personalized Strategies

Captures heterogeneity, identifies biomarkers, enables patient stratification, drives targeted therapies.



Validation and Clinical Studies

1

Initial Validation

Building on autoRECIST. Technical validation in 3 years.

2

Pilot Studies

Year 3-4

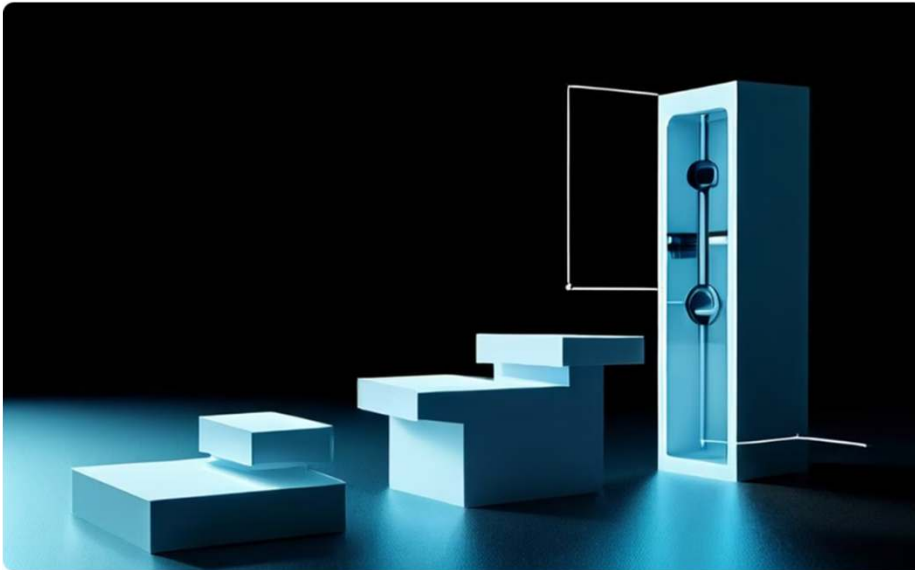
3

Multicenter Trials

Year 4-6, Multi-center clinical trials as autonomous agents.



Regulatory Pathway



510(k) Clearance

Streamlined 510(k) clearance process for the initial pilot stage, demonstrating substantial equivalence to existing devices for a faster path to regulatory approval and revenue generation.



PMA Approval

Following successful clinical validation, we will submit a PMA application to establish the autonomous agent technology as a new class of medical software, paving the way for widespread clinical adoption and reimbursement coverage.



Clinical Validation Is the Bottleneck for AI in Oncology

Despite rapid advances in model architecture and retrospective benchmarking, most AI solutions in oncology have failed to demonstrate **clinical-grade evidence of safety, effectiveness, and real-world utility**. The dominant development paradigm falls short of the evidentiary thresholds required for **regulatory trust, clinical adoption, or payer alignment**.



AI intended to guide clinical decisions must be held to the same evidentiary standard as therapeutics.

1 Randomized controlled trials (RCTs)

Remain the gold standard for demonstrating impact on clinical outcomes

2 Real-World Evidence (RWE)

Can support generalizability, health system integration, and post-market refinement—but is not sufficient for approval of novel endpoints or high-impact models

3 Clinical utility studies

Are critical to secure reimbursement and guideline inclusion

4 Regulatory pathways (510(k), De Novo, PMA)

Require pre-specified validation protocols, transparent data lineage, and generalizability across populations and devices

A close-up photograph of a microchip mounted on a printed circuit board (PCB). The chip is square with numerous pins extending from its edges. The PCB features intricate circuitry with gold-colored traces and various solder points. In the upper right corner, a small, glowing yellow star-like light is visible.

Thank You

