SURROGATE ENDPOINTS: AN INTRODUCTION

Oriana Ciani, PhD

Associate Professor of Practice

CERGAS - Center for Research on Health and Social Care Management

SDA Bocconi

CONFLICT AND DISCLOSURES

- I have participated in AB sponsored by MSD
- I have received research funding from the MRC and the EU Horizon programme
- I have no financial or any other conflicts relating to specific products mentioned in this presentation

OUTLINE

- DEFINITION OF SURROGATE ENDPOINT
- VALIDATION
- USE IN LICENSING AND REIMBURSEMENT

DEFINITION OF SURROGATE ENDPOINT

Disease-centered → Patient-centered

Biomarker
A characteristic that is objectively measured and evaluated as an indicator of normal, pathogenic or pharmacologic responses to a therapeutic intervention

Surrogate endpoint
A biomarker that is intended to **substitute** and **predict** for a final outcome

Target outcome
«A characteristic that reflects how patients feel, function or survive»

e.g., LDL-cholesterol → **Cardiovascular Mortality**

e.g., Bone Mineral Density → **Bone fracture**

SOME ISSUES WITH PROPOSED DEFINITIONS

Source (year)	Definition	Scoping review citations	e-Delphi rating N = 195				Summary of free-text comments
			Median (IQR)	% of rating scores			
				1–3	4–6	7–9	
Prentice (1989)[33]	A response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint.	6 (19%) records	5 (3, 6)	29.6	58.6	**11.7**	Complex and statistical definition with limited usability in trial design—see comments of Definition 3 in Appendix.
Temple (1999)[34]	A laboratory measurement or physical sign that is used in therapeutic trials as a substitute for a clinically meaningful end point that is a direct measure of how a patient feels, functions, or survives and is expected to predict the effect of the therapy	10 (31%)	7 (5, 7)	**11.6**	31.4	57.0	Not inclusive as a surrogate endpoint extends beyond laboratory measurements and signs and their use is beyond therapeutic trials—see comments of Definition 2 in Appendix
NIH Biomarkers Definitions Working Group (2001)[35]	A biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence.	12 (38%)	7 (6,7)	**11.0**	31.8	57.2	Not inclusive as surrogate endpoints extend beyond biomarkers and clinical benefit measured could still be a surrogate endpoint—see comments of Definition 1 in Appendix
BEST (2016)[6]	An endpoint that is used in clinical trials as a substitute for a direct measure of how a patient feels; functions; or survives. A surrogate endpoint does not measure the clinical benefit of primary interest in and of itself; but rather is expected to predict that clinical benefit or harm based on epidemiologic; therapeutic; pathophysiologic; or other scientific evidence.	3 (9%)	8 (7, 9)	**0.6**	7.0	**92.4**	A comprehensive definition although use of 'predict' implies a validated surrogate endpoint—see comments of Definition 4 in Appendix
Ciani et al. (2017)[11]	A biomarker or intermediate outcome used to substitute for a patient or participant relevant final outcome (i.e., severe morbidity; health related quality of life or mortality) and reliably predicts benefit or harm based on epidemiologic; therapeutic; pathophysiologic; or other scientific evidence	Not applicable[a]	8 (7, 8)	**2.3**	14.0	**83.6**	Support for inclusion of intermediate outcome in definition; however, there is limited understanding of 'intermediate outcome'; not all trials seek to evaluate interventions based on severe morbidity, health related quality of life or death; and 'predict' implies a validated surrogate endpoint—see comments of Definition 5 in Appendix
Banff Workshop (2022)[25]	An endpoint replacing a clinical endpoint that constitutes a basis for reliably predicting a treatment effect on the clinical endpoint in a defined context of use.	Not applicable[a]	7 (5.5, 8)	7.8	33.5	58.7	No comments received

NIH: National Institutes of Health; BEST: Biomarkers, Endpoints, and other Tool. [a]Not identified in the scoping review; Bold highlighted; consensus reached; Italic highlighted: consensus not reached.

Table 1: **Surrogate endpoint definitions identified by scoping review and e-Delphi rating.**

Ciani et al. A framework for the definition and interpretation of the use of surrogate endpoints in interventional trials. EClinicalMedicine. 2023 Oct 17;65:102283.

PROPOSED DEFINITIONAL FRAMEWORK

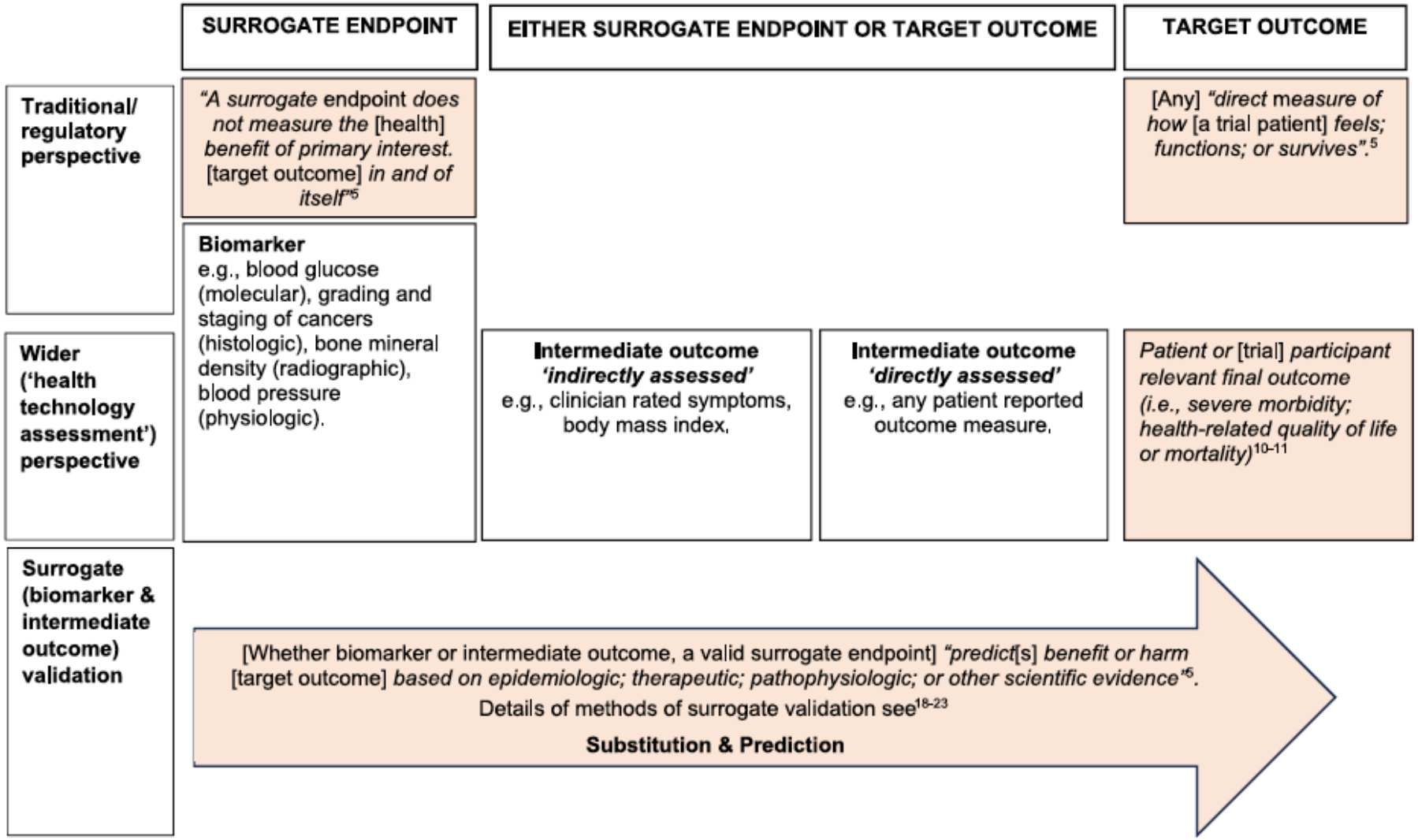


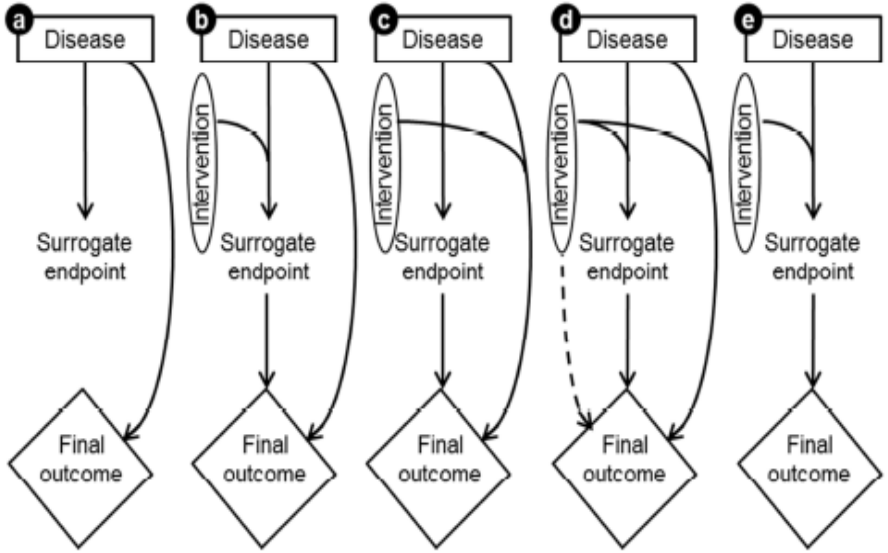
Fig. 3: Proposed framework for the definition and interpretation of surrogate endpoints in interventional trials.

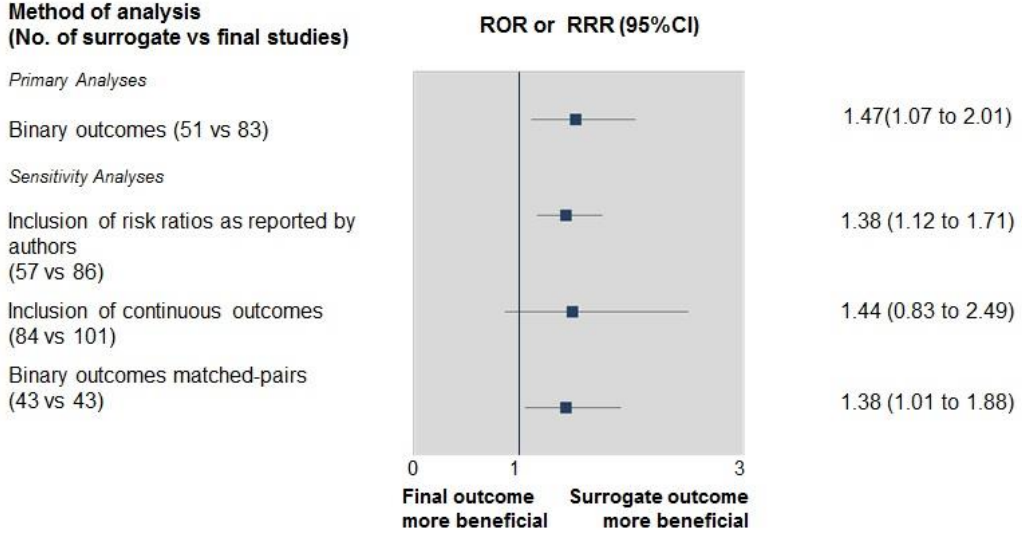
Ciani et al. A framework for the definition and interpretation of the use of surrogate endpoints in interventional trials. EClinicalMedicine. 2023 Oct 17;65:102283.

PROS AND CONS OF THE USE OF SURROGATE ENDPOINTS

- Shorter pivotal trials
- Smaller pivotal trials
- Cheaper pivotal trials
- Ethical reasons

Figure 1.1 Illustrations of different mechanisms for failure of surrogate endpoints





Ciani O, Buyse M, Drummond M, Rasi G, Saad ED, Taylor RS. Time to Review the Role of Surrogate End Points in Health Policy: State of the Art and the Way Forward. Value Health. 2017 Mar;20(3):487-495.

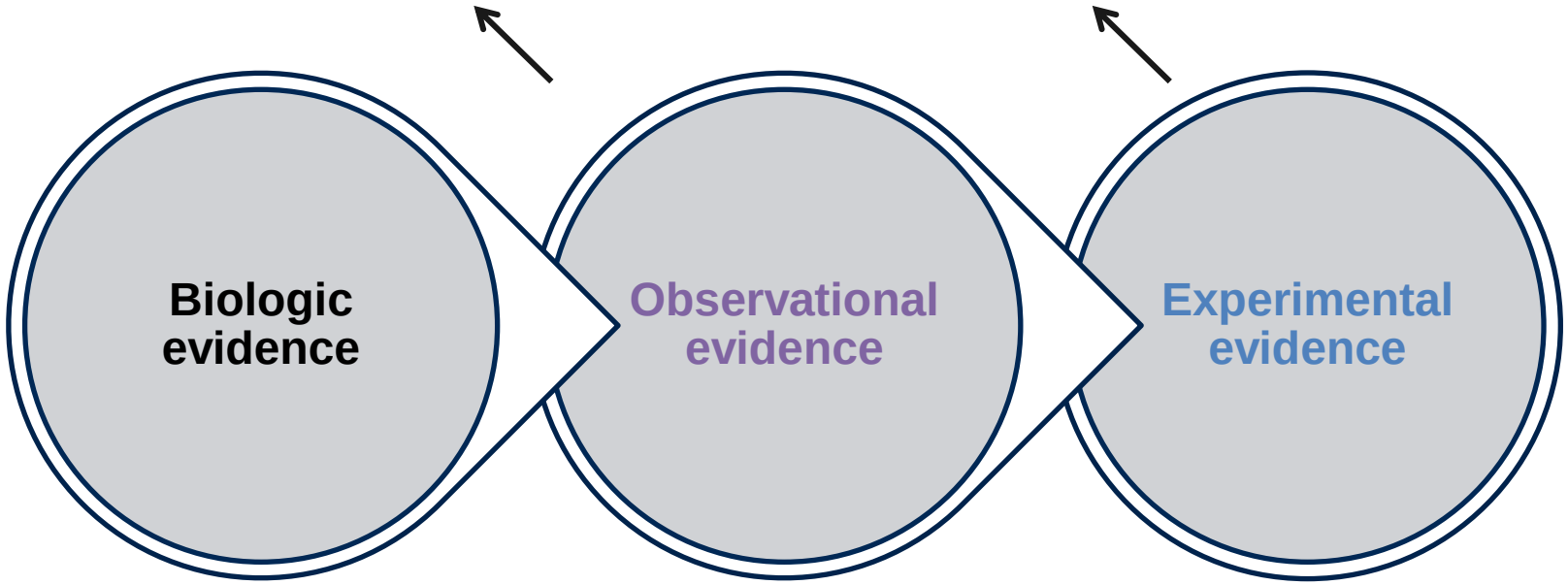
Ciani O, Buyse M, Garside R, Pavey T, Stein K, Sterne JA, Taylor RS. Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. BMJ. 2013 Jan 29;346:f457.

VALIDATION OF SURROGATE ENDPOINTS

- It is fundamental then to establish the **"validity" of a surrogate** i.e., the effect of the intervention on the replacement endpoint reliably predicts its effect on the patient-centered outcome

Association between the surrogate and the target outcome

Association between the treatment-induced change on the surrogate with the treatment-induced change on the target outcome



Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate end points--the challenge of statistical validation. Nat Rev Clin Oncol. 2010 Jun;7(6):309-17.

VALIDATION OF SURROGATE ENDPOINTS

STATISTICS IN MEDICINE
Statist. Med. 2006; **25**:183–203
Published online 26 October 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2319

Statistical evaluation of biomarkers as surrogate endpoints: a literature review

Christopher J. Weir[1,2,*,†] and Rosalind J. Walley[3]

[1]*Division of Cardiovascular and Medical Sciences, University of Glasgow, Gardiner Institute, Western Infirmary, Glasgow, G11 6NT, U.K.*
[2]*Robertson Centre for Biostatistics, University of Glasgow, University Avenue, Glasgow, G12 8QQ, U.K.*
[3]*Nonclinical Statistics, Pfizer Global Research and Development, Ramsgate Road, Sandwich, Kent, CT13 9NJ, U.K.*

Received: 18 January 2022 | Revised: 29 March 2022 | Accepted: 30 March 2022
DOI: 10.1002/pst.2219

WILEY

SPECIAL ISSUE PAPER

Informed decision-making: Statistical methodology for surrogacy evaluation and its role in licensing and reimbursement assessments

Christopher J. Weir[1] | **Rod S. Taylor**[2]

USE IN LICENSING

Regulatory agencies issue approvals for new drugs and biologics that have demonstrated safety and efficacy in “adequate and well-controlled studies”

Pivotal trials are the most critical of these trials. These should provide evidence of **patient benefit derived directly from patient-centered outcomes** (e.g., overall survival or health-related quality of life)

Special regulatory programs codify for special evidentiary standards, such as the use of surrogate endpoints (e.g., Accelerated Approval program at the FDA)

Table 2. Characteristics of Pivotal Trials Supporting New Drugs and Biologics Approved by the US Food and Drug Administration From 1995 to 1997, 2005 to 2007, and 2015 to 2017, Overall and Stratified by Special Regulatory Program Use and Orphan Designation

		% (95% CI)							
				Comparator			End points		
Characteristic	Trials, No.	Randomized	Double-blinded	Active	Placebo	None	Clinical	Scale	Surrogate
Overall									
1995-1997	401	93.6 (90.7-95.8)	79.4 (75.0-83.3)	44.1 (39.2-49.2)	47.4 (42.4-52.4)	8.5 (5.9-11.6)	43.8 (38.8-48.8)	8.0 (5.5-11.1)	48.3 (43.3-53.3)
2005-2007	141	82.2 (74.9-88.2)	67.4 (58.8-75.0)	34.0 (26.3-42.5)	48.2 (39.7-56.8)	17.7 (11.8-25.1)	28.4 (21.1-36.6)	11.3 (6.6-17.8)	60.3 (51.7-68.4)
2015-2017	253	82.2 (76.9-86.7)	67.6 (61.4-73.3)	29.2 (23.7-35.3)	53.0 (46.6-59.2)	17.8 (13.3-23.1)	23.3 (18.3-29.0)	17.4 (12.9-22.6)	59.3 (53.0-65.4)
3-Way *P* value	NA	<.001	<.001	<.001	.17	<.001	<.001	<.001	.004
2-Way *P* value[a]	NA	<.001	<.001	<.001			<.001		
Special regulatory program									
Any									
1995-1997	89	80.9 (71.2-88.5)	74.4 (63.6-83.4)	37.6 (27.4-48.8)	43.5 (32.8-54.7)	18.8 (11.1-28.8)	20.0 (12.1-30.1)	5.9 (1.9-13.2)	75.3 (64.7-84.0)
2005-2007	64	75.0 (62.6-85.0)	56.3 (43.3-68.6)	35.9 (24.3-48.9)	39.1 (27.1-52.1)	25.0 (15.0-37.4)	37.5 (25.7-50.5)	7.8 (2.6-17.3)	54.7 (41.7-67.2)
2015-2017	128	67.2 (58.3-75.2)	53.9 (44.9-62.8)	23.4 (16.4-31.7)	43.8 (35.0-52.8)	32.8 (24.8-41.7)	19.5 (13.1-27.5)	13.3 (7.9-20.4)	67.2 (58.3-75.2)
3-Way *P* value	NA	.02	.004	.02	.92	.02	.71	.07	.32
2-Way *P* value[a]	NA	.004	.003	.03			.22		
None									
1995-1997	316	96.1 (93.3-98.0)	80.7 (75.9-84.9)	45.9 (40.3-51.6)	48.4 (42.8-54.1)	5.7 (3.4-8.9)	50.2 (44.5-55.8)	8.6 (5.7-12.2)	41.3 (35.8-46.9)
2005-2007	77	88.3 (79.0-94.5)	76.6 (65.6-85.5)	32.5 (22.2-44.1)	55.8 (44.1-67.2)	11.7 (5.5-21.0)	22.1 (13.4-33.0)	14.3 (7.4-24.1)	64.9 (53.2-75.5)
2015-2017	125	97.6 (93.1-99.5)	81.6 (73.7-88.0)	35.2 (26.9-44.2)	62.4 (53.3-70.9)	2.4 (0.5-6.9)	27.2 (19.6-35.9)	21.6 (14.7-29.8)	51.2 (42.1-60.2)
3-Way *P* value	NA	.92	.96	.02	.007	.38	<.001	<.001	.014
2-Way *P* value[a]	NA	.44	.83	.02			<.001		
Orphan designation									
Yes									
1995-1997	36	80.6 (62.5-92.5)	61.3 (42.2-78.2)	25.8 (11.9-44.6)	45.2 (27.3-64.0)	29.0 (14.2-48.0)	48.4 (30.2-66.9)	9.7 (2.0-25.8)	41.9 (24.5-60.9)
2005-2007	24	45.8 (25.6-67.2)	29.2 (12.6-51.1)	12.5 (2.7-32.4)	33.3 (15.6-55.3)	54.2 (32.8-74.4)	37.5 (18.8-59.4)	0.0 (0.0-14.2)	62.5 (40.6-81.2)
2015-2017	63	52.4 (39.4-65.1)	39.7 (27.6-52.8)	6.3 (1.8-15.5)	46.0 (33.4-59.1)	47.6 (34.9-60.6)	19.0 (10.2-30.9)	7.9 (2.6-17.6)	73.0 (60.3-83.4)
3-Way *P* value	NA	.02	.09	.009	.80	.13	.003	.94	.004
2-Way *P* value[a]	NA	.009	.05	.02			.009		
No									
1995-1997	372	94.7 (91.9-96.8)	80.9 (76.5-84.9)	45.7 (40.5-50.9)	47.6 (42.4-52.8)	6.8 (4.4-9.8)	43.4 (38.2-48.6)	7.9 (5.3-11.1)	48.8 (43.6-54.0)
2005-2007	117	89.7 (82.8-94.6)	75.2 (66.4-82.7)	38.5 (29.6-47.9)	51.3 (41.9-60.6)	10.3 (5.4-17.2)	26.5 (18.8-35.5)	13.7 (8.0-21.3)	59.8 (50.4-68.8)
2015-2017	190	92.1 (87.3-95.5)	76.8 (70.2-82.6)	36.8 (30.0-44.1)	55.3 (47.9-62.5)	7.9 (4.5-12.7)	24.7 (18.8-31.5)	20.5 (15.0-27.0)	54.7 (47.4-62.0)
3-Way *P* value	NA	.17	.21	.04	.08	.53	<.001	<.001	.12
2-Way *P* value[a]	NA	.22	.26	.14			<.001		

Abbreviation: NA, not applicable.

[a] Two-way *P* value was calculated for differences between 1995 to 1997 and 2015 to 2017 periods.

Zhang AD, Puthumana J, Downing NS, Shah ND, Krumholz HM, Ross JS. Assessment of Clinical Trials Supporting US Food and Drug Administration Approval of Novel Therapeutic Agents, 1995-2017. JAMA Netw Open. 2020;3(4):e203284.

USE IN LICENSING AND REIMBURSEMENT

"…more recent FDA approvals of new drugs and biologics were based on **fewer pivotal trials**, which, when aggregated by indication, had **less rigorous designs** but longer trial durations, suggesting an ongoing need for continued evaluation of therapeutic safety and efficacy after approval"

Pivotal trials inform

Regulatory authorization

Market Access/ Reimbursement

Clinical guidelines

Zhang AD, Puthumana J, Downing NS, Shah ND, Krumholz HM, Ross JS. Assessment of Clinical Trials Supporting US Food and Drug Administration Approval of Novel Therapeutic Agents, 1995-2017. JAMA Netw Open. 2020;3(4):e203284.

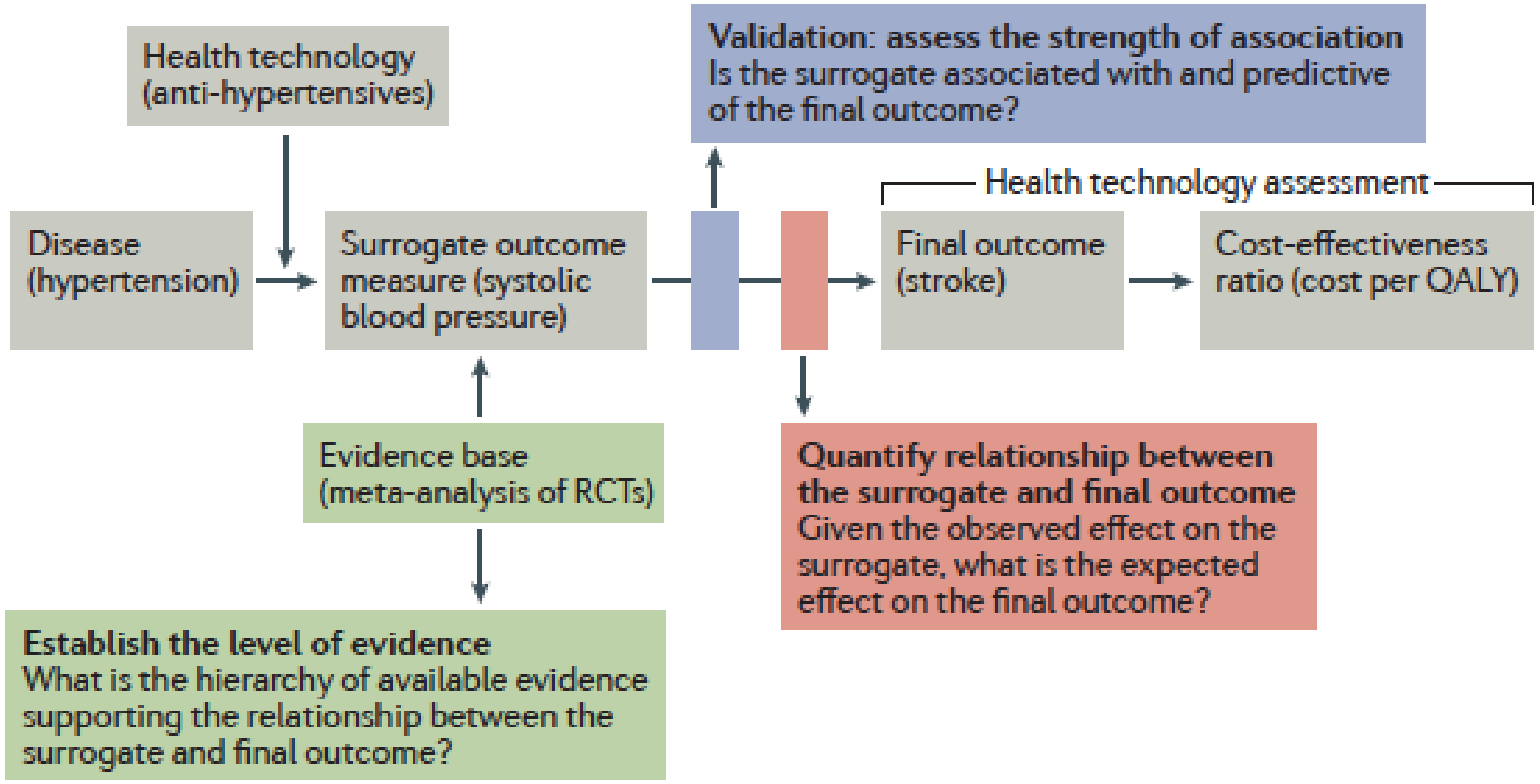
WHAT DO HTA METHODS GUIDELINES CURRENTLY RECOMMEND ON SURROGATES?

44 (98%)	**Argument around use of surrogates in the analysis**	"Surrogate endpoints should be adequately validated: the surrogate–final endpoint relationship must have been demonstrated based on biological plausibility and empirical evidence."*
18 (40%)	**Provide specific examples**	"Example of surrogate endpoints: biomarkers (e.g. cholesterol level, HbA1c); examples of intermediate endpoints: disease-free survival, angina frequency, exercise tolerance"*
13 (29%)	**Give a definition for surrogate endpoint**	"A biomarker can be defined as a characteristic that is objectively (reliably and accurately) measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to an intervention"*
10 (22%)	**Report more detailed methods for the handling of surrogate endpoints**	"currently, there is no systematic, transparent and widely agreed-upon process of biomarker validation…correlation of the effects on the surrogate and the effects on the clinical endpoint based on meta-analyses of several RCTs, as well as the surrogate threshold effect"*
2 (4%)	**Refer to thresholds for validation**	"There is no clear consensus of which correlation values are sufficient to assume adequate surrogacy, but values of between about 0.85 and 0.95 are often discussed"*
3 (7%)	**Specific guidance for disease areas**	Oncology, PFS, treatment intent
3 (7%)	**Specific for MDs**	MTEP, MSAC, State Institute for Drug Control

*Endpoints used in relative effectiveness assessment of pharmaceuticals Surrogate Endpoints, EUnetHTA 2015

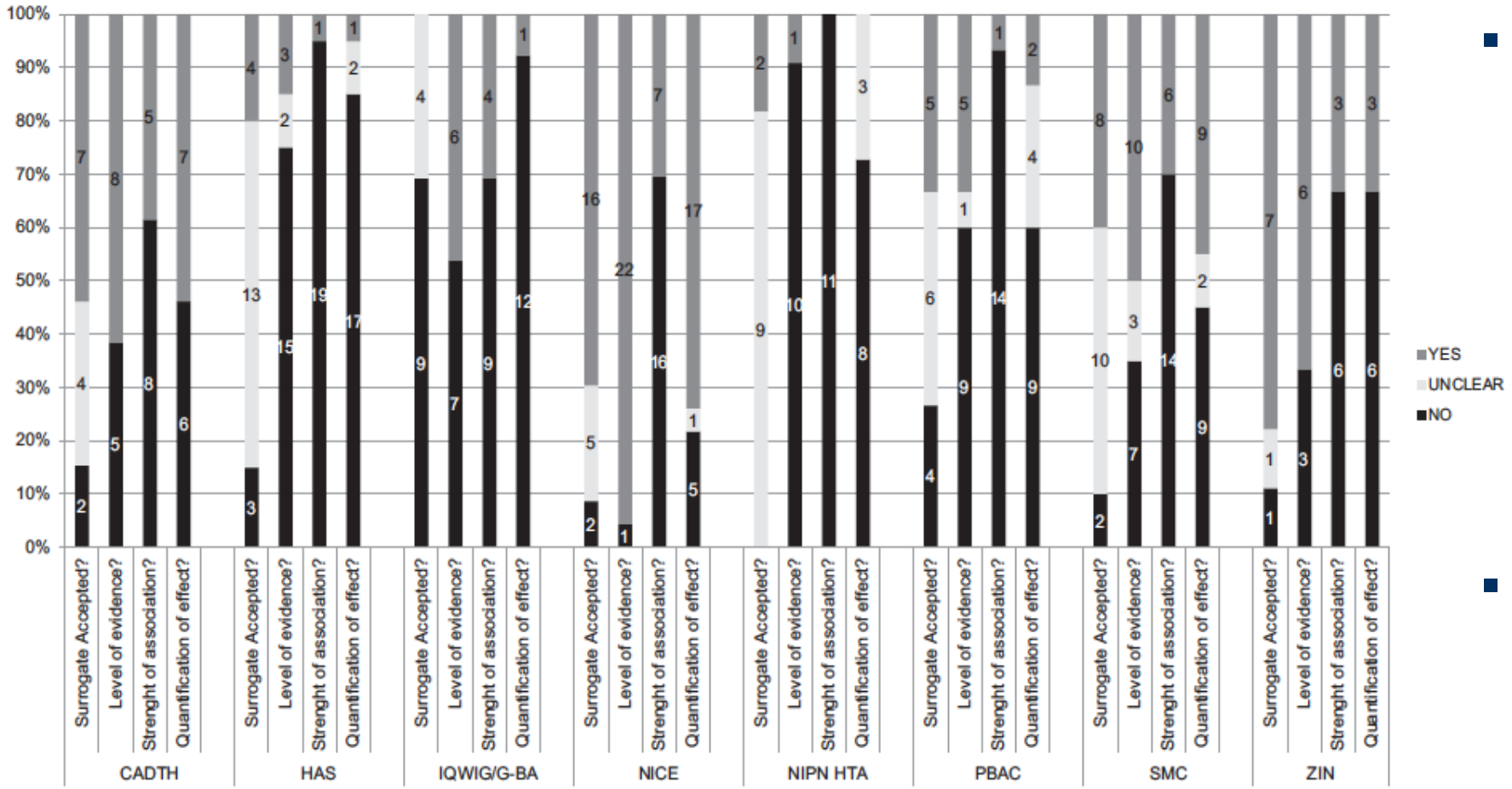
Grigore B et al. Surrogate Endpoints in Health Technology Assessment: An International Review of Methodological Guidelines. Pharmacoeconomics. 2020 Oct;38(10):1055-1070.

PROPOSED APPROACH TO SURROGATE ENDPOINTS IN HTA REPORTS



Ciani O, Buyse M, Drummond M, Rasi G, Saad ED, Taylor RS. Use of surrogate end points in healthcare policy: a proposal for adoption of a validation framework. Nat Rev Drug Discov. 2016 Jul;15(7):516.

HOW IS VALIDATION OF SURROGATE ENDPOINTS EMPIRICALLY ADDRESSED IN HTA REPORTS?



- The different level of scrutiny applied translates into different declared level of acceptability for the same surrogate endpoint, in mostly the same indication, and based on what is theoretically the same evidence available to each appraisal committee
- Overall, the level of agreement across the eight agencies was 0.10 (p = 0.04)

Ciani O, Grigore B, Blommestein H, de Groot S, Möllenkamp M, Rabbe S, Daubner-Bendes R, Taylor RS. Validity of Surrogate Endpoints and Their Impact on Coverage Recommendations: A Retrospective Analysis across International Health Technology Assessment Agencies. Med Decis Making. 2021 May;41(4):439-452.

ACKNOWLEDGEMENT

Reporting of surrogate endpoints in randomised controlled trial protocols: The SPIRIT-Surrogate extension checklist with explanation and elaboration

Anthony Muchai Manyara[1,2+], Philippa Davies[3], Derek Stewart[4], Christopher J Weir[5], Amber E. Young[3], Jane Blazeby[3,6,7], Nancy J. Butcher[8,9], Sylwia Bujkiewicz[10], An-Wen Chan[11,12], Dalia Dawoud[13,14], Martin Offringa[8, 15], Mario Ouwens[16], Asbjørn Hróbjartssson[17, 18], Alain Amstutz[19, 20, 21], Luca Bertolaccini[22], Vito Domenico Bruno[23], Declan Devane[24, 25], Christina DCM Faria[26], Peter B. Gilbert[27], Ray Harris[4], Marissa Lassere[28], Lucio Marinelli[29, 30], Sarah Markham[31], John H. Powers III[32], Yousef Rezaei[33, 34, 35], Laura Richert[36], Falk Schwendicke[37], Larisa G. Tereshchenko[38], Achilles Thoma[39], Alparslan Turan[40], Andrew Worrall[4], Robin Christensen[41], Gary S. Collins[42], Joseph S. Ross[43, 44], Rod S. Taylor*[1,45+], Oriana Ciani[46+]

UKRI Medical Research Council

The research was funded as part of development of SPIRIT and CONSORT extensions by the UK Medical Research Council (grant number MR/V038400/1).

Reporting of surrogate endpoints in randomised controlled trial reports: The CONSORT-Surrogate extension checklist with explanation and elaboration __

Anthony Muchai Manyara[1,2+], Philippa Davies[3], Derek Stewart[4], Christopher J Weir[5], Amber E. Young[3], Jane Blazeby[3,6,7], Nancy J. Butcher[8,9], Sylwia Bujkiewicz[10], An-Wen Chan[11,12], Dalia Dawoud[13,14], Martin Offringa[8, 15], Mario Ouwens[16], Asbjørn Hróbjartssson[17, 18], Alain Amstutz[19, 20, 21], Luca Bertolaccini[22], Vito Domenico Bruno[23], Declan Devane[24, 25], Christina DCM Faria[26], Peter B. Gilbert[27], Ray Harris[4], Marissa Lassere[28], Lucio Marinelli[29, 30], Sarah Markham[31], John H. Powers III[32], Yousef Rezaei[33, 34, 35], Laura Richert[36], Falk Schwendicke[37], Larisa G. Tereshchenko[38], Achilles Thoma[39], Alparslan Turan[40], Andrew Worrall[4], Robin Christensen[41], Gary S. Collins[42], Joseph S. Ross[43, 44], Rod S. Taylor*[1,45+], Oriana Ciani[46+]

Università Bocconi
CERGAS
Center for Research on Health and Social Care Management

SDA Bocconi
SCHOOL OF MANAGEMENT

Surrogacy

THANK YOU

oriana.ciani@unibocconi.it
@OrianaCiani

OUR PANEL TODAY

- The SPIRIT-Surrogate and CONSORT-Surrogate project – *Prof Rod S Taylor*
- The SPIRIT-Surrogate 2023 and the CONSORT-Surrogate 2023 extension checklist with explanation and elaboration – *Dr Anthony Manyara*
- The importance of right communication of the interpretation of the surrogate outcome to patients from a patient perspective – *Dr Ray Harris*
- On factors influencing probability of success for surrogacy validation - *Dr Mario Ouwens*
- Panel discussion and Q&A