A path to surveillance evaluation

Victor J Del Rio Vias

Zurich, 22 June 2017

On technical attributes

Journal of the Royal Statistical Society
SERIES C Applied Statistics

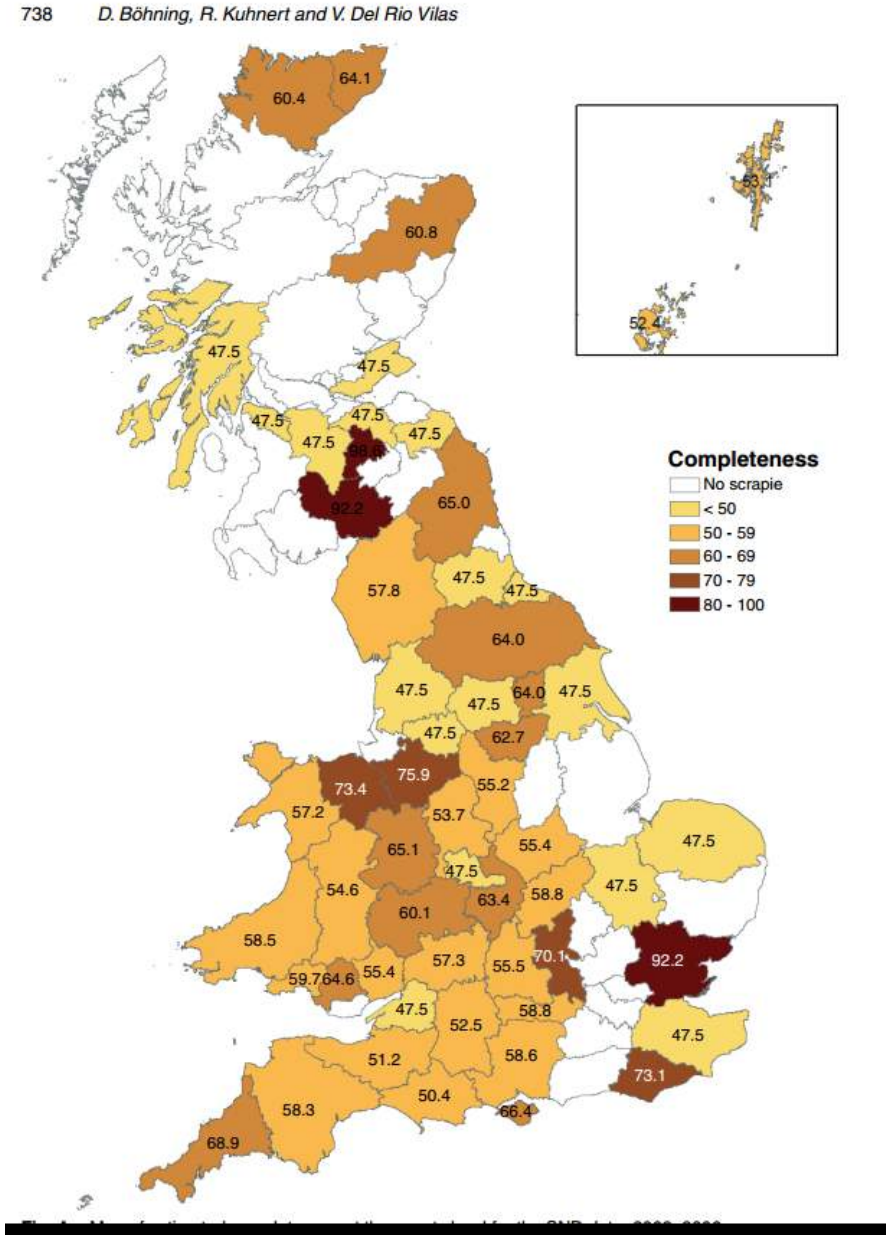
Appl. Statist. (2011)
60, *Part* 5, *pp.* 723–741

Capture–recapture estimation by means of empirical Bayesian smoothing with an application to the geographical distribution of hidden scrapie in Great Britain

Dankmar Böhning,
University of Reading, UK

Ronny Kuhnert
Robert Koch Institute, Berlin, Germany

and Victor Del Rio Vilas
Department for Environment, Food and Rural Affairs, London, UK



By means of comparisons: of sources & countries

Vet. Res. (2008) 39:37
DOI: 10.1051/vetres:2008014

www.vetres.org

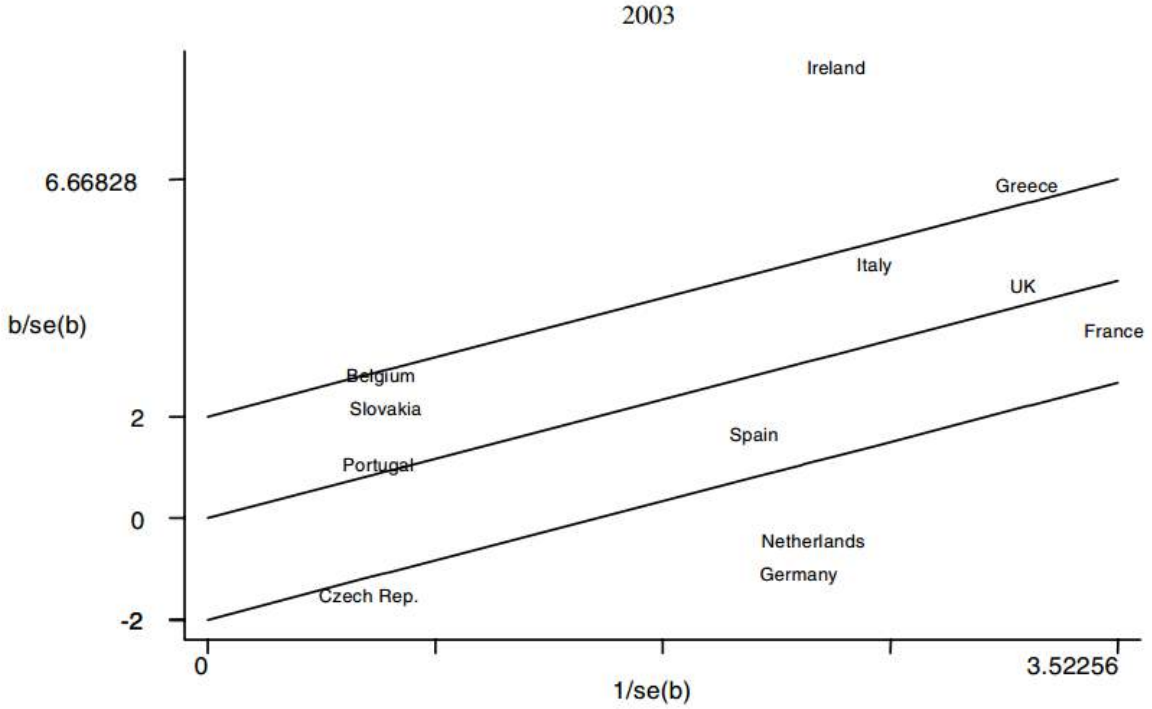
Original article

A comparison of the active surveillance of scrapie in the European Union

Victor Javier DEL RIO VILAS[1]*, Dankmar BÖHNING[2], Ronny KUHNERT[3]

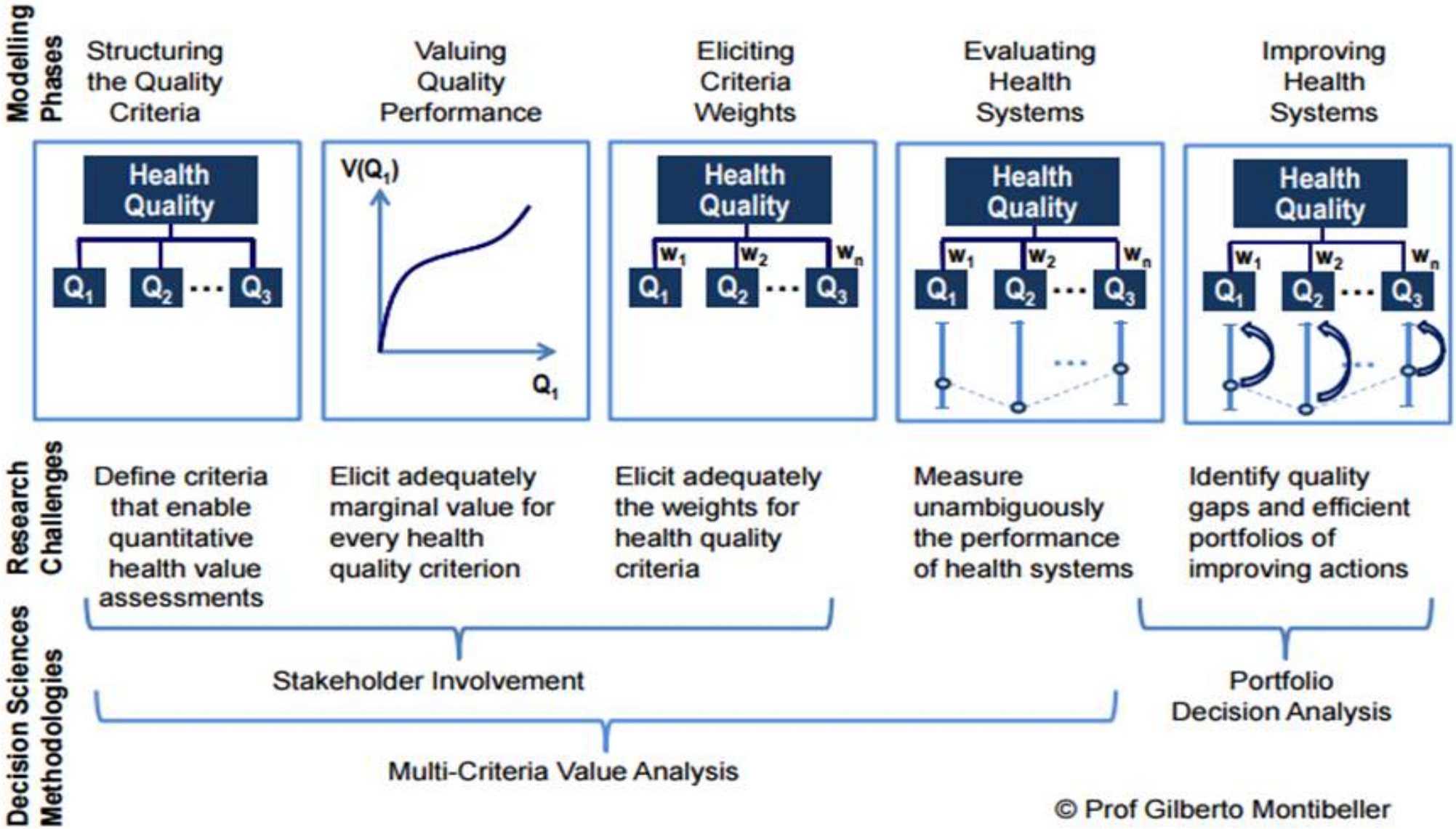
[1] Veterinary Laboratories Agency, New Haw, Addlestone, Surrey, KT15 3NB, UK
[2] Applied Statistics, School of Biological Sciences, The University of Reading, Harry Pitt Building, Earley Gate, Whiteknights road, Reading, RG6 6FN, UK
[3] Department of Epidemiology and Health, Robert Koch-Institute, Seestraβe 10, 13353 Berlin, Germany

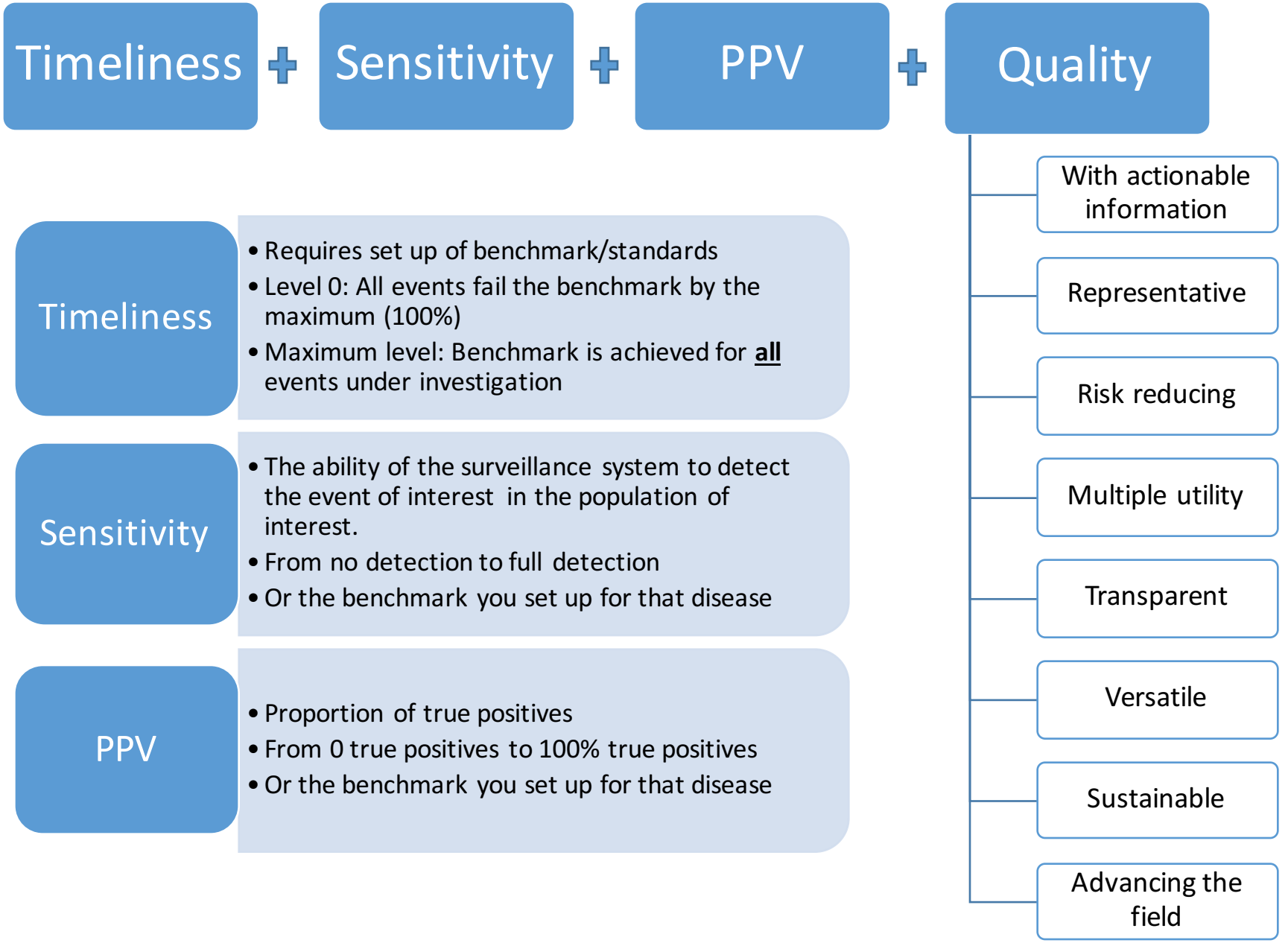
(Received 26 September 2007; accepted 28 February 2008)



Countries	Classical scrapie												Atypical scrapie			
	2003				2004				2005				2004		2005	
	x_{FS}	n_{FS}	x_{AS}	n_{AS}	x_{FS}	n_{FS}	x_{AS}	n_{AS}	x_{FS}	n_{FS}	x_{AS}	n_{AS}	x_{FS}	x_{AS}	x_{FS}	x_{AS}
Belgium	2	496	0	2 376	2	1 516	1	39	1	1 451	0	10	1	0	1	0
Denmark	0	1 320	0	871	0	5 253	0	91	0	4 295	0	97	0	0	0	0
Germany	13	48 629	9	20 116	42	65 488	1	15 628	18	29 550	8	14 894	0	0	0	0
Greece	13	793	49	22 613	17	2 098	4	6 508	34	1 597	13	4 484	0	0	0	0
Spain	7	12 950	13	49 940	7	10 799	7	15 051	17	14 881	14	14 274	1	0	1	1

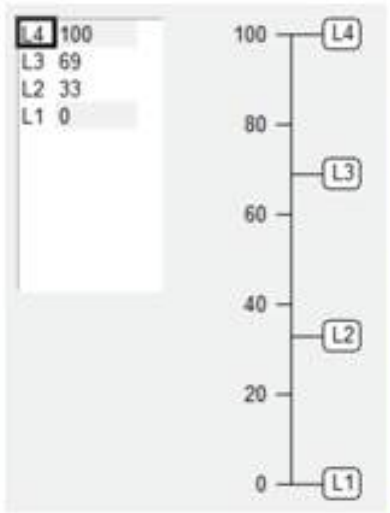
MCDA


Modelling Phases
Structuring the Quality Criteria
Valuing Quality Performance
Eliciting Criteria Weights
Evaluating Health Systems
Improving Health Systems
Health Quality
Q_1 Q_2 ... Q_3
$V(Q_1)$
Q_1
w_1 w_2 w_n
Research Challenges
Define criteria that enable quantitative health value assessments
Elicit adequately marginal value for every health quality criterion
Elicit adequately the weights for health quality criteria
Measure unambiguously the performance of health systems
Identify quality gaps and efficient portfolios of improving actions
Decision Sciences Methodologies
Stakeholder Involvement
Portfolio Decision Analysis
Multi-Criteria Value Analysis
© Prof Gilberto Montibeller

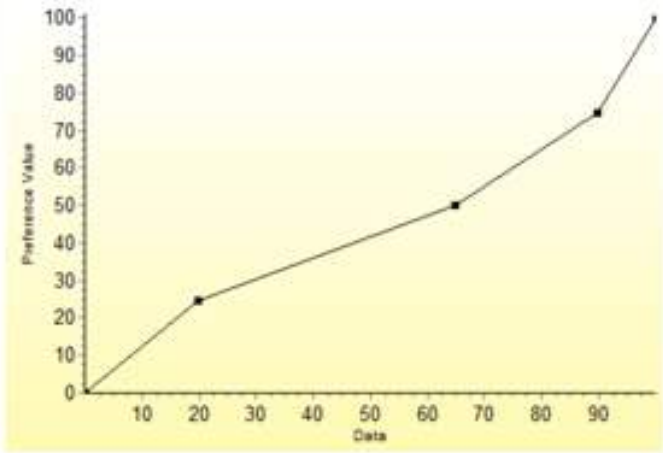

Timeliness
Sensitivity
PPV
Quality
Timeliness
• Requires set up of benchmark/standards
• Level 0: All events fail the benchmark by the maximum (100%)
• Maximum level: Benchmark is achieved for all events under investigation
Sensitivity
• The ability of the surveillance system to detect the event of interest in the population of interest.
• From no detection to full detection
• Or the benchmark you set up for that disease
PPV
• Proportion of true positives
• From 0 true positives to 100% true positives
• Or the benchmark you set up for that disease
With actionable information
Representative
Risk reducing
Multiple utility
Transparent
Versatile
Sustainable
Advancing the field

Code	Type of event	nario (the level of infectiousness & severity of condition & availability of prevention/treatme	Objective
CD1	Communicable event	Highly infectious & severe & no available treatment/prevention	Event detection
CM1	Communicable event	Highly infectious & severe & no available treatment/prevention	Event monitoring
CD2	Communicable event	Highly infectious & severe & with an available treatment/prevention	Event detection
CM2	Communicable event	Highly infectious & severe & with an available treatment/prevention	Event monitoring
CD3	Communicable event	Highly infectious & non-severe & no available treatment/prevention	Event detection
CM3	Communicable event	Highly infectious & non-severe & no available treatment/prevention	Event monitoring
CD4	Communicable event	Highly infectious & non-severe & with an available treatment/prevention	Event detection
CM4	Communicable event	Highly infectious & non-severe & with an available treatment/prevention	Event monitoring
CD5	Communicable event	Low infectiousness & severe & no available treatment condition /prevention	Event detection
CM5	Communicable event	Low infectiousness & severe & no available treatment condition /prevention	Event monitoring
CD6	Communicable event	Low infectiousness & severe & with an available treatment/prevention	Event detection
CM6	Communicable event	Low infectiousness & severe & with an available treatment/prevention	Event monitoring
CD7	Communicable event	Low infectiousness & non-severe & no available treatment/prevention	Event detection
CM7	Communicable event	Low infectiousness & non-severe & no available treatment/prevention	Event monitoring
CD8	Communicable event	Low infectiousness & non-severe & with an available treatment/prevention	Event detection
CM8	Communicable event	Low infectiousness & non-severe & with an available treatment/prevention	Event monitoring
NCD1	Non-communicable event	Severe & no available treatment/prevention	Event detection
NCM1	Non-communicable event	Severe & no available treatment/prevention	Event monitoring
NCD2	Non-communicable event	Severe & with an available treatment/prevention	Event detection
NCM2	Non-communicable event	Severe & with an available treatment/prevention	Event monitoring
NCD3	Non-communicable event	Non-severe & no available treatment/prevention	Event detection
NCM3	Non-communicable event	Non-severe & no available treatment/prevention	Event monitoring
NCD4	Non-communicable event	Non-severe & with an available treatment/prevention	Event detection
NCM4	Non-communicable event	Non-severe & with an available treatment/prevention	Event monitoring

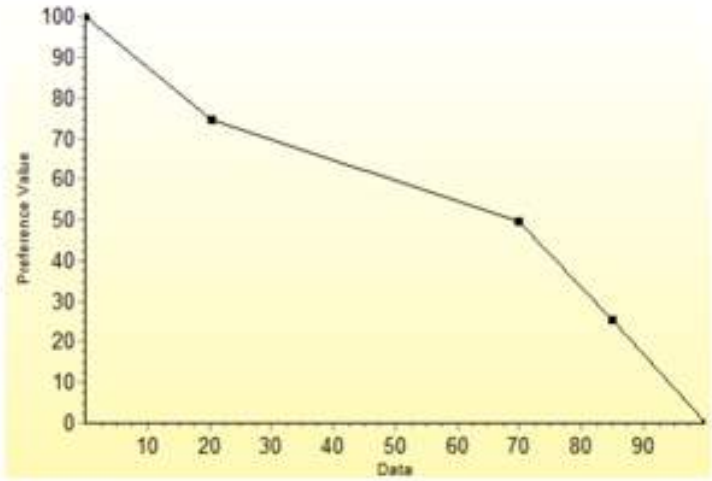
The value functions of timeliness and false positive rate below represent the only CD2 scenario, whereas the value functions of other attributes are valid for all possible scenarios.



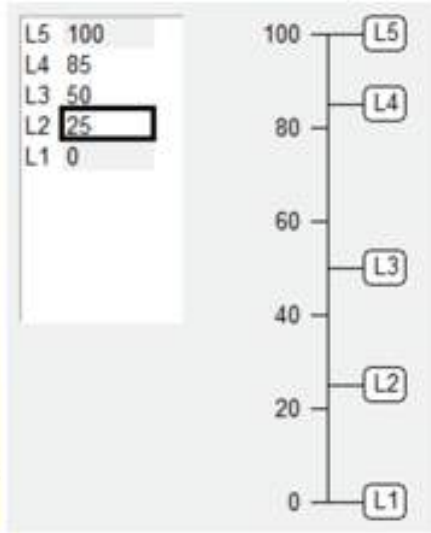
Transparency



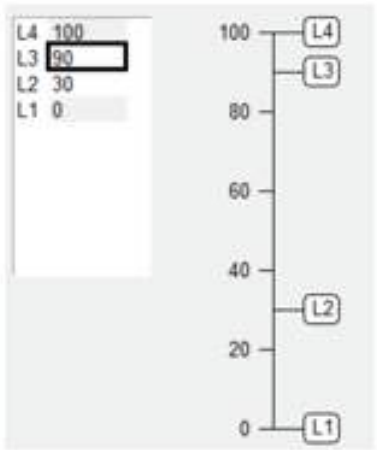
Timeliness



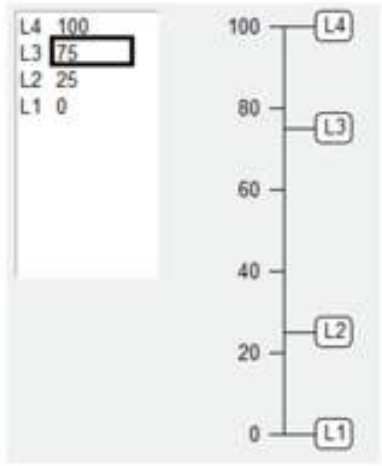
False positive rate



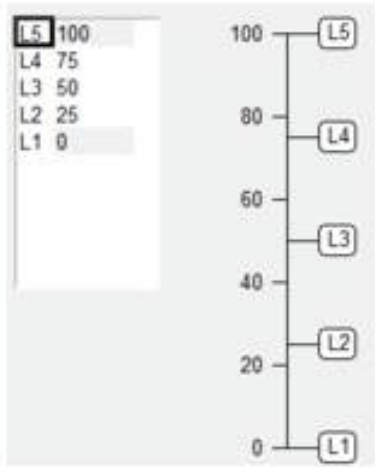
Actionable information



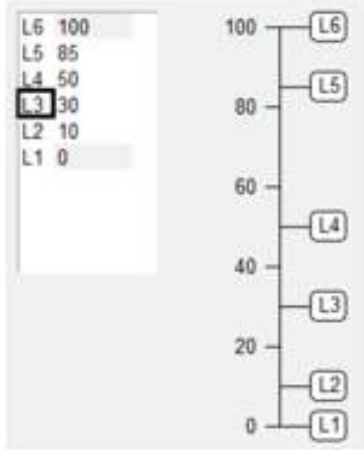
Risk Reducing



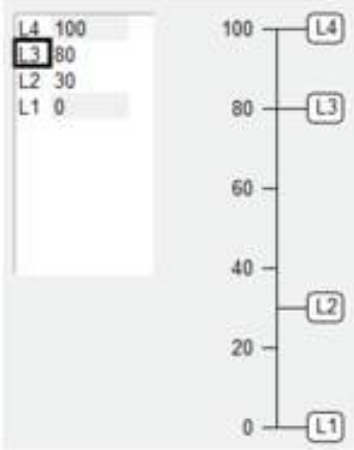
Multiple utility



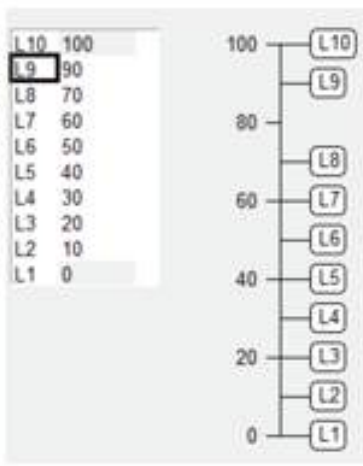
Adv. the field & Innovation



Sustainability



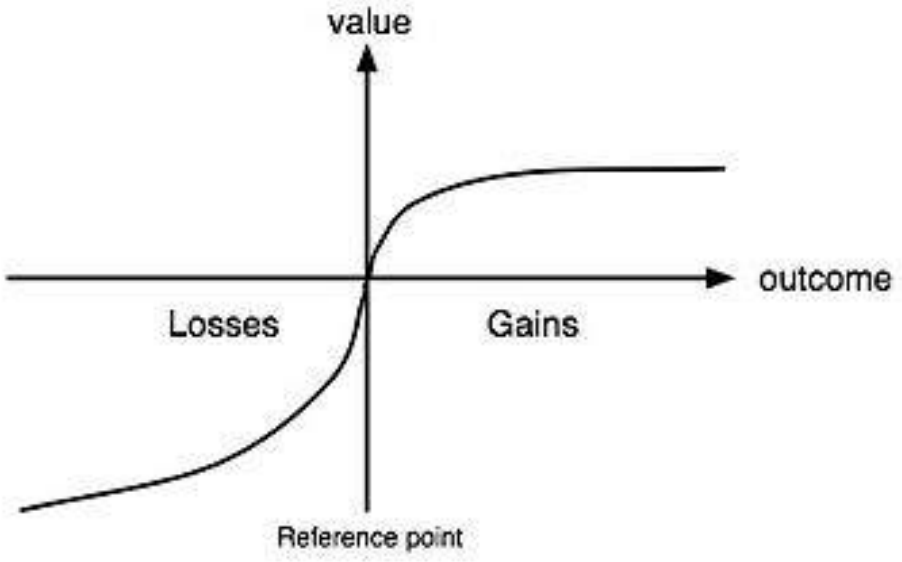
Representativeness



Versatility

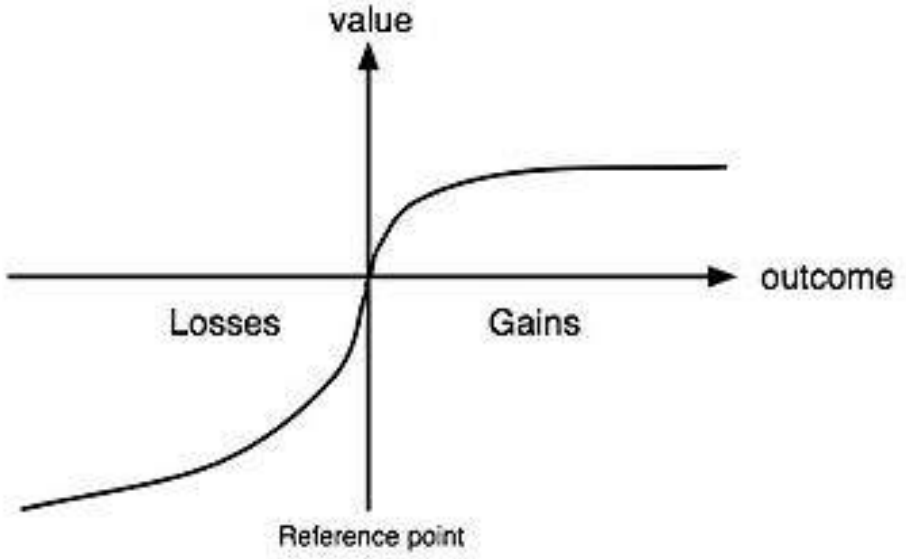
PT

- Several cognitive biases that cause descriptive violations of the expected utility theory, used in health economic evaluations, are described by PT.
- People tend to form reference points (RP) and regard outcomes as deviations from this RP. Hence, people are sensitive to *changes* in outcomes rather than to *final* outcomes.
- People make a distinction between outcomes above the RP (gains) and outcomes below it (losses). They perceive losses to loom much larger than gains of the same absolute magnitude, which results in a higher weight being attached to losses than to gains. This phenomenon is known as *loss aversion*.
- People have difficulties to process probabilities, which they transform nonlinearly into decision weights. This behavior is called *probability weighting* and often causes small probabilities to be overweighed, and large probabilities to be underweighted.



PT

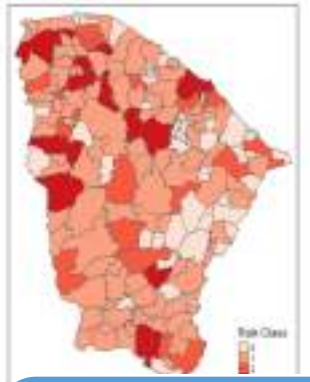
- Derive the target population loss aversion coefficient (and risk aversion) to de-bias interventions effects (surveillance timeliness in our case). Loss averse stakeholders might not invest in surveillance enhancements even if the utility of improved timeliness (the positive prospect) was larger than the disutility of increased false positive/negative results (the negative prospect);
- Elicitation of RPs (from homogenous groups/communities) to frame intervention-specific messages to encourage RP changes in ways that may increase the likelihood of accepting interventions, such as investments to improved timeliness;
- Loss aversion and framing: “lock in losses”, failing to realize losses and keep betting on because people favour risky prospects over sure prospects in the domain of losses.
- Risk aversion and framing: When a procedure is perceived as safe (e.g., sunscreen prevents sunburn and skin cancer), gain-framed messages are predicted to be more effective because people prefer sure prospects to risky prospects in the domain of gains



PT

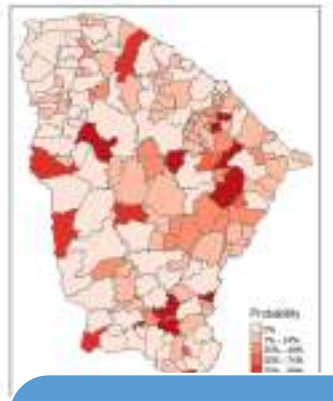
- Compute the amount of diminishing sensitivity of the target population to increasing surveillance timeliness enhancements;
- Compute the probability weighting function that would allow adjustment for the stakeholders' tendency to overestimate (underestimate) unlikely (very likely) extreme outcomes.
- Estimates of the willingness-to-pay by the target population for surveillance enhancements, specifically timeliness;
- INB and ICER adjusting for risk aversion (from our PT models) to compare different surveillance alternatives (with c_i costs and e_i effects (different from "benefits")).

- All models on VL data in Brazil



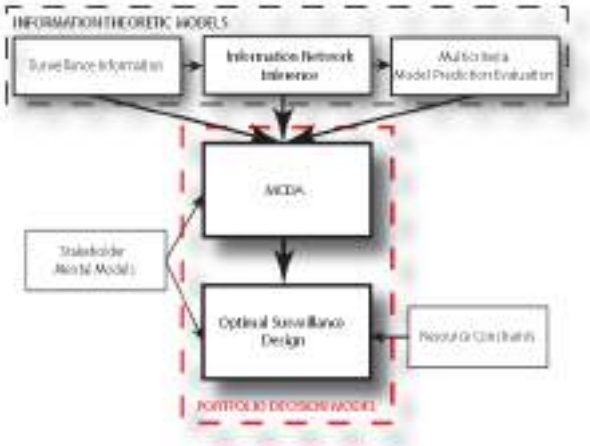
Risks

- Counts
- SIR
- Integrated risk



Capacities

- Ad-hoc
- Evaluated



Decision

- PDA
- Health outcomes of interest

Thanks

Surveillance Evaluation Framework

Thanks

#	Capacity
1	Sensitivity
2	Specificity
3	Timeliness
4	Accuracy
5	[illegible]
6	[illegible] programme
7	[illegible]
8	Sustainability
9	Versatility
10	[illegible] action?
11	Transparent
12	Health promoting
13	Risk reducing
14	[illegible]
15	[illegible]
16	[illegible]
17	Multiple utility
18	Advancing the field/Innovative
19	Easiness of integration

#	Capacity
1	Sensitivity
2	Specificity
3	Timeliness
4	With actionable information
5	Sustainability
6	Versatility
7	Transparent
8	Health promoting
9	Risk reducing
10	Multiple utility
11	Advancing the field/Innovative
12	Easiness of integration

Sensitivity I

- Proportion of cases detected (discretized into 6 levels from 0% to 100%) when surveillance objective is "case monitoring".
- Probability of case detection when surveillance objective is "case detection". The following levels apply:
  - Very unlikely (p<0.2)
  - Improbable (p 0.21 to 0.4)
  - Moderate (p 0.41 to 0.6)
  - Probable (p 0.61 to 0.8)
- Comments:
  - Two scenarios: case detection, and % of cases detected.
    - Just to be accurate (one cannot aim for % of cases detected in a disease free scenario, but for the probability of detection).
    - Two scenarios to capture possible different value functions
  - Discretization of probability into 5-6 classes: only for value elicitation. For users, it will be a 0-100 scale
  - Levels of risk (based on published literature)

Sensitivity II

- Context specific? Yes, surveillance managers will be able to adjust weights to capture contextual trade-offs.
- Definition: "The ability of the surveillance system to detect cases of the event of interest"…too case-specific
- New definition: "The ability of the surveillance system to detect the event of interest (e.g. it could be specific cases of disease, each and every one of them, or it could spikes in time series, or it could be both)".

Specificity

- "The ability of the surveillance system to detect only cases of the condition of interest (i.e. no false positives)".
- Comments
  - The ability of the surveillance system to detect only the event of interest (i.e. no false positives)
  - Not asking practitioners to quantify acceptable alert rates. We are asking to input their actual alerts rates (if they know them…if no, Matteo's)
  - The tension between Se, Sp and timeliness (in Spain we say bueno, bonito and barato (good, beautiful and cheap)…you can only have two.
  - The tool allows, with criteria-specific weights, these trade-offs between functions.

Timeliness

- Comments
  - That classes did not work for last mile situations, to prove freedom, and for prioritisation purposes.
    - Detection is not timely
    - Detection is rarely timely done (spread is certain).
    - Detection is frequently delayed (so spread is common)
    - Detection is somewhat delayed (to minimize spread)
    - Detection is timely (to prevent further spread
      - The rabies last mile/prove freedom.
  - Sp and Timeliness dependence
    - Don't think they are dependent, we always want more Sp independent of the level of the other. No preferential independence
    - The tension between Se, Sp and timeliness (in Spain we say bueno, bonito and barato (good, beautiful and cheap)…you can only have two.
    - The tool allows, with criteria-specific weights, these trade-offs between functions.

Accuracy

- Before it was just about strains/genotyping
  - After comments on “ancillary” and “actionable” data
- It now is “With actionable information”, including strain/genotyping
  - To again, cover not just case but event surveillance
- Levels of performance based on completeness (%) of information on the epidemiology triad?
  - First stage: set up benchmark (i.e what ideal fields of data I want?)
    - Then: 0, 25%, 50%, 75% and 100% of benchmark.

Representativeness

- Comments about Se already covering this
  - If Se 100%, representativeness = 100%
  - If Se < 100%, then representativeness < 100%: double-counting (i.e. for the general situation, whether 100% or not, Se does it)
    - However, if Se < 100%, and always marginalised populations left unserved, don't we want to evaluate this?
    - We want it to be equitable (i.e. marginalised groups will require additional efforts to be reached).
      - Reward surveillance that goes the extra-mile
      - Penalize surveillance that does not
  - Replace representativeness for Equity?

Management of the programme & evaluation of surveillance

- A means to an end.
  - Removed
  - In the case of evaluation of surveillance (by a third external party): moved into transparency

Sustanaibility

- Levels of performance
  - Structures and processes are lean; acceptability by all stakeholders is high. Sustainability is guaranteed in the long term (>3 years).
  - Structures and processes are lean; acceptability by all stakeholders is high; and data provision is properly incentivized to guarantee quality data all the time. Sustainability is guaranteed in the long term (>3 years).
  - What are the definitions of “convoluted”, “lean”, etc…
    - They need a sustainability plan (indeed, with definition of system-specific benchmarks)
    - As we need for versatility (in other words, definition of a number of process related indicators to evaluate attributes such as “flexible, interoperable, and portable” (they are context specific).

Does it lead to policy timely action?

- Tension between “decision quality” principles that contemplate “commitment to action” vs. “controllability).
  - Removed

Transparent

- Definition
  - Need to be open about the limitations of the surveillance evidence, on the use of it/how it informed policy, at all times for all stakeholders. It is a measure of trustworthiness.
  - Generate process scale
- In other words, be open about the four horsemen (of the apocalypse) (that are covered by other criteria and processes)
  - Heterogeneity: by representativeness and Se
  - Bias: by representativeness and Se
  - Uncertainty: inputted
  - Dependences

Health promoting

- Surveillance implementation leads to increased probability of positive health outcomes by means of “health outreach”, and situational awareness (e.g. risk factor surveillance on, say, smoking habits or sexual attitudes).
- Comments: redundant
  - However, it captures situational awareness

Risk reducing/proactive

- Proactive: eliminated (overlap)
- Risk reducing
  - It describes the capacity of the surveillance system to actively seek for the condition, or its triggers (e.g. One Health case study depicting search for viruses in wildlife for EIDs in Rwanda) and in doing so reducing the probability of negative health outcomes (e.g. screening for early detection of cancers or ultrasound examination of children for early detection of hydatid cysts, lead to better prognosis and hence reduces the probability of negative health outcomes).
- Comments about double counting Se?

Efficacy and efficiency

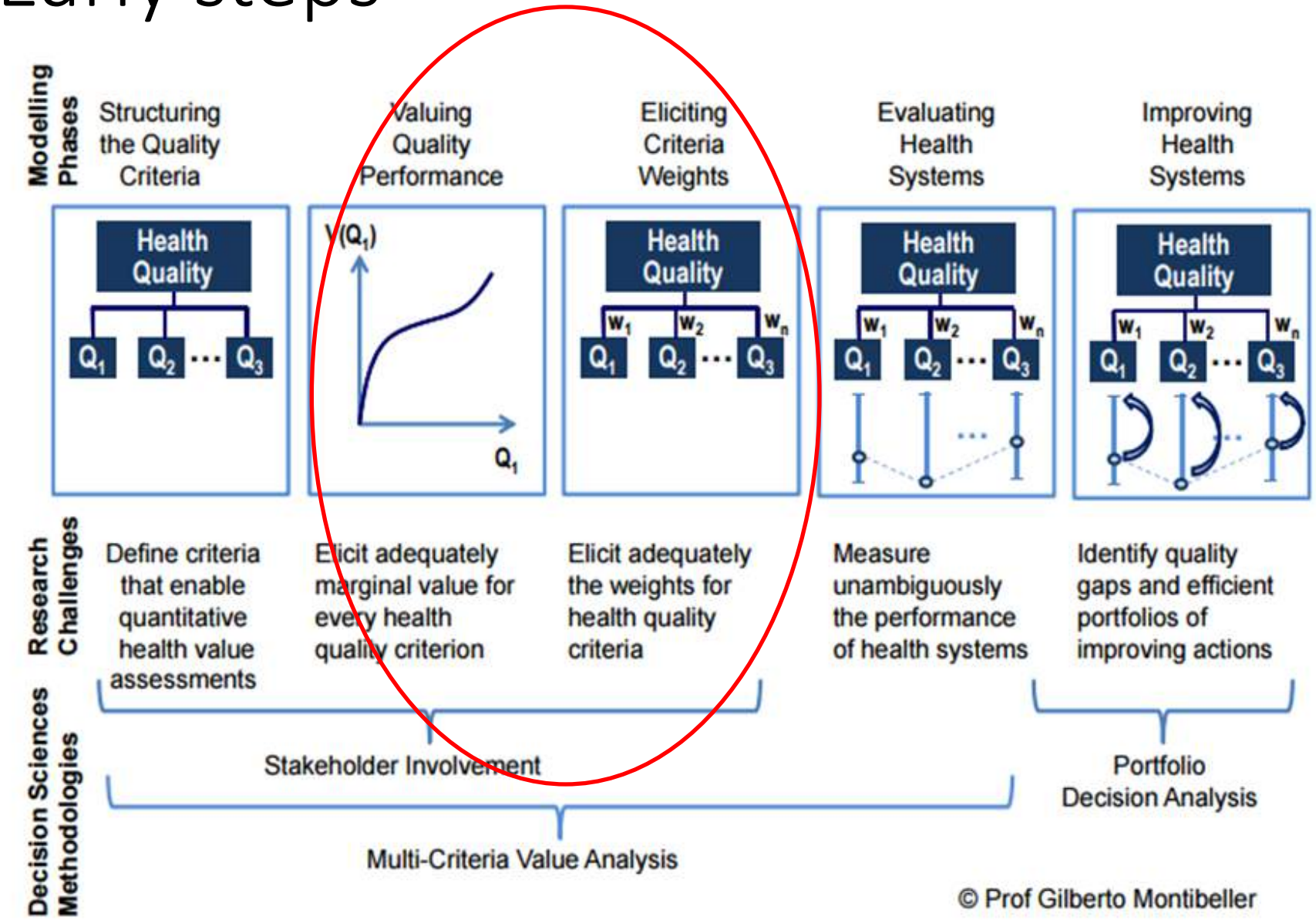
- Removed
- Efficacy: already captured by Se
- Efficiency:
  - We don’t need a cost-effective section because the purpose of this template is to evaluate “value” (the numerator of a value for money model, sort of a cost-utility model). The cost (the denominator) will come from an independent assessment of the surveillance costs (and it will be used in a posterior portfolio decision analysis component).

- Versatile
- Multiple utility
- Ease of integration
  - How important is to keep them separately?
  - If not, double-counting?
  - If merged:
    - how would it remain
    - Scales?

Advancing the field/innovative

- It captures the extra-mile that some surveillance systems deliver.
  - It favours methods and approaches (across the stages of design, implementation, analyses and evaluation) that are innovative and can lead to benefits to other fields, while being scientifically rigorous and accepted by the surveillance community Also these methods should be scientifically rigorous and accepted as being so by the surveillance community

Early steps


Modelling Phases
Structuring the Quality Criteria
Valuing Quality Performance
Eliciting Criteria Weights
Evaluating Health Systems
Improving Health Systems
Health Quality
Q_1 Q_2 ... Q_3
$v(Q_1)$
w_1 w_2 w_n
Research Challenges
Define criteria that enable quantitative health value assessments
Elicit adequately marginal value for every health quality criterion
Elicit adequately the weights for health quality criteria
Measure unambiguously the performance of health systems
Identify quality gaps and efficient portfolios of improving actions
Decision Sciences Methodologies
Stakeholder Involvement
Portfolio Decision Analysis
Multi-Criteria Value Analysis
© Prof Gilberto Montibeller