PRECISION value & health
shift the trajectory

Addressing Uncertainty Around Clinical and Population Health Data in Health Care Decision-Making: A COVID-19 Case Study

Workshop

December 1st, 2021

PRECISIONheor

Virtual ISPOR Europe 2021

30 November - 3 December | In-Person Preconference Summit - 11 November

Panelists

			
Anthony O'Hagan, PhD, Emeritus Professor of Statistics, University of Sheffield	**Shannon Cope, MSc,** Vice President, PRECISIONheor	**Raffaele Vardavas, PhD, MSci,** Senior Mathematician, RAND Corporation	**Marlon Graf, PhD,** Research Economist, PRECISIONheor
Discussant	Discussant	Discussant	Discussion Leader

Expert Elicitation Methods for Quantifying Expert Opinions

Professor Anthony O'Hagan

PRECISIONheor

PRECISION value & health

shift the trajectory

PRECISIONheor

Expert Knowledge as Decision Support

When decisions must be made in rapidly evolving contexts, data are typically sparse

Expert knowledge becomes important and valuable

- e.g. How transmissible will this new variant of the SARS-COV-2 virus be?

- **Experts are used to providing input to such questions, but often only qualitatively**
  - e.g. “I think it will be about as transmissible as (or much more transmissible than) the Brazilian variant”
- **Giving quantitative input is much more valuable**
  - e.g. “I estimate the R_0 for this variant to be 4.5”
- **And acknowledging that there is (typically substantial) uncertainty**
  - e.g. “There is an 80% chance that R_0 will be between 2.5 and 6”
- **Judgements are typically sought from multiple experts**
  - What the decision-maker really needs is an expression of their **combined** knowledge

Elicitation

- **Expert Knowledge Elicitation is the process of:**
  - Representing the knowledge of one or more persons (experts) concerning an uncertain quantity
  - Presented as a **probability distribution** for that quantity

- **Typically conducted as a dialogue between:**
  - The experts – who have substantive knowledge about the quantity (or quantities) of interest – and
  - A facilitator – who has expertise in the process of elicitation

Ideally, it will be conducted face-to-face, but may also be done by video-conference

Subjective but scientific

- **The elicited probability distributions are inevitably subjective**
  - Representing the judgements of the experts involved
- **But must meet the standards of good science**
  - Must be carefully considered
  - Taking full account of the available evidence
  - Should neither overstate nor understate the uncertainty
- **Research has revealed many ways in which experts make poor judgements**
  - Particularly, when questioned in naïve ways
- **Judgements are influenced, and potentially biased, by**
  - How questions are phrased
  - The order in which they are asked
  - Interactions between experts

Examples of potential bias

- **A political party seeks to know if people are in favour or against a proposed policy**
  - They commission an opinion poll
  - “Are you in favour or against …” will get more ‘in favour’ than “Are you against or in favour …”
  - The effect is stronger the nicer the interviewer
  - Politicians know this!
- **People make judgements about the number M (in millions) of Muslims in England**
  - A training exercise I often use
  - They are asked for two probabilities, Prob(M > 2) and Prob(M > 8)
  - One group get M > 2 question first, the other group get M > 8 first
  - Average probabilities from 100+ respondents in each group

	2 million first	8 million first
P(M > 2)	0.693	0.795
P(M > 8)	0.292	0.391

Elicitation protocols

- **Many pitfalls for naïve elicitation**
  - Poor elicitation misleads decision makers, resulting in bad decisions
- **Elicitation should follow a well-constructed protocol to minimise potential for bias**
  - Three leading protocols: SHELF, Cooke, Delphi
  - Designed by experienced practitioners in the field
  - They differ in various ways, but particularly in regard to interaction and combination
- **Interaction**
  - Should experts discuss their judgements?
  - Allows sharing of information and aids understanding of different opinions
  - But introduces other potential biases
    - Dominant experts
    - Experts too ready, or too unwilling, to accept the arguments of others
- **Combination**
  - How should differing judgements of a group of experts be combined into a single distribution?
    - To represent their combined knowledge

The SHELF protocol

- **Preparation**
  - Assemble a dossier of available evidence
  - Recruit experts and get their commitment to the project
  - Train the experts in the kinds of probabilistic judgements they will be required to make
- **Facilitated elicitation workshop**
  - Usually 4 to 8 experts
  - Conduct practice runs to familiarise experts with the process and to reinforce preparatory training
  - Make private individual judgements first to establish each expert's initial position
  - Group discussions to understand reasons for differences in their individual judgements
  - Make group judgements from the perspective of a Rational Impartial Observer (RIO)
  - Feedback and validity checking throughout
- **Reporting**
  - SHELF templates provide auditable record of the process

Other leading protocols

- **Delphi**
  - Ask experts for judgements by questionnaire
  - Feed back summary of judgements and invite revision
    - Iterate if resourcing allows to aim for some convergence
  - Traditional Delphi only asks for estimates
    - Does not result in a probability distribution
  - Probabilistic Delphi has more complex questionnaire to elicit probabilistic judgements
  - Interaction: Minimal – experts don’t meet and are anonymous, but some exchange via feedback
  - Combination: After final questionnaire, simply average across the experts
- **Cooke (aka Classical)**
  - Experts make judgements about quantity of interest (QoI) and also 10 or more ‘seed variables’
  - Facilitator knows true values of seed variables, which must be as similar to QoI as possible
  - Interaction: None – experts interviewed separately
  - Combination: Weighted average of the experts’ judgements, with weights derived from their performance on the seed questions

The role of the facilitator

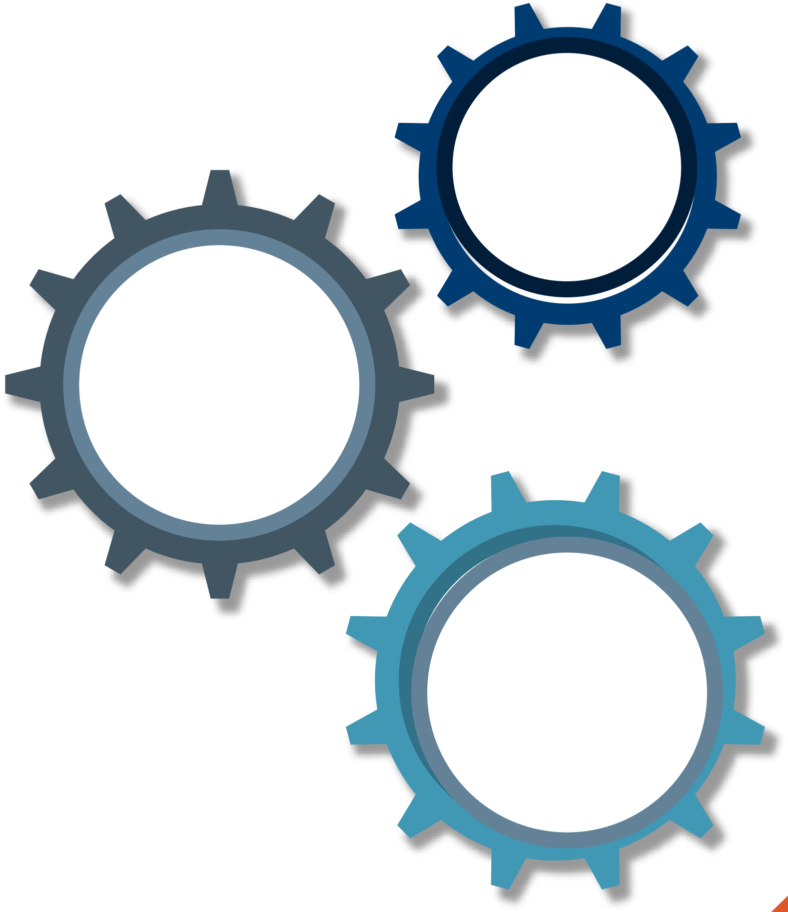
- **A crucial role in the SHELF protocol**
  - Less so, but still important, in other protocols
- **Guides the experts**
  - Helping them to make judgements that accurately represent their opinions
- **Manages the discussion**
  - Ensures all opinions are given due weight and consideration
  - Focuses discussion on areas of divergence and prompts experts to explain their judgements
- **Directs the experts in making group judgements**
  - From the RIO perspective
  - Gives feedback to check consistency of judgements
  - Fits an agreed probability distribution

Summary

- **Expert judgement is an important input to decision-making**
  - Particularly in rapidly-changing, data-poor contexts like emerging threats, pandemics, etc.
  - But also wherever data is inconclusive or conflicting

- **Expert judgements need to be elicited with great care**
  - Poor judgements are worse than useless
  - Elicitation is not an easy option

- **Elicitation should be conducted by a skilled facilitator**
  - Following a recognised protocol such as SHELF
  - Should be fully-documented



What does this mean for Covid-19 decision making?

- **Emerging threat**
  - Vital early decisions had to be made with very little data
  - Great uncertainty about parameters of even simple epidemiological models
  - Rigorous expert elicitation would surely have helped to quantify those uncertainties realistically
- **Why was it not used?**
  - Insufficient awareness among decision makers
  - Shortage of skilled practitioners/facilitators
  - Not much time to train experts in making probabilistic judgements
- **My list of lessons for future emerging threats (and new variants)**
  - Decision makers must be made aware of the value of elicitation
    - And trained in how and when to deploy it
  - Train panels of experts from relevant disciplines (virology, epidemiology, pharmacology, …)
    - So they are ready to be called on at short notice
  - We also need to train many more skilled practitioners
  - Elicitation should be employed routinely, wherever data are sparse or conflicting
    - To build facilitator expertise, and to refresh regularly the training of expert panels

PRECISIONheor

Shannon Cope

Addressing Uncertainty Around Clinical and Population Health Data in Health Care Decision-Making: A COVID-19 Case Study

Applications for health economics: Extrapolation of survival curves for HTA reviews

October 21, 2021

PRECISIONheor

Expert elicitation long-term survival case study

Acknowledgements: Dieter Ayers and Harlon Campbell

Cope et al. BMC Medical Research Methodology (2019) 19:182
https://doi.org/10.1186/s12874-019-0823-8

BMC Medical Research Methodology

RESEARCH ARTICLE Open Access

Integrating expert opinion with clinical trial data to extrapolate long-term survival: a case study of CAR-T therapy for children and young adults with relapsed or refractory acute lymphoblastic leukemia

Check for updates

Shannon Cope[1*], Dieter Ayers[1], Jie Zhang[2], Katharine Batt[3] and Jeroen P. Jansen[4]

Lifetime horizon requires fully extrapolated survival

Health technology assessment (HTA): Evaluating cost-effectiveness of new interventions

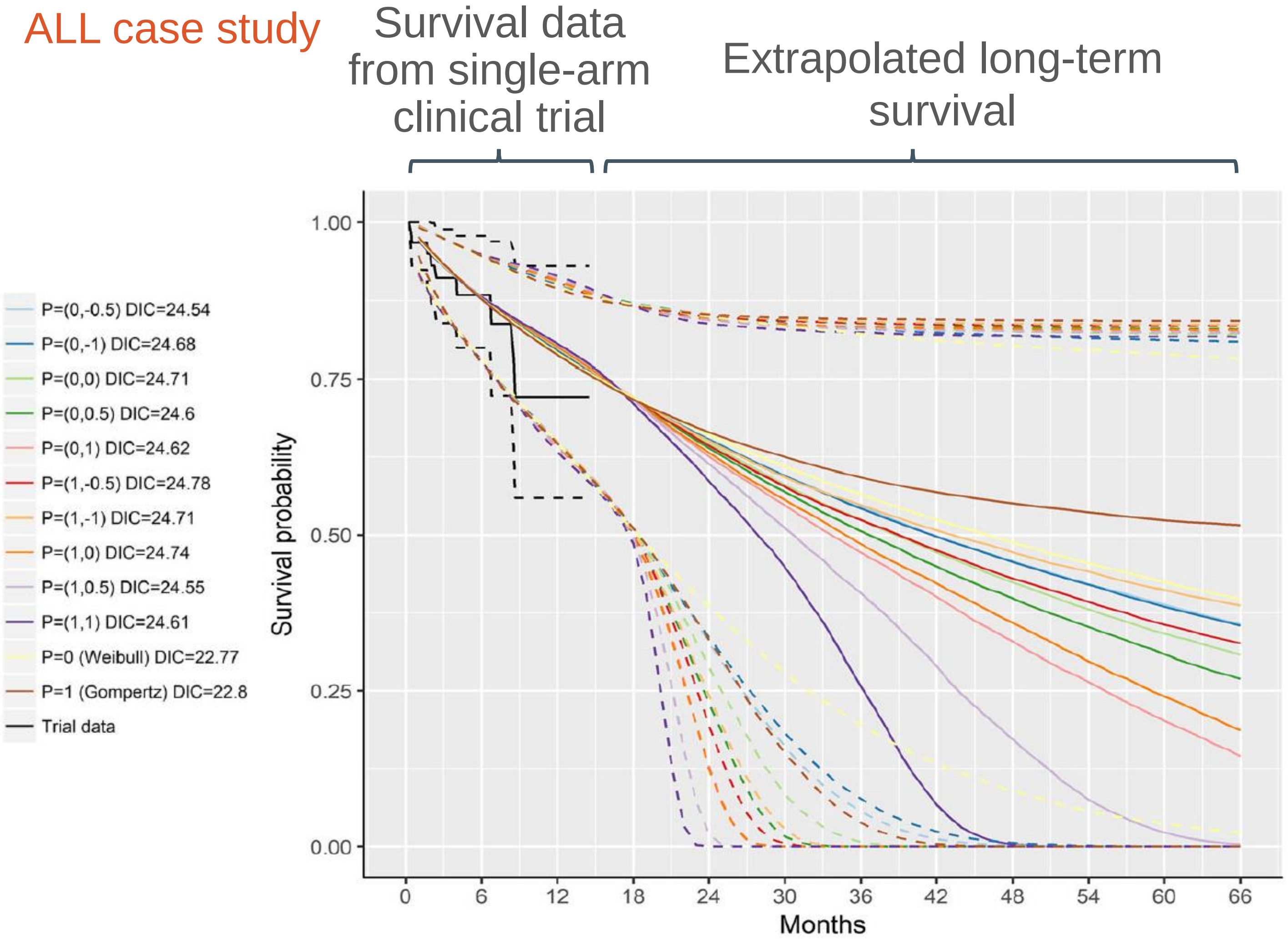
- Health technology assessment (HTA) decision-makers evaluate the cost-effectiveness of new interventions over a **lifetime horizon** in order to inform reimbursement recommendations
- In the absence of long-term data from clinical trials evaluating new interventions, survival beyond the observed follow-up data needs to be **fully extrapolated**
- National Institute for Health and Care Excellence (NICE) recommend:
  1. Fit alternative parametric survival models to the observed survival from clinical trial
  2. Assess suitability of survival models:
     - Visual inspection
     - Log cumulative hazard plots
     - Goodness of fit (AIC/BIC)
     - Assess plausibility using clinical validity ~~and external data~~

How?

Sources: Latimer, N. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011. Available from http://www.nicedsu.org.uk; Rutherford, MJ., Lambert, PC., Sweeting, MJ., Pennington, R., Crowther, MJ., Abrams, KR., Latimer, NR. NICE DSU Technical Support Document 21. Flexible Methods for Survival Analysis. 2020 [Available from http://www.nicedsu.org.uk]

Variability in extrapolated survival across models

ALL case study

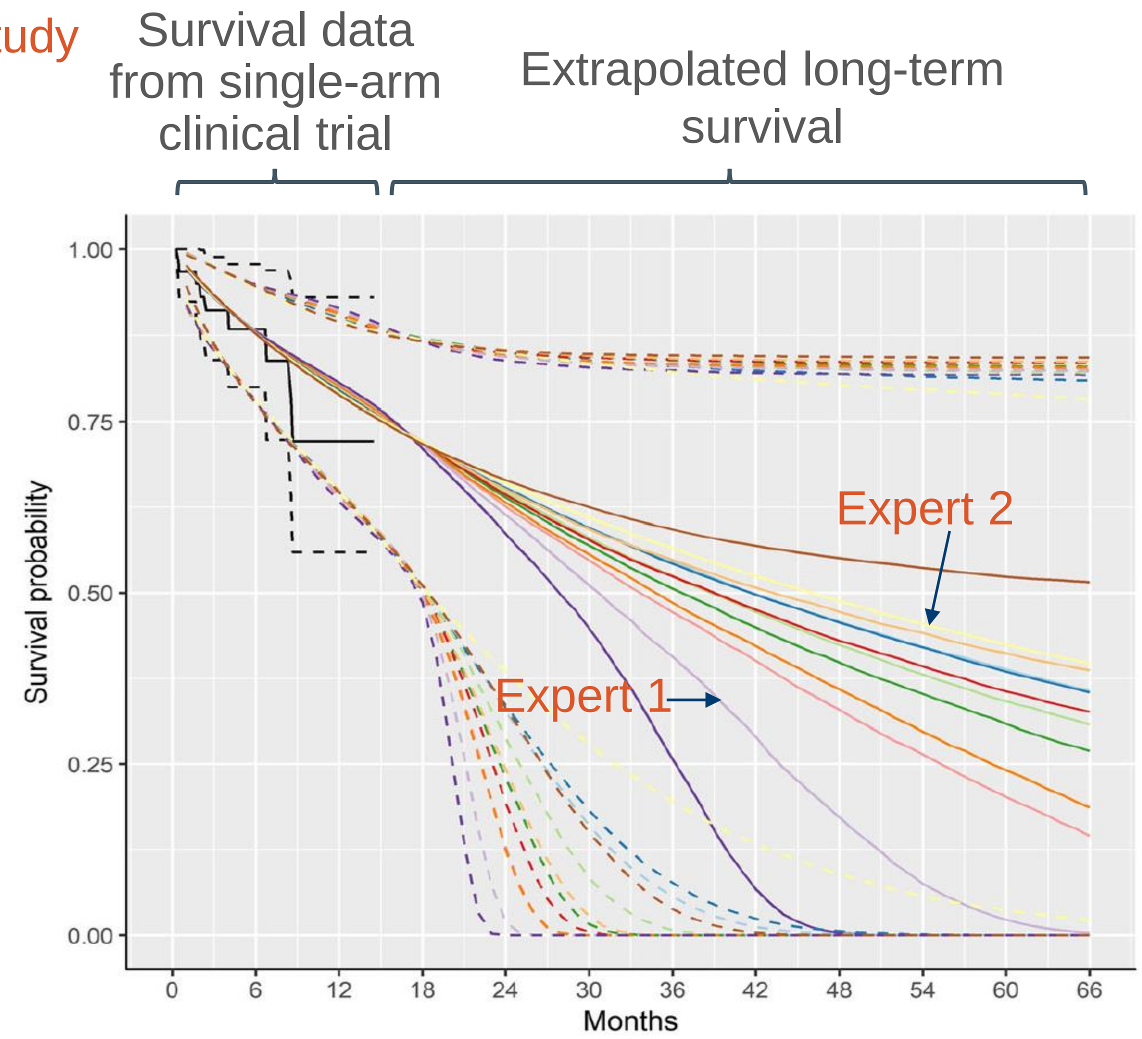


- Fractional polynomial models had comparable fit the observed data (AIC)
- However, considerable **variability** in extrapolated survival: 0% to 50% at 66 months

Source: Cope, S., Ayers, D., Zhang, J., Batt, K. and Jansen, J.P., 2019. Integrating expert opinion with clinical trial data to extrapolate long-term survival: a case study of CAR-T therapy for children and young adults with relapsed or refractory acute lymphoblastic leukemia. *BMC medical research methodology*, *19*(1), p.182.

Plausibility of extrapolations: Informal expert opinion

ALL case study

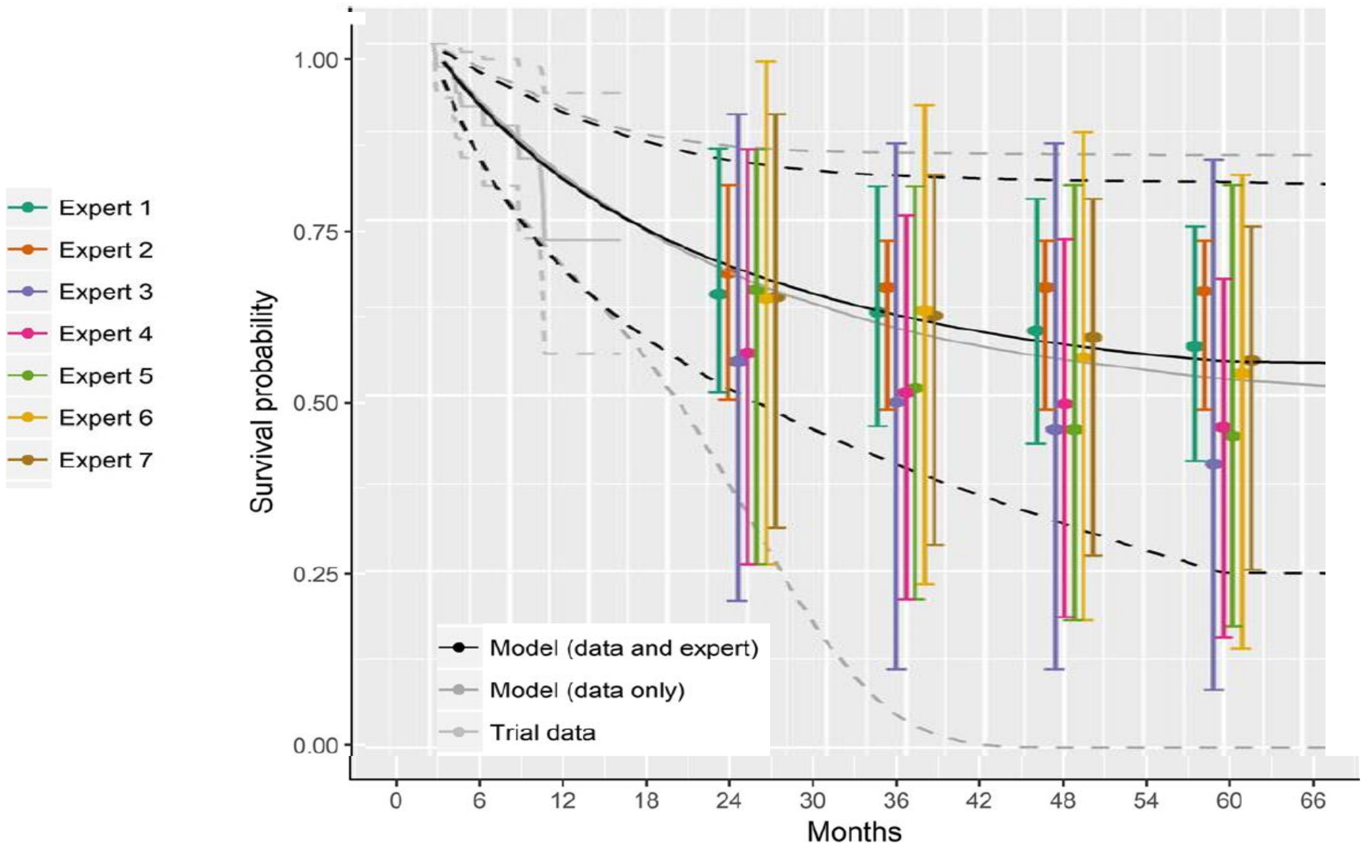


- **Informal** process with one (or small number of) clinical expert(s)
- Often **presented** with alternative extrapolated survival curves
- Asked to identify the most *plausible* 1 or 2 models
- **Qualitative** summary of feedback support model choice
- No quantitative estimates of long-term survival or uncertainty in those estimates
- No formal integration of expert feedback in model

Source: Cope, S., Ayers, D., Zhang, J., Batt, K. and Jansen, J.P., 2019. Integrating expert opinion with clinical trial data to extrapolate long-term survival: a case study of CAR-T therapy for children and young adults with relapsed or refractory acute lymphoblastic leukemia. *BMC medical research methodology*, *19*(1), p.182.

Structured expert elicitation of long-term survival

Provides more reliable representation of expert opinion including quantitative estimates of uncertainty

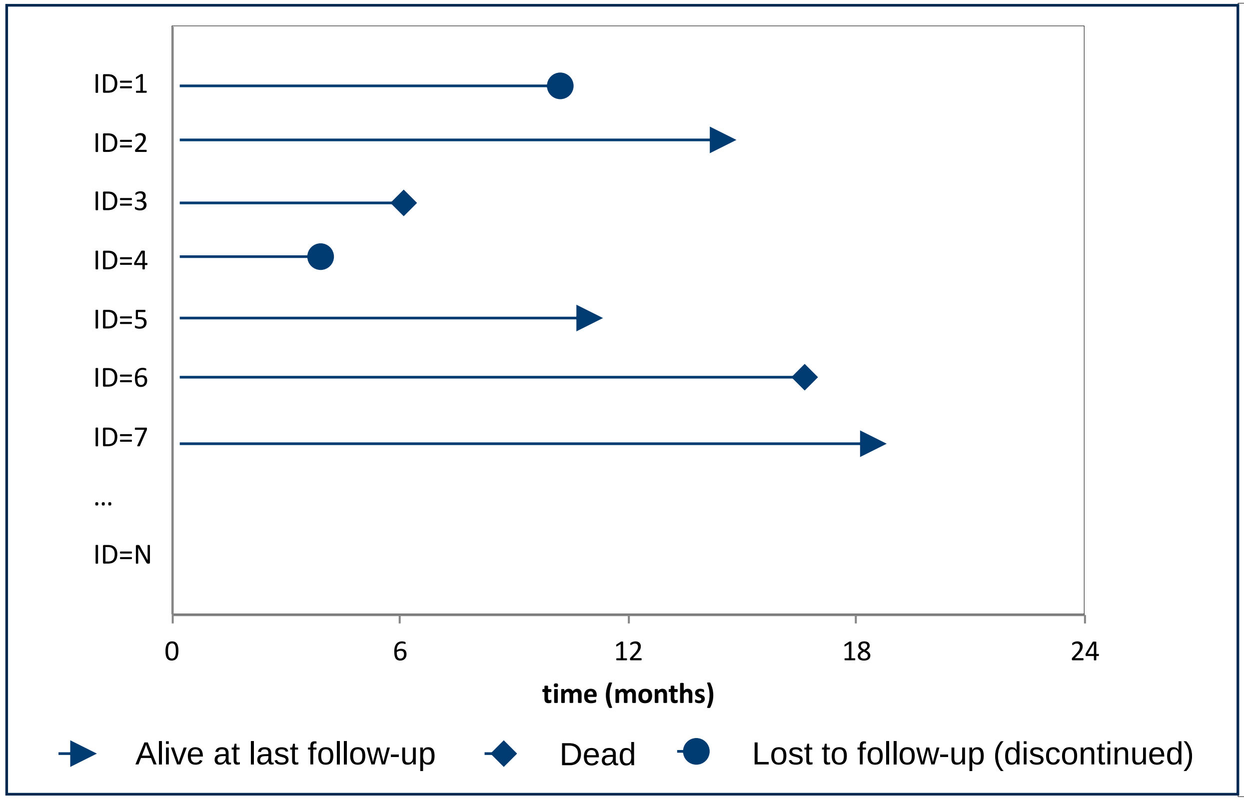


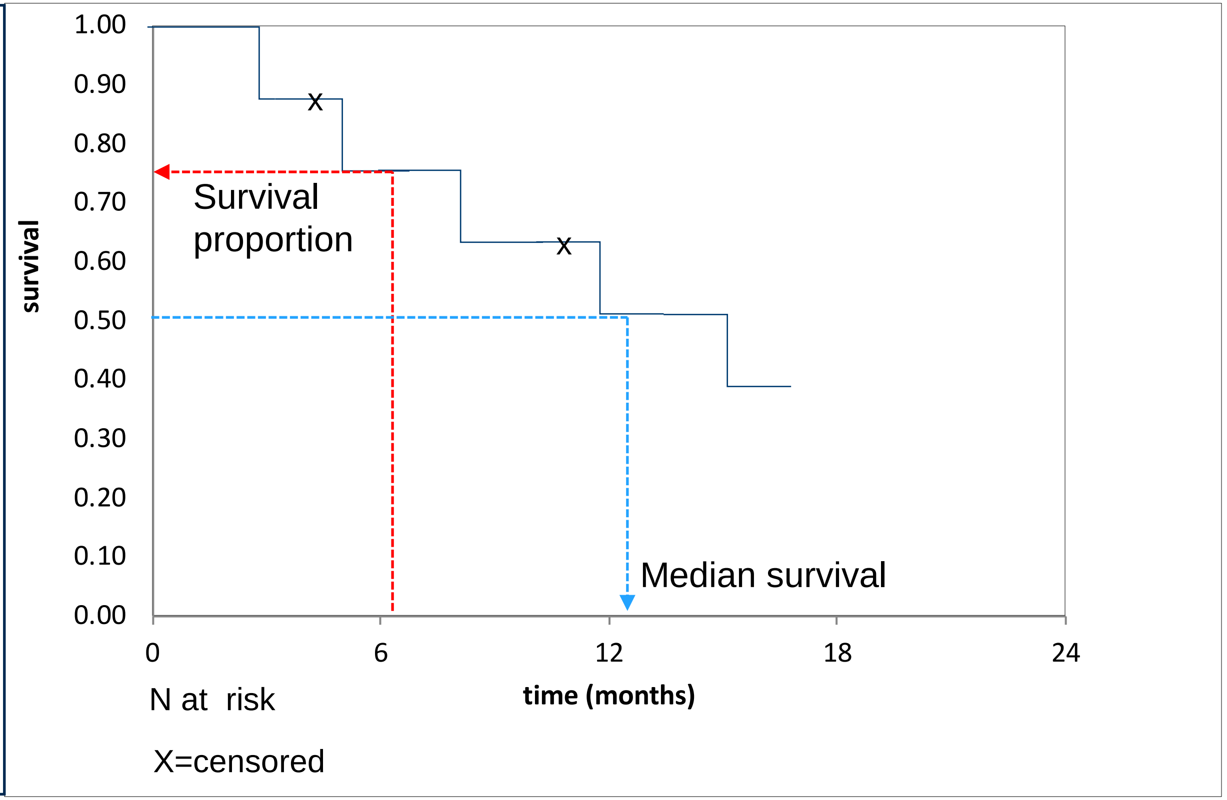
- **Formal** process with multiple clinical experts
- Not **presented** with alternative extrapolated survival curves
- Asked to estimate survival % at specific time points, including uncertainty
- **Quantitative** summary across experts
- Allows for formal integration of expert feedback in model

Source: Cope, S., Ayers, D., Zhang, J., Batt, K. and Jansen, J.P., 2019. Integrating expert opinion with clinical trial data to extrapolate long-term survival: a case study of CAR-T therapy for children and young adults with relapsed or refractory acute lymphoblastic leukemia. *BMC medical research methodology*, *19*(1), p.182.

Quantity of interest for time-to-event outcomes

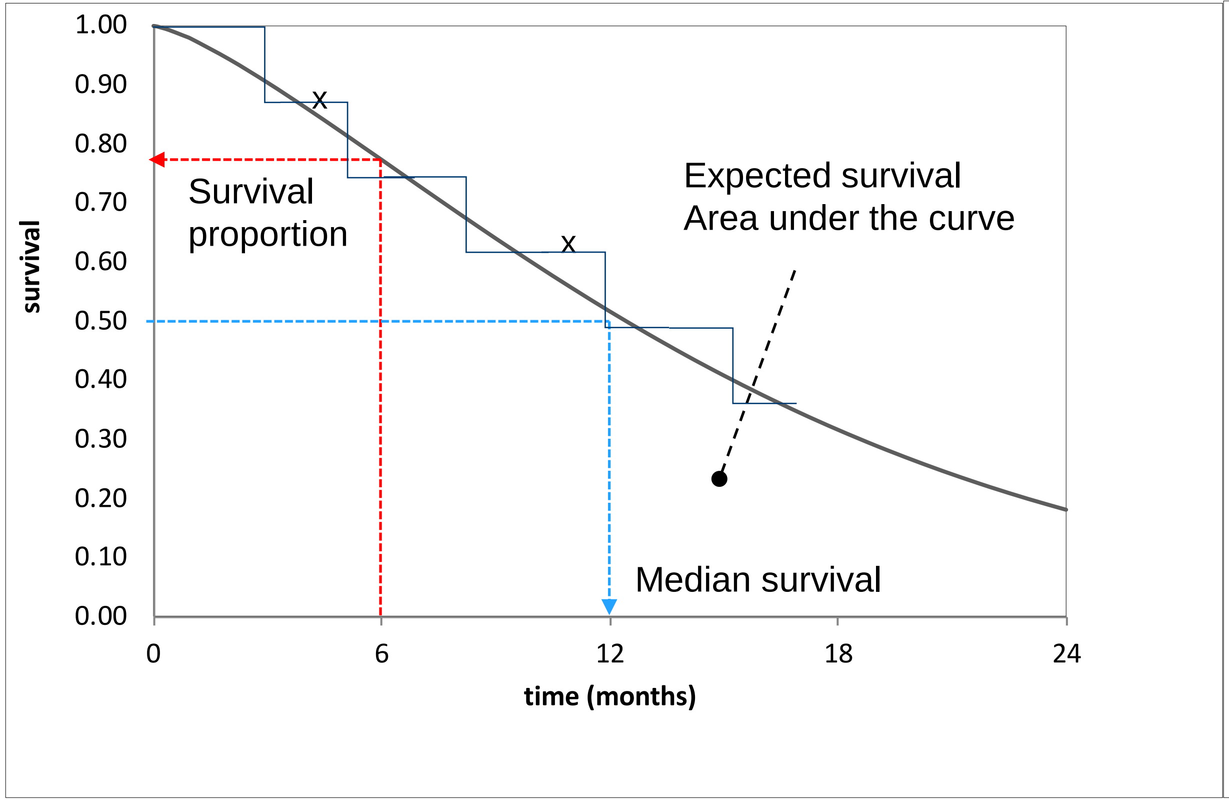
What is the observed data and how is it summarized?

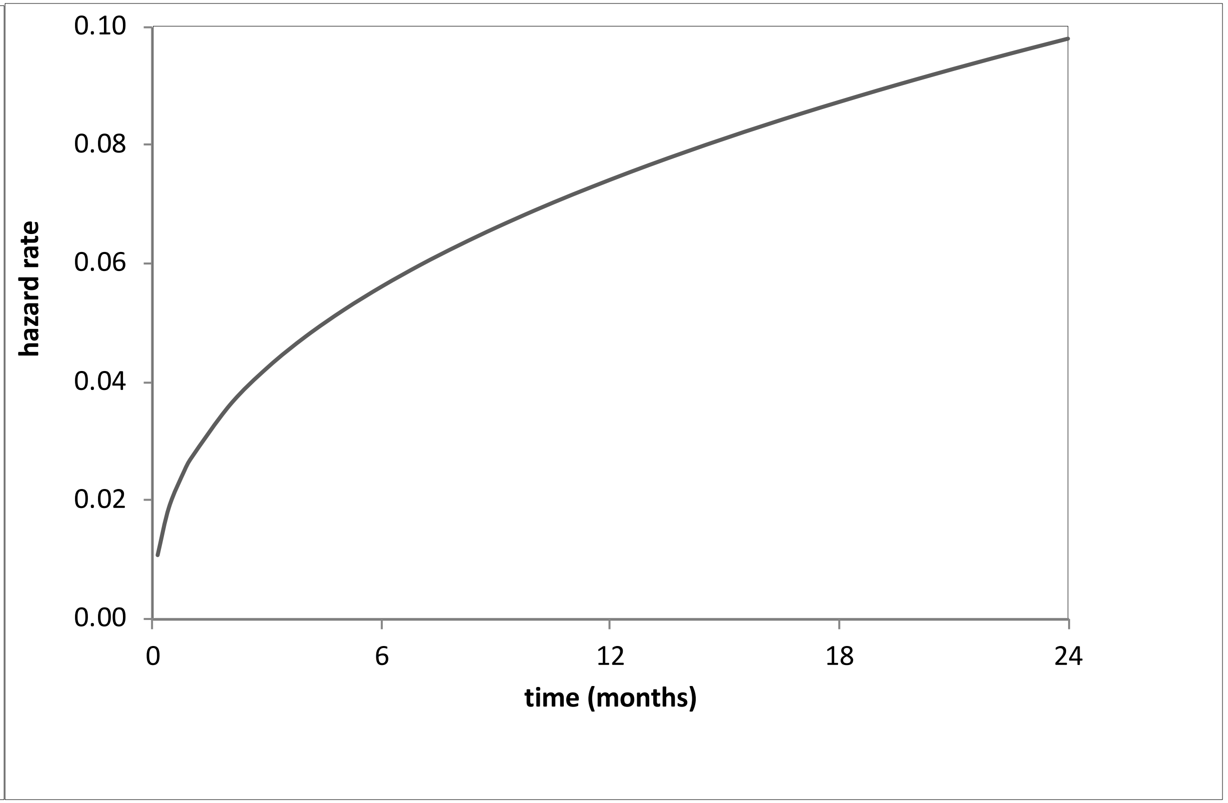




Quantities of interest for time-to-event outcomes

What is the observed data and how is it summarized?





What measures should be elicited to model survival?

Literature review by Bojke et al. 2021

Measure	Interpretation	Cumulative effect	Advantage	Disadvantage	Refs from Bojke 2021
Time to events per individual	Time when each patient has event (or censor)	No	Relates to directly observable events	Not practically feasible to ask about each patient	--
Survival % at one time point	Proportion patients alive at specific time point	Yes	Clinically relevant - directly relates to KM curves in trials;	Not sufficient information to inform more flexible models	52, 65, 66
Conditional probabilities at specific time points	Proportion patients alive at specific time point, conditional on being alive at end of last interval	Yes	Clinically relevant - directly relates to KM curves in trials; Coherent with model (can inform changes in hazard over time)	Requires decision regarding specific time points of interest	53, 59
Median survival	Survival time when 50% patients alive	Yes	Commonly used and clinically relevant; Reflects cumulative effect Could inform more simplistic models	Ignores what happens after 50% of subjects have experienced the event Not informative for long-term extrapolation	70
Expected survival (average)	Area under curve when all patients have died	Yes	Directly relevant for cost-effectiveness analysis Reflects cumulative effect	Rarely observed or clinically relevant	58
Hazard at specific time points	Instantaneous probability of death at time u conditional on not having experienced the event prior to time u	No	Directly informative for model extrapolation and may emphasize changes over time	Does not reflect the cumulative effect Not directly observable, may be challenging to understand	--
Parameters of survival distribution	Depends on assumed distribution	NA	Directly informed model	Not observable or understandable clinically Require decision on model	187

Source: Bojke et al. 2021. Developing a reference protocol for structured expert elicitation in health-care decision-making: a mixed-methods study. Health Technology Assessment Vol 25(37). DOI 10.3310/hta25370

Quantities of interest

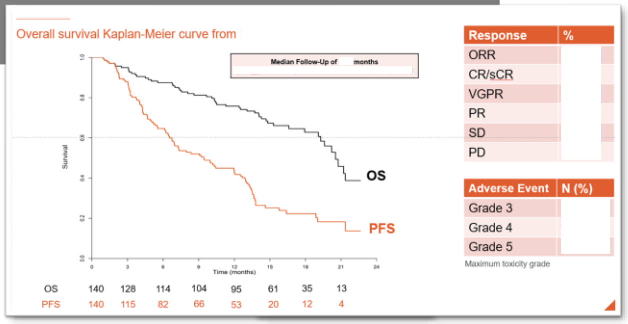
- The quantities of interest could be defined as:
  - The expected proportion of patients in the trial that are alive at 2 years, **given the observed Kaplan-Meier (based on the available follow-up data)**
  - The expected proportion of patients in the trial that are alive at 3 years, given the estimated proportion alive at 2 years
  - The expected proportion of patients in the trial that are alive at 4 years, given the estimated proportion alive at 3 years
  - The expected proportion of patients in the trial that are alive at 5 years, given the estimated proportion alive at the earlier time points

- Population is from pivotal trial so fit for purpose
- Time points were justified based on clinical context (piloted with clinician); minimized to avoid expert fatigue

Expert elicitation using SHELF

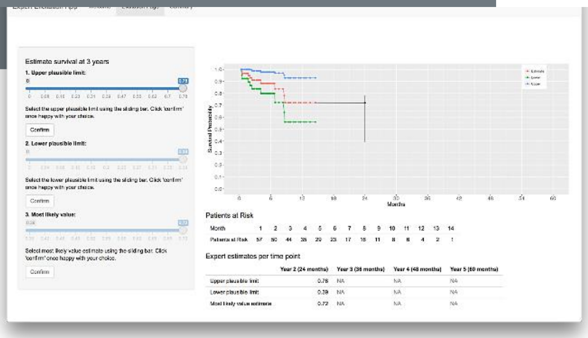
1. Evidence Dossier Review

- Experts complete survey regarding experience and conflicts of interest
- Experts review Evidence Dossier prior to individual expert interviews (provide feedback regarding evidence)
- Expert review Training and Practice Exercise



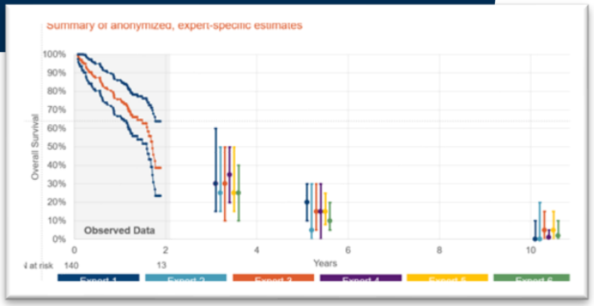
2. Individual Expert Interviews

- Facilitator review Evidence Dossier with experts
- Facilitator review Training Exercise with experts
- Facilitator elicited plausible limits and most likely values using web-based application with guidance from facilitator



3. Expert Consensus Meeting

- Review anonymized results from individual interviews
- Discuss points of divergence among individual estimates guided by facilitator
- Elicit group consensus on plausible limits and most likely values from perspective of Rational Impartial Observer

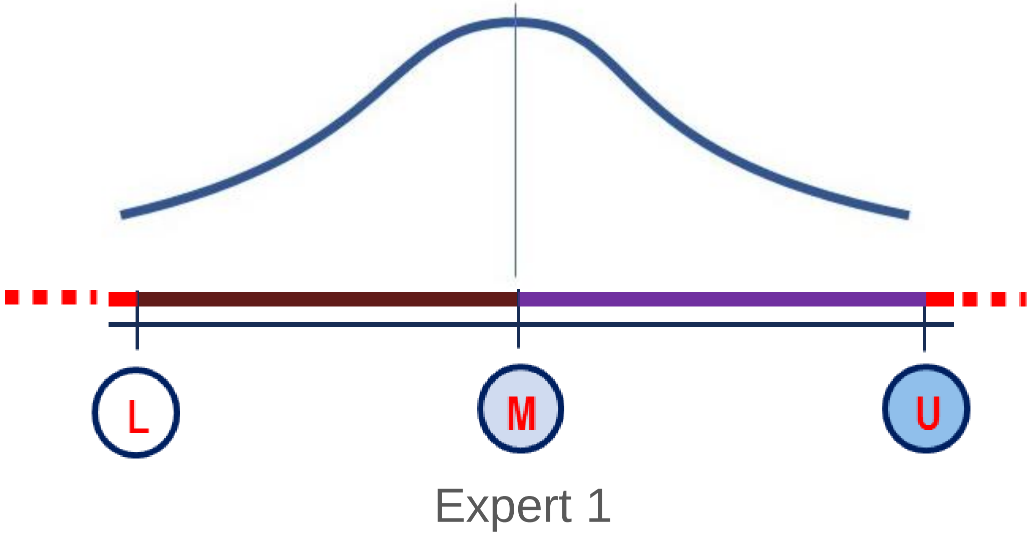


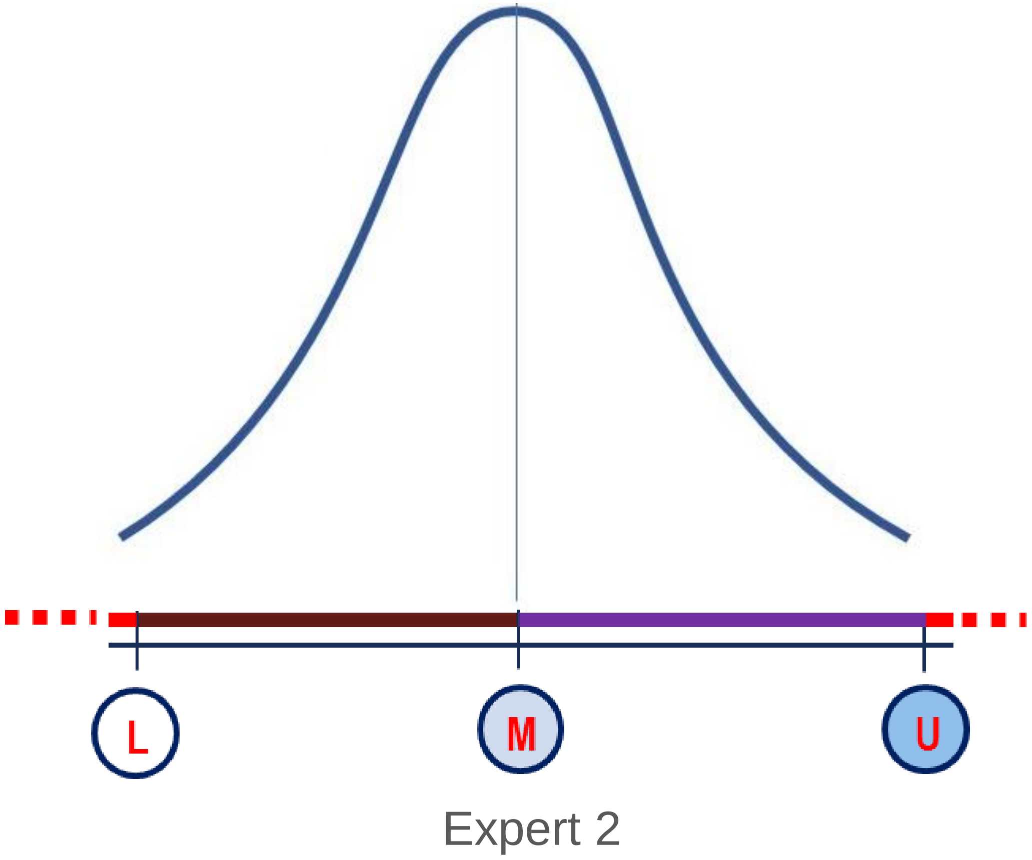
Expert survival probabilities assumed to be normal

Expert estimates regarding proportion patents alive at time point j

$$S(u_j) \sim Normal(\theta_j, \sigma_j^2)$$

- Probabilities bounded by zero and one
- Can use truncated normal
- Could explore beta distribution

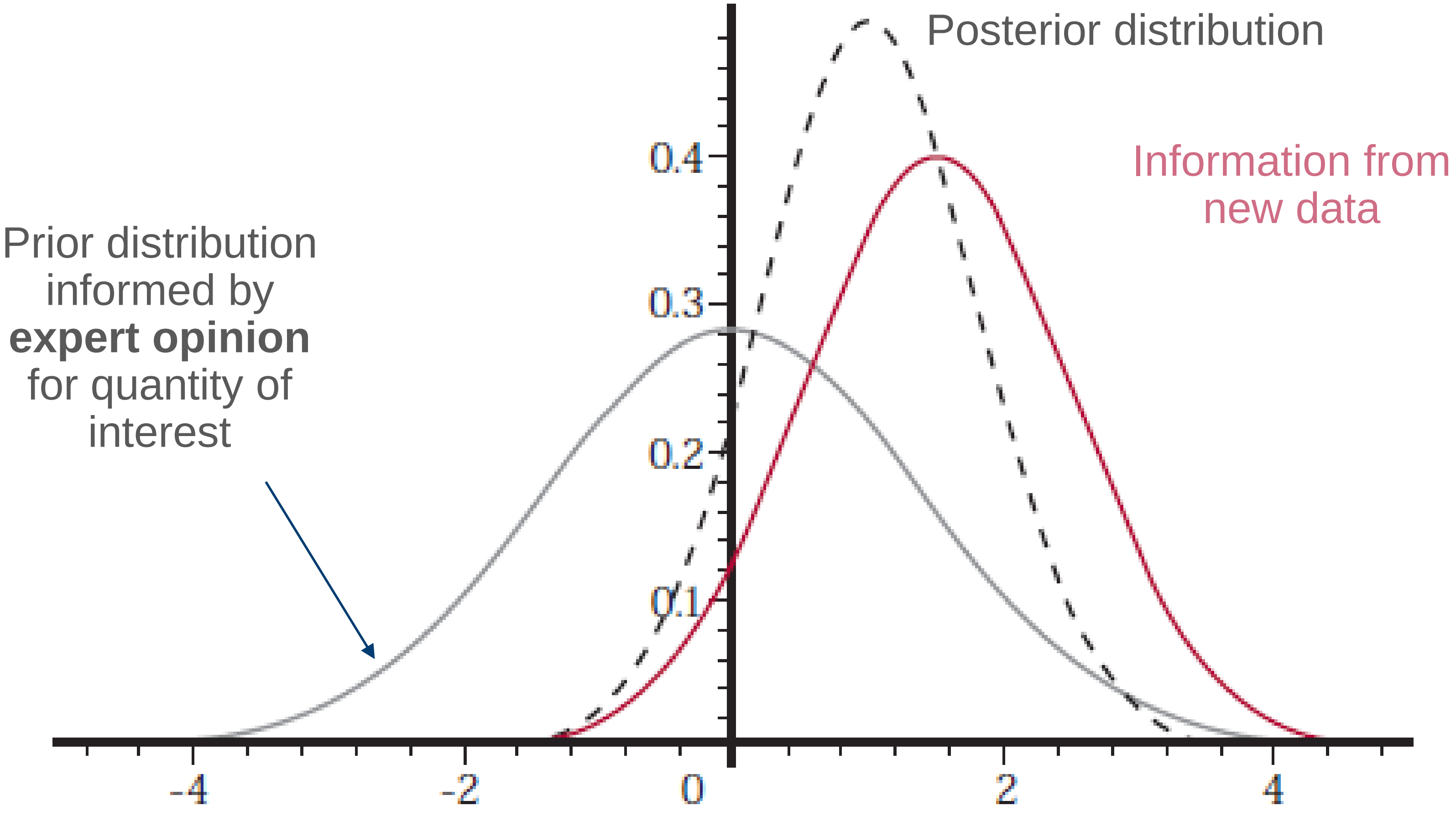




PRECISIONheor

Expert opinion integrated using informed prior distribution

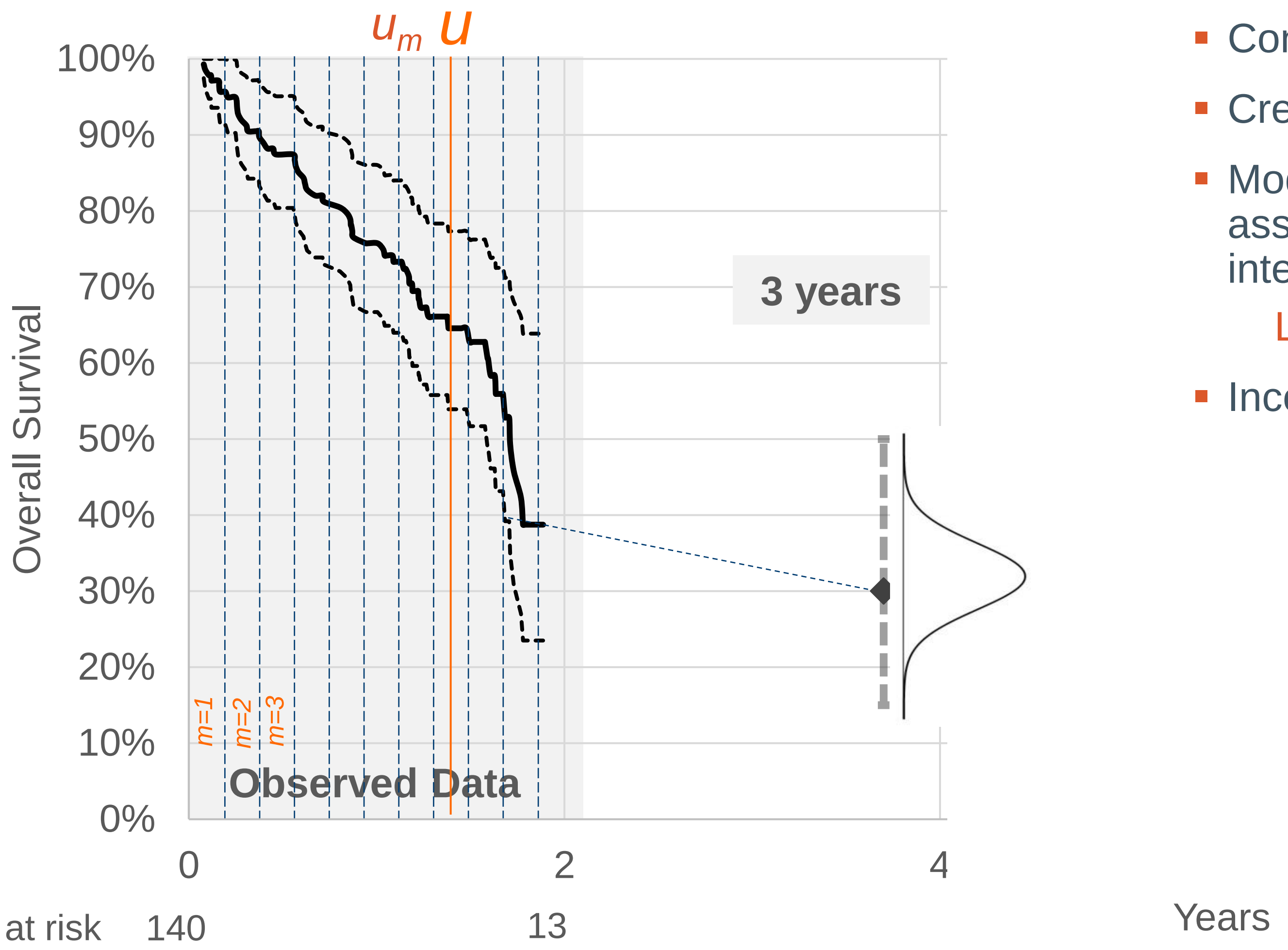
In most cases for other types of outcomes, integrated using prior distribution



Source: O'Hagan A, Luce B. A primer on Bayesian Statistics in Health Economics & Outcomes Research. Centre for Bayesian Statistics in Health Economics

Integrating expert opinion with clinical data

Triple class exposed relapsed refractory case study



- Construct interval data (m) from Kaplan Meir
- Create dataset with discrete hazards over time
- Model discrete hazard data using fractional polynomial assuming binomial likelihood for number of patents died per interval

  Likelihood: $r_m \sim \text{Binomial}(p_m, n_m)$

- Incorporate expert opinion at each time point j:

$$S(u_j) \sim Normal(\theta_j, \sigma_j^2)$$

$$p_{m+1} = 1 - \frac{S(u_{m+1})}{S(u_m)}$$

$$h_m = \frac{-\ln(1 - p_m)}{\Delta u_m}$$

Source: Ayers, Cope, Dhanda, Towle, Mojebi, Delforge, Rodriguez Otero, Trudel, Weisel, Zamagni, Hari. Expert elicitation of long-term survival for patients with relapsed/refractory multiple myeloma treated with idecabtagene vicleucel (ide-cel, bb2121) in the KarMMa phase 2 trial. Presented at the 18th International Myeloma Workshop; September 8–11, 2021, Vienna, Austria. Presentation Number 1074971

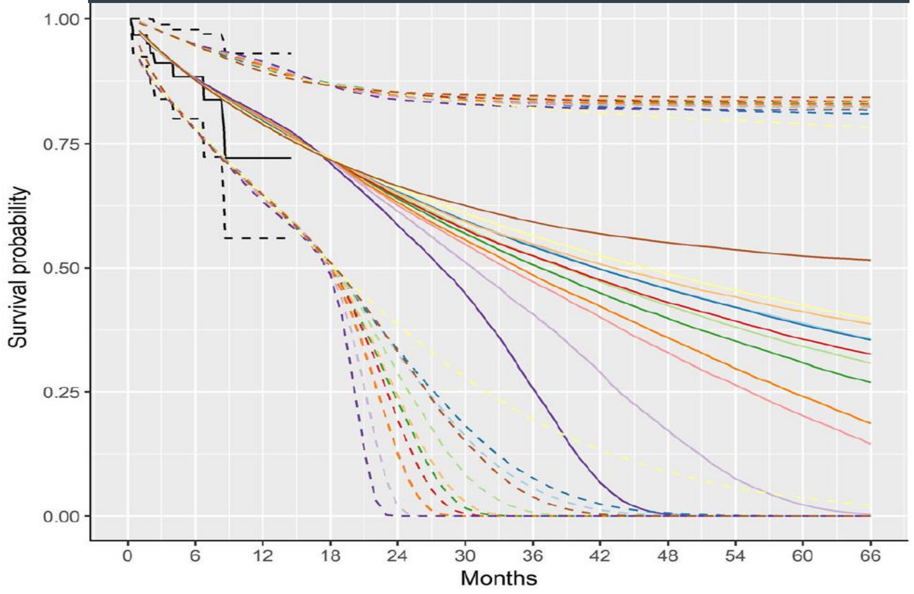
PRECISIONheor

Integrating expert opinion & clinical trial data

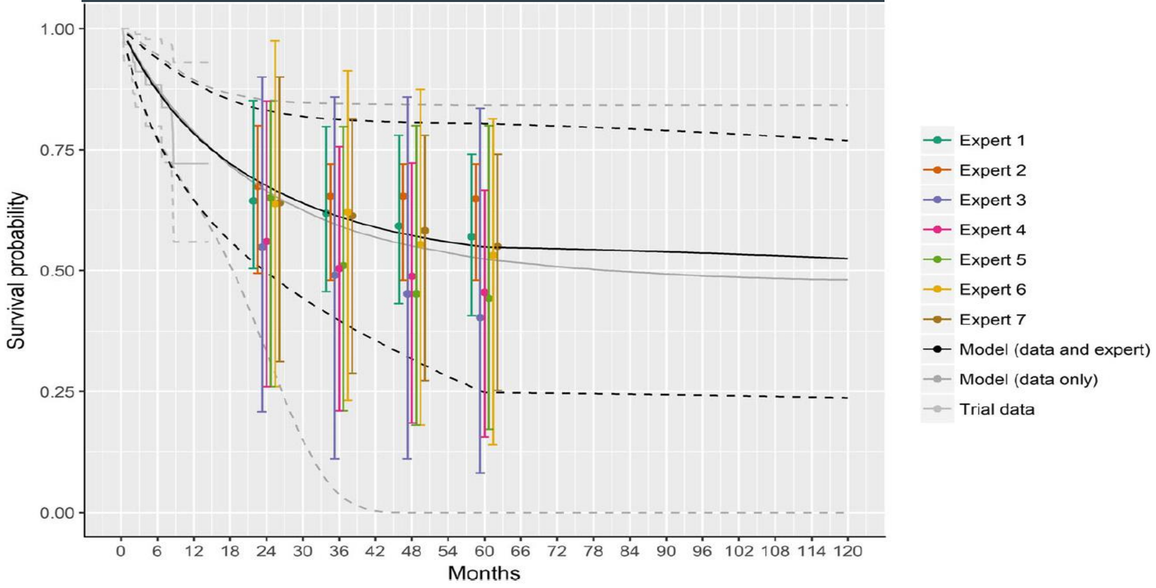
Provides a coherent method to integrate expert opinion regarding long-term survival

- May improve the clinical **plausibility** of long-term survival extrapolations
- May reduce the uncertainty in long-term survival extrapolations
- Estimates can be directly incorporated into the cost-effectiveness model

Large uncertainty in estimates **without** expert opinion



More credible estimates **with** expert opinion



Source: Cope, S., Ayers, D., Zhang, J., Batt, K. and Jansen, J.P., 2019. Integrating expert opinion with clinical trial data to extrapolate long-term survival: a case study of CAR-T therapy for children and young adults with relapsed or refractory acute lymphoblastic leukemia. *BMC medical research methodology*, *19*(1), p.182.

Acceptability of SHELF

HTA perspective

- SHELF was identified in a review of guideline recommendations and interviews with NICE decision-makers
- European Food Safety Authority identified SHELF as 1 of 3 methods that represented best practice in expert elicitation.
- SHELF aligned with principles outlined by Medical Research Council good practice reference guide for structured expert elicitation in healthcare decision making*:
  - Transparency,
  - Fitness-for-purpose,
  - Consistency,
  - Reflection of uncertainty,
  - Recognizing and addressing biases,
  - Recognizing between-expert variation,
  - Collecting individual estimates prior to group consensus, and
  - Promoting high performance.

Source: Peel A, Jenks M, Choudhury M, et al. Use of Expert Judgement Across NICE Guidance-Making Programmes: A Review of Current Processes and Suitability of Existing Tools to Support the Use of Expert Elicitation. Applied Health Economics and Health Policy. 2018;16(6):819-836; European Food Safety Authority (EFSA). Guidance on expert knowledge elicitation in food and feed safety risk assessment. EFSA Journal. 2014;12(6):3734.; Bojke L, Soares MFO, Fox A, et al. Developing a reference protocol for expert elicitation in healthcare decision making. Health Technology Assessment Reports. 2019.

Reference protocol (Bojke 2021)

Table 15. A reference protocol for HTA

Element	Reference method suggested
Experts	• ≥5 experts with substantive expertise/ experience who represent full range of valid experts' beliefs • Minimize and record conflicts of interest
Quantities elicited	• Use simple observable quantities (not complex parameters) • Use clear wording (decomposed when fits with experts' mental models) • Capture dependence between variables - express dependent variables in terms of independent variables if limited normative skills
Approach to elicitation	• Use individual elicitation, even if group interaction follows (suggest group discussion for early technologies or complex quantities) • Explore between-expert variation
Method	• Variable interval method fixed interval method (FIM) work well (FIM if training not face-to-face)
Aggregation	• Fit distributions to experts' individually elicited judgements • Use linear pooling to summarize individual distributions (equal weighting of experts) • Adjust to improve coherence and consistency, not to reduce variability • Internal and external review can be used to assess validity
Delivery	• Face to face if possible (especially for training of experts); practical constraints may dictate remote delivery • Provide feedback to experts during the SEE • Provide opportunity to revise their distributions
Training and piloting	• Training is crucial on how to avoid bias and express uncertainty • Undertake pilot
Rationales and documentation	• Collect rationales for how experts made judgements • Justify and document methodological choices

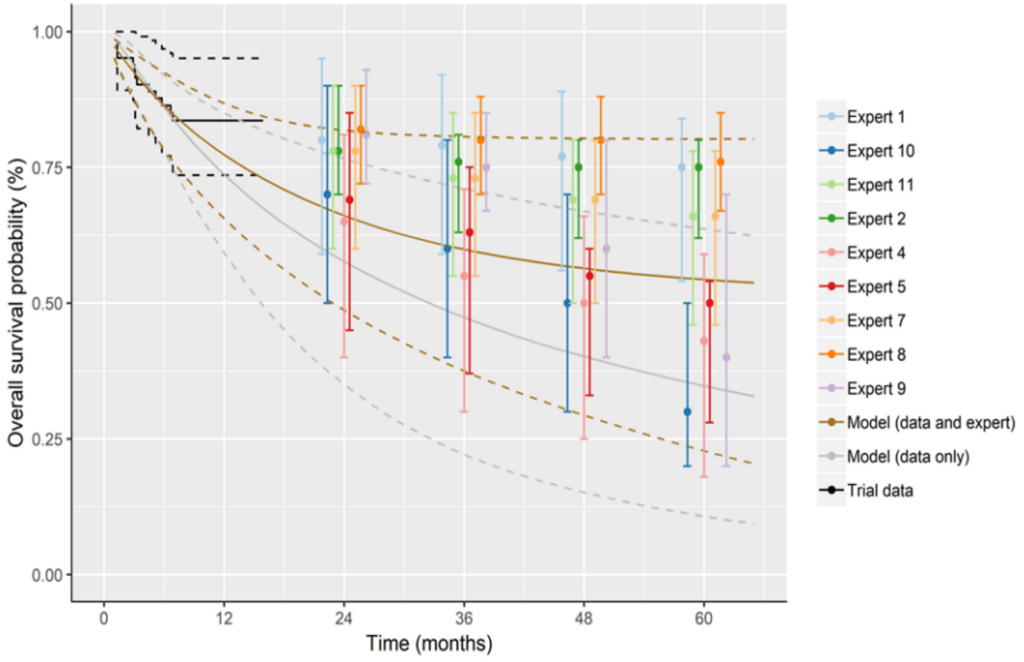
Source: Bojke et al. 2021. Developing a reference protocol for structured expert elicitation in health-care decision-making: a mixed-methods study. Health Technology Assessment Vol 25(37). DOI 10.3310/hta25370

Method has been used to support reimbursement

NICE submission for cemiplimab (LIBTAYO®) in cutaneous squamous cell carcinoma

Case study	Support reimbursement of cemiplimab in the UK for advanced cutaneous squamous cell carcinoma
Challenge	‘Cemiplimab trial data is promising but uncertain’ (immature single-arm clinical trials)
Study	Performed an expert elicitation exercise to supplement survival estimates (2–5 years) with clinical opinion using the SHELF
Feedback from NICE	‘The expert elicitation was **clearly reported** and appears to have been **well-conducted**’

Figure 26: Expected outcomes based on best fitting model with and without expert information for cemiplimab from the Phase II, EMPOWER-CSCC 1 trial



Abbreviations: NICE, National Institute for Health and Care Excellence; UK, United Kingdom
Source: NICE. Single Technology Appraisal: Cemiplimab for treating metastatic or locally advanced cutaneous squamous cell carcinoma [ID1367]. 2019. https://www.nice.org.uk/guidance/ta592/documents/committee-papers

Applications for COVID-19

Limited time to act

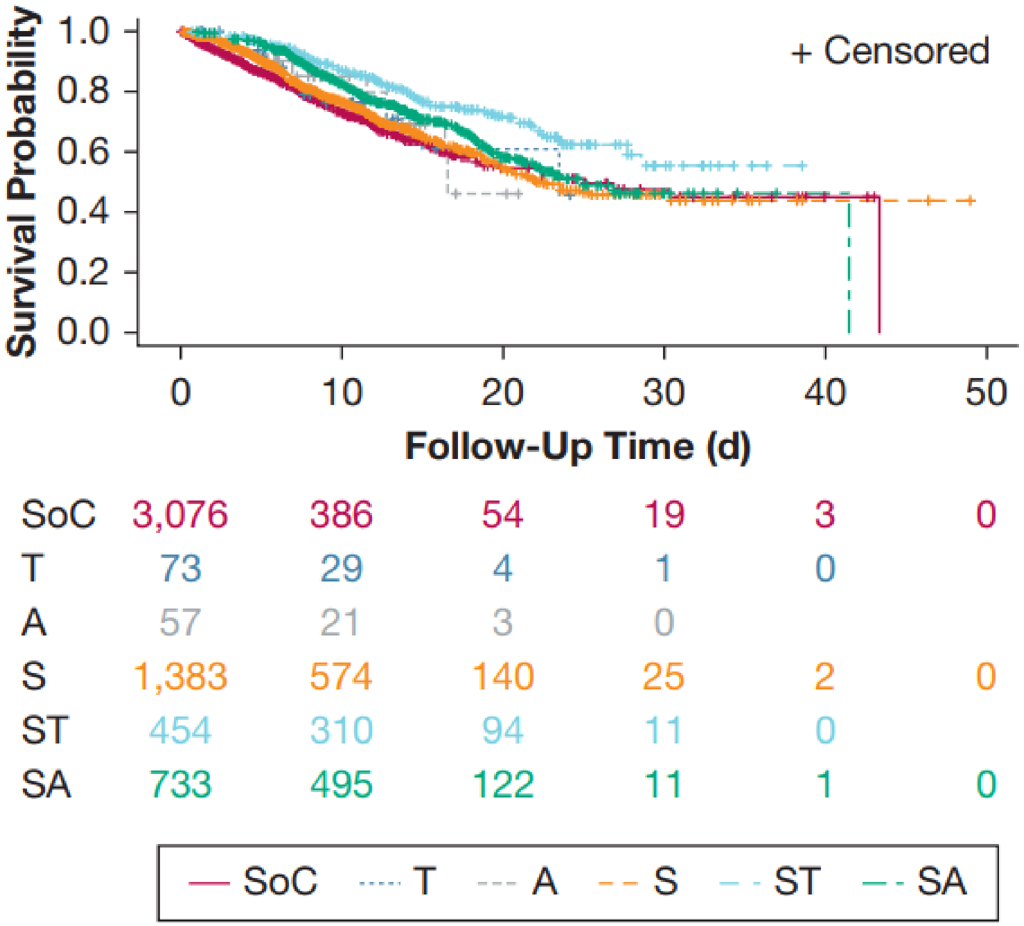


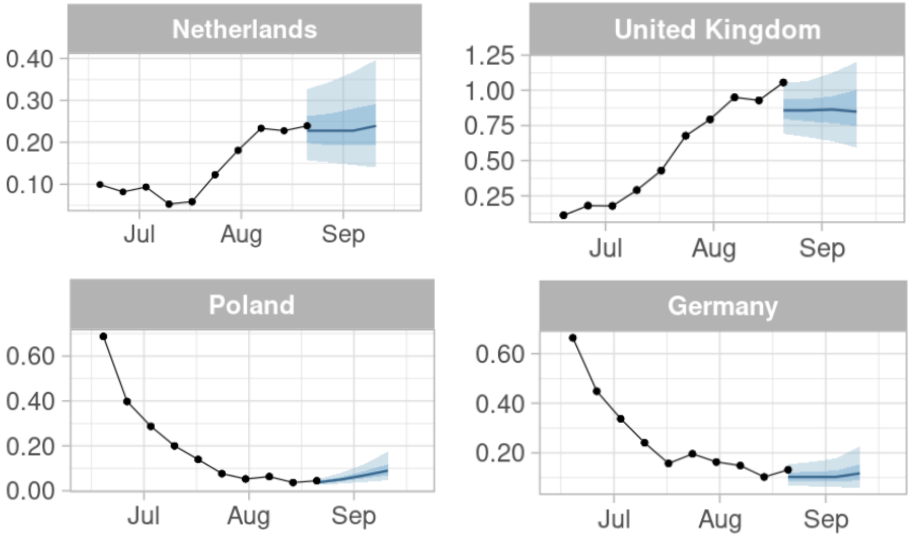
Figure 2 – *Model-based Kaplan-Meier plots showing treatment groups (adjusted for covariates). This figure represents the unadjusted Kaplan-Meier plots for treatment groups with number of patients at risk (ie, patients who remained admitted at the hospital at that time point). The treatment groups are as follows: A = anakinra only; S = steroid only; SA = steroids plus anakinra; SoC = standard of care; ST = steroids plus tocilizumab; T = tocilizumab only.*

- Proposed approach allows us to integrate expert opinion in systematic, structured, and transparent manner, which can be continually updated based on new information

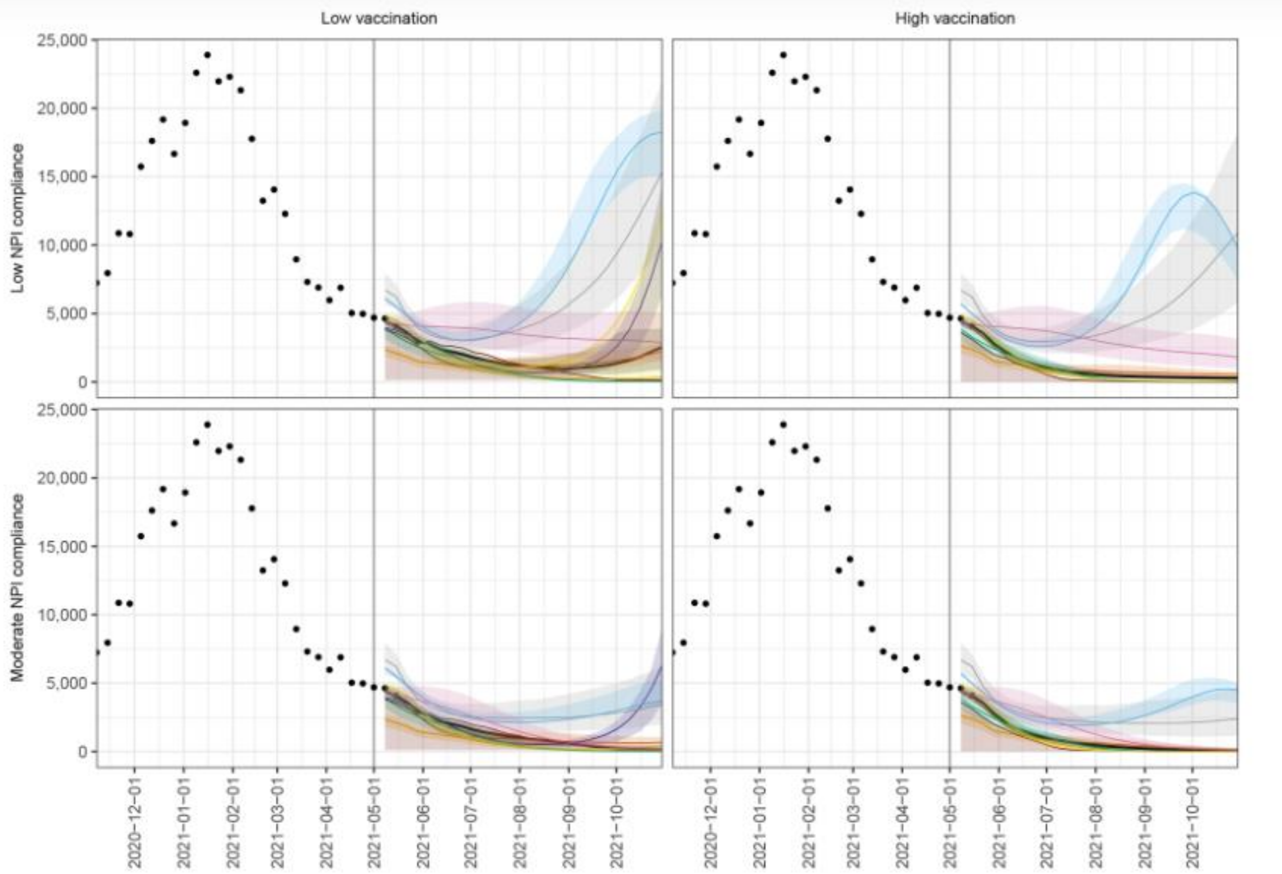
Source: Narain et al. Northwell COVID-19 Research Consortium. Comparative Survival Analysis of Immunomodulatory Therapy for Coronavirus Disease 2019 Cytokine Storm. Chest. 2021 Mar;159(3):933-948. doi: 10.1016/j.chest.2020.09.275.

Applications for COVID-19

European COVID-19 Forecast Hub: True and predicted (Aug 16, 2021) deaths per week per 100,000



Scenario Modeling Hub: death over time



Source: https://www.technologyreview.com/2021/05/28/1025478/covid-ensemble-model-forecast-trustworthy/ https://covid19forecasthub.eu/

Applications for COVID-19

COVID-19 reopening strategies at the county level in the face of uncertainty: Multiple Models for Outbreak Decision Support

- **COVID-19 Forecast Hub** and **European COVID-19 Forecast Hub**
  - Performance of individual models was highly variable
  - Most accurate short-term predictions from many models '*ensemble model*'
  - Accuracy reduced and uncertainty grows for longer term predictions
- **COVID-19 Scenario Modeling Hub**
  - Scenario models (what if?) longer term predictions (4-6 months) for exploring hypotheticals
  - Identify sources of uncertainty that may be relevant to decision-making (i.e. locations with high vaccine hesitancy, variant prevalence, duration of immunity from natural infection, state-level management practices)
  - **Formal expert elicitation methods** applied to multiple models to generate, retain, and synthesize valuable individual model ideas and share insights in group discussions, while minimizing various cognitive biases.
  - **Decision-theoretic framework** to account for within- and between-model uncertainty to assess actions in a timely manner

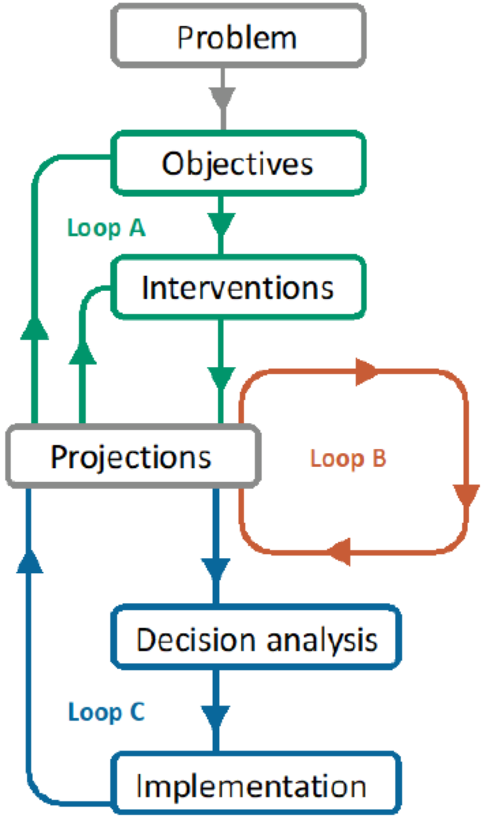


Fig. 1: Multiple Models for Outbreak Decision Support (MMODS) framework

Source: https://covid19forecasthub.org/ https://covid19forecasthub.eu/ https://covid19scenariomodelinghub.org/viz.html https://www.medrxiv.org/content/10.1101/2020.11.03.20225409v1.full.pdf+html
https:Shea, K., Runge, M.C. Pannell, D., Probert, W.J.M. Li S.-L., Tildesley, M., Ferrari, M. (2020) Harnessing multiple models for outbreak management. Science 368: 577-579

PRECISIONheor

Increasing use of expert elicitation related to clinical trials

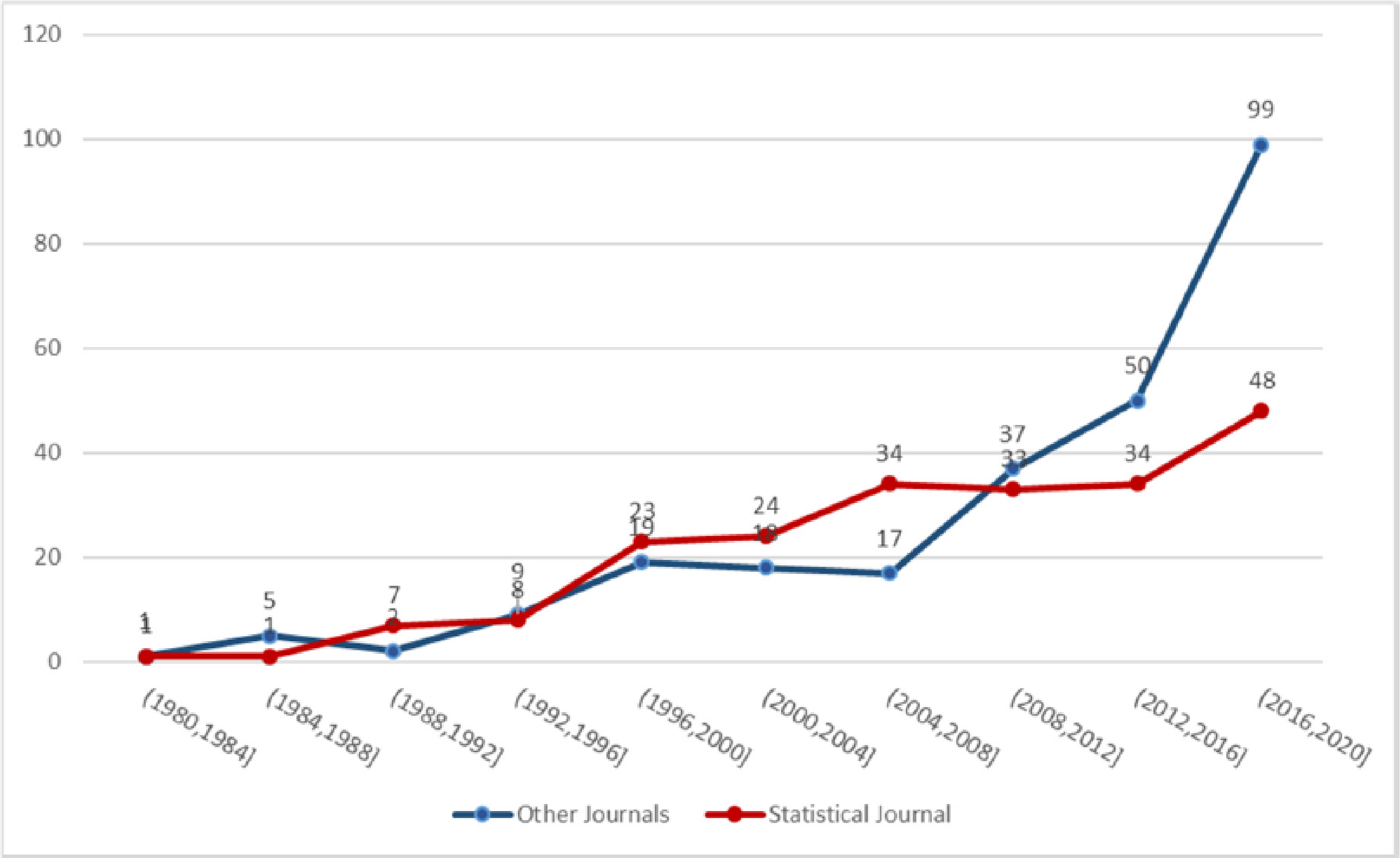


Figure 2. Articles pertinent to Prior Elicitation (n = 470) according to Journal type and year.

Source: Azzolina, Berchialla, Gregori Baldi. Prior Elicitation for Use in Clinical Trial Design and Analysis: A Literature Review. Int. J. Environ. Res. Public Health 2021, 18(4), 1833; https://doi.org/10.3390/ijerph18041833

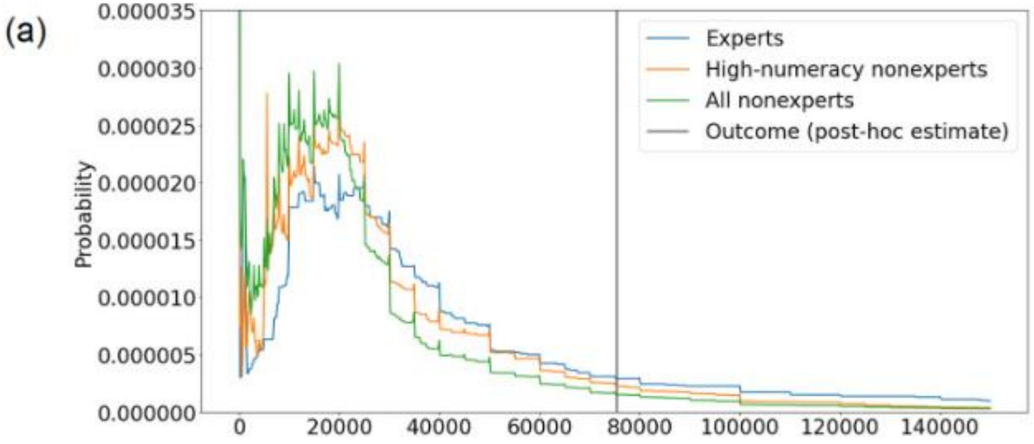
PRECISIONheor

Applications for COVID-19

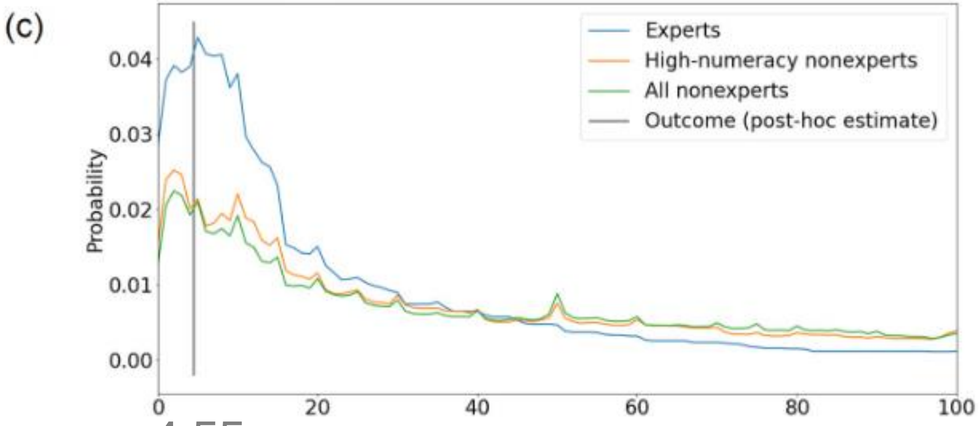
How accurate are experts in their predictions regarding survival?

	Question 1
Question	How many people in the country you're living in do you think will have died from COVID-19 by December 31st 2020?
How true outcome estimate was derived	Total number of "deaths within 28 days of positive test" having a date of death earlier than 1 Jan 2021

75,346



	Question 3
Question	Out of every 1000 people who will have been infected by the virus worldwide, how many do you think will have died by December 31st 2020 as a result?
How true outcome estimate was derived	1000 multiplied by the age-specific infection fatality rates estimated by the Imperial College COVID-19 response team in Oct 2020, weighted by worldwide age distribution



4.55

Source: Recchia G, Freeman ALJ, Spiegelhalter D(2021) How well did experts and lay people forecast the size of the COVID-19pandemic? PLoSONE16(5):e0250935.https://doi.org/10.1371/journal.pone.0250935

Challenges and additional research

- Bojke et al. 2021: ‘Area future research: Survival parameters - how to elicit parameters of survival models, in particular uncertainty relating to these’
- Extending model to include traditional parametric models used for CEA
- Extending models to use exact time-to-event times (events and censor) rather than discrete hazards
- Exploring alternative model formulations to incorporate the dependency of expert opinion on the observed data from clinical trial
  - Including parameters to explicitly weight expert opinion vs. observed data
  - Using trial as prior and modeling expert information
  - Using hierarchical model to explicitly capture between-expert heterogeneity
- Considering how dependent expert opinion is on the observed data
  - Size of study and N at risk over time
  - Provide no information on trial vs. different hypothetical data sets
- Model averaging, weighted by expert opinion?

RAFFAELE VARDAVAS

Mathematical Models of Disease Transmission and Model Parameter Uncertainties

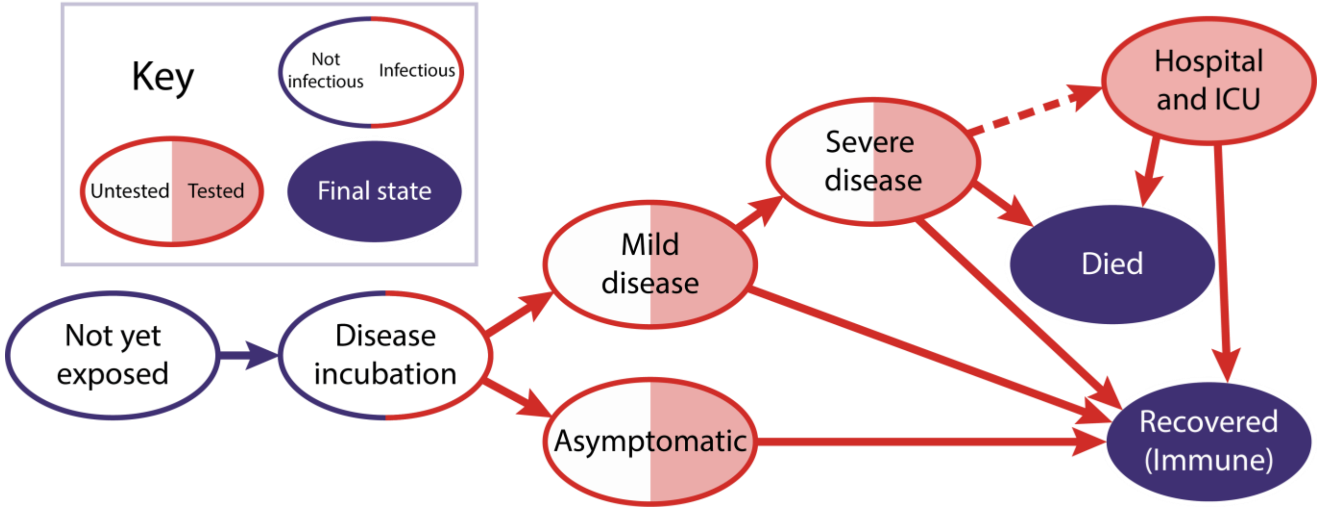
RAND CORPORATION

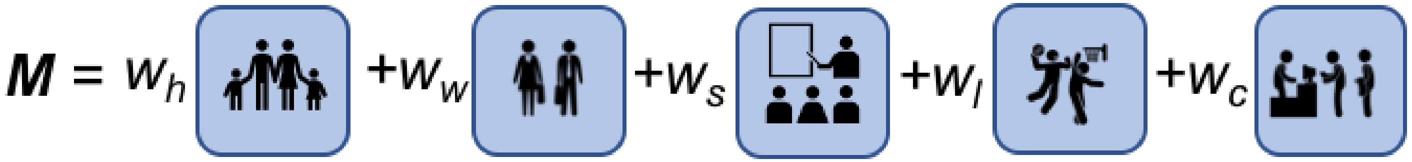
Types of models

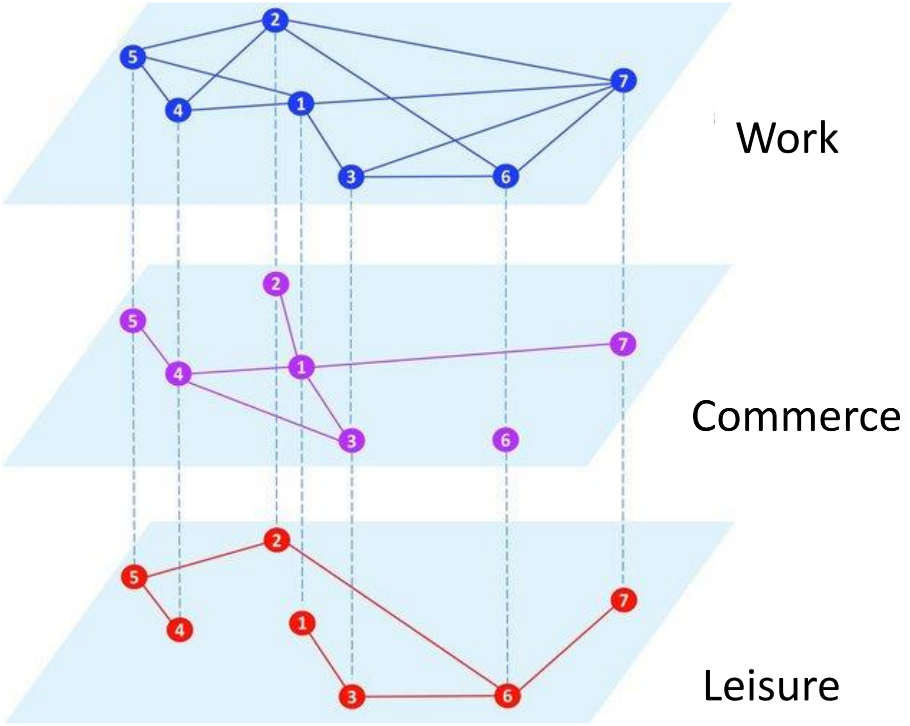
- Statistical-based models: Regression-based, Curve-fitting, Machine-learning.
- Theory-based model: Based on modeling the causal mechanisms of disease transmission and progression. Think SEIR-type models.
    - Equation-Based/ Population-Based models (PBM): Compartmental stock and flow models formulated using coupled ordinary differential equations. Top-down approach. Fast.
    - Microsimulation models: Individual-level, Bottom-up, Allows for greater heterogeneity, stochastic, transition rates are usually external inputs.
    - Agent-Based models (ABM): Similar to Microsimulation models, many transition rates are informed by internal causal rules; They can better incorporate space and network influences.

Manheim, David, Margaret Chamberlin, Osonde A. Osoba, Raffaele Vardavas, and Melinda Moore, Improving Decision Support for Infectious Disease Prevention and Control: Aligning Models and Other Tools with Policymakers' Needs. Santa Monica, CA: RAND Corporation, 2016. https://www.rand.org/pubs/research_reports/RR1576.html.

RAND's COVID-19 PBM Model Structure





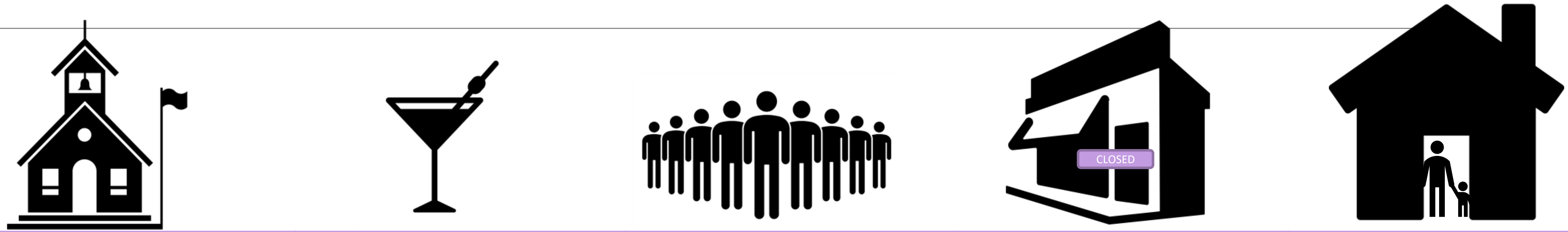


5 age groups

2 health statuses

Raffaele Vardavas, Pedro Nascimento de Lima, and Lawrence Baker
Could periodic nonpharmaceutical intervention strategies produce better COVID-19 health and economic outcomes? Journal on Policy and Complex Systems, Volume 7, Number 1 Spring 2021

Social Distancing Policy Portfolios



	Close schools	Close bars and restaurants	Ban large events	Close nonessential businesses	Shelter-in-place (except essential workers)
Level 1	✓				
Level 2	✓	✓			
Level 3	✓	✓	✓		
Level 4	✓	✓	✓	✓	
Level 5	✓	✓	✓	✓	✓

Modeling the Interventions

Nonpharmaceutical Intervention (NPI) Portfolio	$w_{household}$	w_{work}	w_{school}	$w_{commercial}$	$w_{recreational}$	w_{other}
Baseline	18.40%	32.8%	16.10%	8.98%	4.23%	19.47%
Level 1: Close schools	25.70%	23.0%	0.00%	8.98%	5.07%	17.93%
Level 2: Close schools, bars, and restaurants; and ban large events	22.00%	23.0%	0.00%	6.29%	3.80%	11.68%
Level 3: Close schools, bars, and restaurants; ban large events; and close nonessential businesses	18.40%	6.56%	0.00%	3.59%	2.96%	5.84%
Level 4: Close schools, bars, and restaurants; ban large events; close nonessential businesses; and quarantine the most vulnerable	Above the scenario for nonvulnerable groups, below the scenario for vulnerable groups (65 or older or those with a chronic condition)					
Level 5: Close schools, bars, and restaurants; ban large events; close nonessential businesses; and quarantine everyone but essential workers (shelter-in-place)	18.40%	6.56%	0.00%	2.69%	0.85%	1.95%

Validation and Calibration

- Validity: Does prediction agrees with the observed reality?
  - Face Validity: Check the validity of both the predictions and the assumptions. Do assumptions make sense at its face value, in view of the existing knowledge and our common sense?
  - Cross Validation: Can the model reproduce observations or data that was not used to inform it?

- Calibration:
  - Sample combinations of parameter values using a Latin Hyper Cube approach (LHS).
  - Generate hundred thousands of cases each with different combinations of parameter values.
  - At disease invasion phase we linked the growth rate to the R_0 expressed by a mathematical combination of model parameters found using the Next Generation Method.
  - We applied Incremental Mixture Approximate Bayesian Computation (IMABC) approach.
  - Filtered cases to match hospitalizations and deaths timeseries based on the applied NPI level

Uncertainties

- Well-characterized: those for which historical or clinical data provides information.
  - Examples: Length of disease states, magnitude of the seasonal effect on mixing.
  - These parameter values are chosen the the IMABC calibration.
  - Provide a subset of LHS cases of parameter value combinations giving the "Calibrated Set"

- Deep: those for which calibration or existing clinical evidence provides little information.
  - Examples: Vaccine efficacy to prevent transmission, changes in transmissibility, immunity duration.
  - These Parameters are further sampled using a LHS and combined with the Calibrated Set to explore the impact of policies across these deep uncertainties and select policies that perform well and are Robust to these deep uncertainties.

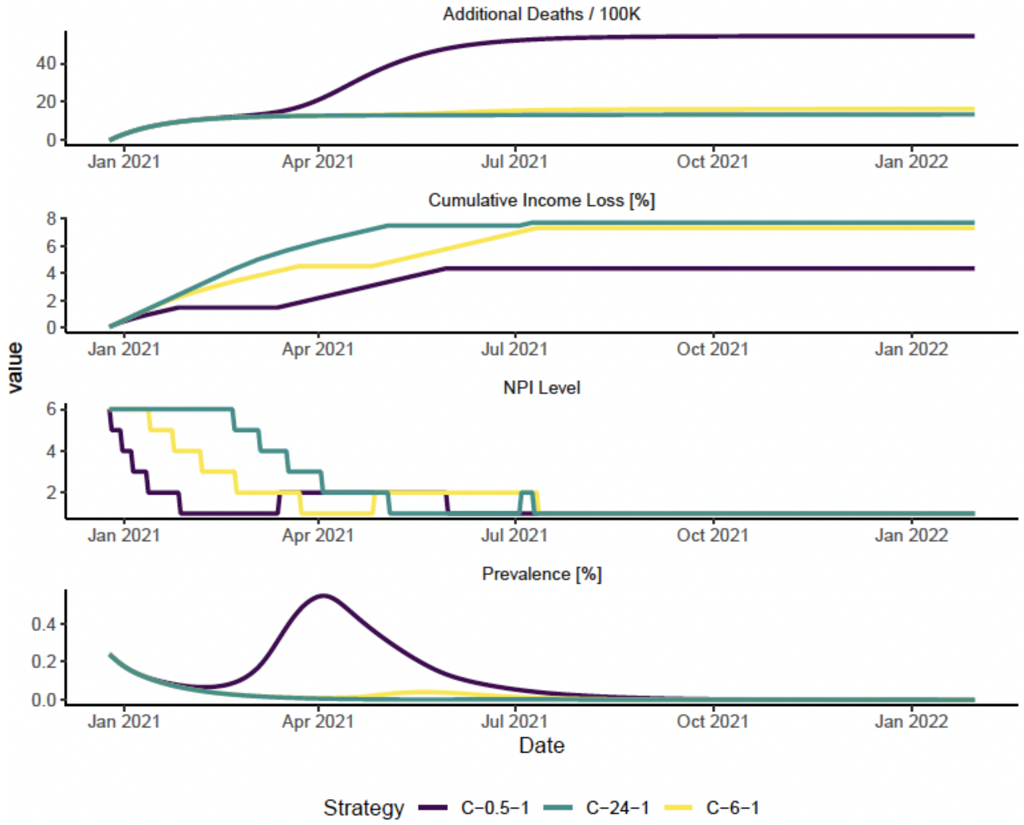
Robust Decision-Making Framework

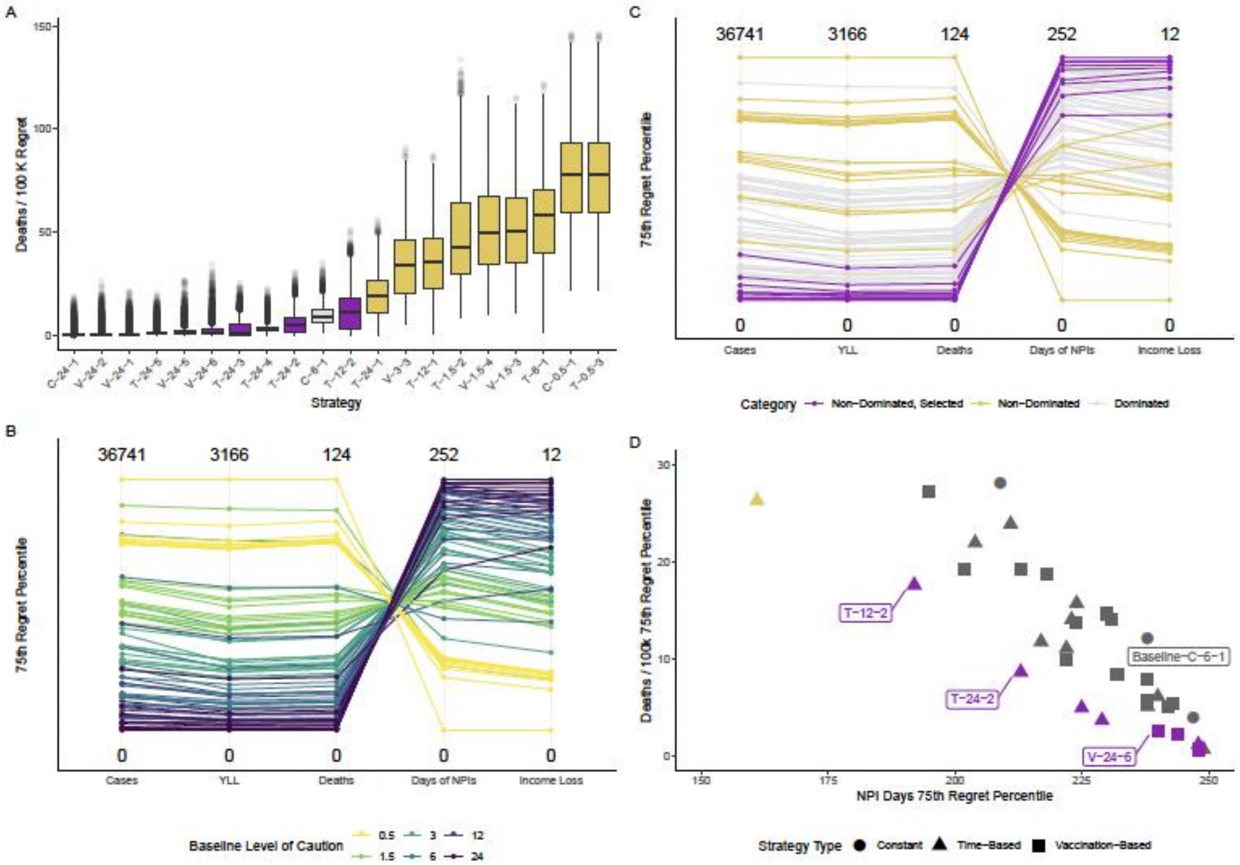
X - Uncertainties	**L - Policy levers**

• Vaccine efficacy to prevent transmission
• Loss of immunity
• Behavioral response to vaccination
• Willingness to vaccinate
• Changes in transmissibility (i.e., induced by variant strains)
• Actual vaccination Rate
	• Baseline level of caution x_b
	• NPI strategy $s \in \{C, T, V\}$
	• Time-based strategies $s = T$
	  – Level of caution factor α
	  – Transition date T_α
	• Vaccination-based strategies $s = V$
	  – Vaccination reference point V_{mid}
	  – Relaxation rate k_c

R - Relationships (models)	**M - Metrics**
Meta-population deterministic ODE [10, 33] Computable general equilibrium model [36]	75^{th} Regret percentile of deaths / 100 k people, years of life lost, cases, income loss, and days under NPIs

Example Results





Lima, Pedro Nascimento de, Robert J. Lempert, Raffaele Vardavas, Lawrence Baker, Jeanne S. Ringel, Carolyn M. Rutter, Jonathan Ozik, and Nicholson Collier,
Reopening California: Seeking Robust, Non-Dominated COVID-19 Exit Strategies. Santa Monica, CA: RAND Corporation, 2021. https://www.rand.org/pubs/working_papers/WRA1080-2.html.

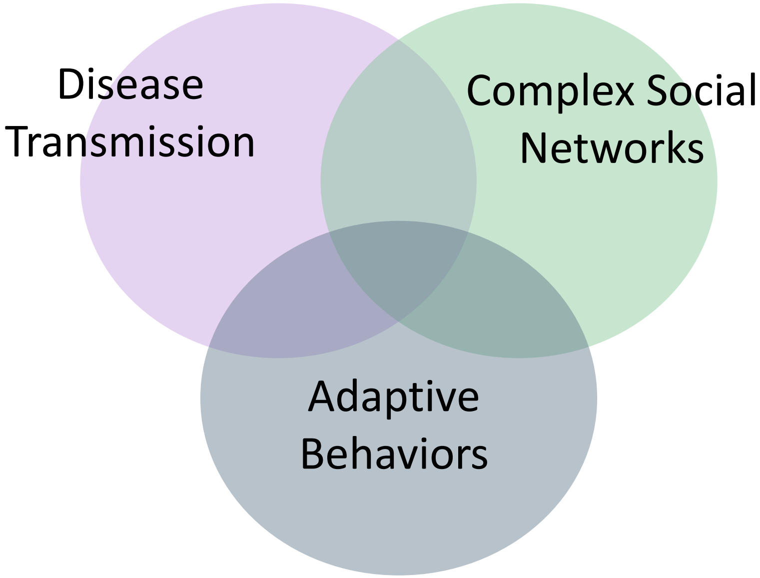
Modeling Infectious Behaviors

- Very few models directly include adaptive behavioral effects and how risk perceptions influence transmission dynamics.

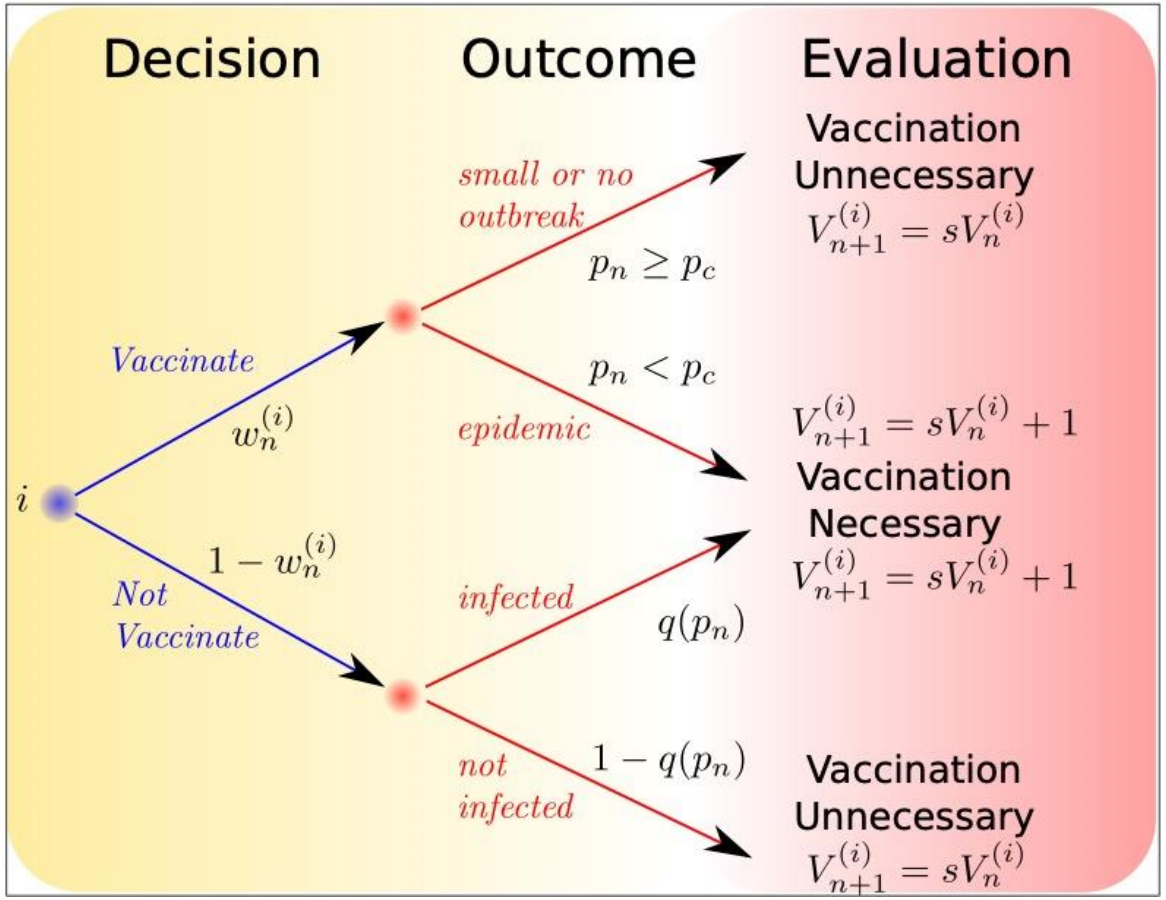
- PBMs can include feedback loops to describe at the aggregated level behavioral responses.

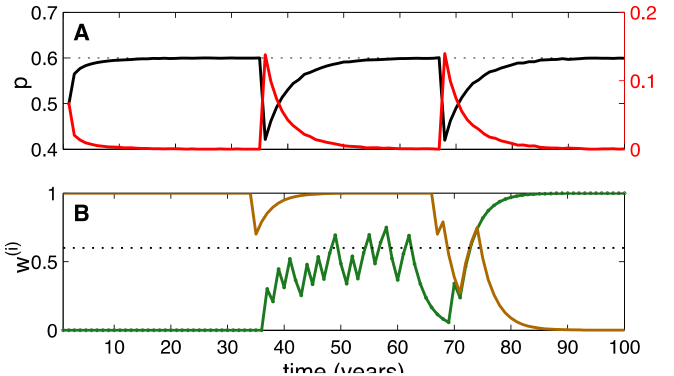
- ABM allow explicit specification of micro-level behavioral mechanisms that affect public risk perceptions and individual protective behaviors and attitudes.

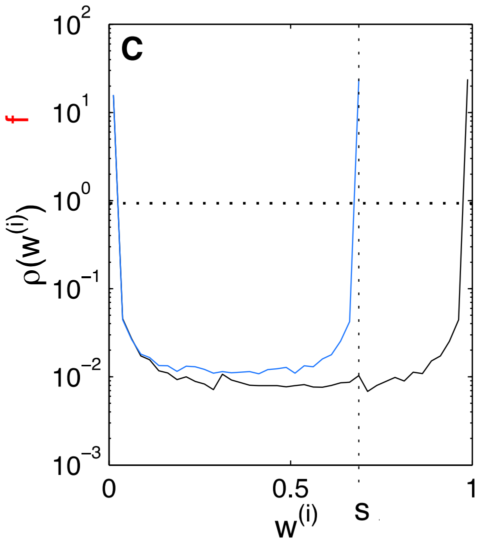
- ABM allow to include complex social networks which provide preferential transmission paths and determine influences in behaviors.

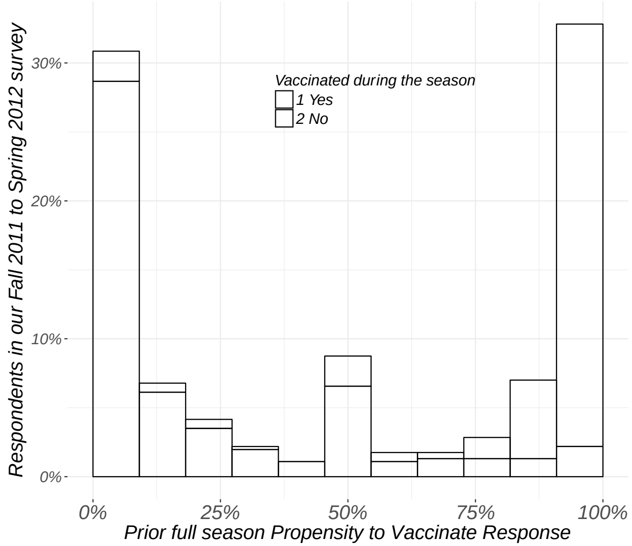


Inductive Reasoning in Flu Vaccination

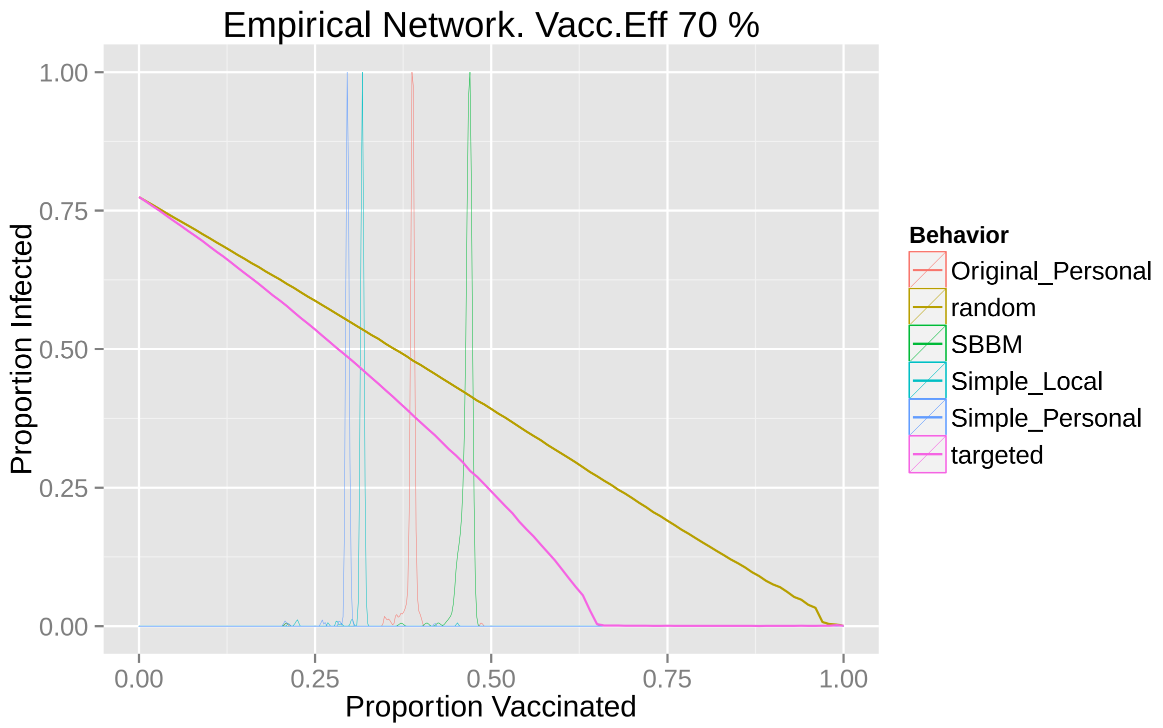


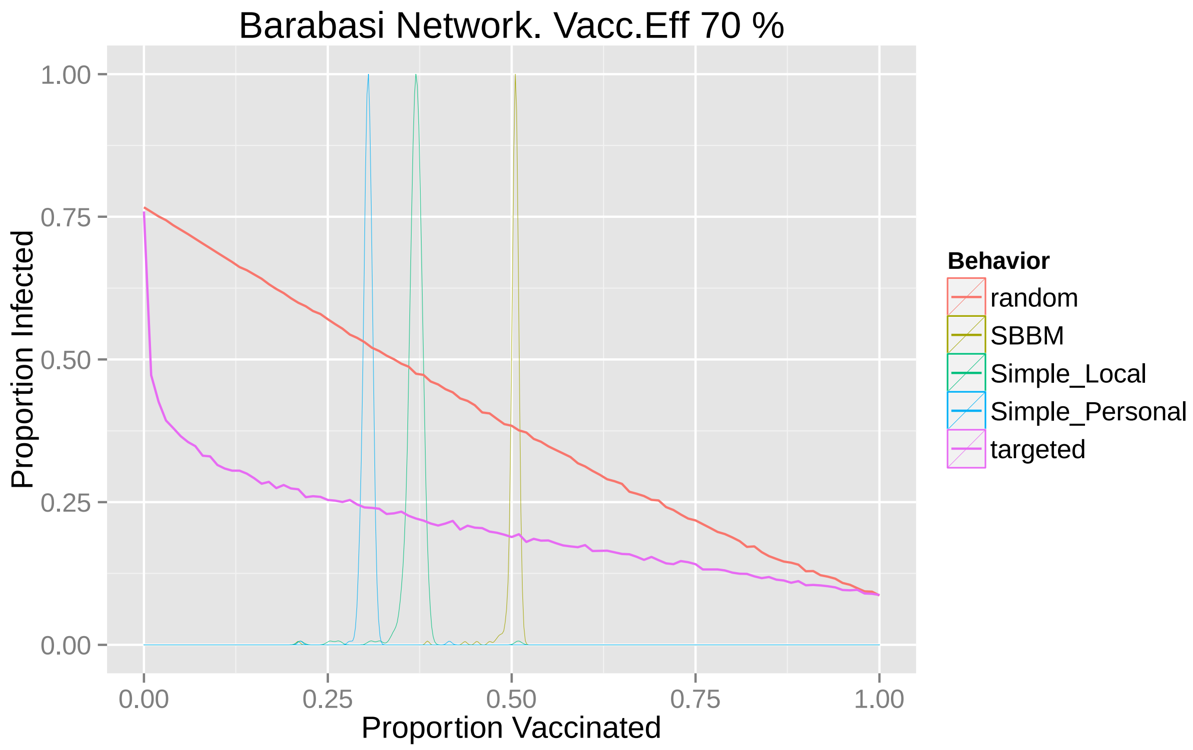






Network Effects





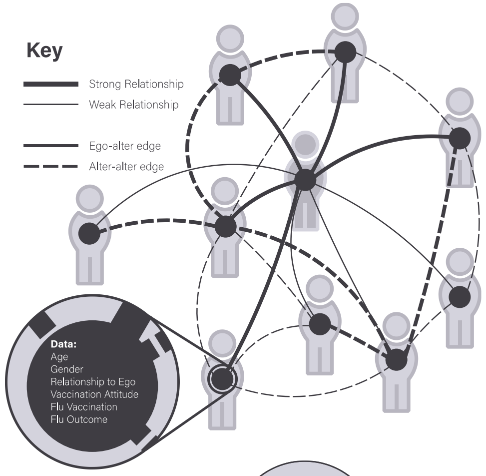
Tailored Survey to inform ABM

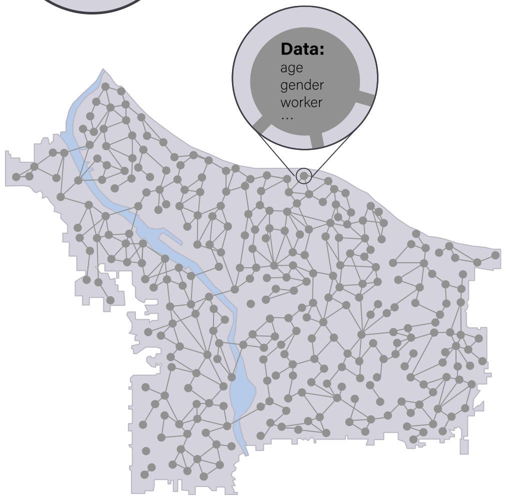
- Design and field longitudinal surveys tailored to inform simulation models.

- Include Egocentric Social Network data which can be fused using new ML-methods to Sociocentric Networks.

- RAND's American Life Panel – FluPaths: Helps inform a new ABM of influenza vaccination decision.

- New COVID-19 ABMs should be developed to with adaptive behaviors informed by Surveys and based on Cognitive Models such as Adaptive Control of Thought - Rational (ACT-R) framework.





Acknowledgements

Dr. Andy Parker

Pedro Nascimento de Lima

Lawrence Baker

Dr. Rob Lempert

Dr. Matt Walsh

Dr. Paul Davis

NIH

National Institute of Allergy and Infectious Diseases

Feel free to reach out with any additional questions!

			
Anthony O'Hagan, PhD, Emeritus Professor of Statistics, University of Sheffield	**Shannon Cope, MSc,** Vice President, PRECISIONheor	**Raffaele Vardavas, PhD, MSci,** Senior Mathematician, RAND Corporation	**Marlon Graf, PhD,** Research Economist, PRECISIONheor
a.ohagan@sheffield.ac.uk	shannon.cope@precisionvh.com	rvardava@rand.org	marlon.graf@precisionvh.com

THANK YOU

We look forward to staying in touch.

PRECISIONheor

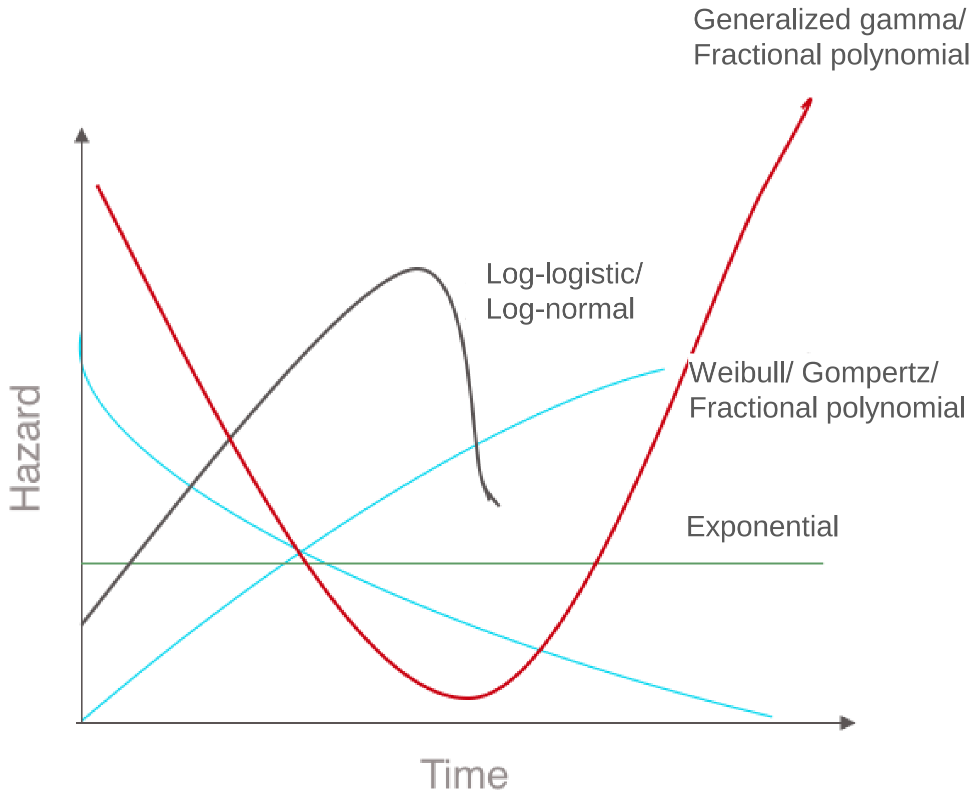
a precision value & health team

PRECISIONheor

HTA: Lifetime horizon requires fully extrapolated survival

Models used to extrapolate time-to-event outcomes for clinical trial for cost-effectiveness analysis

Models	Distribution	Characteristic of hazard function
TRADITIONAL	Exponential	- Constant over time
	Weibull/ Gompertz	- Can decrease or increase monotonically, but cannot change direction (i.e., no turning points)
	Log-logistic/ Log normal	- Can be non-monotonic
	Gamma/ Generalized gamma	- Can simplify to Weibull, exponential, and log normal distributions - Can be non-monotonic
FLEXIBLE	Fractional polynomials (second-order)	- More flexible extensions of Weibull and Gompertz models allowing arc- and bathtub shaped hazard functions
	Piecewise	- Fit alternative parametric functions over different segments of time
	Splines (hazard, odds, normal scales)	- Piecewise polynomials joined at knots
	Other: landmark, and mixture (cure) models	- Landmark on point of response - Mixture models to capture subgroups



Latimer, N. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011. Available from http://www.nicedsu.org.uk
Rutherford, MJ., Lambert, PC., Sweeting, MJ., Pennington, R., Crowther, MJ., Abrams, KR., Latimer, NR. NICE DSU Technical Support Document 21. Flexible Methods for Survival Analysis. 2020 [Available from http://www.nicedsu.org.uk]

Prompts during the process

Ensure plausible ranges are wide enough to capture uncertainty in estimates

- Upper plausible range (U)
  - Suppose in the future somebody tells you that the true value of the QoI has now been established quite accurately, and that its value is greater than your U. What would be your reaction?
    A. Accept that this is plausible and admit that your judgement was flawed?
    B. Strongly suspect that the method used to determine this value was flawed?
    - If you selected A, revise your U upward

- Most likely value (MVL)
  - Suppose we offer a large prize if you can guess correctly whether the true value of the QoI is above or below your most likely value (you don't lose anything if you choose incorrectly)
    A. Which way would you guess? Would you prefer 'above' or 'below'?
    B. Make sure you really don't have a preference for guessing 'above' or 'below'

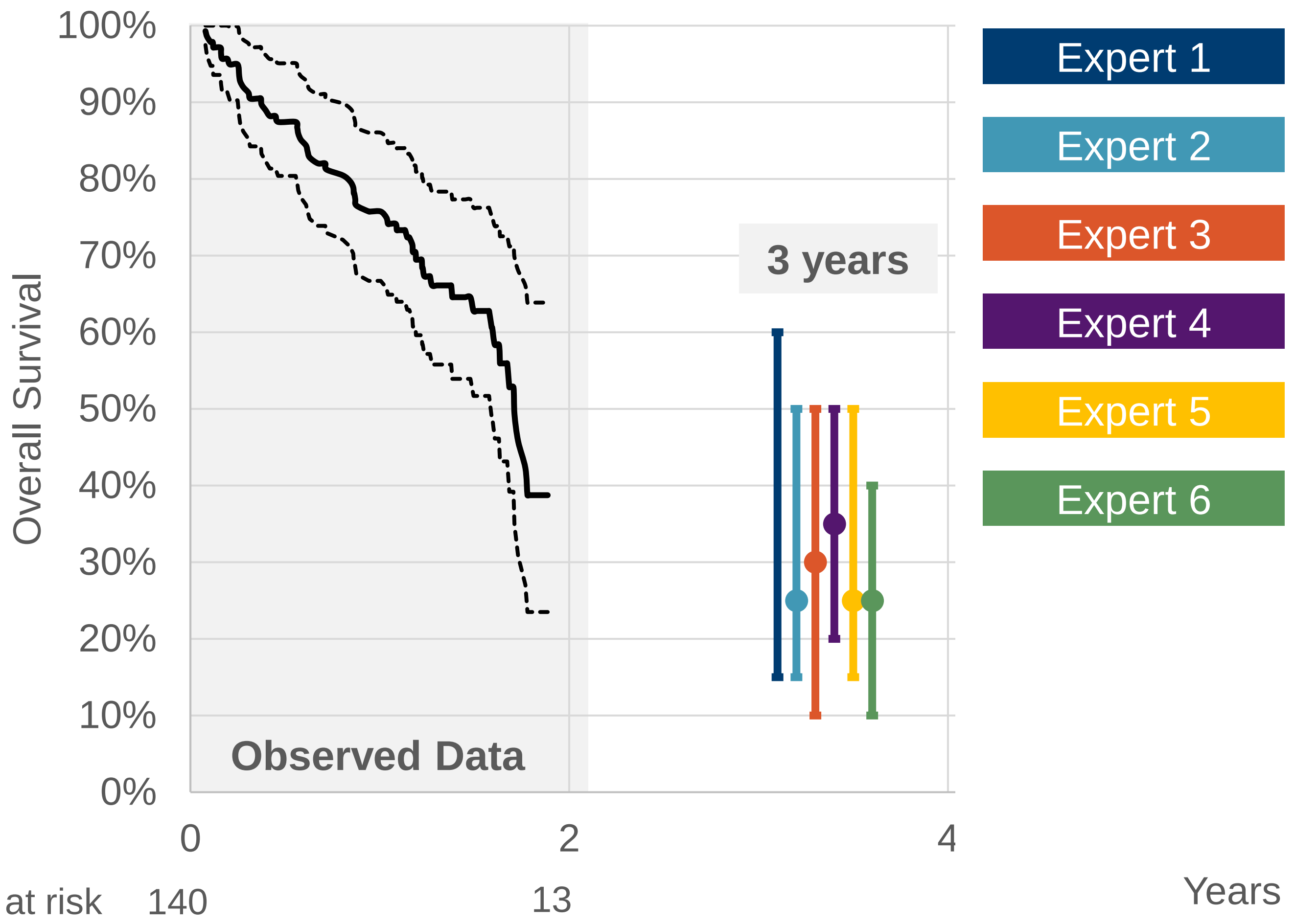
Consensus Meeting

Rational Impartial Observer (RIO) perspective

- Consensus represents a synthesis of the available collective knowledge and judgement
- Due to different clinical experiences, expertise, and interpretations of the trial/study data, we do not expect experts to reach complete agreement
- Rather, we ask experts to judge what a rational impartial observer might reasonably believe
- A *Rational Impartial Observer* represents an external observer (such as an objective decision-maker) who has observed your individual estimates, has listened to your discussion, and understands your arguments
  - They may find merit in all differing opinions and will likely give weight to the persuasiveness of each argument when ultimately considering judgement

Consensus by time point

Triple class exposed relapsed refractory case study



Discussion Topics

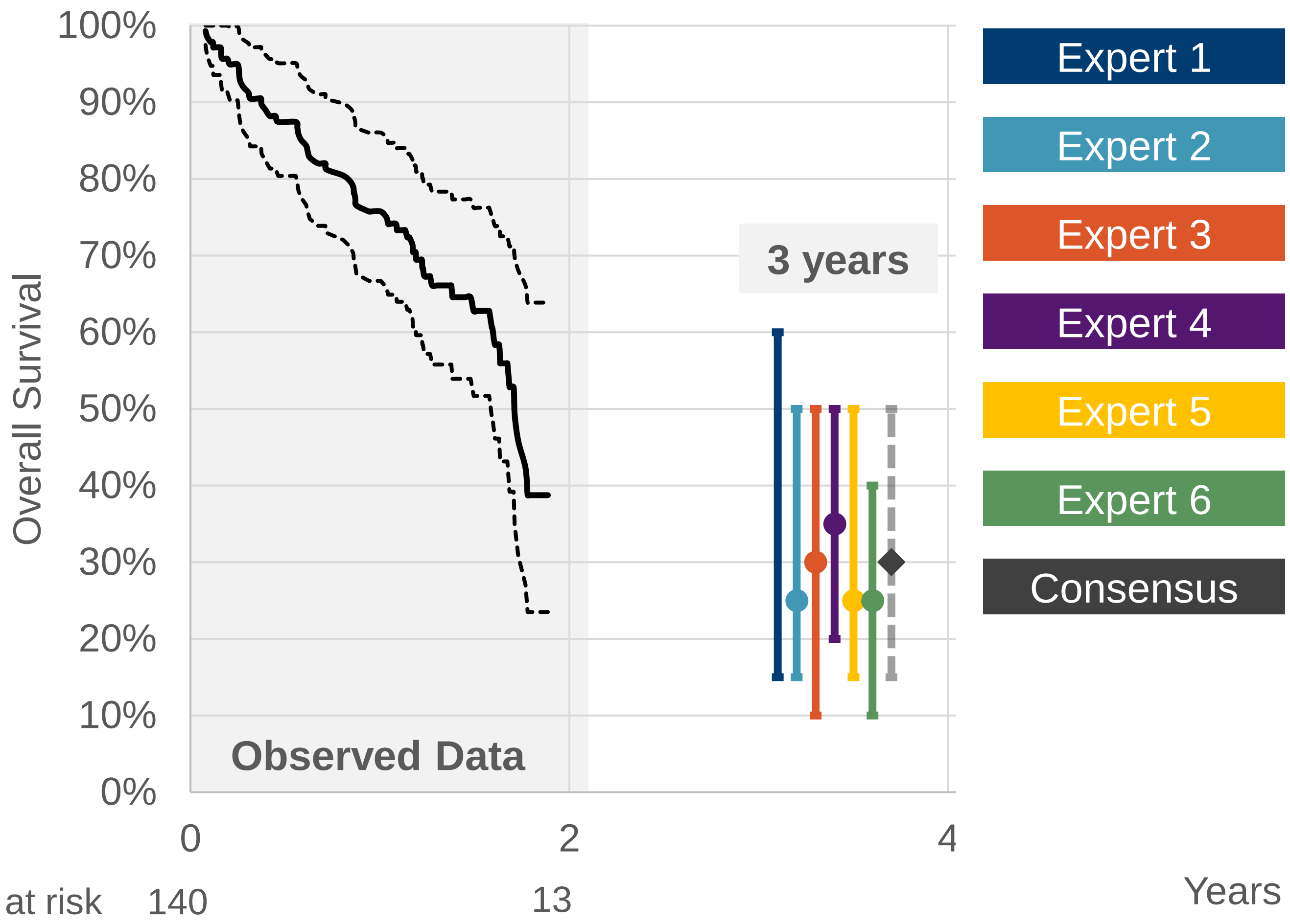
1. Experts **1** and **6**, do you have insights on upper plausible limits?
2. Experts **3**, **4**, and **6**, do you have insights on lower plausible limits?
3. All experts, do you have insights on slight variability in most likely values?
4. Any other unique insights to share?

Ayers, Cope, Dhanda, Towle, Mojebi, Delforge, Rodriguez Otero, Trudel, Weisel, Zamagni, Hari. Expert elicitation of long-term survival for patients with relapsed/refractory multiple myeloma treated with idecabtagene vicleucel (ide-cel, bb2121) in the KarMMa phase 2 trial. Presented at the 18th International Myeloma Workshop; September 8–11, 2021, Vienna, Austria. Presentation Number 1074971

PRECISIONheor

Consensus by time point

Triple class exposed relapsed refractory case study



Discussion Topics

1. Experts **1** and **6**, do you have insights on upper plausible limits?
2. Experts **3**, **4**, and **6**, do you have insights on lower plausible limits?
3. All experts, do you have insights on slight variability in most likely values?
4. Any other unique insights to share?

Ayers, Cope, Dhanda, Towle, Mojebi, Delforge, Rodriguez Otero, Trudel, Weisel, Zamagni, Hari. Expert elicitation of long-term survival for patients with relapsed/refractory multiple myeloma treated with idecabtagene vicleucel (ide-cel, bb2121) in the KarMMa phase 2 trial. Presented at the 18th International Myeloma Workshop; September 8–11, 2021, Vienna, Austria. Presentation Number 1074971

Time-to-event model for single-arm clinical trial

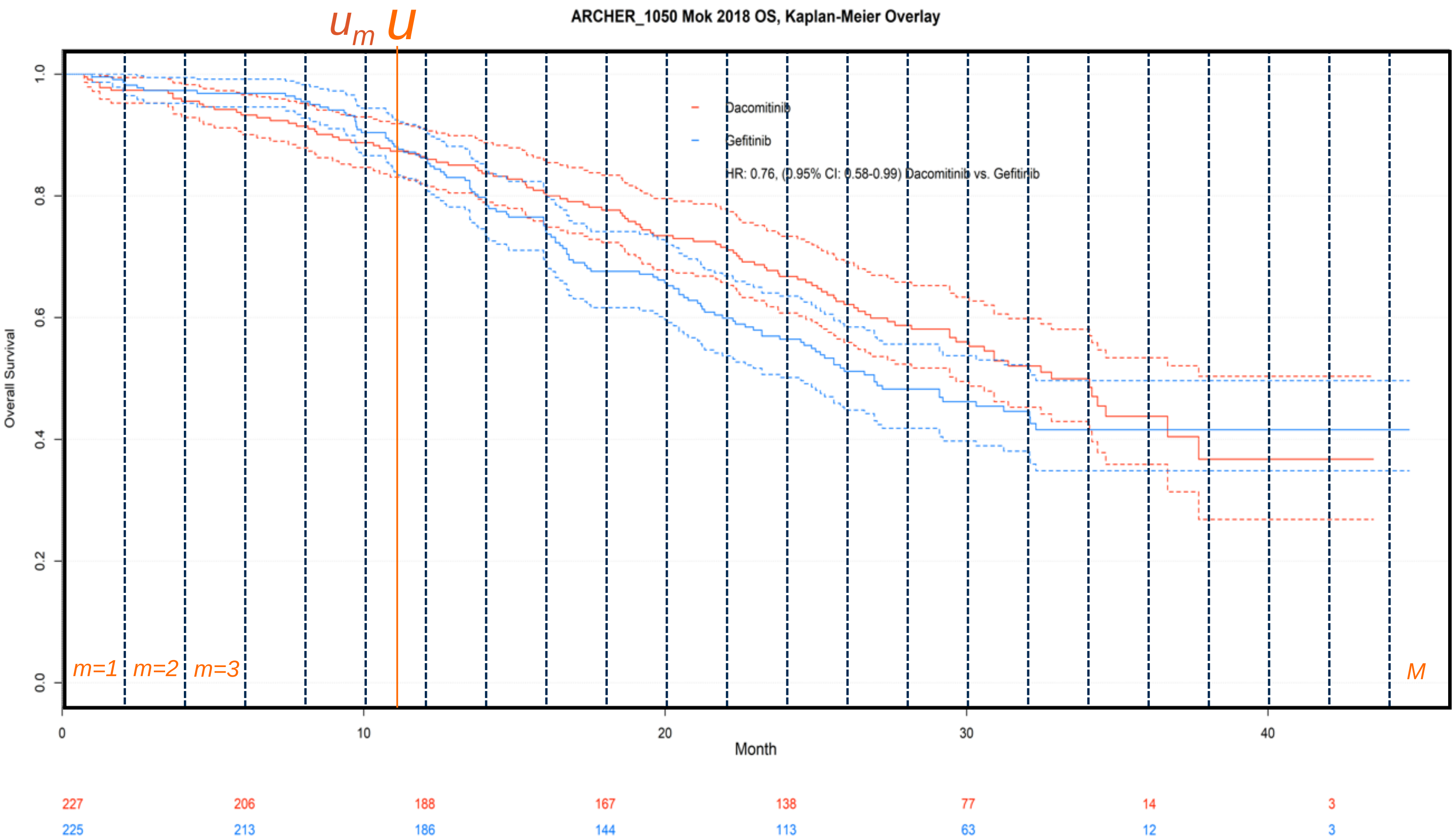
Bayesian fractional polynomial model for survival

- Construct interval data (m) from Kaplan Meir curve from single-arm clinical trial
- Create dataset with discrete hazards over time
  - Number of patients died during interval (r_m)
  - Number of patients at risk at beginning of interval (n_m)
  - Duration of interval (Δu_m)
- Model discrete hazard data using fractional polynomial assuming binomial likelihood for number of patents died per interval
- Use vague priors for all parameters

PRECISIONheor

Discrete hazards over time

Successive non-overlapping intervals provide evidence for conditional survival probabilities



Likelihood

$$r_m \sim \text{Binomial}(p_m, n_m)$$

Conditional survival probability at time *u*

Data

r	*n*
4	125
4	121
2	117
5	114
2	109
3	107
2	104
4	94
4	90
3	81
4	78
3	61
5	58
1	48
2	47
3	41
0	38
3	29
3	26
2	18

Time-to-event model for one single-arm study

Model: $\ln\left(h(u)\right) = \begin{cases} \alpha_0 + \alpha_1\, u^{p_1} + \alpha_2\, u^{p_2} & p_1 \neq p_2 \\ \alpha_0 + \alpha_1\, u^{p} + \alpha_2 u^{p} \ln(u) & p = p_1 = p_2 \end{cases}$ with $u^0 = \ln(u)$

Likelihood for trial data: $r_m \sim \mathrm{Binomial}(p_m, n_m)$

Link discrete hazard to cumulative hazard:

$$p_m = 1 - e^{-\Delta u_m h_m}$$

$$h_m = \frac{-\ln(1 - p_m)}{\Delta u_m}$$

Defining the quantity of interest for time-to-event outcomes

Challenges in eliciting time-to-event outcomes

- Alternative possible quantities of interest to elicit
- Need to consider censored events
- Need to elicit quantities at multiple time points
- Dependencies in quantities over time
- Expectation that uncertainty should fan out over time
- Models based on derived estimates from observed events
- Interest in using flexible survival models (multiple parameters)
- Limited understanding of time-to-event summary statistics by experts
- Possible biases specific to this context

PRECISIONheor

Reference protocol (Bojke 2021)

Quantities to elicit

- Clearly relates to decision problem so quantities are fit for purpose to pivotal trial population
- Target quantities decomposed into quantities that are simpler (reflect what experts are likely to observe)
- Elicited quantities coherent with model parameters
- Use dynamic graphical displays
- Includes pilot stage
- *Considers what experts may not observe (censoring)*

Bojke et al. 2021. Developing a reference protocol for structured expert elicitation in health-care decision-making: a mixed-methods study. Health Technology Assessment Vol 25(37). DOI 10.3310/hta25370

Applications for COVID-19

Limited evidence, especially in specific populations

Received: 21 April 2020 | Revised: 16 May 2020 | Accepted: 18 May 2020

DOI: 10.1111/apa.15373

A DIFFERENT VIEW

ACTA PÆDIATRICA
NURTURING THE CHILD
WILEY

Efficacy, safety and cost-effectiveness of hydroxychloroquine in children with COVID-19: A call for evidence

Carlos E. Rodríguez-Martínez[1,2] | Ricardo M. Fernandes[3,4] | Daniel B. Hawcutt[5,6] | Ian P. Sinha[7] | Rafael Leite Pacheco[8]

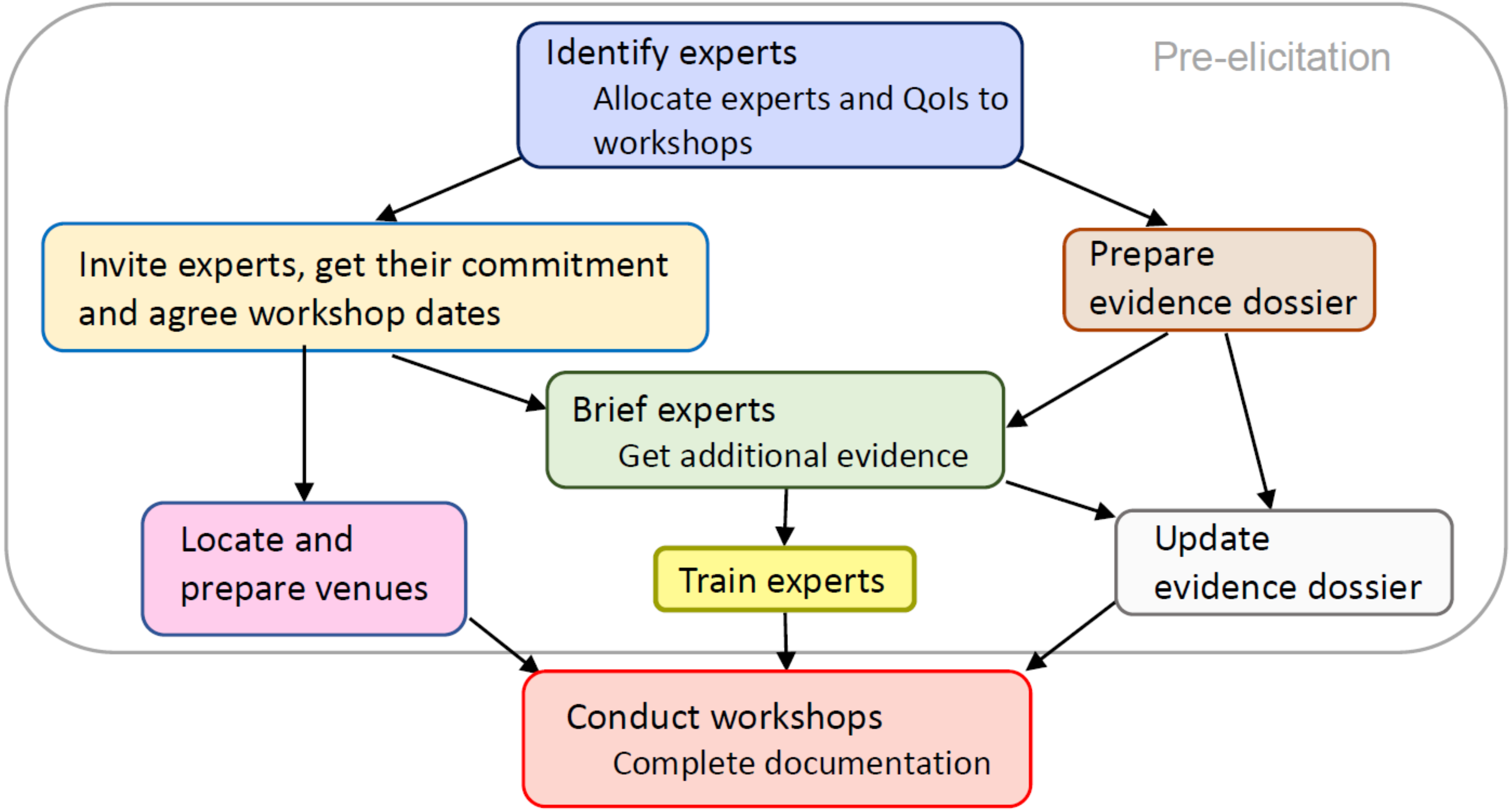
[1]Department of Pediatrics, School of Medicine, Universidad Nacional de Colombia, Bogota, Colombia

[2]Department of Pediatric Pulmonology and Pediatric Critical Care Medicine, School of Medicine, Universidad El Bosque, Bogota, Colombia

[3]Department of Pediatrics, Santa Maria Hospital, Centro Hospitalar Universitário Lisboa Norte, Lisboa, Portugal

Rodriguez-Martinex, Fernandes, Hawcutt, Sinha, Leite Pacheco (2020). Efficacy, safety and cost-effectiveness of hydroxychloroquine in children with COVID-10: A call for evidence. Acta Paediatrica. DOI: 10.1111/apa.15373

Overview SHELF



Source: Oakley J. E. and O'Hagan, A. (2019). SHELF: the Sheffield Elicitation Framework (version 4.0). School of Mathematics and Statistics, University of Sheffield, UK. (http://tonyohagan.co.uk/shelf)