



Measuring the effectiveness tobacco control policies

Ron Borland PhD

Workshop, 1st Asian SRNT Meeting, Bangkok, October 2008

Sponsors

- Roswell Park Cancer Institute Transdisciplinary Tobacco Use Research Center (US NIH funded)
- US CDC (Centers for Disease Control)
- IARC: International Agency for Control of Cancer



Who is this workshop for?

- Those interested in systematically improving public policy
- Those interested in how science works
- Those hoping to do better science
- Those wanting to use the outputs from science more effectively



Faculty

- Ron Borland PhD, Cancer Council Victoria, AUSTRALIA
- Geoffrey T Fong PhD, University of Waterloo, CANADA
- Anne Quah PhD, University of Waterloo
- Hua Yong PhD, Cancer Council Victoria
- Buppha Sirirassamee PhD, Institute of Population and Social Research, THAILAND
- Rahmat Awang PhD, National Poisons Center, MALAYSIA



Today	28 th Oct, 2008 Tuesday
09:00 – 9:30	Workshop overview ? Introduction from CDC rep?
09:30 – 10:30	Introduction to the IARC Handbook (Ron Borland)
10:30 – 10:45	Tea break
10:45 – 12:00	Models of evaluation (Geoff Fong)
12:00 – 13:00	Lunch break
13:00 – 14:30	Intro to cognitive testing (Anne Quah) Intro to mediational models (Hua Yong)
14:30 – 14:45	Tea break
14:45 – 16:15	Cognitive Testing workshop (Jim Thrasher & others)
16:15-16:30	Review of the day



Tomorrow	29 th Oct, 2008 Wednesday
09:00 – 10:30	Mediational models workshop (Hua Yong & others)
10:30 – 10:45	Tea break
10:45 – 12:00	Panel discussion (All) Summarize & conclude (Ron Borland)
Noon	End



Workshop aims

- Better understand how to use science to improve public policy
 - *Appreciate what we actually know*
- Appreciate criteria determining the power of a study design
- Understand dynamic relationship between theory and evidence
- Be able to build useful theoretical models
- Understand how to best equate questions across cultures and contexts



What can science do?

For any intervention:

- Find out whether it can work
 - *Prediction for the future*
- Demonstrate that it is working
 - *Documentation of the present*
- Identify ways of making it better or functions it does not serve
 - *Preparing for the future*



Research stages

- Understanding mechanisms
- Developing interventions
- Efficacy trials
 - *That they can work*
- Effectiveness trials
 - *That they do work (in particular contexts)*
- Dissemination trials
 - *How to get them used optimally*
- That they produce net population change
 - *In cost effective ways*



Policy initiatives

From:

- That they might work
 - *Contain potent components*

To

- Whether they work or not
 - *Context by context*
- How to make them work most effectively?



What do policy makers want?

Before committing

- Simplest possible fix for the problem
 - *Preferably fits into existing structures*
 - *Consistent with government's values and priorities*
- Confidence that it will work
- Confidence that it will deliver value for money
- Confidence it won't have adverse unintended effects
 - *Especially those predicted by opponents*



What do policy makers want? After implementation

- Evidence that it has worked
- Evidence that it delivers value for money
- Evidence that it delivers equity of outcomes
- Evidence that it is appreciated
- Reassurance that it has not had adverse unintended effects



Components of intervention effects

- Inherent potency (efficacy)
- Degree of compliance
- Expectancy beliefs
- Contextual factors
- Uptake rates



Intervention effects

A product of:

- **Intervention effectiveness (potency)**
 - *How much help per individual*

and

- **Intervention reach**
 - *How many helped*

Policy interventions typically high reach , but low potency



Evidence-based medicine

- Focus on randomised controlled trials
 - *Good for drug testing*
 - *Less useful in other areas*
- Generally assumes high degree of transferability from trials to real world
 - *Jumps from the answer to “Can it work?” to assume that it does work*



Limitations of RCTs

- Impractical when can't randomize
 - *Compromises threaten validity more than understood*
- Compartmentalizes knowledge
 - *Ignores learnings from related areas*
- Favours intense, costly interventions which target individuals
- Ignores who benefits



The ideal RCT

- Random allocation to condition
- Double blinding
 - *Participants don't know*
 - *Investigators (therapists) don't know*
- No communication between participants
- Intervention elements are independent
- Representative samples



Non-random allocation

- Usual in policy evaluations
 - *Except for analogue studies*
- Only as good as the comparability of the groups at baseline
 - *Requires pre-measures to assess*
- No changes other than the target one.



Compartmentalization of knowledge

- Need to move beyond the model developed for pharmacotherapies
- Most new drug therapies are unique
 - *Most reasonable to generalize from similar drugs*
 - *Unless mechanism fully understood*
 - Eg flu vaccine



We need a new approach, one that is designed to the needs of evaluating complex interventions, often implemented in conjunction with others.

The goal of this workshop is to introduce such a model



Measuring the Effectiveness of Tobacco Control Policies

IARC Handbook

Editors: R Borland & KM Cummings

Technical Editor: M Leon-Roux

Group of scientists from:

Universities, Cancer Research Organizations,

WHO, CDC, and IARC

Handbook content finalized

March 2007, Lyon, France



The Handbook

- Compendium of methods to most effectively evaluate policies
- Focus on
 - *countries or large sub-units within them*
 - *policies large enough, or disseminated enough, to have population impact*



Key areas covered

- **Methods for better evaluation**
- **Review of potential measures**
- **Sources of data that are collected**
- **Issues in key policy domains**



Specific Goals of Handbook

- Develop a common framework and language for evaluation
- Review strengths of possible research designs
- For each policy domain
 - *Use theory to identify core constructs*
 - *Identify measures of constructs*
 - *Assess utility of routinely collected data*
 - *Design evaluation structure*



Forms of tobacco

- **Cigarettes:**
 - *Major focus of tobacco control*
 - Most harmful form
 - Low capacity to reduce toxicity
- **Other smoked products**
 - *Inhaled forms as dangerous as cigarettes*
 - *Others irrelevant for big policy issues*
- **Smokeless tobacco**
 - *Less harmful than smoked*
 - Can reduce their toxicity



What are policies?

- Decisions of governments (jurisdictions)
- The manifestation of those decisions
 - *Rules, and their enforcement*
 - *Taxes and obligations*
 - *Programs, including systems for delivering them*
 - *Absence of rules etc.*



What policies do we consider?

- Those we know about
- Those that might occur
 - *As soon as possible before they become a reality*
- This means that the set of policies must remain open while the problem remains



Tobacco control – key outcomes

- Behaviour (eg smoking)
- Product characteristics
- Exposure
 - *To toxins*
 - *To psychoactive bits (nicotine)*
- Health consequences
 - *More immediate*
 - *Longer term: disease and death*

NB: Exposure and Health only important when product is involved



Goals of tobacco control interventions

- Reduce use of tobacco
 - *Reduce uptake of use*
 - *Increase cessation of use*
- Reduce harm from tobacco use
 - *Protect others from passive smoking*
 - *Encourage moves to less harmful products*



Policies to reduce use

- Smoking prevalence the key outcome
 - *Consumption less important*
 - *Use – non-use easier to measure than levels of use*
 - Probably more valid indicator of risk
- Self report usually sufficient
 - *Some areas need biomarkers*
- **Main outcome discussed here**



Policies to reduce harm of use

- *Re-emerging area*
- Here prevalence is NOT enough
- Consumption important
 - *Measures of toxic load*
- Measures further down causal chain
 - *Biomarkers of effects*
 - *Disease and death*
 - Often too far in future to be useful for initial policy decisions, but critical as it comes on-line



Challenge

- To agree on and document a model for policy evaluation
 - *Overcomes the problems of generalizing from the RCT model of evaluating effects*
 - *Designed for the realities of complex systems*



Understanding complex systems

- Analogous to model used to infer causes of diseases
- Encourages triangulation of evidence
 - *From different sources*
 - *From different levels of analysis*
 - Micro to macro
- Need for a data-base of evidence from which to draw inferences
 - *Backwards as well as forwards!*



Foci of evaluations

- **Jurisdictional rational for policy**
 - *Must be able to demonstrate effects to maintain credibility*
- **Theoretical effects of a policy**
 - *May be wider than justification*
 - Eg Smoke-free to protect workers also protects customers and reduces tobacco consumption
- **Possible alternative effects**
 - *If alternative model of effects is right*
 - *If industry counteraction*



Key design elements

- Cross-sectional comparisons
- Trends over time
- Plotting of mediational pathways
- Testing of alternative explanations
- Looking for incidental effects
- Representative samples or population data



Evaluation in complex systems

1. What is the goal?

- *Tobacco use*
- *Product modification, or harm reduction*

2. Model of how interventions work

- *Targeted effects*
- *Other effects*

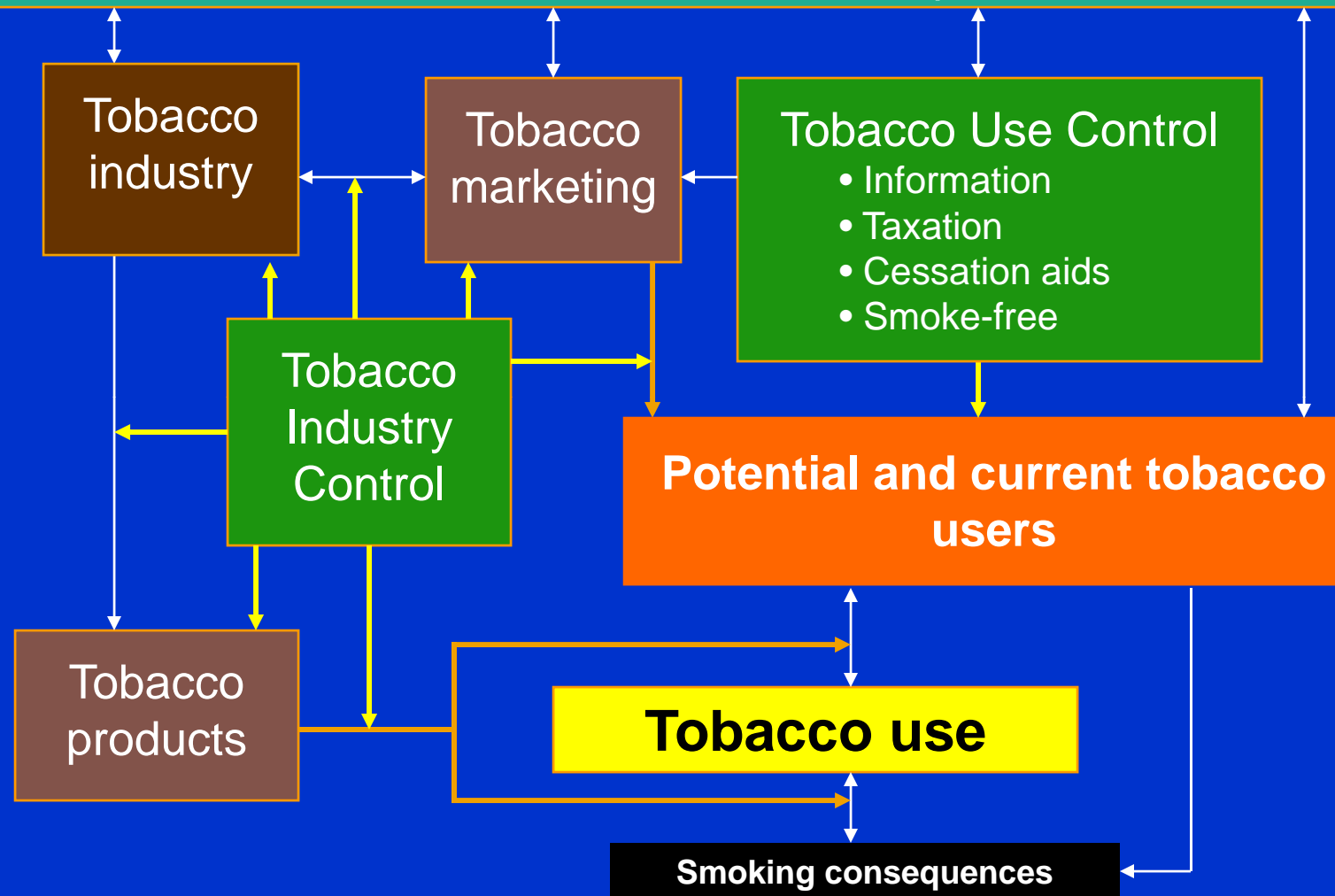
3. Consider possibilities suggested by alternative models

4. Monitor changes in the environment

- *Other possible determinants of effects*
 - *ie, tobacco industry counteractions*



Environment: Physical, institutional, communication, policy, legal, scientific, cultural, social & inter-personal



A model of tobacco control

Evaluation in complex systems (cont)

- Understanding how all the elements fit together
- The interventions we make
- The activities of the tobacco industry
- Effects of other aspects of society
 - *Constitutional constraints*
 - *Level of economic prosperity*
 - *Equity issues*



Knowing what is happening

- Good surveillance system
- Representative surveys
 - *Adult tobacco use (GATS or equivalent)*
 - *Youth uptake (GYTS)*
- Repeated regularly to determine trends
- Limited value for understanding how change occurs



Theory – mediational pathways

- Key task is to spell out what these are:
- Need to look both ways
 - *From policy to possible consequences (outcomes)*
 - May be more than one possible causal pathway to an outcome
 - *Back from observed changes in outcomes to possible causes*



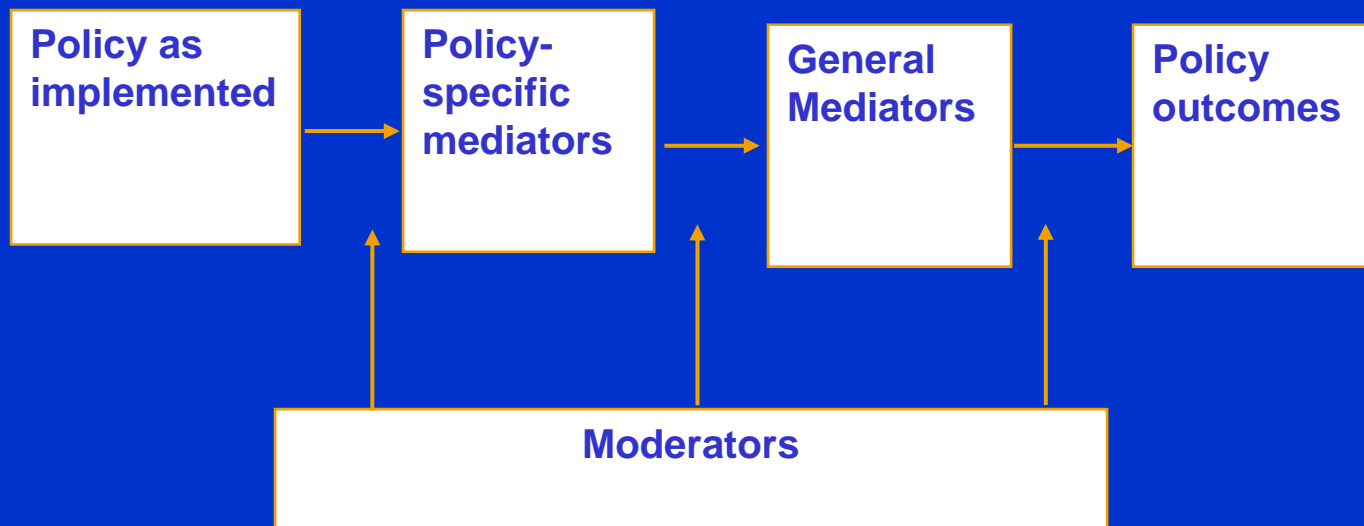
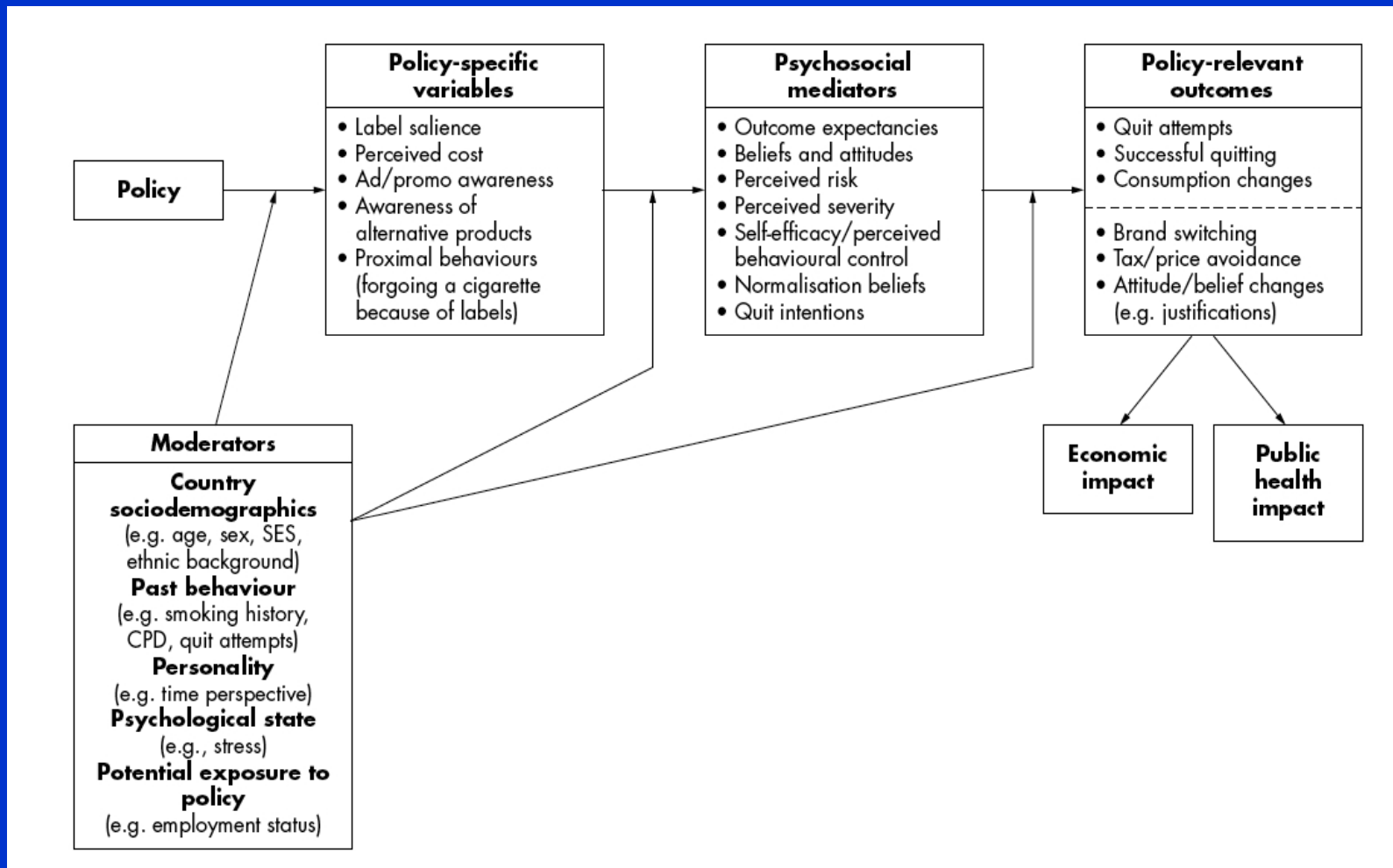


Figure 1: A generalized model of mediation making allowance for moderator effects





Tobacco use behaviours

- **Policy**

- *Knowledge of*
- *Beliefs about*
- *Environmental effects*

Proximal
Mediators

- **Beliefs**

- **Intentions**

General
Mediators

- **Behaviour – Target outcomes**



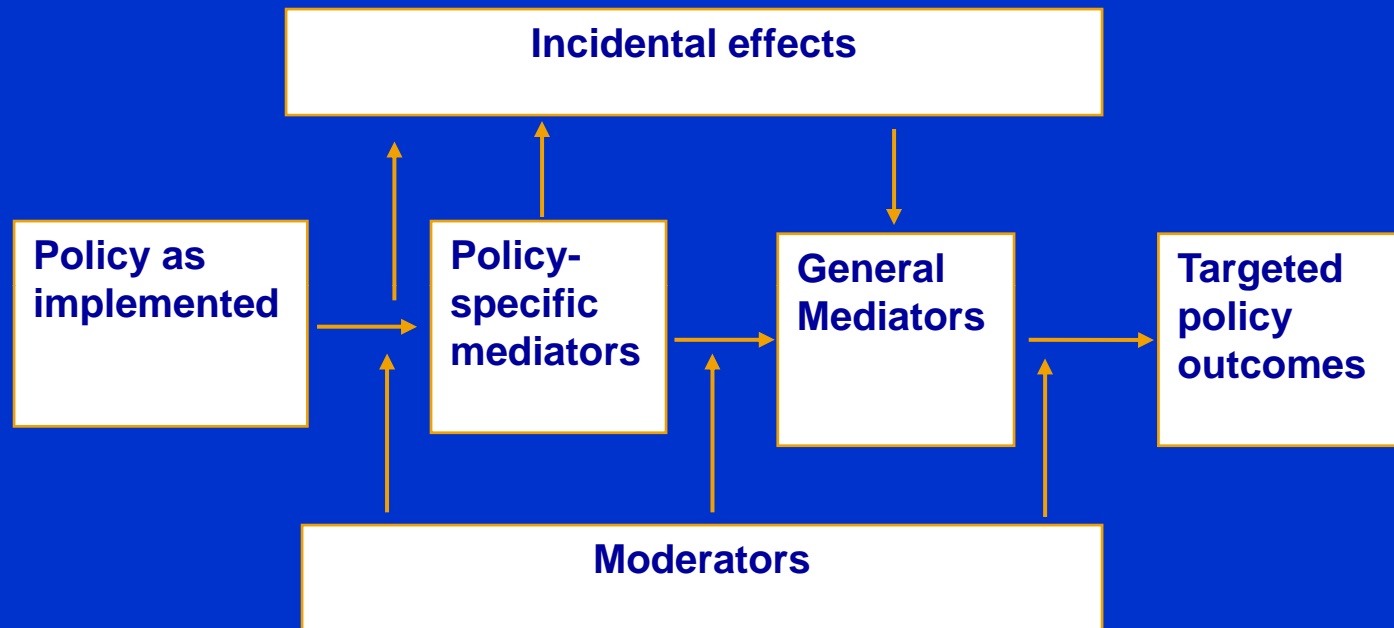


Figure 1: A generalized model of mediation making allowance for both moderator and incidental effects



Value of understanding mechanisms

- **Strengthens causal inferences**
 - *Can help differentiate effects of multiple interventions with different modes of action*
- **For diagnosis when problems**
 - *Where did things go wrong?*
 - *Who did they go wrong for and why?*
- **Understanding mechanisms can stimulate new interventions**



What do we know?

- Breaking down the compartmentalization of knowledge
- Guidelines for generalizing
 - *Including from other behaviours/issues*
- Often asking “When does it work (best)?” rather than “Does it work?”



Study validity determined by

- Internal validity
 - *Component with greatest measurement error*
 - Question ambiguity
 - Socially desirable responding
 - Sampling error
 - Impact of measuring on outcomes
- Design effects



Inferring causal links

- Testing models of effect
- Consider mediational pathways
- Consider moderators
- Include both intended and incidental effects



Validity of causal inference

- Cause precedes effect
- Cross-sectional comparisons
 - *The more the better*
 - *The more similar, the better*
- Trends over time within the population
 - *The more data points the better*
 - From pre-post to interrupted time series
 - *Usually stronger than between-group comparisons*
- Combined even stronger



Value of cohorts over repeat cross-sectional

- Can assess change at level of individual
 - *rather than group*
- Increases power to detect effects
- Harder to maintain representativeness
 - *drop out*
 - *“time in sample” effects*
- **Stronger for determining causation**
- **Weaker for estimating population effects**



An Example: The ITC study design

- Parallel studies in multiple countries
- Representative as possible
- Cohorts of target population
- Measures of all major policy initiatives
- Measures of mediational pathways
- Measures of some possible moderators
- Allows us to model the impact of policies



Twin needs for evaluation

- **Inferential power**
 - *Number of sets of observations*
 - *Number of comparison groups*
 - *Cohorts*
- **Generalizability**
 - *Representative*
 - *Enough representatives of all sub-groups*



The ideal

- **Something like ITC cohorts**
 - *But with useful biomeasures*
 - *Validation of behavior reports*
- **Linked to gold-standard surveillance survey with large samples**
 - *With data for many years*
- **Focussed studies, including RCTs where possible**
- **Integrated together in theory-driven comprehensive reviews**



The bottom line

- How much of the societal trends in tobacco use can we explain?
- Documented effects of our interventions
- Counter-interventions from big tobacco
- Broader societal trends
- Can we do better?



A new model for meta-analysis

- Not of studies using same methods
 - *No corpus of good enough RCTs for this*
 - *These studies also miss key questions*
- Use as diverse a set of studies as possible
 - *Different measures and methods*
 - *Different validity threats*
- When combined, more than the sum of its parts
 - *Convergent reduction in threats*



Benefits of the new approach

- Stronger designs can
 - *Both prospectively and retrospectively increase the utility of weaker designs*
- Combined data allows for stronger and more wide-reaching conclusions
 - *Effect sizes*
 - *Generalisability of effects*
 - Across populations
 - Across sub-groups
 - *Grounded in reality*
 - *Not in artificial trials*



An example

- Consider a post-only measure of perceived effect
 - *Warning labels make me more likely to quit*
- Do they really?
- If longitudinal study shows that this question actually predicts quitting
 - *Increased strength of inference from original study*
- If not,
 - *shows original study to be misleading*



Prospective predictors of quit attempts

Predictor	W2→W3	W3→W4	W4→W5
Cognitions	1.57	1.44	1.46
	(1.28)	(1.14)	(1.16)

All analyses control for sociodemographics (including country) and cigarettes per day; plus other Warning label variables.

Figures in brackets below are after controlling for intention to quit.

Evaluation of established interventions

- Is it working here as well as it should?
- Consider
 - *Form of the intervention*
 - *Quality of implementation*
 - *Characteristics of targeted populations*
- Compare effects to those found elsewhere
 - *If different, follow-up*



Measures and data

- **Total population**
 - Eg National consumption as estimated by tax receipts
 - *Indicates what is happening*
- **Intermediate**
 - *Sub-units by which population data is collected*
 - *Summations from individual level*
- **Individual level**
 - *Most of what we need*
 - *Critical for understanding how things happen*



Direct vs indirect measurement

- Direct measures should have low error
- Most constructs of interest are not directly measurable
 - *Require indirect measures*
 - *Issues of validity critical*
 - *Often measure aspects of related concepts*
- For some constructs of interest, there may be known or theorised biases



Validity of measures

- *Only has meaning in reference to a construct and the precision required*
- **Face validity**
 - *Critical for volitional constructs*
- **Convergent validity**
 - *Related to things it should be*
- **Criterion validity**
 - *Relationship to a gold standard*
- **Predictive validity**
 - *Predicts outcomes of interest*



Example of dependence

- **Concept is quite clear**
 - *It is the core of what makes quitting difficult*
- **Construct uncertain**
- **Variety of measures**
- **Therefore Convergent and Predictive validity focus of efforts**
 - *Genetic studies, search for gold standard*



Example of cotinine

- Measure of nicotine metabolites in body
 - *Assumptions about metabolic pathways*
- Indicator of intakes
 - *Function of:*
 - Intake
 - Body size
 - Rate of metabolism
- Often incorrectly used as measure of intake



Criteria for selecting constructs to measure

- Outcome of interest
- Key mediator of outcome
- Demonstrates link to policy
- Critical for refuting alternative explanations
- Helps assess differential effects in sub-populations of interest
- **Quality of evidence it will do the job**



Limits of self-report

- Social desirability
- Response style
 - *Eg strength of agreement*
- Memory biases
- Extent issue attended to
- Context of asking
- Problems of comprehension
- Variation in interpretation of questions
 - *Different assumptions*
 - *Different frame of reference*
 - Eg smoking vs quit
 - *Different nuances of meaning*
 - Eg the word addiction



Self-report measures

- Experiences and beliefs
 - *Clarity of concept critical*
 - *No external means of validation*
- Behavior
 - *Reports in principle verifiable*
 - *Memory and attention affect responses*
- Environmental conditions
 - *Externally verifiable, in principle*
 - *Sensitivity bias*
 - Affects use at individual level



When should we allow use of different wording for the same question?

- An assumption needs to be stated
 - *For some populations we might need to explain what something is before asking, while for others where it is widespread, this would be insulting*
- When frame of reference changes
 - *Eg quit, smoking*
- When a language culture does not have a directly comparable concept



Question wording

- Sometimes measurement is robust to varied ways of asking
 - *Often for self-reported behaviour*
- Other times question wording is critical to measuring a construct
 - *Particularly for attitudes*



Question wording

- **Some constructs are very sensitive to question form**
 - *Esp attitudes*
- **The precision of a question may need to vary by context**
 - *EG, Need to explain what NRT is*
 - *Where well known explaining it may be seen as condescending*
 - *However, where not well known, failure to explain may create similar problems*



Complementarity of measures

- Where two or more measures are possible, and each has different limitations, it can be worth pointing out the value of assessing both to see if conclusions are equivalent
 - *Eg, for consumption, relating population sales (or tax) data to estimates from self-report survey data*



When different studies give different answers

- *Inevitable when using samples and/or*
- *When there is error of measurement*
- For description, calibration of measures
- For outcomes, reviews including meta-analyses



Factors determining confidence in inferences of causality

- **Size of effects** (relative to natural variation)
- **Plausible alternative explanations**
- **Demonstration of mediational pathways**
- **Replication in diversity of contexts**
- **Evidence from RCTs**



Temporal issues

- Processes work on different time scales
 - *Affects optimum frequency of surveying*
 - *Affects time period to ask about*
- No optimal time frame for all aspects of tobacco use

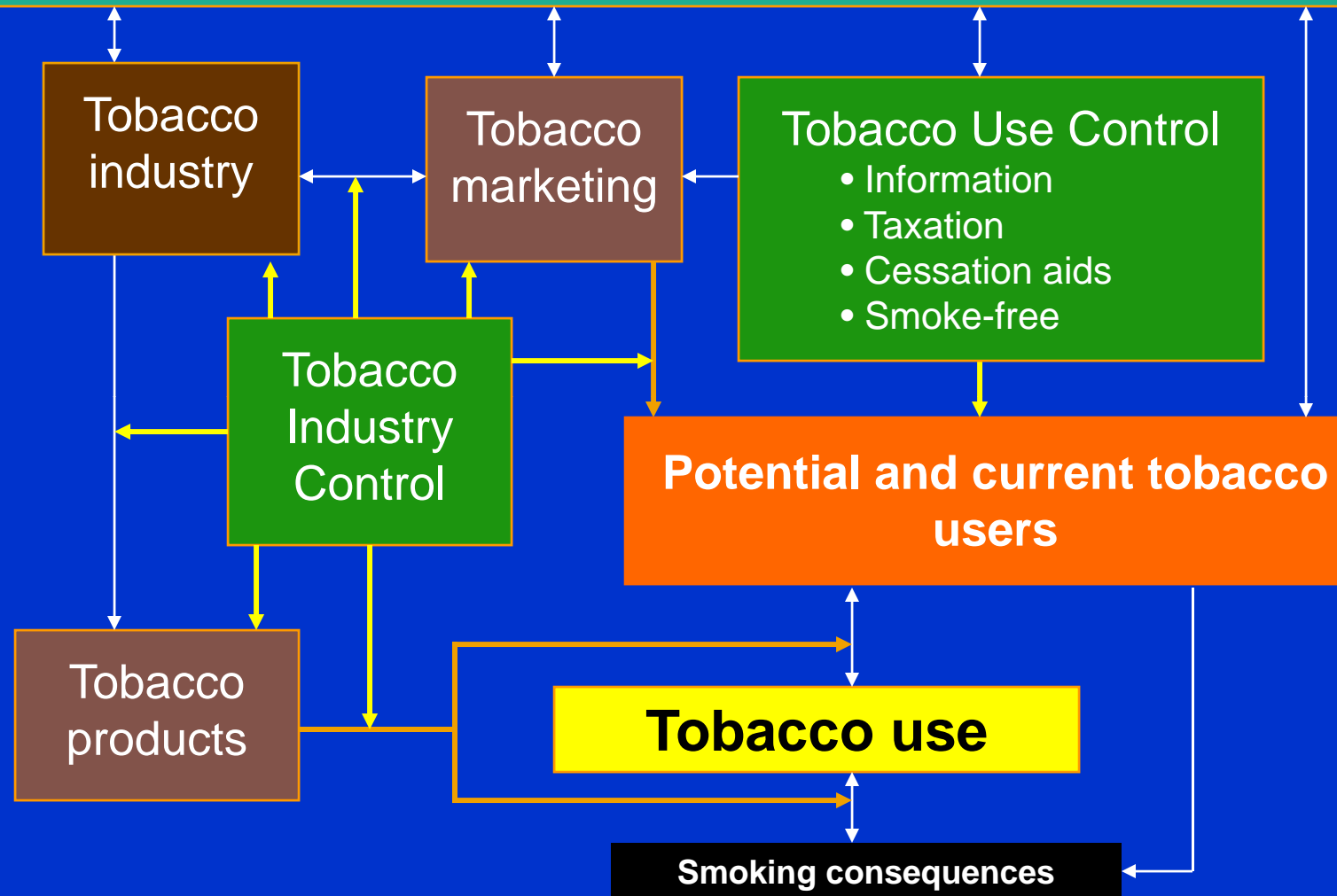


Equity issues

- Both between and within countries
 - *Critically important*
- What is needed to optimize the intervention for disadvantaged groups?
 - *Availability and access*
 - *Tailoring of communication*



Environment: Physical, institutional, communication, policy, legal, scientific, cultural, social & inter-personal



A model of tobacco control

Single post only design

- Uses respondent attributions of effect
- Weakest design
- Better when indicators validated with stronger designs



Pre-post comparison

- Stronger than independent comparison when
 - *Changes over time less than changes between groups*
 - Likely to be the usual situation
- No control for secular trends
- Assumes no other sources of effect



Adding more observation points

- Increased specificity of intervention with respect to trends
- Allows duration and shape of effect to be better estimated



Independent Comparison Group

- **Utility increased with**
 - *Number of comparators*
 - *Comparability pre-intervention*
 - *Lack of changes in other potential influences*



Adding extra units

- *Either to intervention or control*
- *Or varying levels of interaction*
- **Increases inferential power**
 - *Controls (somewhat) for unique differences*



Policy implementation

- Often cannot,
- Often should not,
 - *wait for the evidence*



Evidence-based policy making is not just about acting where there is evidence

- It includes acting where evidence is lacking
 - *Cannot or should not wait*
- But acting in ways that maximise the knowledge gained from implementation



Levels of evidence

1. The policy can work, and does work in comparable jurisdictions
2. The policy can work, and does work in very different jurisdictions
3. Analogue evidence of policy effectiveness
4. Lack of evidence, but a good theoretical rationale, and no credible alternative model suggesting adverse effects
5. As for 4, but with a credible alternative model predicting adverse effects



How do we determine policy effects?

- Identify key features of implemented policy
- Seek a common conceptual frame to organise understanding
- Consider how design features might reduce inferential validity threats



What are the core constructs?

- Which are policy-specific and which are general?
- How good are our measures?
- How well do the constructs translate across cultures?



Are there policy moderators?

- *Stable factors which affect outcomes*
- What are they?
- How well can we measure them?
- Some moderators can be changed by the interventions
 - *These changes may mediate other effects*



What outcomes should we seek?

- Disease and death
 - *Rarely as effects often long after intervention*
 - *Need consistency with what we know*
- Tobacco use
 - *Most common outcomes to use*
 - *Use – non-use easier to measure than levels of use*
 - Probably more valid indicator of risk



Perspective

- Study of the ways in which tobacco use is organised in a society, from the perspective of government, the ultimate arbiter of what is permitted and what is encouraged
- How well government actions are working?
- What problems remain?



Major themes

- Study designs
- The roles of theory
- Measures
 - *Questions*
 - *Observations*
 - *Recordings*
- Sources of information
 - *Documents*
 - *Existing collections*
 - *Purpose-built data collections*



Evaluation of quality and/or utility

- **Ranking systems**
 - *Methods*
 - *Measures*
 - *Outcomes*
- **Levels of evidence used for clinical endeavors not appropriate**
- **We need our own**



Questions to address (examples)

- How do we determine the effects of a policy?
- How does a policy have its effects?
 - *How to assess implementation?*
 - *What are barriers to optimal use?*
 - *What affects compliance?*
- What role can this data source have in policy evaluation?
- What are key measures?



Study designs

- RCTs rarely possible
 - *Often lack ecological validity*
- Value of multiple methods
 - *Understanding complex systems*
- Key elements that are useful
 - *Cross-sectional comparisons*
 - *Trends over time*
 - *Plotting of mediational pathways*
 - *Testing of alternative explanations*
 - *Representative samples or population data*



Most of the outcomes we seek are multiply determined

- Consider relative contribution of each intervention
- Interactive effects
- Moderator effects



Some terminology

- **Concept vs Construct**
 - *Concept - Theoretical notion*
 - *Construct – Operationally defined concept; ie, the thing measured*
 - E.g., Smoker could be defined as somebody who smokes at least weekly
 - *Could be daily*
 - *Could be monthly*



Types of data

- **Total population**
 - *Eg National consumption as estimated by tax receipts*
- **Intermediate**
 - *Sub-units by which population data is collected*
 - *Summations from individual level*
- **Individual level**



Centrality of measure*

- Key, should always be measured
- Good composite – recommended when relevant component measures cannot be used
- Component measures – recommended measures to elaborate pathways of effect
- Optional

* when construct is focal



Measures

- Provide 3 levels of assessment
 - *Concensus view without evidence*
 - *Some evidence of utility*
 - *Adequate evidence of utility*



Recommendations with insufficient evidence

- These will unfortunately be common
- Useful to identify the kind of research that would validate them

NB If group feels some variables should be measured purely on face validity, please say so



Limitations of data – your tasks

- **Measure or indicator specific**
 - *Discuss when considering each measure*
 - *Include variations from general*
- **General to this type of data**
 - *Collate in general chapters and Introductory chapters*
 - Note new ones and refer to Editors or group covering relevant Chapter



Logic of choice of concepts to measure

- If we find a change in outcome of interest, does this help us link that outcome to the policy?
- If we find no change: does this help us understand whether this is because
 - *The policy does not affect this outcome*
 - *The evaluation is underpowered to differentiate effects of the size we might expect*



Levels of evidence

1. No evidence
2. Possible effect
3. Probable effect
4. Established effect

Nature of effect

- *Negative*
- *Null*
- *Positive*

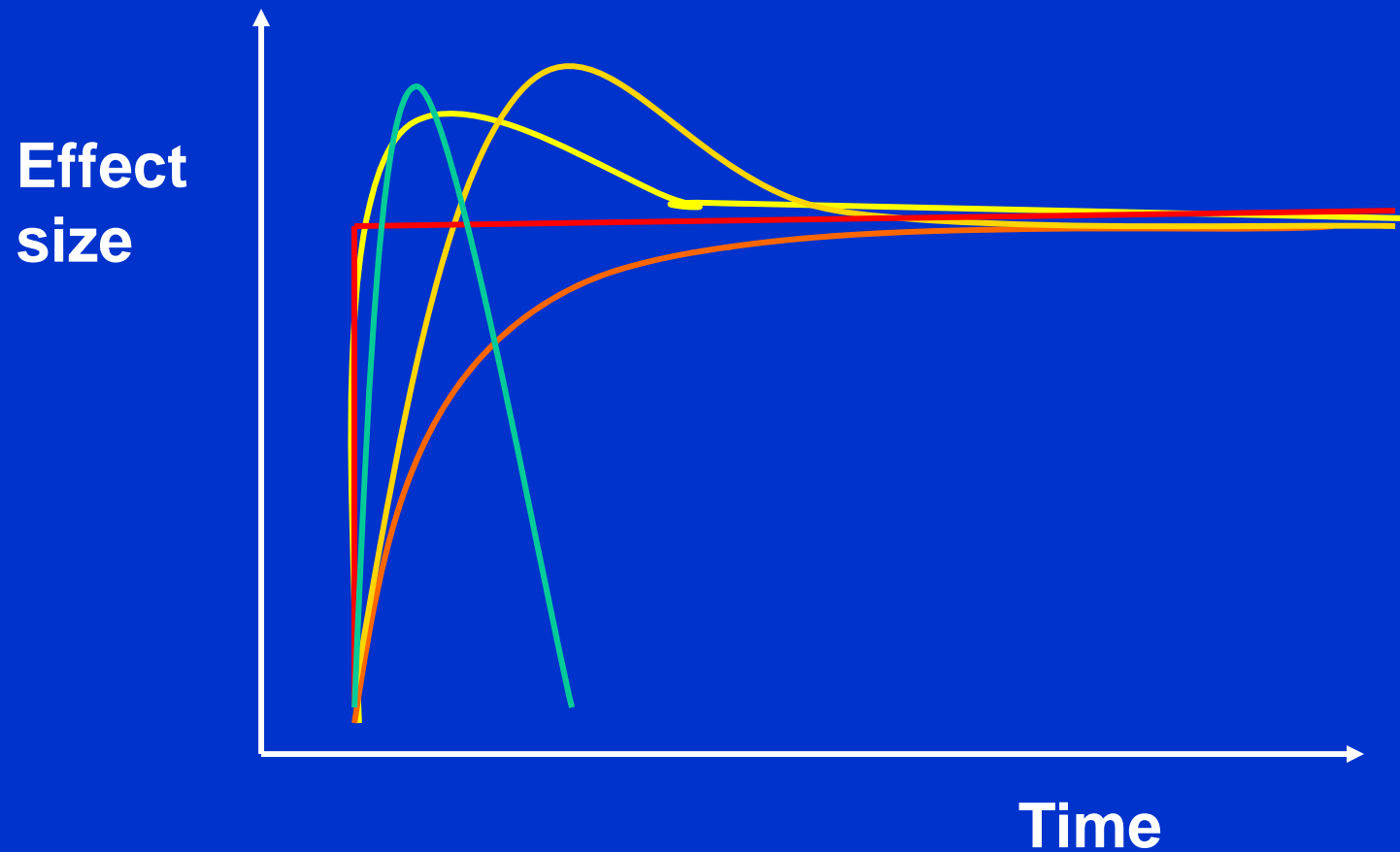


Incidental effect

- Need to be included
 - *Direction and magnitude of effects critical*
- Can occur to
 - *Targetted individuals*
 - *Other individuals*
 - *Organisations (eg businesses)*



Nature of effects



Generality of effects

- **Generality**
 - *Demonstrated*
 - *Probable*
 - *Assumed*
 - *Status uncertain*
- **Identified limits or variation in effect size**
 - *Possible*
 - *Probable*
 - *Established*



Dealing with gaps

- **Includes**
 - *Areas we have forgotten*
 - *Areas where research is needed*
- **Specific**
 - *Add them into the relevant chapter*
- **General**
 - *Point them out to the editors*
 - *Expand the general chapters*
 - *Incorporate into summary chapter*

