Evaluating Complex Whole System Programmes: Overview and Approaches

NHS Health Scotland

Neil Craig

Principal Public Health Advisor

Evaluation Team

NHS Health Scotland

Outline

1. Definitions

   - complexity

   - evaluation

2. Methods

3. An example: Monitoring and Evaluating Scotland's Alcohol Strategy (MESAS)

What makes an intervention complex?

- Number of interacting components
- Number and difficulty of behaviours required by those delivering or receiving the intervention
- Number of groups or organisational levels targeted by the intervention
- Number and variability of outcomes
- Degree of flexibility or tailoring of the intervention permitted

Source: Craig P et al, *BMJ 2008; 337 doi:*
http://dx.doi.org/10.1136/bmj.a1655

But is complex more than just complicated?

Complex systems:

- Have feedback loops

- Adapt themselves to the environment

- Go through non-linear changes e.g. tipping points

- Are sensitive to initial conditions

$\Rightarrow$ Outcomes emerge from interactions between multiple components

$\Rightarrow$ Outcomes include unexpected and unintended effects

Evaluation

- Assessing whether programmes went as planned in terms of processes or outcomes

But what for?

- Accountability?

- Generalisability?

- Programme improvement?

- Equality monitoring?

=> Causation
=> Is it working, how and for whom?

Challenges

- Selection biases
- Confounders
- Complicated and variable programmes

=> Experiment?

Limits to experimentation

- Practicality
- Ethics
- Political feasibility
- Standardisation undermines the essence of the programme

Experimental evaluation designs

- Cluster randomisation

- Stepped wedge designs

- With or without standardisation of the programme (and careful documentation either way)

Story so far…

- There is more scope to experiment that sometimes assumed

- Evaluators need to think about how to address the issues experimental methods are designed to address

- Evaluations should draw on different methods to cope with complexity

Non-experimental designs

- Natural experiments
- Longitudinal studies
- Cross-sectional surveys
- Secondary analysis of routine data
- Case studies
  - quantitative data
  - qualitative data

Causal inference

Three components of causal inference:

1. **Congruence with the program theory** – are the results consistent with the program theory?

2. **Counterfactual comparisons** – what would have happened without the intervention?

3. **Critical review** – what are plausible alternative explanations for the results?

Patricia Rogers & Jane Davidson www.betterevaluation.org

Contribution Analysis

IF

1. There is a plausible and evidence based **theory of change**
2. **Planned activities** were implemented
3. **Expected results** occurred
4. **Other factors** influencing the outcomes have been assessed and accounted for

THEN

- It is reasonable to conclude that the programme is making a difference i.e. it is contributing to (influencing) the desired outcomes

Source: Mayne, J. Contribution Analysis: Coming of age? *Evaluation*, 2012, 18(3): 270-280.

Programme theory

- A *simplified* model of how an intervention is understood to contribute to a chain of intended outcomes

- To guide planning, implementation, monitoring and evaluation

$\Rightarrow$ To make evaluation tractable, some simplification is necessary

$\Rightarrow$ Does this mean we are not adequately evaluating *complex* systems?

1. Congruence

Are the results consistent with programme theory?

Use various qualitative and quantitative methods to:

- Compare implementation with the achievement of intermediate and final outcomes
- Compare predictions with actual outcomes
- Disaggregate outcomes for different groups
- Compare timing of outcomes with programme theory
- Compare dose-response patterns with programme theory

Using program theory to interpret findings

Programme delivered	Students involved in programme activities	Student behaviour changes	Health outcomes improved	Interpretation
✖	✖	✖	✖	Implementation failure
✔	✖	✖	✖	Engagement failure
✔	✔	✖	✖	Theory failure (early causal link)
✔	✔	✔	✖	Theory failure (later causal link)
✔	✔	✔	✔	Consistent with theory
✔	✔	✖	✔	Theory failure (diff causal path)

Adapted from Sue Funnell & Patricia Rogers (2011) Purposeful Program Theory.

2. Counterfactual comparisons
What would have happened without the intervention?

- Use a control group or comparison group
- Make comparisons across different cases/areas
- Compare the trajectory before and after the intervention

3. Critical review

Are there other plausible explanations of the results?

- Identify alternative explanations and see if they can be ruled out
- Identify and explain exceptions
- Compare expert predictions with actual results
- Ask participants and other key informants

Case study: MESAS

Monitoring & Evaluation of Scotland's Alcohol Strategy

Alcohol Strategy Theory of Change
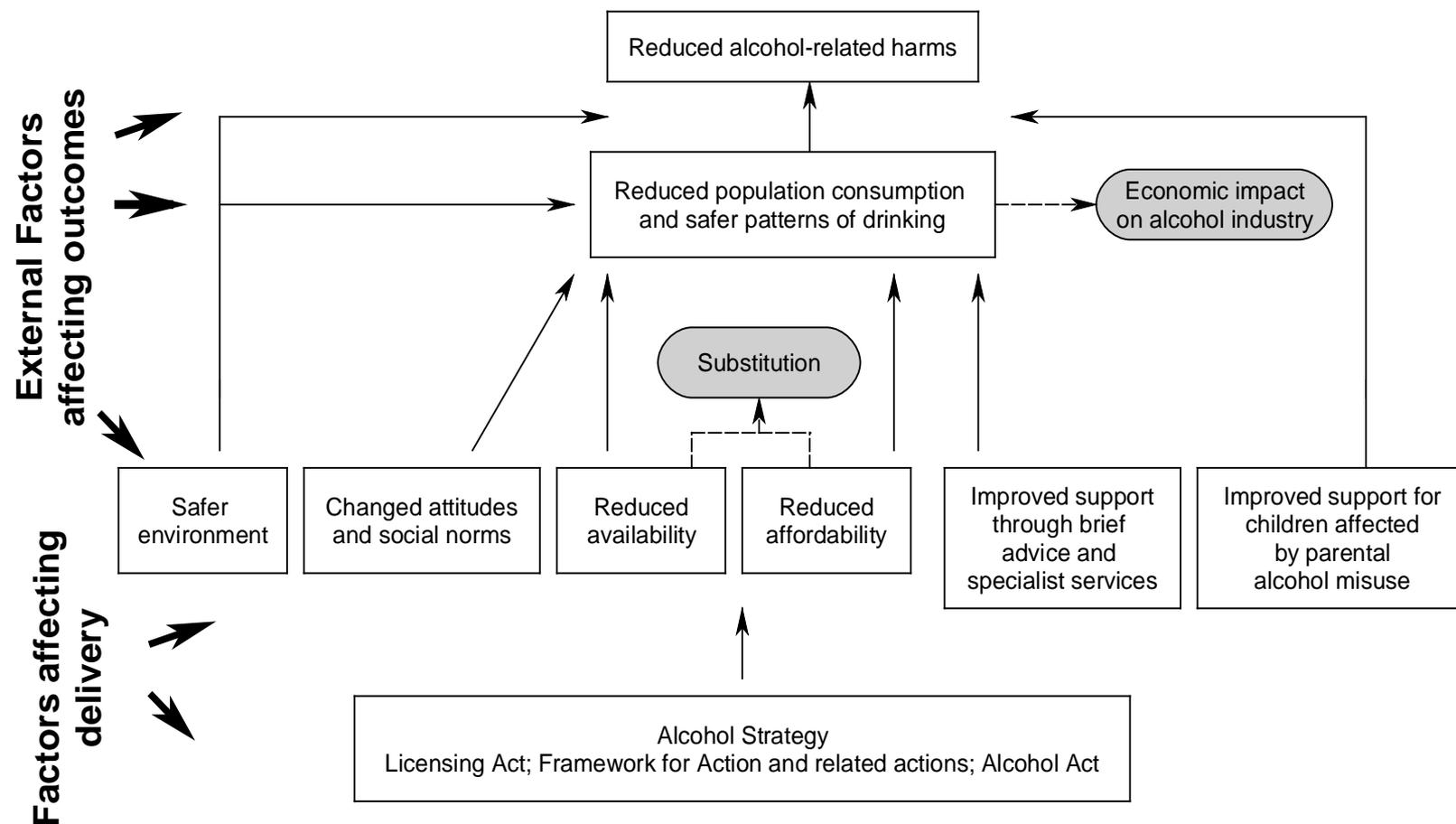
Causal inference 1: Congruence with programme theory

The MESAS evaluation is looking at the extent to which intended outcomes have been achieved and whether the routes to those outcomes were as envisaged in the programme theory

Congruence with programme theory

- Evaluation of implementation:
  - Alcohol brief interventions
  - Licensing Act
  - Treatment and care study

- Evaluation of outcomes
  - monitoring of attitudes
  - monitoring of harms
  - monitoring of consumption
  - monitoring of price

- Unintended outcomes

Causal inference 2: Counterfactual

- What would have happened in the absence of the intervention?

- Not much scope to set up control groups for a national strategy
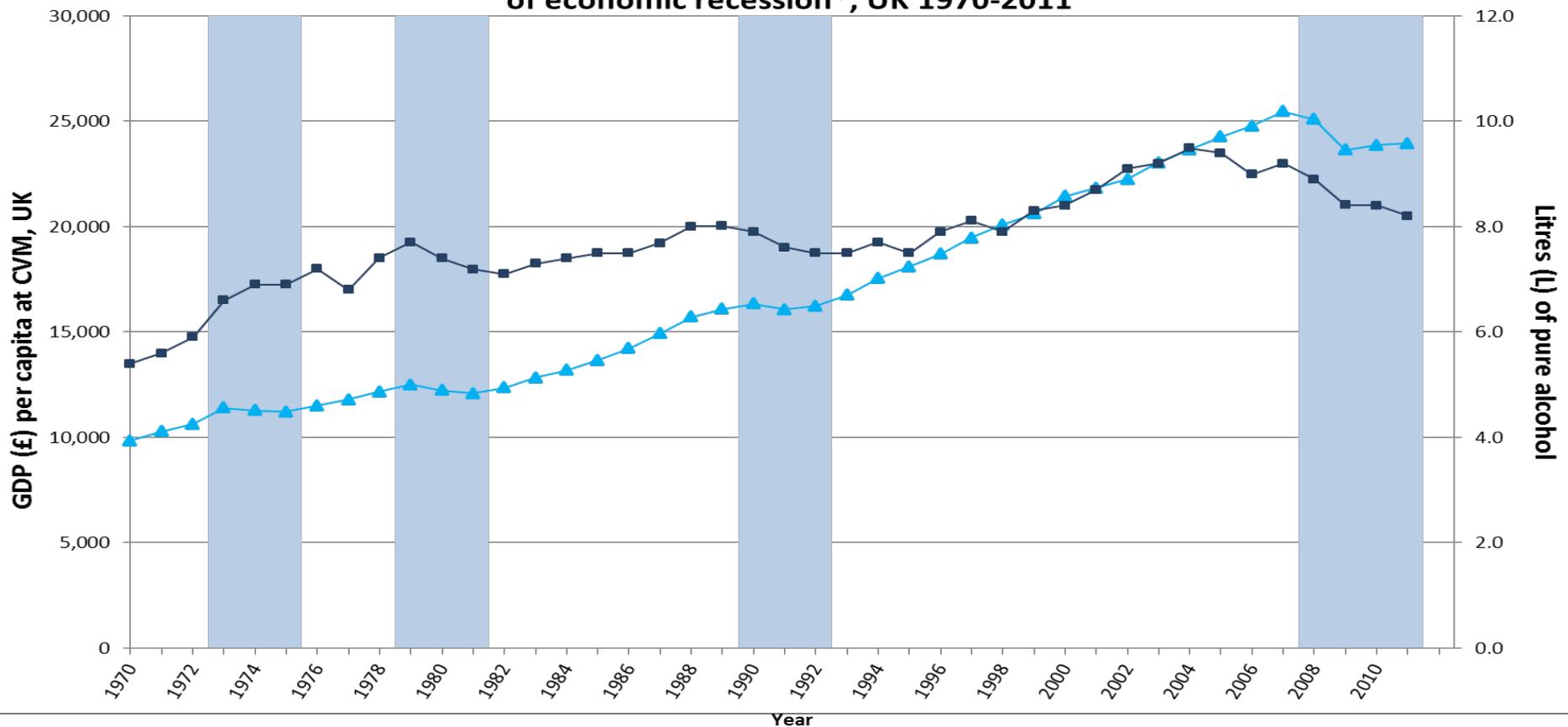
- But 'natural experiments' between countries?

- Evaluation of impact of ban on multi-buy discounts in Scotland
  - policy reduces the incentive to buy more than intended
  - bigger falls in Scotland than over the same period in the rest of the UK

Causal inference 3: Critical review

- Other plausible explanations?

- In Scotland we have seen a period of falling alcohol sales and falling alcohol harms since the alcohol strategy *A Framework for Action* was implemented.
- But this has also been a period of recession and falling incomes, which may in turn have caused the fall in sales and harms.
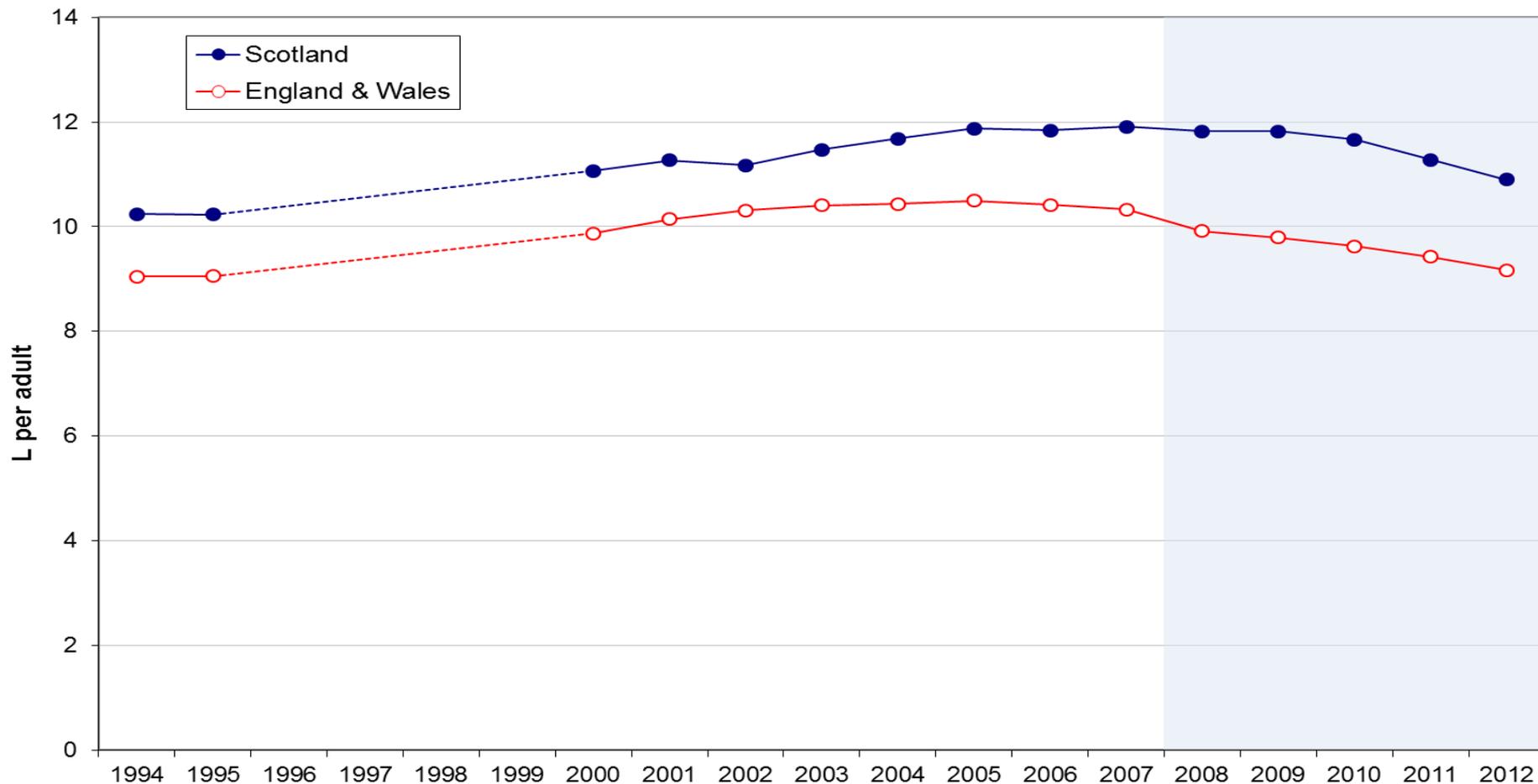
Sales fall during recessions - usually

Per capita sales of pure alcohol, per capita Gross Domestic Product (GDP) and periods of economic recession*, UK 1970-2011
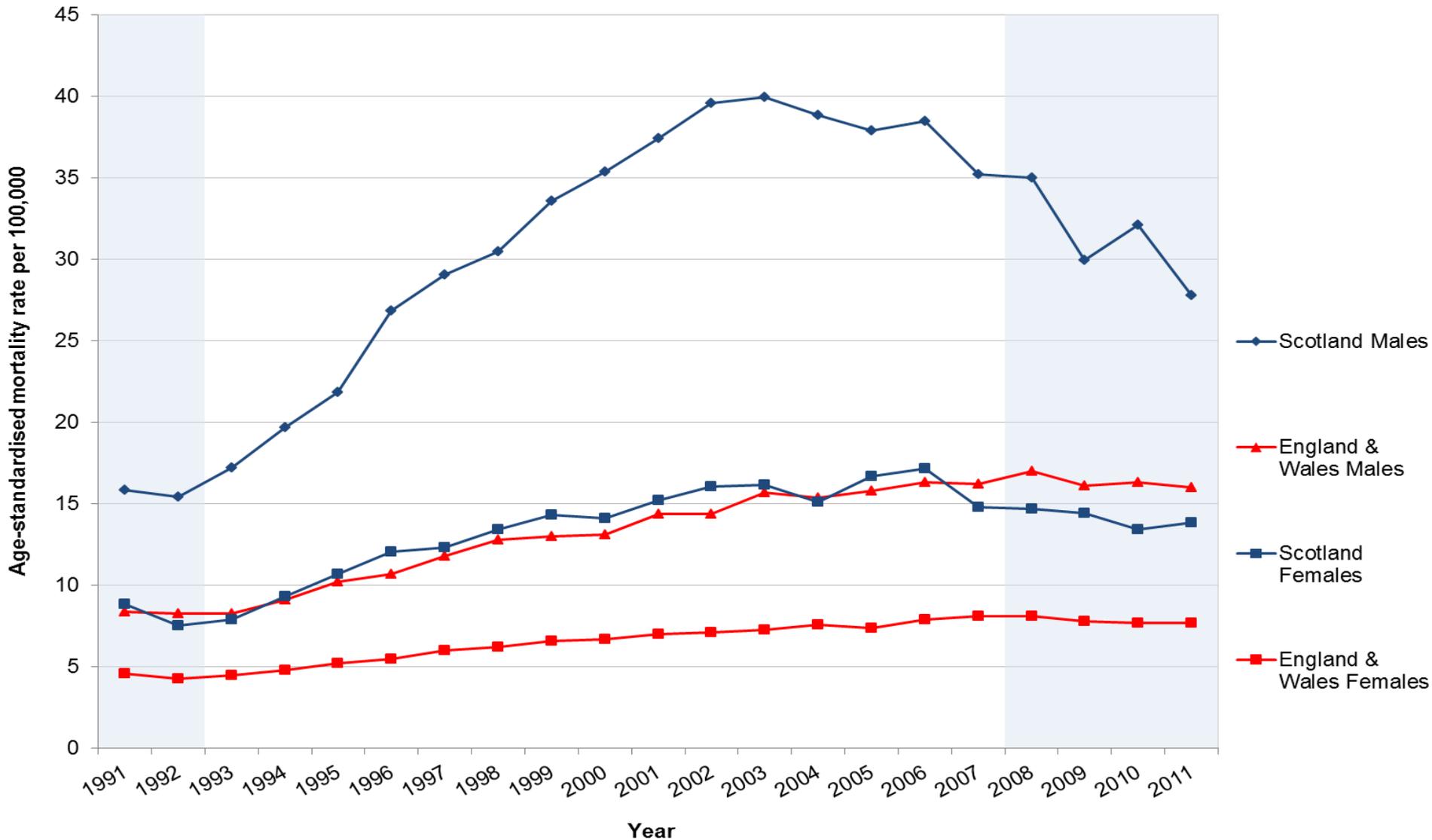
Litres of pure alcohol sold per adult (aged ≥16 years) in Scotland and England & Wales, 1994-2012

Alcohol-related deaths, Scotland, England & Wales, EASR, by gender, 1991-2011

Conclusions 1

- Developing and evaluating a complex intervention is multi-stage and iterative

- Experimental designs reduce selection bias and confounding but are not always practicable

- Understanding processes is important but should not replace evaluation of outcomes

- Complex interventions may work best if tailored to local contexts - evaluation needs to adapt accordingly

- Reports of studies should include a detailed description of the intervention

Source: adapted from Craig P et al, *2008, 337 doi:*
http://dx.doi.org/10.1136/bmj.a165

Conclusions 2

- Evaluation of complex programmes is messy

- Particularly messy where the scope to standardise the intervention and set up a counterfactual is limited

- But theory-based approaches using range of types of evidence…

- … together with critical interpretation can lead to more plausible claims about the *contribution* of complex programmes to outcomes

Useful references

Craig P et al, *Developing and evaluating complex interventions: the new Medical Research Council guidance BMJ, 2008, 337 doi:* http://dx.doi.org/10.1136/bmj.a165

Funnell S & Rogers P. *Purposeful Program Theory, 2011.*

Mayne J. Contribution Analysis: Coming of age? *Evaluation*, 2012, 18(3): 270-280

Beeston C, Reid G, Robinson M, Craig N, McCartney G, Graham L and Grant I (on behalf of the MESAS project team). *Monitoring and Evaluating Scotland's Alcohol Strategy*. Third Annual Report. Edinburgh: NHS Health Scotland, 2013

Pawson R & Tilley N. Realistic Evaluation, 1997.