College of Policing

college.police.uk

BetterEvidence forBetterPolicing™

Evaluation masterclass

Early Intervention Foundation Leadership Academy
Ryton, 2 November 2016

College of Policing

college.police.uk

BetterEvidence forBetterPolicing™

Welcome and introduction

Paul Quinton and Steph Waddell

Why evaluate?

- Learning from experience
- Contributing to the wider evidence base
- Developing a business case
- Providing accountability
- Three examples…
  - Mentoring
  - Moving to Opportunity
  - Troubled Families

Carey Oppenheim

"First, evaluation remains vital; it is an essential component of effective and transparent policymaking and public investment. We need to know what works, for whom and in what context, and if it is not working. The disappointing outcome of this first evaluation of the Troubled Families programme is not a strike against the value of evaluation in public policy…

Second, data and evidence (positive and negative) need to be shared openly and efficiently, and used to inform policy.

Third, and relatedly, the purpose of evaluation is to help us **improve**, not to **prove**."

Morning agenda

When	What
9:15 – 9:45	Impact evaluation
9:45 – 10:15	Logic models
10:15 – 10:30	Discussion and self-reflection
10:30 – 10:45	Refreshment break
10:45 – 11:15	Process evaluation
11:15 – 11:45	Economic evaluation
11:45 – 12:00	Discussion and self-reflection
12:00 – 12:45	Evaluation case study
12:45 – 13:30	Lunch

Afternoon agenda

Time	Session
13:30 – 14:00	Commissioning research
14:00 – 14:30	Quality assuring research
14:30 – 14:45	Discussion and self-reflection
14:45 – 15:00	Refreshment break
15:00 – 15:45	Open research surgery
15:45 – 16:00	Concluding remarks and close

Self-reflection Thinking how to apply what you have heard when evaluating your action plan

Surgery Asking for advice and guidance and opening up your ideas to peer review

Who’s who from the College

- **Paul Quinton**
  Evidence and Evaluation Advisor
  Uniformed Policing Faculty

- **Levin Wheller**
  Research and Analysis Standards Manager
  Knowledge, Research and Practice Unit

- **Sarah Colover**
  Senior Research Officer
  Knowledge, Research and Practice Unit

- **Will Finn**
  Senior Research Officer
  Knowledge, Research and Practice Unit

College of Policing

college.police.uk

BetterEvidence forBetterPolicing™

Impact evaluation

Levin Wheller

college.police.uk

Logic models

Levin Wheller

College of Policing

college.police.uk

BetterEvidence forBetterPolicing™

Discussion and self-reflection

Everyone

College of Policing

college.police.uk

BetterEvidence forBetterPolicing™

Process evaluation

Sarah Colover

This section will cover….

- Value and purpose of process evaluations
- Evaluation, implementation and theory failure
- Programme fidelity / implementation quality
- Qualitative data

Process evaluation – what it is and why it matters…

I keep six honest serving men
(they taught me all I knew);
Their names are **What** and **Why** and **When**
and **How** and **Where** and **Who**.

Rudyard Kipling

Process evaluation

Understanding success and failure…

All about monitoring the implementation of you intervention…

What is the organisational and wider political/ social context in delivering your new intervention?

What barriers were there to implementation?

What facilitated intervention?

What lessons were learned?

What went well/ what would you do differently?

What were the possible impacts of these issues on the effect/ outcome of your intervention?

Use new data or routinely collected information

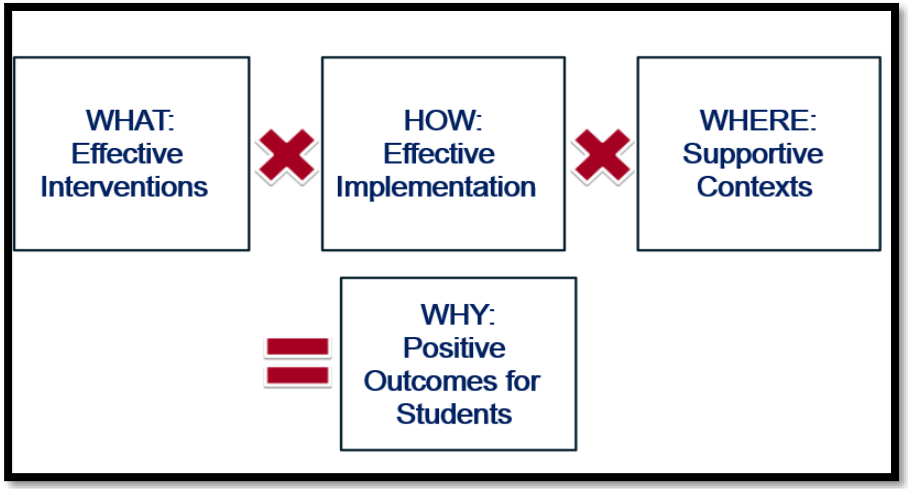
Implementation

- How the intervention was run can be a source of failure in itself
- Programme fidelity: did you implement what you intended to?

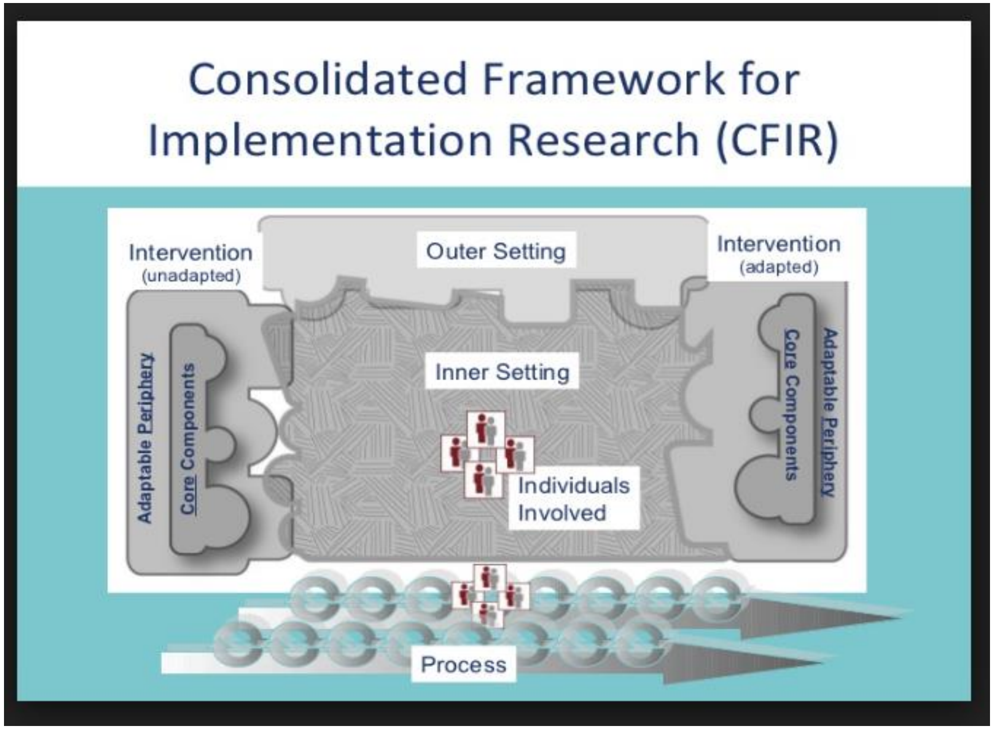
- Evaluation failure: e.g. having a small sample size
- Theory failure: having a bad idea to start with

- Important to be honest and upfront with what the problems were. This can be difficult, BUT evaluation is about finding out what works/or doesn’t and why in order to improve and change.
- Context : think about replication and why it wouldn’t work due to differences at:
    - National level
    - Local
    - Within force differences
    - Individuals/ power relationships

Examples of implementation models

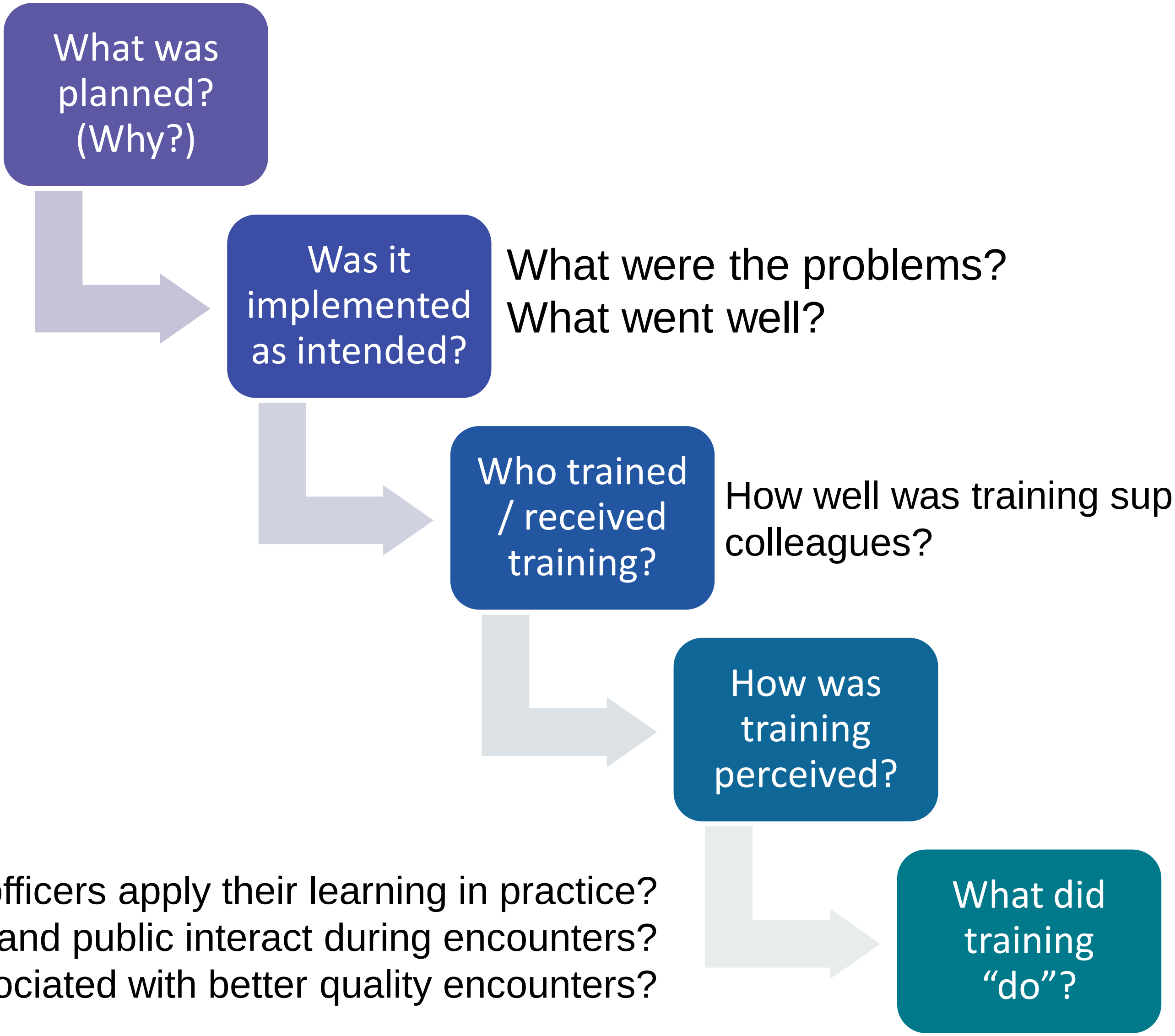


WHAT: Effective Interventions × HOW: Effective Implementation × WHERE: Supportive Contexts = WHY: Positive Outcomes for Students



Consolidated Framework for Implementation Research (CFIR)

Case study 1: Stop and Search Training Experiment



What was planned? (Why?)

Was it implemented as intended?

What were the problems?
What went well?

Who trained / received training?

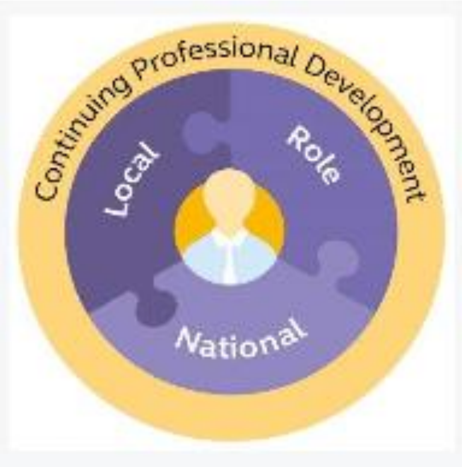
How well was training supported by colleagues?

How was training perceived?

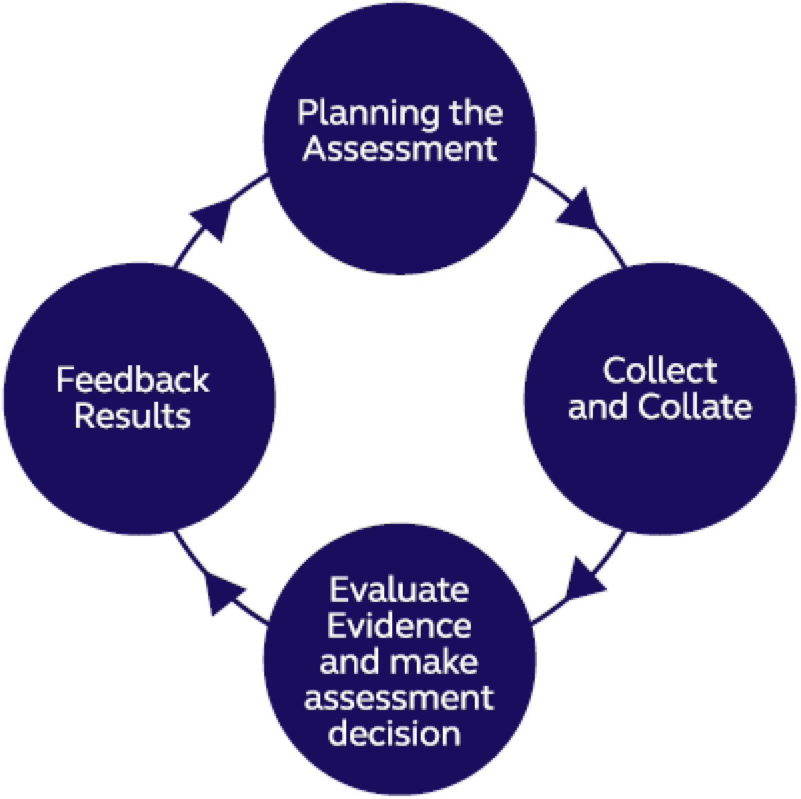
What did training “do”?

Did officers apply their learning in practice?
How did the police and public interact during encounters?
What factors were associated with better quality encounters?

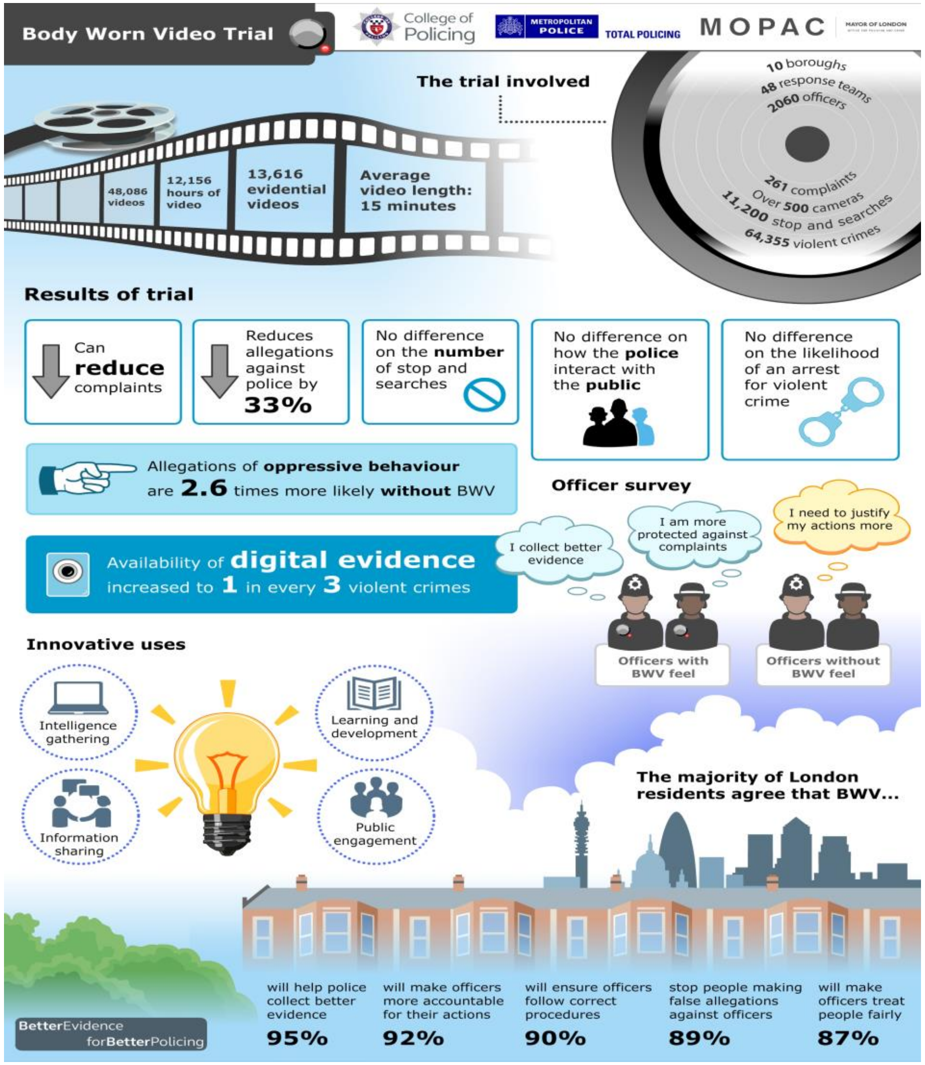
Case study 2 – Defining and Assessing Competence (DAC) Pilot Evaluation



Assessment cycle



Case study 3 – Body Worn Video (RCT)



Body Worn Video Trial

College of Policing · Metropolitan Police · TOTAL POLICING · MOPAC · MAYOR OF LONDON

The trial involved

10 boroughs
48 response teams
2060 officers
261 complaints
Over 500 cameras
11,200 stop and searches
64,355 violent crimes

48,086 videos
12,156 hours of video
13,616 evidential videos
Average video length: 15 minutes

Results of trial

- Can **reduce** complaints
- Reduces allegations against police by **33%**
- No difference on the **number** of stop and searches
- No difference on how the **police** interact with the **public**
- No difference on the likelihood of an arrest for violent crime

Allegations of **oppressive behaviour** are **2.6** times more likely **without** BWV

Availability of **digital evidence** increased to **1** in every **3** violent crimes

Officer survey

Officers with BWV feel:
- I collect better evidence
- I am more protected against complaints

Officers without BWV feel:
- I need to justify my actions more

Innovative uses

- Intelligence gathering
- Learning and development
- Information sharing
- Public engagement

The majority of London residents agree that BWV...

will help police collect better evidence	will make officers more accountable for their actions	will ensure officers follow correct procedures	stop people making false allegations against officers	will make officers treat people fairly
95%	**92%**	**90%**	**89%**	**87%**

BetterEvidence forBetterPolicing

Systematic/Qualitative methods

Method
Focus groups
Interviews
Observations
Reflective diaries
Action research
Case study
Longitudinal
Document analysis

Q. Can you think of anymore methods?

Q. What are the pros and cons of each one?

BUT process evaluation is not just qualitative, using quantitative monitoring data can sometimes be all you need (e.g. numbers of people attending a parenting session and attrition in attendance).

Consider the logic/consider the data: Triangulation

- Issues with complicating factors and reporting artefacts might be solved by using **triangulation**
- Triangulation is essentially cross-checking – utilising more than one source/type of information to understand an issue
- Using multiple sources/types of data can reduce the uncertainty associated with a conclusion based on just one source (Webb et al., 1966)
- If only one source of data is used, conclusions may be misleading, and actions based on those conclusions may waste money and time

Triangulation: competing evidence

- When triangulating, alternative sources of evidence may not always agree
- For example, qualitative information about “what is really going on” from people on the ground may differ from the impression provided by quantitative data
- If this happens, it is important to carry on investigating in order to arrive at a picture that is most supported by the available information

Thank you for listening
Any questions?

sarah.colover@college.pnn.police.uk

college.police.uk

Economic evaluation

Paul Quinton

Thinking economically…

- You are awarded a grant from the Home Office Innovation Fund to set up a street triage team with the aim of 'providing a better initial response to mental health incidents and improving clinical outcomes'
- You use the Home Office grant to:
  - purchase a new street triage vehicle
  - second an NHS mental health practitioner to work alongside a police officer for 6 months
- During the funding period:
  - the street triage team responds to 400 mental health incidents
  - the officer uses her s136 powers a total of 75 times

The Home Office urgently require you submit a return-on-investment report – what issues do you need to consider?

Thinking economically…

Investment?

- Purchase price of the vehicle
- Fuel and vehicle maintenance
- Salary and on-costs of the NHS practitioner
- Time spent by the police officer on street triage duties
- Opportunity costs of the NHS practitioner and police officer
- Broader context (eg, budget cuts and reduced services)

Return?

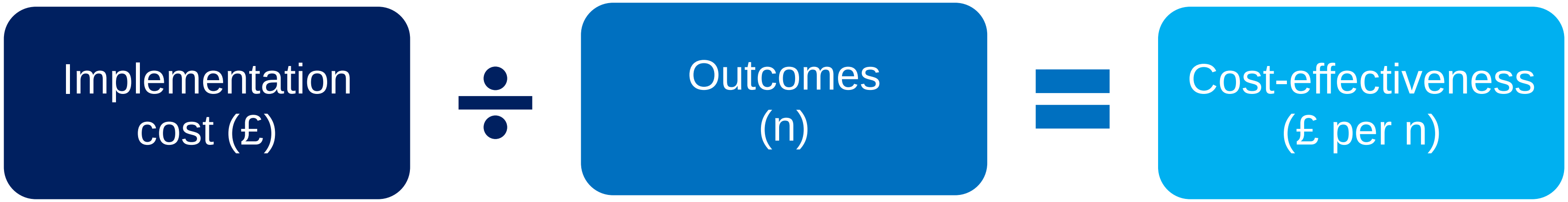
- The counter-factual (eg, no of incidents, s136 use)
- Reduction in demand
- Downstream time savings (eg, custody)
- Measurable clinical outcomes and other social benefits
- Additional costs incurred by police and partner agencies (eg, increased hospital admissions)

Building a business case

- **Strategic case**
  Rationale for the project, including context and the case for change
- **Economic case**
  Assessment of the expected costs and benefits of the project
- **Commercial case**
  Procurement and contractual steps required to deliver the project
- **Financial case**
  Sources of funding and provisions for liabilities or cost over-runs
- **Management case**
  Arrangements for project delivery, governance and performance monitoring

1. Assessing cost-effectiveness

- Did implementation of the intervention deliver value-for-money in terms of its costs and outcomes?
- Only inputs are costed
- Typical result of cost-effectiveness analysis
  - Cost per unit of outcome



Implementation cost (£) ÷ Outcomes (n) = Cost-effectiveness (£ per n)

1. Assessing cost-effectiveness

- **Costs**
  - Intervention A = £120,000
  - Intervention B = £100,000

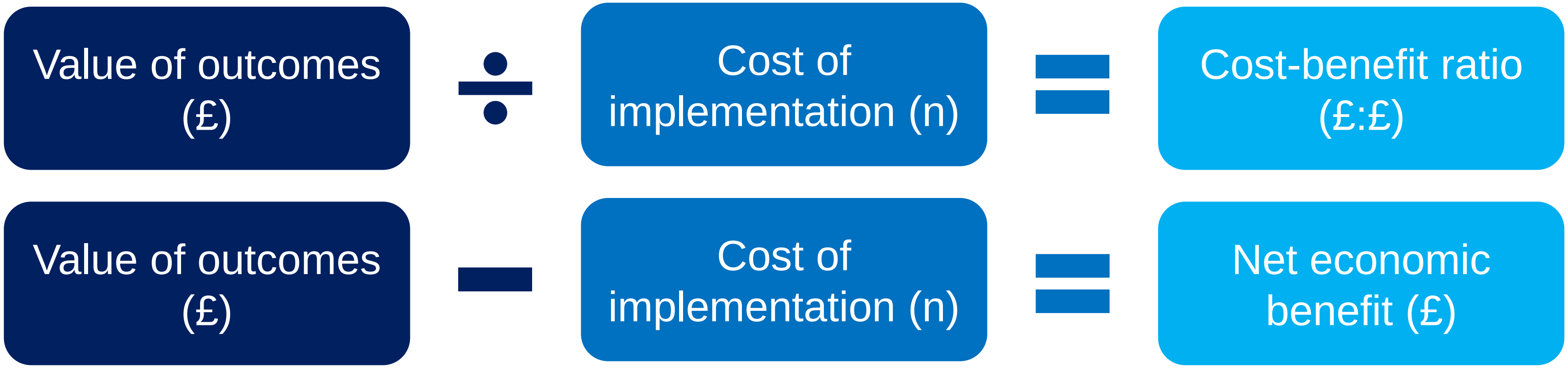
- **Outcomes**
  - Intervention A = a reduction in 100 burglaries
  - Intervention B = a reduction in 60 burglaries

- **Cost-effectiveness**
  - Intervention A = £120,000 / 100 = **£1,200 per burglary reduced**
  - Intervention B = £100,000 / 60 = **£1,667 per burglary reduced**

Adapted from Dhiri and Brand 1999

2. Assessing cost-benefits

- Was the cost of implementing an intervention outweighed by the monetary value of its outcomes?
- Both inputs and outcomes are costed
- Typical results of cost-benefit analysis:
  - Cost-benefit ratio
  - Net economic value



Value of outcomes (£) ÷ Cost of implementation (n) = Cost-benefit ratio (£:£)

Value of outcomes (£) − Cost of implementation (n) = Net economic benefit (£)

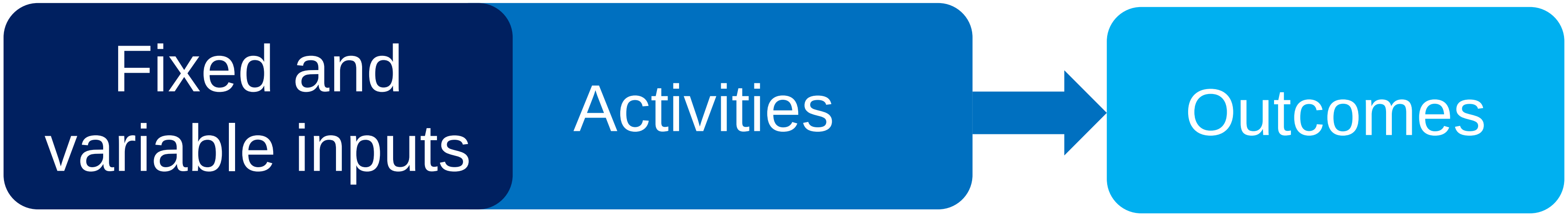
2. Assessing cost-benefits

- **Costs**
  - Intervention A = £120,000
- **Outcomes**
  - Intervention A = 100 reduced burglaries
- **Monetary value of outcomes**
  - Intervention A = 100 x £1,500 = £150,000
- **Cost-effectiveness**
  - Cost-benefit ratio = £150,000 / £120,000 = **1.25:1**
  - Net economic benefit = £150,000 – £120,000 = **£30,000**

Intervention B: cost-benefit ratio = 0.9:1, net economic benefit = -£10,000

Adapted from Dhiri and Brand 1999

An economic evaluation framework



- Staff costs
- Training
- Premises
- Running costs
- Equipment
- Research
- Communications
- Partners

- Time-use
- Outputs

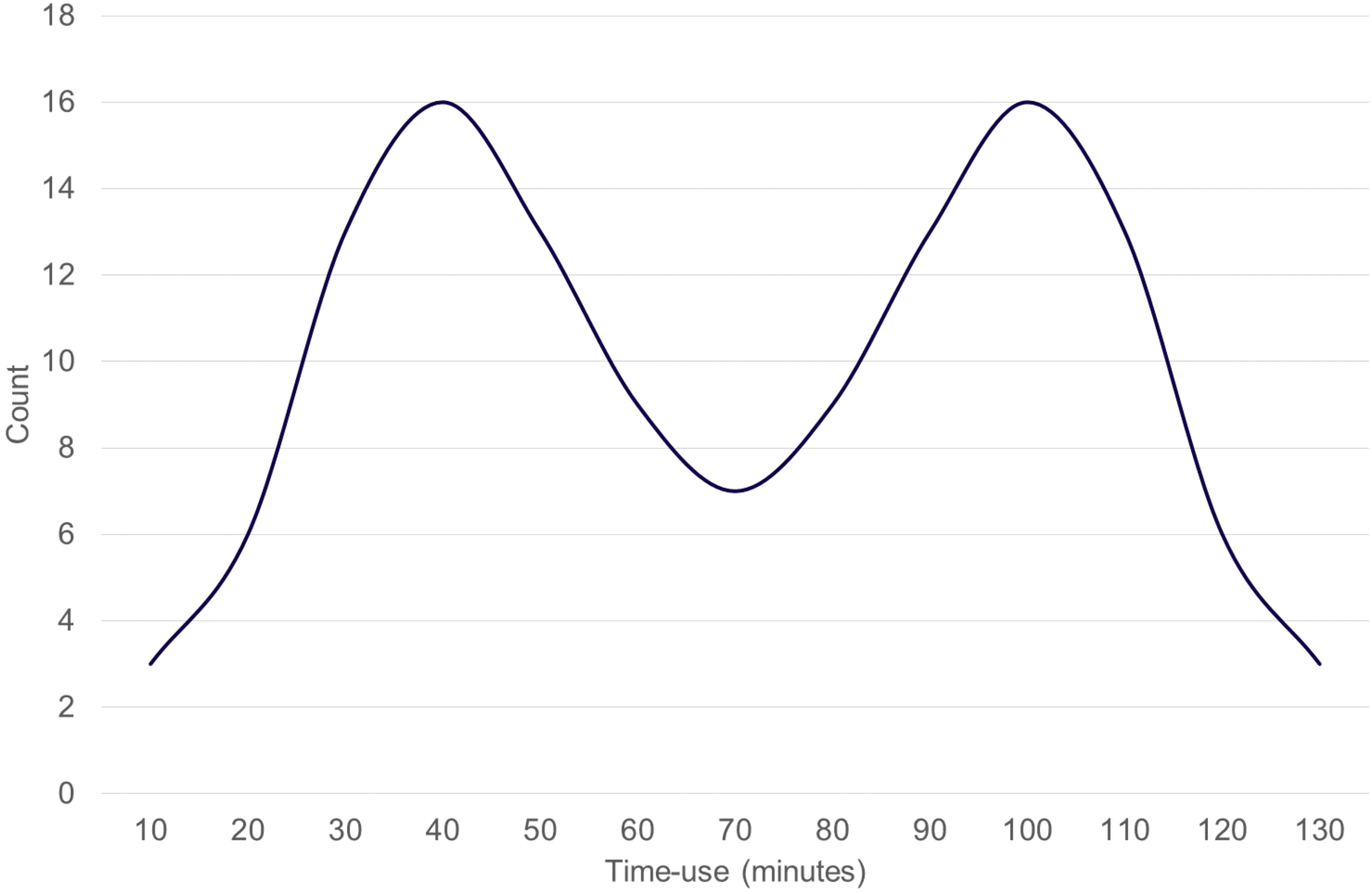
- Intended effects
- Side effects
- Effect on demand
- Wider social impacts

Measuring time-use

Method	easure
Observation	bjective
Systems data	bjective
Activity diaries	ubjective
One-off surveys	ubjective

- Consideratio
  - The Hawt
  - Inter-rater
  - Sampling
  - Bureaucra
  - Perceived
  - Messines

Average time-use?



Being realistic

- Average time spent by officers responding to mental health incidents

  = 1,800 minutes (30 hours per day)

- Cost of a police officer (average starting salary = £23,259)

  = £13.50 per hour

- Cost of dealing with mental health incidents

  = £405 per day

- Total cost of dealing with mental health incidents

  = £147,825 per annum

Being realistic

- Average time spent by officers responding to mental health incidents

  = 1,800 minutes (30 hours per day)

This figure looks very specific…

- Data source?
- Sampling?
- Margins of error?
- Variability in data?
- Seasonality?

- Cost of a p e officer (average starting salary = £23,259)

  = £13.50

- Cost cidents

  = £40

- Total th incidents

  = £14

Being realistic

- Average time spent by officers responding to mental health incidents

  = 1,800 minutes (30 hours per day)

- Cost of a police officer (average starting salary = £23,259)

  = £13.50 per hour

- Cost of dealing with me alth incidents

  = £405 per day

- Total cost of dealin

  = £147,825 per ann

This figure may not be representative…

- Data source?
- Mean average?
- Average of all response officers?
- Other employment costs (eg, employer contributions and training)

Being realistic

- Average time spent by offic alth incidents

  = 1,800 minutes (30 hours p

- Cost of a police officer (ave ge starting salary = £23,259)

  = £13.50 per hour

- Cost of dealing with mental health incidents

  = £405 per day

- Total cost of dealing with mental health incidents

  = £147,825 per annum

This figure looks unrealistic…

- Working days?
- Leave?
- Training?
- FTE?

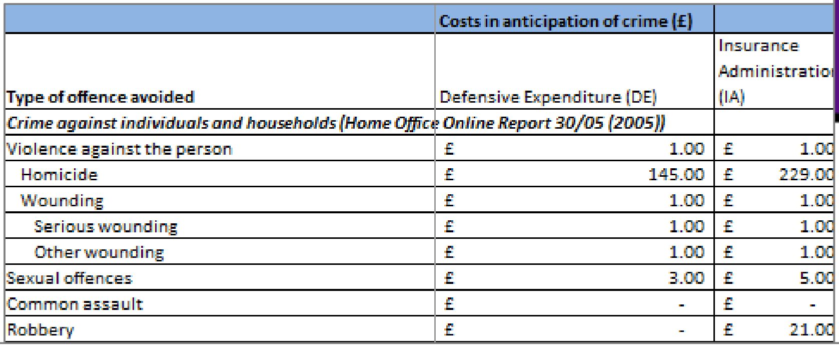
Be

This figure looks very certain…

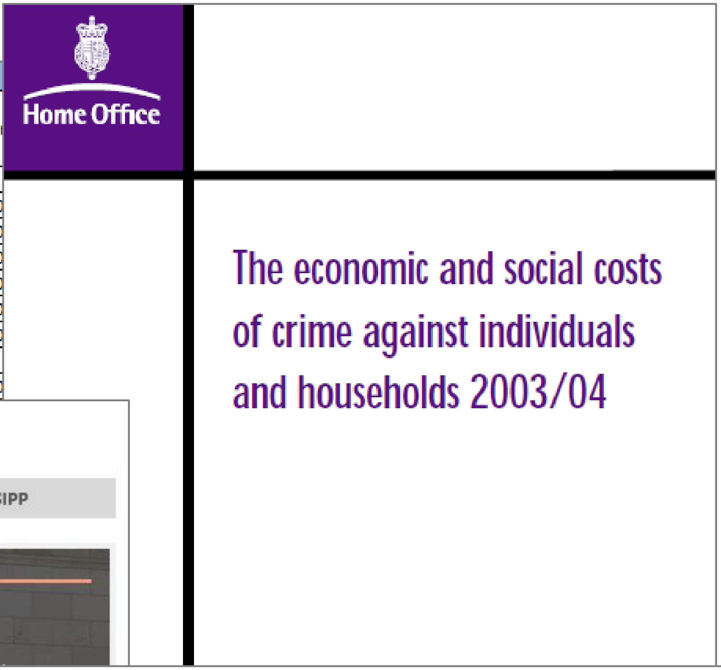
- Questionable averages?
- Unrepresentative costs?
- Unrealistic multipliers?
- Re-deployable resources?

- A sponding to mental health incidents
  = y)
- Cost o ce officer (average starting salary = £23,259)
  = £13.5 hour
- Cost of aling with mental health incidents
  = £405 p er day
- Total cost of dealing with mental health incidents
  = £147,825 per annum

Placing a monetary value on outcomes



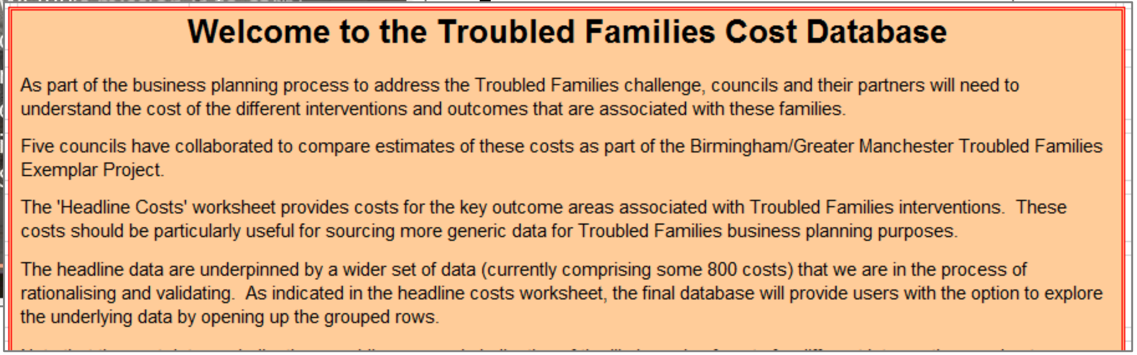
	Costs in anticipation of crime (£)	
Type of offence avoided	Defensive Expenditure (DE)	Insurance Administratio (IA)
Crime against individuals and households (Home Office Online Report 30/05 (2005))		
Violence against the person	£ 1.00	£ 1.00
Homicide	£ 145.00	£ 229.00
Wounding	£ 1.00	£ 1.00
Serious wounding	£ 1.00	£ 1.00
Other wounding	£ 1.00	£ 1.00
Sexual offences	£ 3.00	£ 5.00
Common assault	£ -	£ -
Robbery	£ -	£ 21.00



Home Office

The economic and social costs of crime against individuals and households 2003/04

Washington State Institute for Public Policy



Welcome to the Troubled Families Cost Database

As part of the business planning process to address the Troubled Families challenge, councils and their partners will need to understand the cost of the different interventions and outcomes that are associated with these families.

Five councils have collaborated to compare estimates of these costs as part of the Birmingham/Greater Manchester Troubled Families Exemplar Project.

The 'Headline Costs' worksheet provides costs for the key outcome areas associated with Troubled Families interventions. These costs should be particularly useful for sourcing more generic data for Troubled Families business planning purposes.

The headline data are underpinned by a wider set of data (currently comprising some 800 costs) that we are in the process of rationalising and validating. As indicated in the headline costs worksheet, the final database will provide users with the option to explore the underlying data by opening up the grouped rows.

Other things to consider

- Counterfactual
- Optimism bias
- Timescales for implementation and changes in outcome
- Cashable and non-cashable savings
- Reduced services and increased demand
- Beneficiaries
- Knock-on costs
- Uptake and attrition
- Scaling-up implementation

College of Policing

college.police.uk

BetterEvidence forBetterPolicing™

Discussion and self-reflection

Everyone

College of Policing

college.police.uk

BetterEvidence forBetterPolicing™

Evaluation case study

Professor Stuart Kirkby, UCLan

Lunch

College of Policing

college.police.uk

BetterEvidence forBetterPolicing™

Commissioning research

Paul Quinton

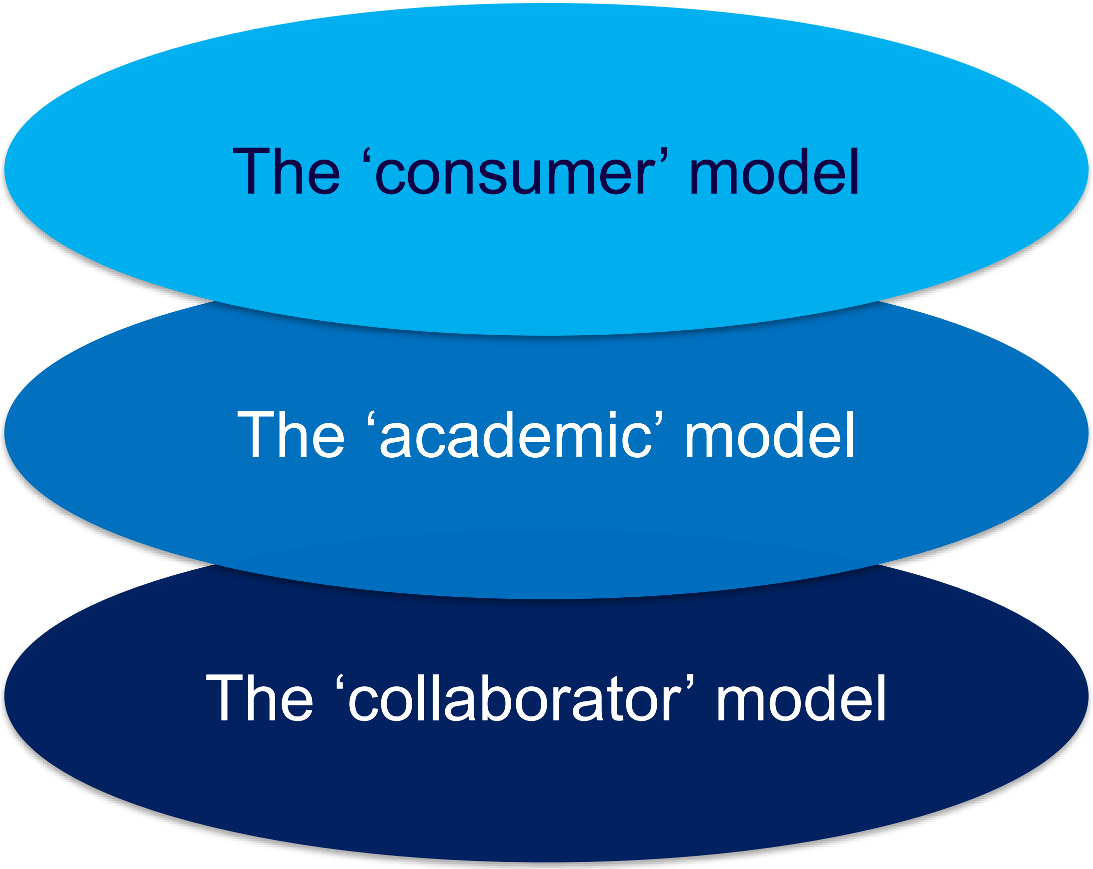
Resourcing an evaluation

- Police and partner analysts
- Interns – ESRC internships, CASE studentships
- Officers and staff doing dissertations – fast-track schemes, further education, College bursary scheme
- Established police/academic partnerships – Police Knowledge Fund
- Unfunded police/academic collaborations – access, publications
- Academic funding – ESRC, charitable trusts
- Home Office funding – Innovation and Transformation Funds
- Top-sliced funding – 10% of implementation spend

Deciding on the scope

- What can you afford?
- What’s feasible?
- What data do you have access to?
- What statements do you want to make about the impact and implementation of your intervention?
- Who do you need to convince of what and what will convince them?
- Are you clear on what the aims of the intervention are?
- Do you have a design for the evaluation?
- Do you want a supplier to propose a design for the evaluation?
- What relationship do you want to have with the supplier?

Differing approaches



The ‘consumer’ model

The ‘academic’ model

The ‘collaborator’ model

Inviting tenders for evaluation

- Procurement thresholds
  - Single tenders
  - Quotes and framework agreements
  - Full open competition
  - European open competition
- Unfair competitive advantage
- Expressions of interest exercises
- Timing and timescales for invitations to tender
- Revealing budget details
- Revealing assessment criteria
- Post-tender negotiations

Assessing tenders for evaluation

- How many work days will be delivered?
- How much time is being spent on different activities?
- Who’s doing what and what value does it add?
- What are their daily rates?
- How much flexibility is there in the proposal?
- Have they proposed anything unnecessary?
- Are you paying for capital spend?
- Is the T&S charged in addition?
- Do they have a good track record and right expertise?
- Is there evidence of good project management?

Appointing a supplier

- Weighting assessment criteria
- Intellectual property
- Impartiality and conflicts of interest
- Vetting – data and site access
- Break clauses
- VAT on research
- Payment in arrears
- Budget roll-over
- Working at risk
- Final payments

Working with a supplier

- Start-up meeting
- Frequency and type of contact
- Payment milestones
- Transparency and quality of analysis and write-up
- Strength and tone of conclusions
- Independent academic advice – College research surgeries
- Peer review
- Publishing research

College of Policing

college.police.uk

BetterEvidence forBetterPolicing™

Quality assuring research

Will Finn

A group exercise

- Review the two tables of results from a hypothetical evaluation and answer the following questions…

❑ What might be wrong with the approach they have outlined?

❑ What improvements can you make to the approach?

- Five minutes discussion followed by feedback on what you identified

Table 1

" There is a perception that call handlers have a tendency to err on the side of caution resulting in minor incidents being misgraded and inappropriate deployments. A statistically valid dip sample of 382 call logs during November 2010 shows that a high proportion of grade 2 incidents have been misgraded. This prompts the force to consider a series of refresher training sessions with call handlers designed to remind them about call grading policy and reduce the proportion of calls that are misgraded. Grade three calls do not require a deployment therefore there is potential time saving to be made by ensuring calls are appropriately graded."

Table 1: Grading of calls before and after re training of call handlers.

	Pre intervention (No.s of calls)	Pre intervention % of total)	Post intervention (No.s of calls)	Post intervention % of total)	Change from pre to post	% change from Pre to post
Grade one	93.0	24.3%	90.0	23.6%	-3.0	-0.8%
Grade two	186.0	48.7%	125.0	32.7%	-61.0	-16.0%
Grade three	103.0	27.0%	167.0	43.7%	64.0	16.8%
Total	382.0		382.0			

(1) Pre data was obtained in November 2010 from a self completion form distributed to call handlers. The form asked them to record the number of calls they received and how they had graded them.
(2) Post data is from a statistically valid dip sample of 382 call cards over one week period immediately after completion of the training.

Answers: Table 1

" There is a perception that call handlers have a tendency to err on the side of caution resulting in minor incidents being misgraded and inappropriate deployments. A statistically valid dip sample of 382 call logs during November 2010 shows that a high proportion of grade 2 incidents have been misgraded. This prompts the force to consider a series of refresher training sessions with call handlers designed to remind them about call grading policy and reduce the proportion of calls that are misgraded. Grade three calls do not require a deployment therefore there is potential time saving to be made by ensuring calls are appropriately graded."

Table 1: Grading of calls before and after re training of call handlers.

	Pre intervention (No.s of calls)	Pre intervention % of total)	Post intervention (No.s of calls)	Post intervention % of total)	Change from pre to post	% change from Pre to post
Grade one	93.0	24.3%	90.0	23.6%	-3.0	-0.8%
Grade two	186.0	48.7%	125.0	32.7%	-61.0	-16.0%
Grade three	103.0	27.0%	167.0	43.7%	64.0	16.8%
Total	382.0		382.0			

(1) Pre data was obtained in November 2010 from a self completion form distributed to call handlers. The form asked them to record the number of calls they received and how they had graded them.
(2) Post data is from a statistically valid dip sample of 382 call cards over one week period immediately after completion of the training.

'Statistically valid' this statement has no meaning in statistical terms

The amount of time in November is not specified. It should be for the same amount of time as the 'post' sample.

This is an inappropriate comparison. Different data sources have been used here, e.g. self completion forms to collect the pre data and call cards to collect the post data.

This is not a good time to measure the success of an intervention.

A more accurate comparison would be to take a sample from the same time of year or a time of year with a similar demand profile as the before sample.

Table 2

The time saved as a result of reducing these inappropriate deployments is shown to make a significant saving to the force.

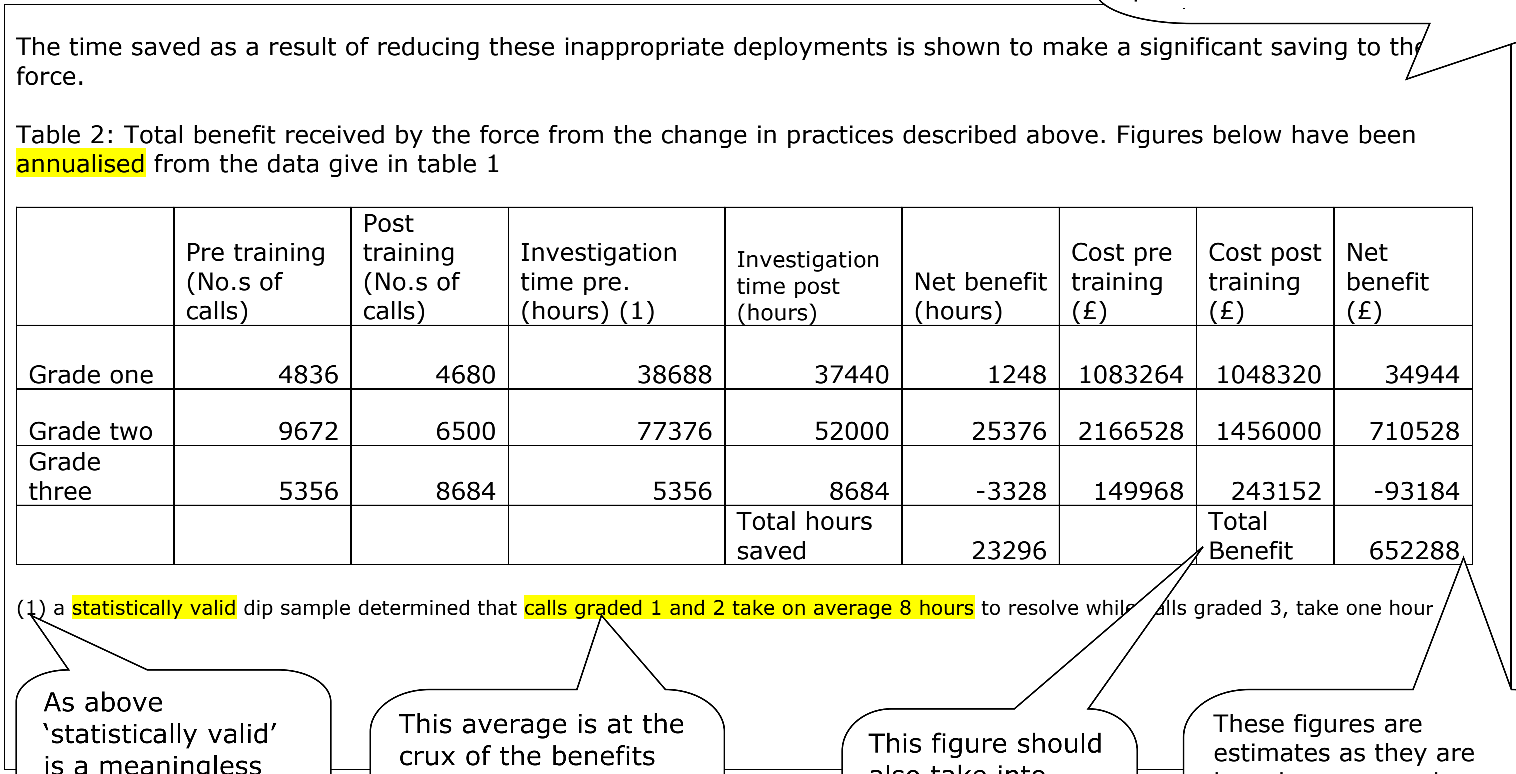
Table 2: Total benefit received by the force from the change in practices described above. Figures below have been annualised from the data give in table 1

	Pre training (No.s of calls)	Post training (No.s of calls)	Investigation time pre. (hours) (1)	Investigation time post (hours)	Net benefit (hours)	Cost pre training (£)	Cost post training (£)	Net benefit (£)
Grade one	4836	4680	38688	37440	1248	1083264	1048320	34944
Grade two	9672	6500	77376	52000	25376	2166528	1456000	710528
Grade three	5356	8684	5356	8684	-3328	149968	243152	-93184
				Total hours saved	23296		Total Benefit	652288

(1) a statistically valid dip sample determined that calls graded 1 and 2 take on average 8 hours to resolve while calls graded 3, take one hour

Answers: Table 2

‘Annualised’ simply means multiplied out to apply to a whole year. In this case the one week sample has been multiplied by 52. Therefore any problems with the one week sample



The time saved as a result of reducing these inappropriate deployments is shown to make a significant saving to the force.

Table 2: Total benefit received by the force from the change in practices described above. Figures below have been annualised from the data give in table 1

	Pre training (No.s of calls)	Post training (No.s of calls)	Investigation time pre. (hours) (1)	Investigation time post (hours)	Net benefit (hours)	Cost pre training (£)	Cost post training (£)	Net benefit (£)
Grade one	4836	4680	38688	37440	1248	1083264	1048320	34944
Grade two	9672	6500	77376	52000	25376	2166528	1456000	710528
Grade three	5356	8684	5356	8684	-3328	149968	243152	-93184
				Total hours saved	23296		Total Benefit	652288

(1) a statistically valid dip sample determined that calls graded 1 and 2 take on average 8 hours to resolve whil alls graded 3, take one hour

As above ‘statistically valid’ is a meaningless statement.

This average is at the crux of the benefits calculation. Does this average come from reliable data?

This figure should also take into account the cost of the training.

These figures are estimates as they are based on one week samples. Therefore they should be presented as such.

What we'll cover

- Key considerations for critical appraisal
- Frequent pitfalls of research
- Samples
- Confidence intervals
- Appropriate comparisons
- Summary

Questions to ask yourself

- Does the report answer my question? Is the method used appropriate?
- Is the rationale for the method explained? If not, then ask why.
- Do I have sufficient information to replicate the study?
- What other factors could have influenced the result, and have they been controlled for as much as possible?
- Are you happy the research is reliable enough to base your decisions on?

The Usual Suspects

Here are a few of the most common ways that study results can be misrepresented:

- Proxy measures – often necessary but remember, they are just proxies.
- Cherry picking & data dredging – does the analysis match the agreed aims of the research?
- Quantifying qualitative data – it's indicative not representative
- Why might a researcher exaggerate the effect? Publication bias

- Being specific when you shouldn't be – question precise results!
- Comparing apples with oranges – are the comparisons appropriate?

Sampling

Have you sampled the right population?	
If quantitative data is collected to measure impact	• has the data been collected so that it is representative of the wider population? • did everyone in population of interest have an ‘equal chance’ of participating? If not, why not? • **“random sampling”** or **“stratified sampling”** • Beware of **“quota sampling”**
If qualitative data is collected to evaluate implementation	• has the data been collected so a diversity of views/experiences are captured? • **“purposive sampling”**

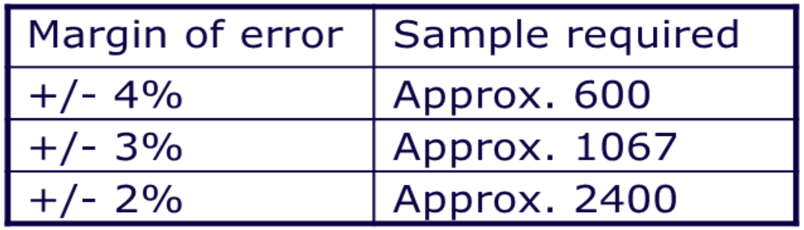
Is your sample big enough?

- The sample should large enough to detect any so real differences that exist between your groups (or time periods)
- This is because average measures from a sample are *estimates* and have *margins of error (*also known as a confidence intervals).

Confidence Intervals

- Expect estimates – alarm bells should ring if you're presented with precise measures of effect.
- As evaluations will take a 'sample' of your population of interest, any calculations made based on observing that sample will be an estimate of the **true effect**.
- Social science usually uses 95% confidence intervals. In simple terms, this means that if you were to take 20 different samples, the results from one of these would fall outside of the range due to chance.
- Usually there is a trade-off between accuracy and cost.

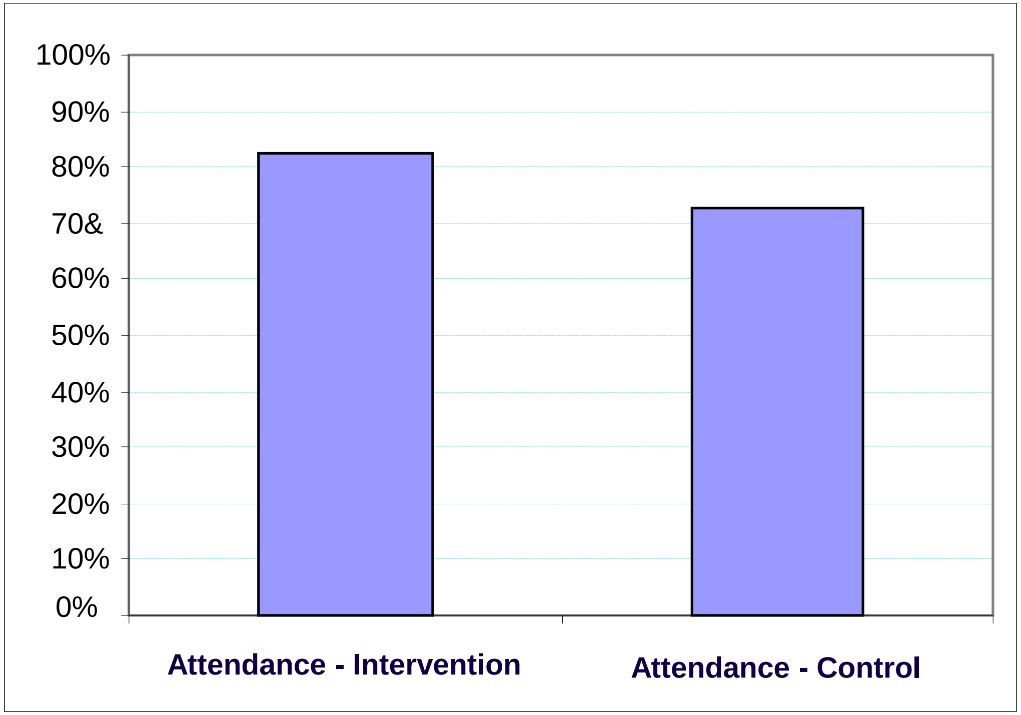
- If you are using survey data or data that can be similarly divided into proportions then this table provides details of the sample required for different margins of error.



Margin of error	Sample required
+/- 4%	Approx. 600
+/- 3%	Approx. 1067
+/- 2%	Approx. 2400

The impact of confidence intervals

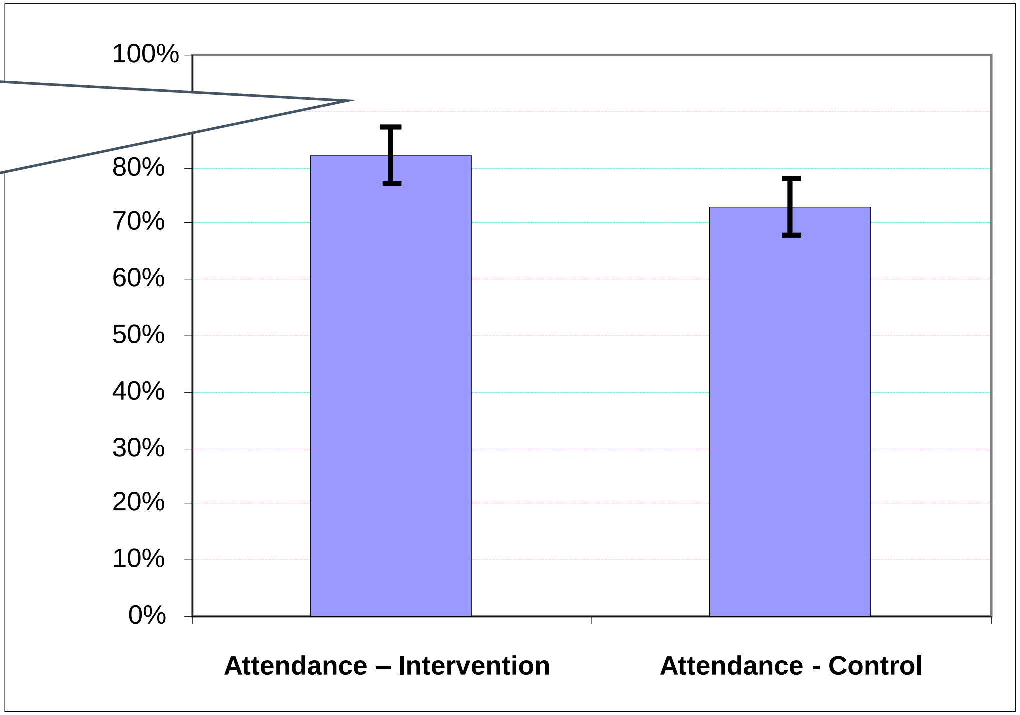
Look at this chart – it seems average attendance in school was higher for pupils who received the intervention compared to those who didn't. But is this so when we take into account the confidence intervals?



The impact of confidence intervals

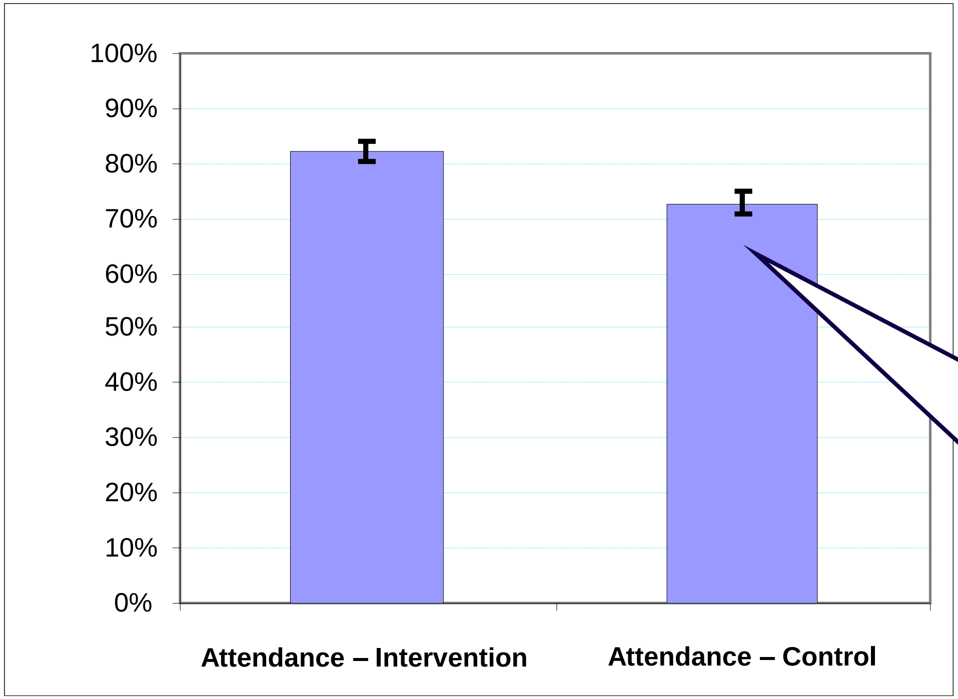
This chart shows the confidence interval associated with the previous result. Notice that the interval bars seem to ‘overlap’ (i.e. cover the same range of %). This means we cannot be confident that the difference between the two sets of figures is statistically significant.

These ‘**error bars**’ indicate 95% confidence intervals (i.e. on only 1 in 20 occasions would the averages be expected to lie outside this range as a result of chance).



The impact of confidence intervals

If we had trialled the scheme in more schools and monitored 2,500 pupils instead of 400, we would have calculated a confidence range of plus or minus 2%. If we had achieved the same attendance rate, we could have said that a significant difference had been seen – since the ranges of the two confidence intervals would not have overlapped.



Remember diminishing returns once you have sampled about 400 people. Consider how much extra accuracy you need, and how much it will cost to increase the sample

In general, the more people in your sample, the smaller the confidence interval.

An appropriate comparison

Type of comparison	Question to ask
If the allocation to the intervention was random…	how was the randomisation process conducted?
If the allocation was not random…	how was it ensured the groups were comparable? You should expect some analysis of the groups to ensure they’re sufficiently similar (e.g. demographics, individual characteristics)
If evaluation compares one group over time…	are the time periods being compared the same? (e.g. same time of year, length of time and avoiding significant events) **– Beware of DIP samples!!**
In all circumstances…	• is the data collected in the same way for both groups/time periods? • has enough time elapsed for any effect be a ‘true’ effect? - The Hawthorne effect

Summary

1. Does the report answer my question?

2. Does the study use a suitable comparison to estimate the effect of the intervention?

3. Is there an adequate description of, and rationale for, the sample used and the methods for how the sample was identified and recruited?

4. Is there an adequate description of the methods used to collect *and* analyse the data?

5. Are the concluding statements supported by the evidence?

Overall: Do you have enough reliable evidence to be able to say whether something worked and what actually worked?

Discussion and self-reflection

Everyone

Refreshment break

College of Policing

college.police.uk

BetterEvidence forBetterPolicing™

Open research surgery

Everyone

College of Policing

college.police.uk

BetterEvidence forBetterPolicing™

Concluding remarks and close

Paul Quinton and Steph Waddell

Contact details

paul.quinton@college.pnn.police.uk
07595 007 421
@pkquinton