

The Third Wave of the Evidence Revolution: Using global evidence to inform local policy and practice

Howard White

CEO, Campbell Collaboration



@campbellreviews @HowardNWhite

The three waves of the evidence revolution

Wave one: New public management and the results agenda

- Origins US, UK, Australia and New Zealand
- Adopted Clinton and Blair governments
- Shifted focus from monitoring inputs (how much money we spend)...



- to outcomes (long-term unemployed returned to work, families lifted out of poverty, women empowered etc.



UK Modernizing Government (aka 'the results agenda')

The Government wants to ensure the effectiveness of the services the public receive. That is what makes a difference to the quality of people's lives. The way to do this is through Public Service Agreements (PSAs).

UK Cabinet Office

Performance targets

In particular DFID will ensure that:

(i) at least 75% of bilateral country resources are directed at low income countries by 2002, compared to 67% currently.

In the 30 largest recipients of British aid, DFID aims to make a major contribution to the achievement of:

(ii) an annual 1.5% increase in GDP per capita, from the current average of 1.0%; (Objectives 1 to 5)

(iii) a reduction of under 5 and maternal mortality rates from 74 to 70 per 1000 live births and from 324 to 240 per 100,000 live births respectively by 2002; (Objectives 1 to 5)

(iv) an increase from 61% to 91% of children in primary school by 2002. (Objectives 1 to 5)

IIA

103^d CONGRESS
1ST SESSION

H. R. 826

IN THE SENATE OF THE UNITED STATES

MAY 26 (legislative day, APRIL 19), 1993

Received; read twice and referred to the Committee on Governmental Affairs

AN ACT

To provide for the establishment of strategic planning and performance measurement in the Federal Government, and for other purposes.

1 *Be it enacted by the Senate and House of Representa-*
2 *tives of the United States of America in Congress assembled,*

3 **SECTION 1. SHORT TITLE.**

4 This Act may be cited as the "Government Perform-

Government Results and Performance Act, 1993

- USAID: six strategic development goals
- E.g. “broad-based economic growth and agricultural development encouraged”
- For each goal defined outcome indicators at both country and global levels
- E.g. “average annual growth rates in real per capita income above 1 per cent”



USAID
FROM THE AMERICAN PEOPLE

- FY 2000 performance report states that “nearly 70 per cent of USAID-assisted countries were growing at positive rates in the second half of the 1990s, compared with 45 per cent in the early part of the decade”

But: ‘one cannot reasonably attribute overall country progress to USAID programs’

GAO: ‘so broad and progress affected by many factors other than USAID programmes, [that] the indicators cannot realistically serve as measures of the agency’s specific efforts’

And so...

USAID abandoned the use of strategic indicators as performance measures (retaining them as 'Development Performance Benchmarks')

This does not mean should NOT do monitoring... but know what it can and cannot do

TABLE 6: PERFORMANCE AGAINST THE HEALTH SERVICE OUTCOME TARGETS

Indicator	Performance 2014/15	Performance 2015/16		HSDP Target 2015/16
		Achievement	Disaggregation	
ART Coverage	56%	88%		57%
HIV+ pregnant women not on HAART receiving ARVs for eMTCT during pregnancy, labour, delivery and postpartum	72% (2013/14)	68.3%		85%
TB case detection Rate (all forms)	80% (2014/15)	NA		83%
IPT ² doses coverage for pregnant women	53.4% (2014/15)	55%		58%
IPT ³ doses coverage for pregnant women	NA	NA		93%
In Patient malaria deaths per 100,000 persons per year	30 (2013/14)	22	M – 20 F – 23	13
Malaria cases per 1,000 persons per year	460 (2013/14)	400	M – 365 F – 480	198
Under five vitamin A second dose coverage	26.6% (2013/14)	28%	M – 27% F – 28%	66%
DPT ¹ -Hib-Hep ³ Coverage	102.4% (2014/15)	103%	M – 105% F – 99%	95%
Measles coverage under 1 year	90% (2014/15)	96%	M – 96% F – 93%	90%
Bed occupancy rate (Hospitals & HC IVs)	NA	83%	RRH	
	50% (2013/14)	62%	GH	62%
	59% (2013/14)	52.2%	HC IV	55%
Average length of stay (Hospitals & HC IVs)	NA	4	NRH	4
	NA	4	RRH	4

Monitoring

GOVERNMENT ANNUAL REPORT

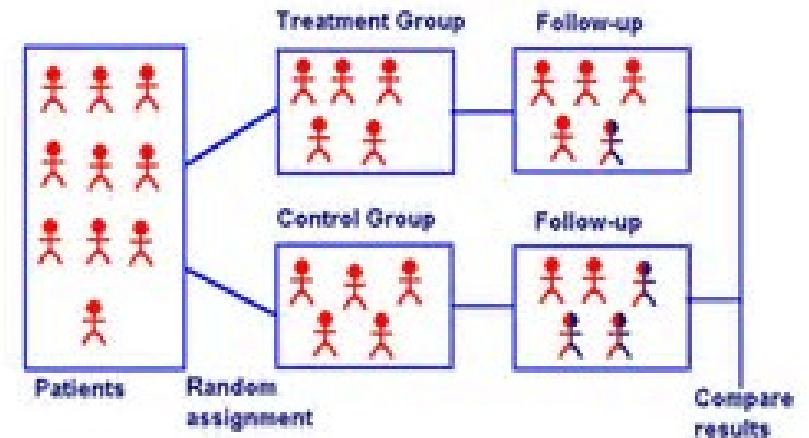

 FINANCIAL
VOLUME 1

 NATIONAL
OFFICE OF THE COMMISSIONER

So how do we measure what difference a programme makes, i.e. impact?

By using rigorous impact evaluations with a valid comparison group to control for selection bias, preferably a randomized controlled trial

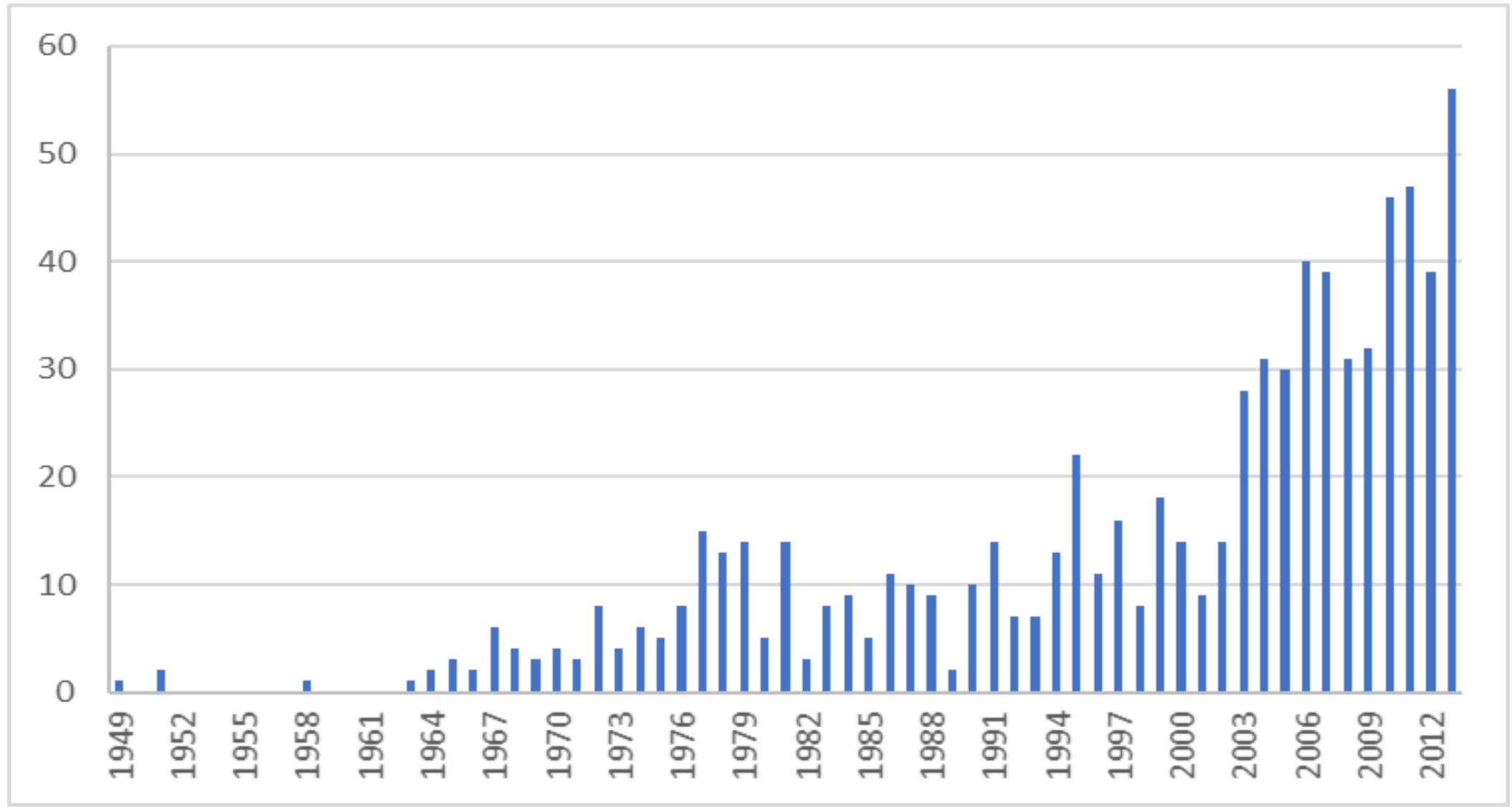
Randomized Controlled Trials



Examples include: (aspirin & streptokinase), (simvastatin & vitamins)

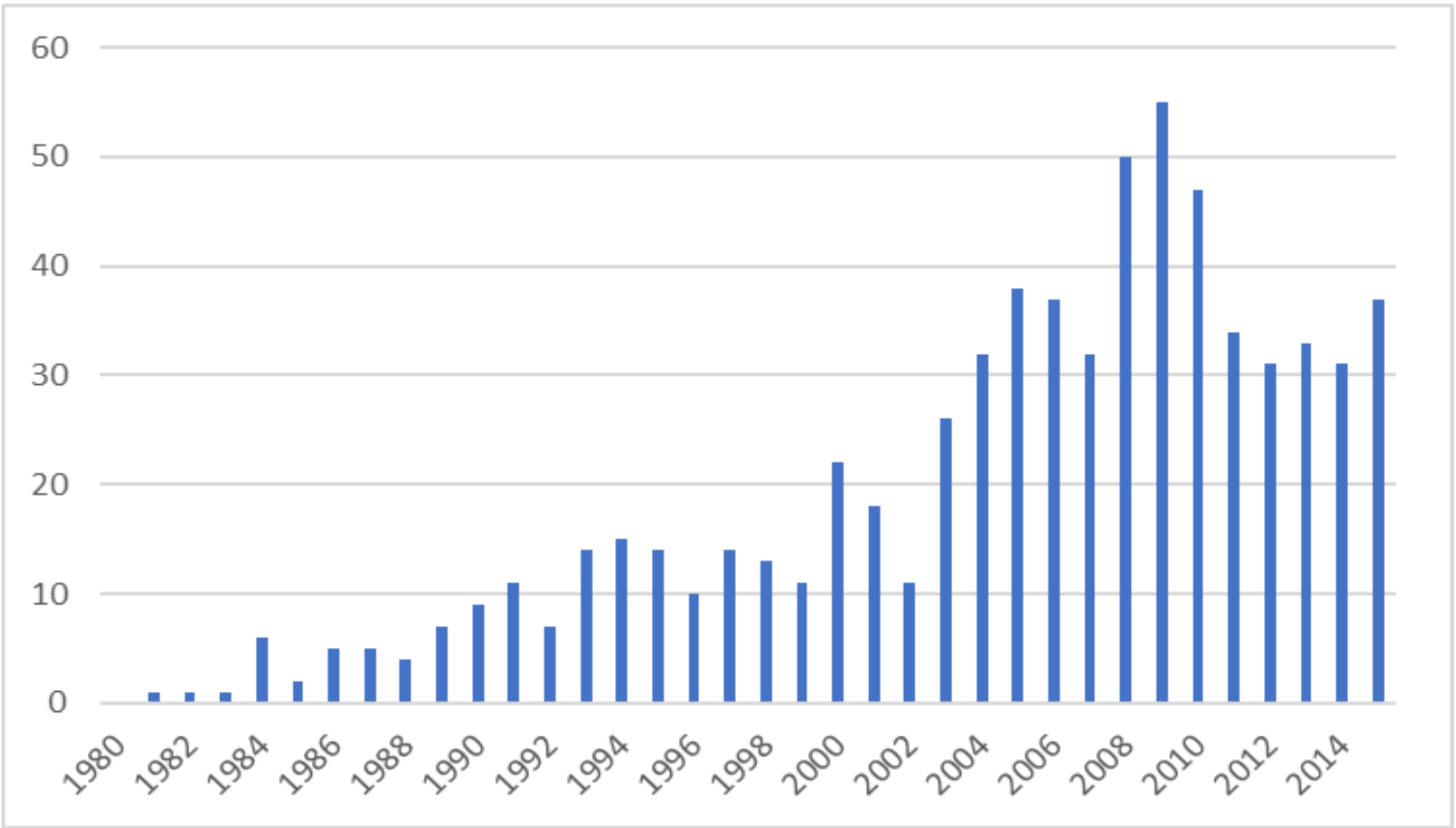
Wave Two: The Randomization Revolution

Number of social work RCTs published by year

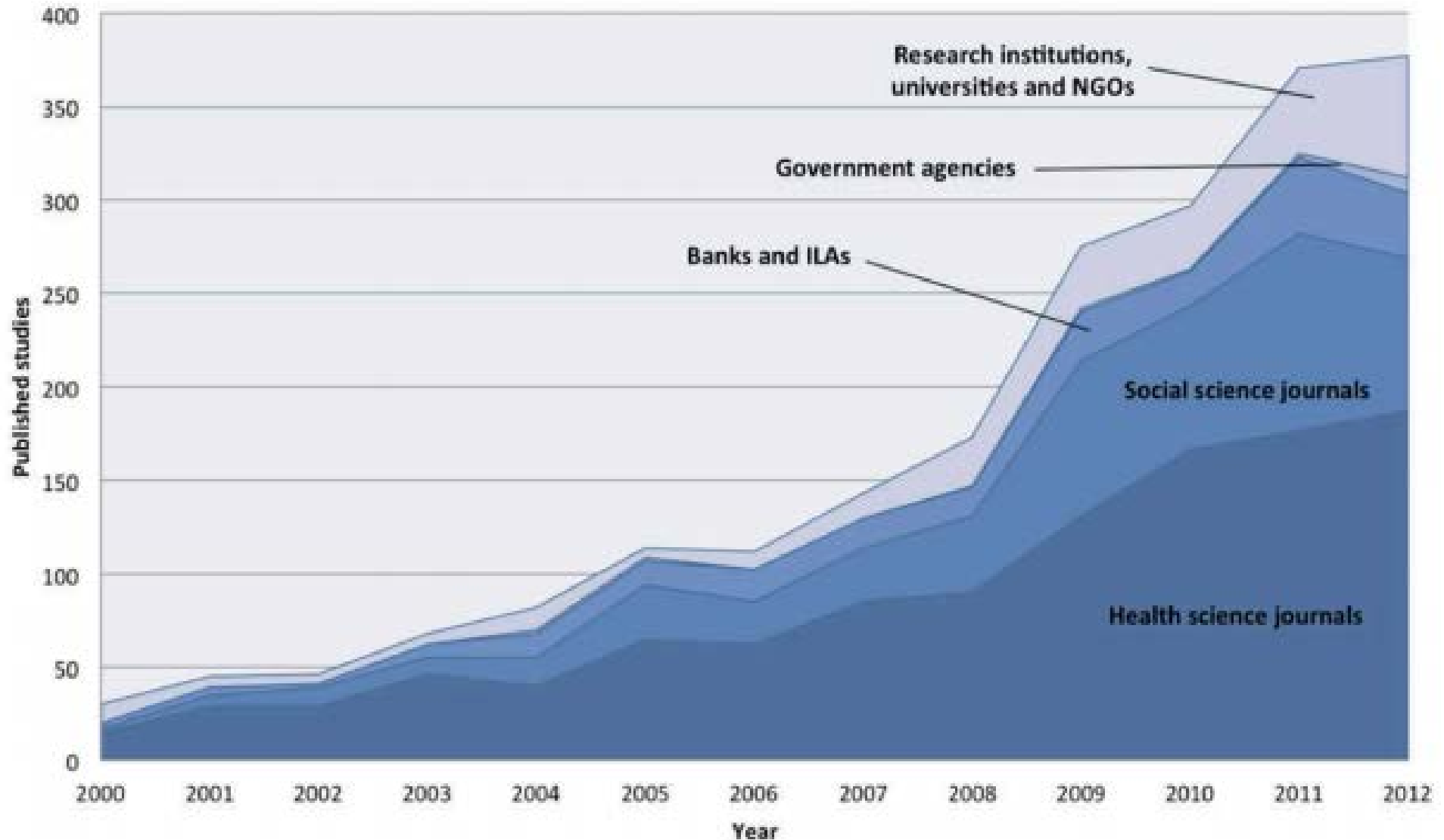


Source: Calculated from Bruce Thyer 'A Bibliography of Randomized Controlled Experiments in Social Work (1949–2013): Solvitur Ambulando' Research on Social Work Practice 2015, Vol. 25(7) 753-793

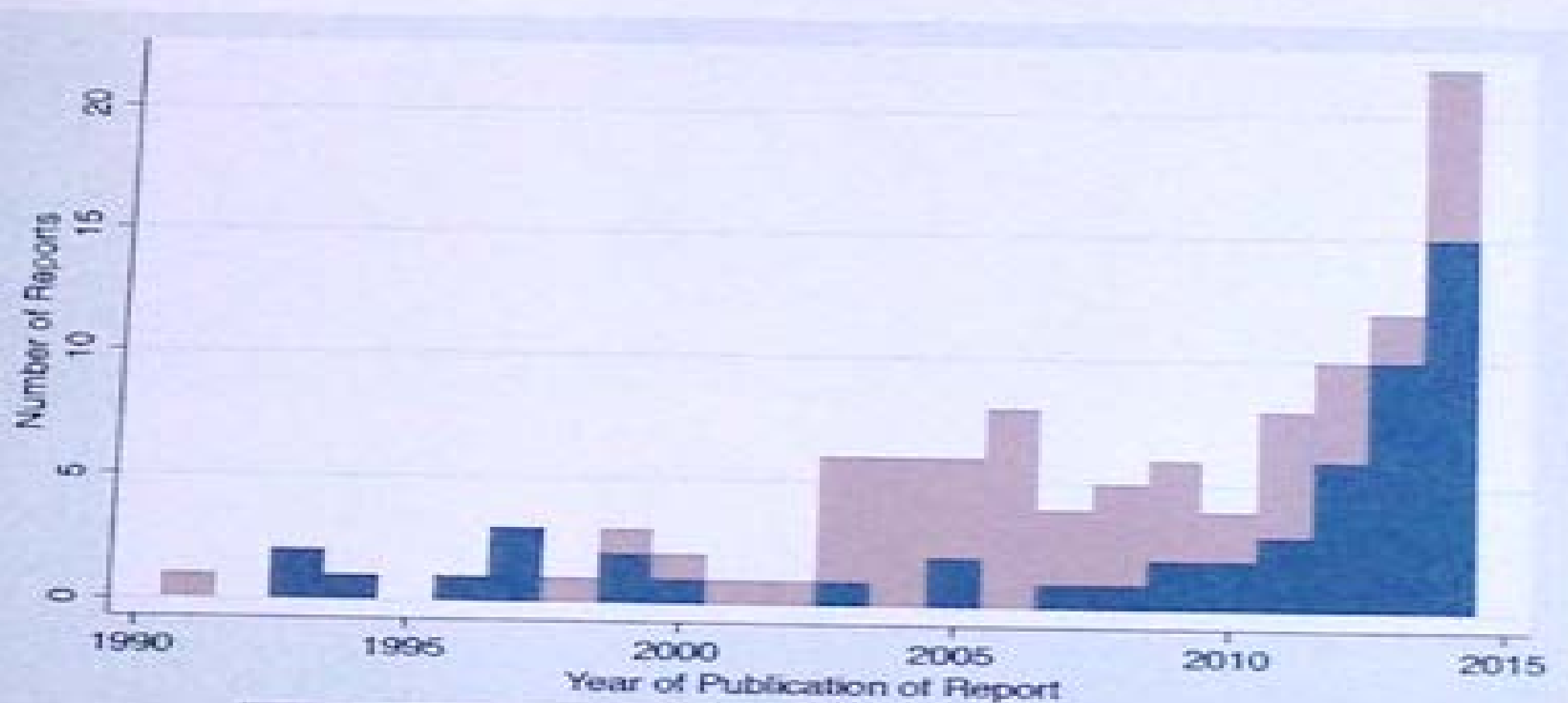
Education RCTs in IES database by year of publication



Annual publication rigorous impact evaluations in international development



Total number of reports and reports relying on RCTs by year of publication



But we should not use single studies to inform global policy

- Mandatory arrest
- Nurse Family Partnership
- Deworming



The screenshot shows the website for the Australian Nurse-Family Partnership Program. The header includes the program's name, a logo, and contact information for the National Program Centre at (07) 3114 4667. A navigation menu lists 'Home', 'News', 'Careers', 'Contact Us', and 'Program Staff'. Below this is a search bar and a main menu with categories: 'ABOUT', 'FOR MOTHERS', 'HOME VISITING TEAM', 'IMPLEMENTING SERVICES', and 'PROVEN RESULTS'. The 'PROVEN RESULTS' section is active, showing a breadcrumb trail: 'Home > Proven Results > Evidence of Program Effectiveness'. A sidebar on the left contains a 'Listen' button and a list of links: 'Proven Results', 'Evidence of Program Effectiveness', and 'Publications'. The main content area is titled 'Evidence of Program Effectiveness' and contains the following text: 'The ANFPP is based on the Nurse-Family Partnership home visiting program developed by Professor David Olds in the United States. Informed by rigorous research, the program has developed over more than three decades. The positive impact of the program has been demonstrated through three separate well-designed and r...ted randomised controlled trials. These occurred at: Imira, New York in 1977 which targeted white low-income women located at semi-rural population, Memphis, Tennessee in 1987 which involved low-income urban African-American mothers, and, and Denver, Colorado in 1994 which included large number of Hispanic families.



Wave Three: The Rise of Rigorous Evidence Synthesis

What is a systematic review?

- Systematic reviews
 - A systematic approach to summarizing the results from all existing studies on a specific question
- Meta-analysis
 - A statistical technique for combining effect sizes (impact estimates) into a single average treatment effect and examining sources of variation in effect sizes
- Primary studies
 - The individual studies which are included in a systematic review
- Impact evaluation
 - Studies include counterfactual analysis of effects. Randomized controlled trials (RCTs) are the most common impact evaluation design, but reviews can include valid non-experimental designs

Steps in the *systematic* review process

- Setting the question (the PICO)

Systematic

- Search strategy

Systematic

- Screening

Systematic

- Coding

Systematic

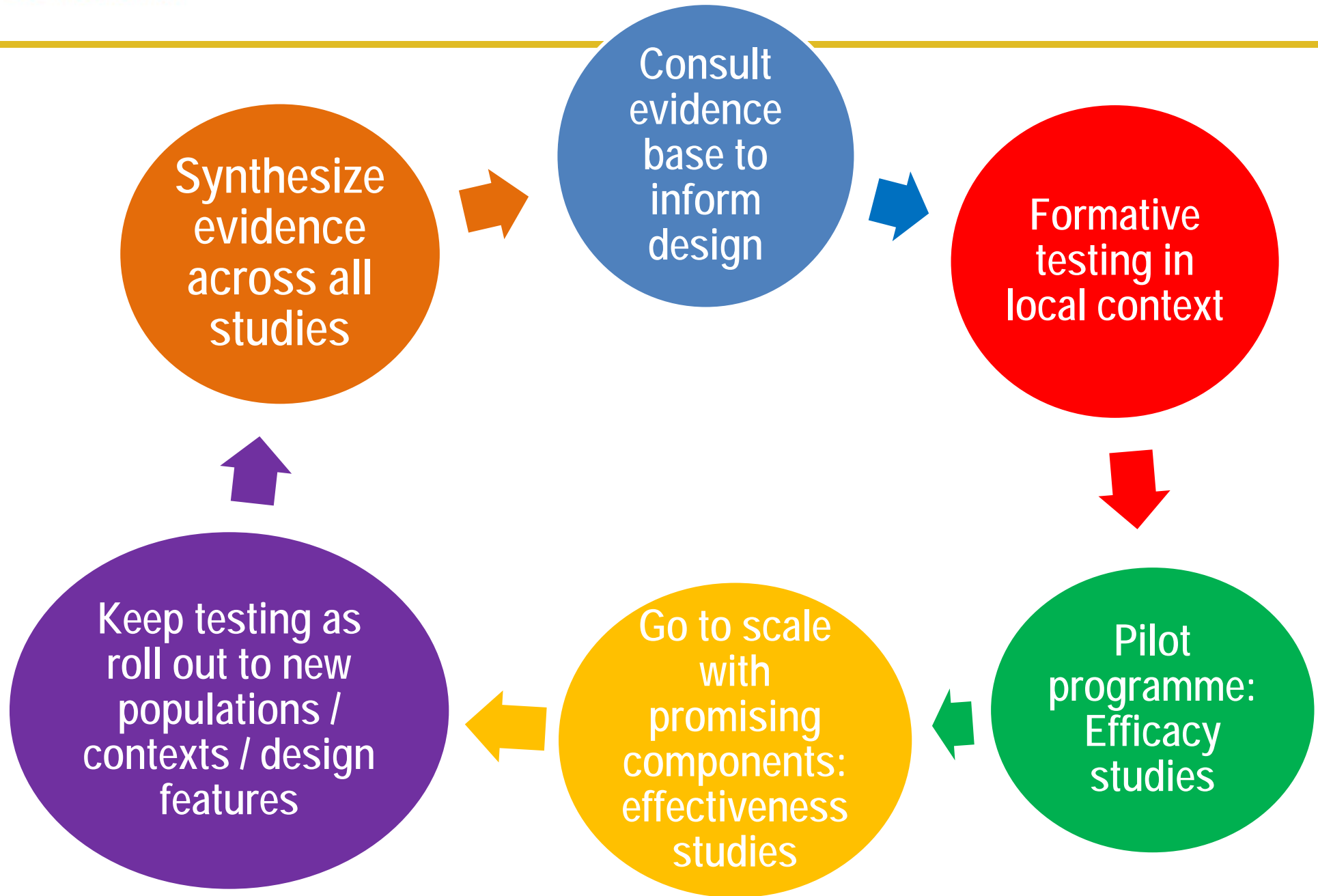
- Synthesis

Systematic

- Reporting

- Engagement with policy and practice

Evidence-driven project cycle



Institutionalization
of use of evidence is
taking place

Institutionalisation of the use of evidence: health

The World Health Organization (WHO) follows a guideline development process, described in detail in the **WHO Handbook for Guideline Development (2nd edition)**, overseen by the Guidelines Review Committee (GRC) established by the Director-General in 2007. The WHO Guidelines Review Committee ensures that WHO guidelines are of a high methodological quality, developed using a transparent and explicit process, and are **informed on high quality systematic reviews** of the evidence using state-of-the art systematic search strategies, synthesis, quality assessments and methods.

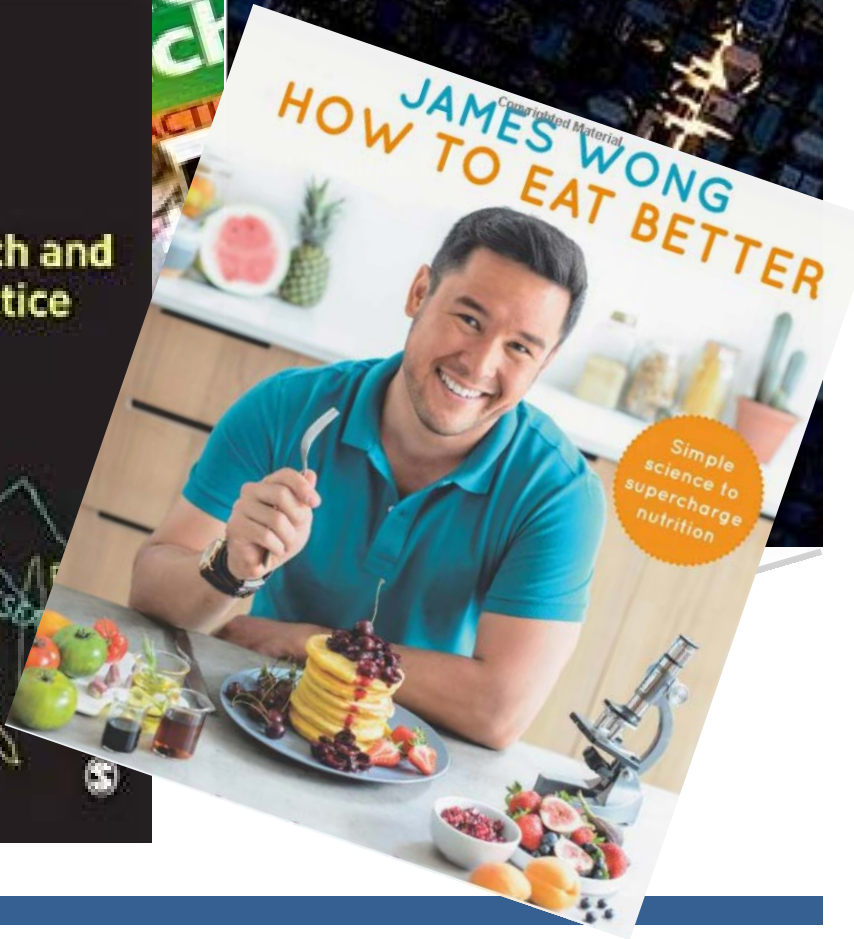
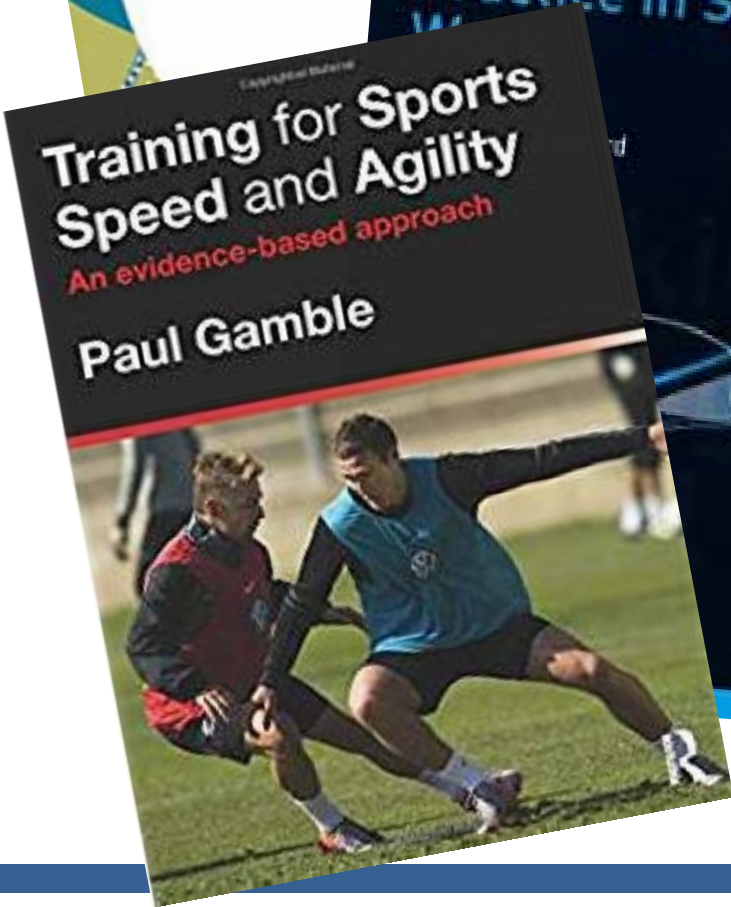
National Institutes Health Research (NIHR):

- Provides infrastructure support to 21 Cochrane Groups
- NIHR Cochrane Programme Grant Scheme funds reviews of relevance to NHS
- NIHR Cochrane Incentive Awards to accelerate reviews

National Institute for Clinical Excellence (NICE), Use systematic reviews for:

- Guideline production
- Eligibility for NHS resources

Language of evidence spreading to all sectors



What is happening? The UK model

- What Works Centres
- Funded by government and Big Lottery
- Commission reviews, largest also primary studies



E
E
F
ducation
ndowment
oundation

Funding > 500 trials in > ¼
primary schools in UK

Evidence portal

E.g. Pupil premium: in 2015
64% used Teaching and
Learning Toolkit compared to
36% in 2012. But 77% use
funds on programmes for all
pupils

What
Works
Network



What Works
Wellbeing

What
Works
Crime Reduction



CENTRE FOR
AGEING
BETTER

what works centre for
local economic growth

WHAT
WORKS
SCOTLAND



Early Years Toolkit

An accessible summary of educational research for early years teaching

Example of an evidence portal

Filter Toolkit

Toolkit Strand 

Cost 

Evidence Strength 

Months Impact 

Filter results by keywords



Cost



Evidence



Months Impact

Reset 

Communication and language approaches

Moderate impact for very low cost, based on very extensive evidence.



+6

Digital technology

Moderate impact for moderate cost, based on limited evidence.



+4

Earlier starting age

Moderate impact for very high cost, based on very limited evidence.



+6

Early literacy approaches

Moderate impact for very low cost, based on extensive evidence.



+4

- History since early seventies (e.g. negative income tax)
- What Works Clearing House well-established in education
- More recently ‘Moneyball for government’



- Moneyball for gov programmes
- Eg, Head Start, Nurse Family Partnership

But

- Single studies
- Possible COI



Evidence for What Works in Education

We review the research on the different programs, products, practices, and policies in education.

Then, by focusing on the results from high-quality research, we try to answer the question “What works in education?”

Our goal is to provide educators with the information they need to make evidence-based decisions.



- Core funding to government research agencies to produce systematic reviews
- Priorities agreed through annual consultation exercise
- Evidence used for funding decisions and guidelines

- Knowledge Centre for Education (Norway)



- Example: school dropouts



Five general lessons from systematic review literature

Most things don't work: the 80% rule

- 80% of businesses fail in the first five years
- Business: Over 13,000 RCTs of new products/strategies conducted by Google and Microsoft, 80- 90% no significant effects.
- Education: 90 interventions evaluated in RCTs by IES - 90% had weak or no positive effects.
- Employment/training: Department of Labor-commissioned RCTs 75% weak or no positive effects

And when things do work, effect sizes are small

Systematic reviews find smds of around 0.05 – less than 0.2 is considered small

E.g. Review of welfare to work programs covering 46 RCTs: employment at follow up 61% treatment vs 58% control; number needed to treat = 33; Share on welfare after one year: 65% treatment versus 72% control

Unrealistic expectations of effects create underpowered studies



The power of bias in economics research

John P.A. Ioannidis^{*}, T.D. Stanley^{**} and Hristos Doucouliagos^{***}

October 2015

Abstract

This paper investigates two critical dimensions of the credibility of empirical economics research: statistical power and bias. We survey 159 empirical economics literatures that collectively draw upon 64,076 estimates of economic parameters reported in more than 6,700 empirical studies. Using this extensive quantitative survey of empirical economics, we calculate statistical power and likely bias. We find that half of the areas of economics research assessed have nearly 90% of their results under-powered. The median statistical power is 18%, or less. A simple weighted average of those reported results that are adequately powered (power $\geq 80\%$) reveals that nearly 80% of the reported effects in these empirical economics literatures are exaggerated: typically by a factor of two and with one-third inflated by a factor of four or more.

Mergers and Acquisitions

80% of M&A bad for bottom line

Cisco tripled profits through 60

M&A

Exploited heterogeneity

Had data on **9,000** cases

Other e.g. Nurse Family
Partnership

More research needed!



IZA DP No. 8193

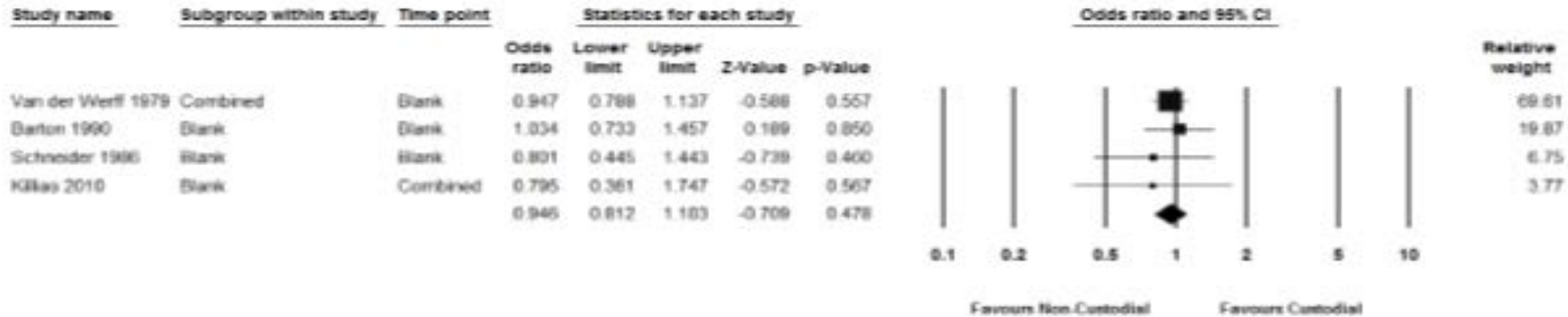
Do Interventions Targeted at Micro-Entrepreneurs and Small and Medium-Sized Firms Create Jobs? A Systematic Review of the Evidence for Low and Middle Income Countries

Michael Grimm

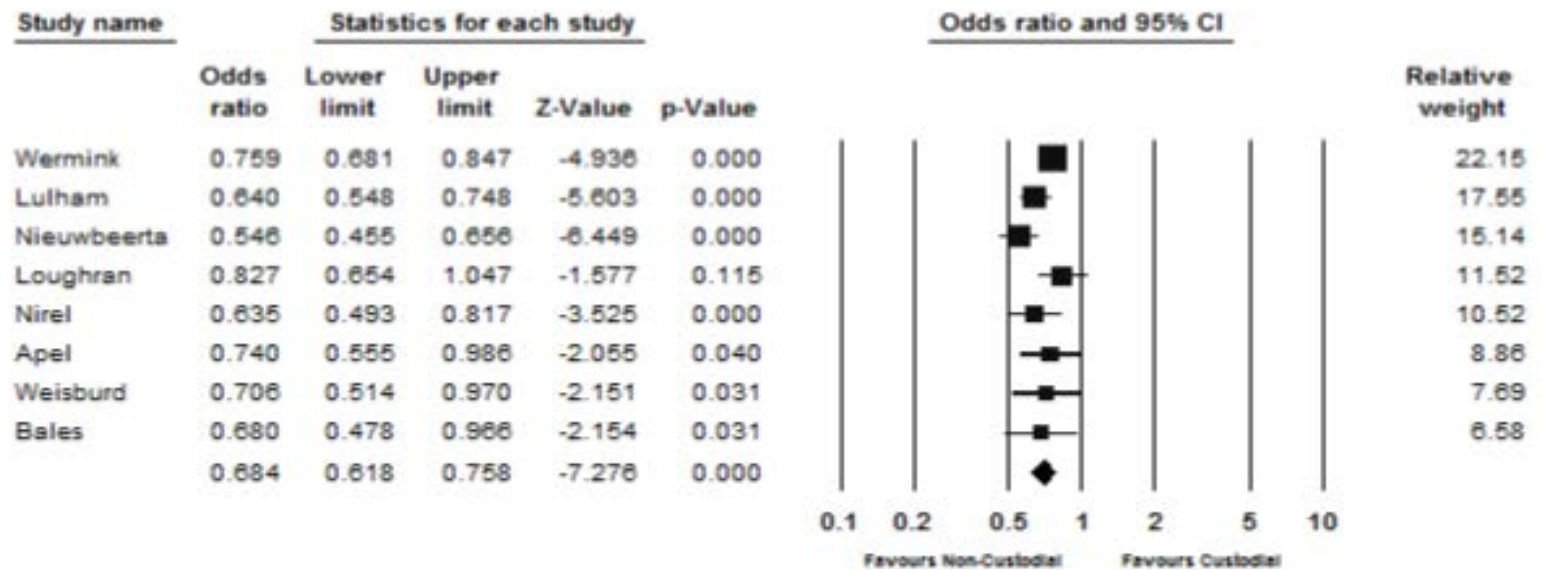
Anna Luisa Paffhausen

“randomized controlled trials find systematically smaller effects than quasi-experimental studies”

RCTs

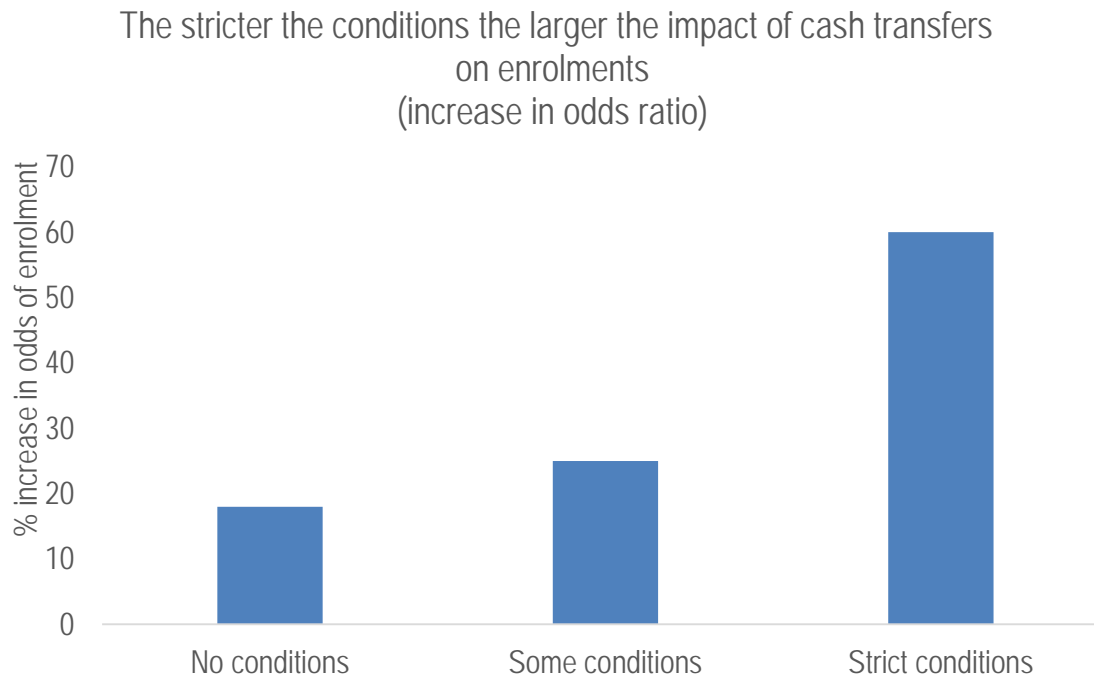


Non-experimental (mostly PSM)



Conditional cash transfers

- Secondary, bigger less frequent payments, and conditions matter



Food supplementation

- Improves nutrition
- Bigger effect if
 - Targeted
 - Supervised



Source: [unintelligible] Honduras [CC BY 2.0 (<http://creativecommons.org/licenses/by/2.0>)], via Wikimedia Commons

Use reviews (or evidence and gap maps) to identify research priorities

Especially long-run effects

Youth and transferable skills evidence and gap map

	Learning and behaviour						Academics, employment, livelihoods and demography					Institutions			Cross cutting themes				
	Individual knowledge	Individual beliefs and attitudes	Observed transferable skills	Social participation and interaction	Health and safety behaviours	Livelihoods and employment behaviours	Criminality	Demography and health	Academic and schooling outcomes	Employment	Wages, income and assets	Other livelihoods measures	Educational institutions	Private sector	Societal and political	Measurement of long-term outcomes	Gender-specific analysis	Cost effectiveness analysis	Early school leavers
Formal education	Teacher training programmes and curriculum reform	●	●	●	●		●	●	●	●	N/A	N/A		N/A	N/A	●			
	Teacher networking and support									N/A	N/A	N/A		N/A	N/A				
	Teacher incentives									N/A	N/A	N/A		N/A	N/A				
	Skills courses at school	●●	●●	●●	●	●●		●								●●	●●	●	●
	Institutional management and capacity building	N/A	N/A								N/A	N/A	N/A		N/A	N/A			
Extra-curricular activities	Student clubs, groups and associations	●	●	●	●	●						●	N/A		●	●	●		
	Career counselling and job fairs							N/A					N/A		N/A				

- More primary studies: doing the same thing as someone else already did is a good thing
- Need more and better reviews
- Scope for methods development in review production

- Use rigorous evidence of effectiveness
- Use high quality reviews
- Assess global evidence, test locally
- Build institutions for use of evidence

Thank you

Visit www.campbellcollaboration.org

Sign up for our newsletter