Evidence-based Medicine & Evidence-based Policy:

The world's most perfectly developed method
& the 79-pound weakling?

Ray Pawson

LEEDS JAN 2017

The Standard Narratives

EBM

Pharmaceutical drug trials represent the gold standard in causal analysis - the supreme opportunity to mount a bias-free RCT.

- Randomisation of individual subjects.
- Blinding of subjects, practitioners and researchers.
- The manufactured treatment is perfectly reproducible.
- Objective clinical outcomes and biological surrogate measures untrammelled by human interpretation.

And so on.

EBP

There is no grand narrative. The field is considered quarrelsome and overloaded with paradigms. EBP is likened to a tree with multiple methodological roots and branches (Alkin, 2004). Hence:

- Experimental evaluation*
- Emancipatory evaluation
- Formative evaluation
- Developmental evaluation
- Goal-free evaluation
- Theory-driven Evaluation
- Realist Evaluation*

And so on. And so on. And so on.

A similar contrast

The 79-pound weakling (EBP)

The world's best developed man (EBM)

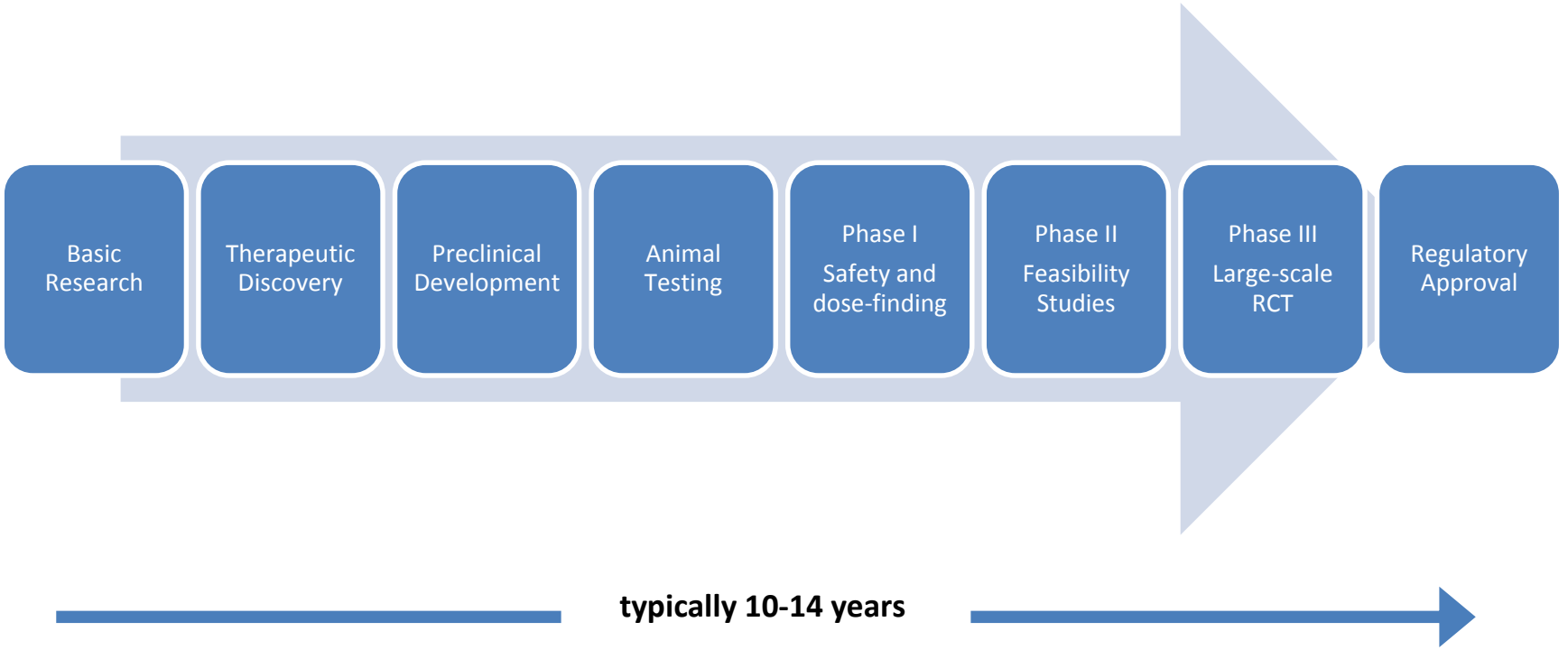
The triumph of the weakling method

Thesis 1: Pharmaceutical RCTs are only possible and meaningful because of prior basic science eliciting, testing and refining the drug's mechanism of action (*RCTs serve an auditory and regulatory function.)*

Thesis 2: Because they lack this grounding in generative explanation, RCTs used in the evaluation of social programmes are *improvised* and never definitive.

Thesis 3: Realist evaluation and synthesis are based on eliciting, testing and refining programme mechanisms (*and thus grab the mantle of science within EBP.)*

The drug development ‘pipeline’



Step 1: Basic Biological Research

Healthy cells co-operate:

i) with each performing its own specialised function

ii) by observing the other rules imposed upon it by the rest of cellular society

iii) by accepting signals from other parts of the same organ

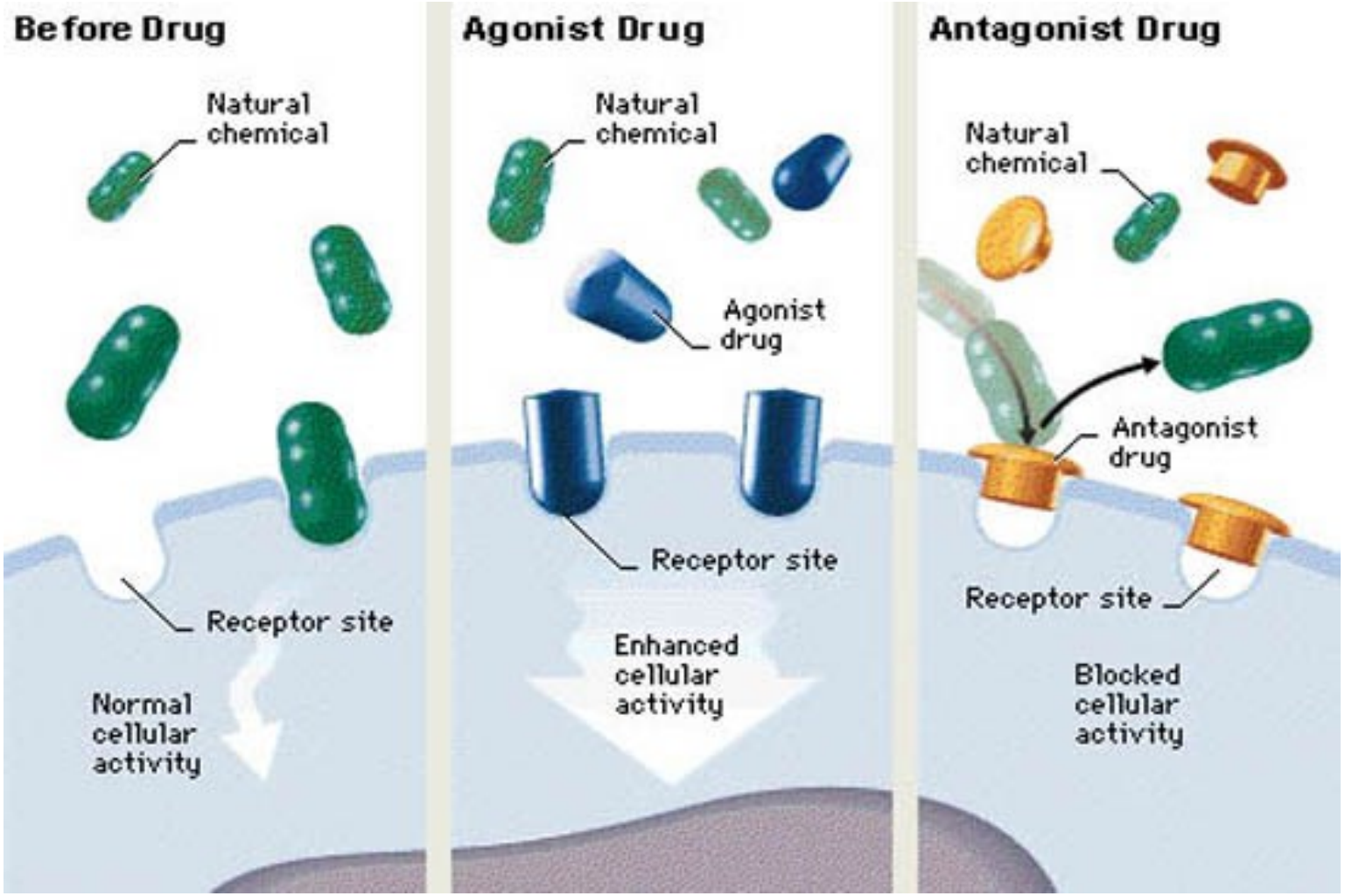
iv) by interpreting those signals correctly

v) by responding to those signals and by transmitting its own signals to other cells accurately.

The first goal is to reach an understanding of derailments in the cellular systems that result in the initiation and maintenance of a disease.

The second objective is to raise hypotheses on potential biological mechanisms of action (MOAs) that could modify these malfunctions sufficiently to halt, circumvent or slow the development of the disease.

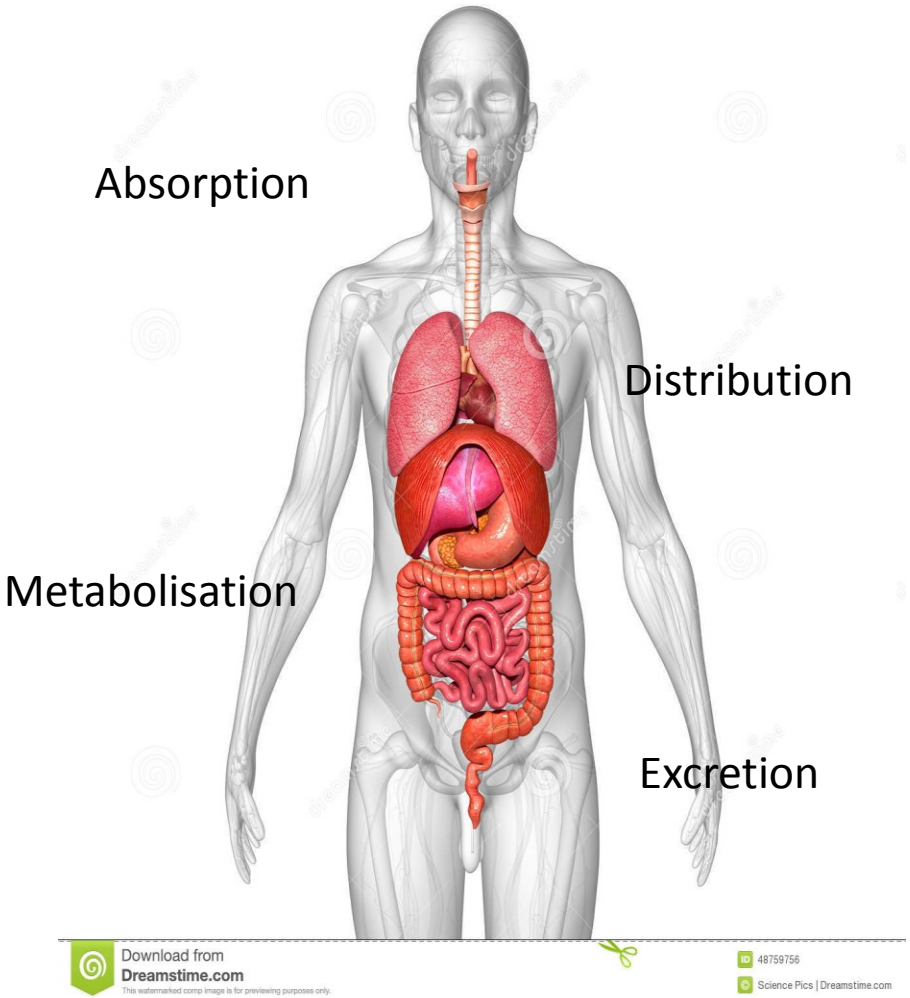
The two fundamental MOAs:
'agonists' (activators) and 'antagonists' (blockers)



Step 3: Preclinical Development

- ‘Prescribers are often unaware that the drug industry spends perhaps a quarter of its research budget for a new drug on pharmaceutics, i.e. devising the right *presentation* to ensure the drug is effectively absorbed, properly distributed, and remains at its site of action long enough to produce an effect. (McGavock, 2011).
- ‘Presentation’ is an equivalent of the implementation pathway in social programmes.
- An entire discipline is devoted to this investigation drug throughput – namely, human pharmacokinetics. It is often defined as the study of *‘what the body does to drugs*’ and contrasted frequently with pharmacodynamics, namely *‘what drugs do to the body’.*

ADME – the key pharmacokinetic parameters



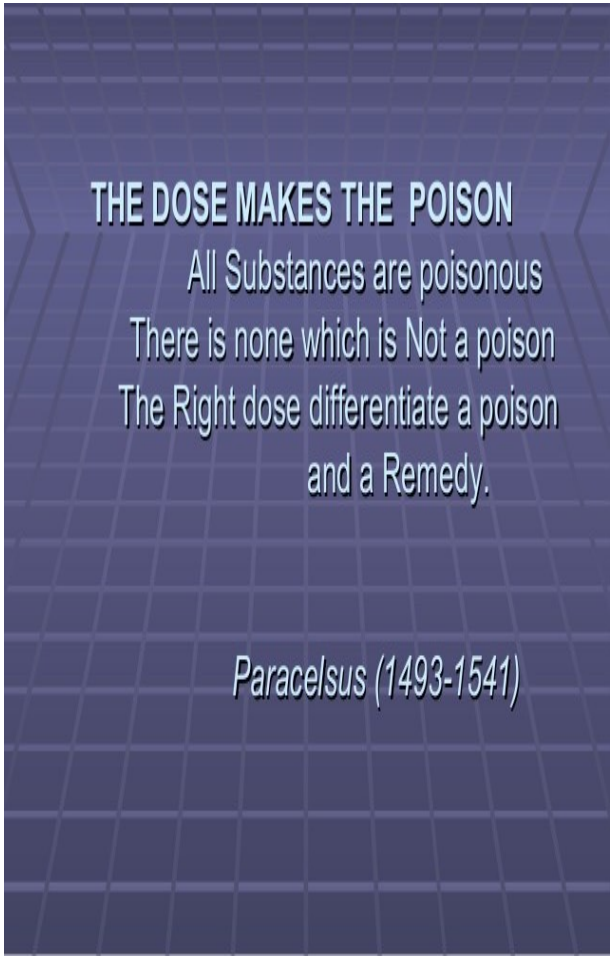
A. Absorption is the release of a drug into the body. Human cells have linings, some of them with pores allowing relatively free passage of drug molecules, some considerably less. Pharmacokinetics finds the chemical means to enhance this process.

D. Distribution refers to the onward dissemination of substances throughout the body. It aims to target the intended amount of drug to its intended site of action in the intended time period. Pharmacokinetics estimates dosages and dosage intervals to achieve a 'steady state'.

M. Drug metabolism is the breakdown of drugs by the body's enzymes. When a drug (or poison) enters the body it tries to modify the chemical structure of the 'foreign compound'. The rate of metabolisation determines the duration and intensity of a drug's action.

E. Excretion is the removal of substances from the body; in this instance to prevent the irreversible accumulation of drugs in the body. Defective liver and kidney functions are unwanted consequences of drug usage. Accurate dosaging is required to prevent aggregations.

Step 5: Safety and Dose-finding Studies



All drug treatments work up to a point and in moderation. The golden rule is to find a dosage that maximises the requisite biologic activity but minimises the toxic risk.

Dose-finding uses a research strategy known 'dose escalation'.

1. The first cohort of patient volunteers is treated at a starting dose that is considered to be safe based on animal toxicological data.
2. Subsequent cohorts are treated in increasing numbers at increasing dosage. Escalation continues until a significant number of patients experience dose-limiting toxicities.
3. The recommended dose for subsequent trials is conventionally defined as the dose level below this toxic limit.

10-14 years later ... the phase III trial

We reach the classic experimental vs. control group RCTs on large patient populations. N.B. *But before its application - EBM has utilised other methods to establish*:

- How and why the treatment works
- How to present the drug so that the body adapts
- The appropriate dosage to balance efficacy and safety
- The eligible patient-sub population which is most likely to respond
- A bank of biomedical tests and measures to gauge disease progression

In other words:

- EBM works to the larger remit of 'what works for whom in what circumstances, in what respects *and* why'
- Causal explanation is established prior to rather than during the RCT
- The RCT is essentially corroborative. It functions to establish a statistically significant net effect in a known population. Size is what matters.

What constitutes 'regulatory proof'?

FDA's MANDATE: To establish "*Substantial Evidence*" of drug safety and effectiveness:

CONGRESS REPORT: 'FDA has interpreted this term to mean that the manufacturer must provide at least two adequate and well controlled Phase III clinical studies, each providing convincing evidence of effectiveness'. (Thaul 2012)

THE RATIONALE:

Simply put, the Agency adopts an *empiricist approach* to the fundamental regulatory questions of safety and effectiveness. Theories about mechanism of action of a drug or disease mechanisms play important parts in drug development and approval, but they are *entirely subsidiary* to the fundamental questions that must be answered in the course of drug approval; namely, is a drug effective, and is it safe in use. These questions can only be answered ... by direct examination of the question in a well designed and conducted clinical experiment.' (Katz, 2004b 316).

End of story? What happens post-approval?

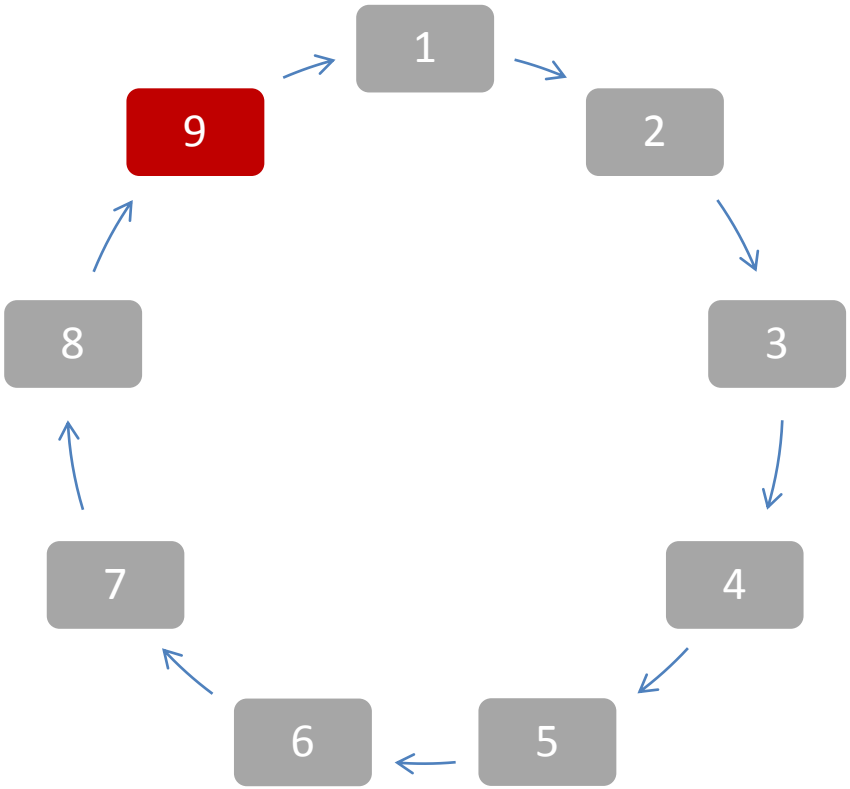
"New chemicals are 'put on probation' for up to two years and labelled with a **black triangle** to ensure prescribers are aware of the need to monitor them carefully. This information helps to build up a broader picture of how the treatment works in the general population". (MHRA)

1. Study participants generally represent a select (and possibly healthier) subset of eventual recipients
2. Participants within a trial may receive increased attention compared with 'real life' care
3. Study drugs will generally be given for shorter durations than in post-approval applications
4. 'Compliance' with treatment may worsen in the general population
5. The limited co-morbidities of patients in the RCT will not represent those of drug recipients outside the trial setting.

End of story? Even the best treatments work ***partially***

- 'Mounting evidence suggests that there is frequently considerable variation in the risk of the outcome of interest *in clinical trial populations*. These differences in risk will often cause clinically important heterogeneity in treatment effects (HTE) across the trial population, such that the balance between treatment risks and benefits may differ substantially between large identifiable patient subgroups' (Kent et al, 2010).
- 'When HTE is present, the modest benefit ascribed to many treatments can be misleading because modest average effects may reflect a mixture of substantial benefits for some, little benefit for many and harm to a few' (Kravitz et al, 2004).

The real gold standard: the everlasting cycle of drug development



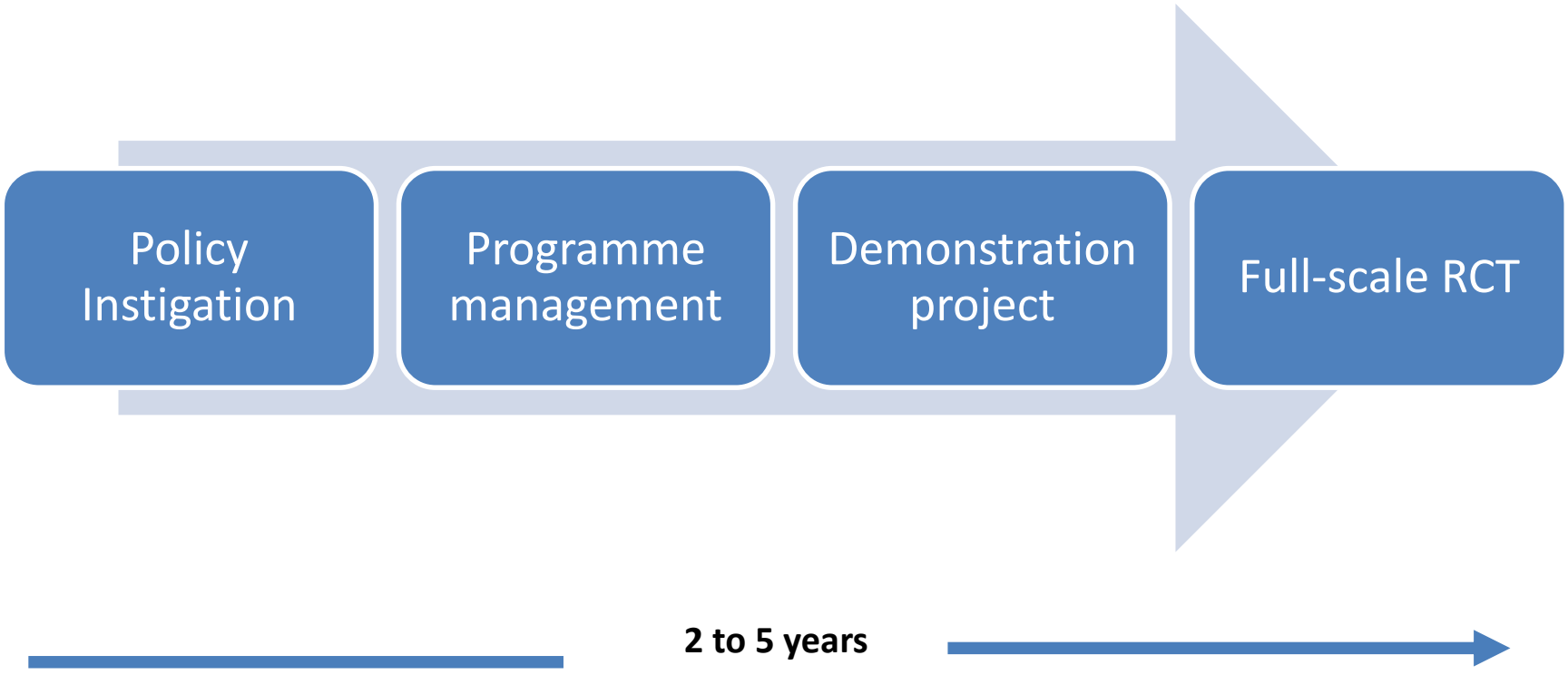
Legend:

1. Basic biological research
2. Therapeutic discovery
3. Preclinical development
4. Animal testing
5. Safety and dose-finding
6. Feasibility trial
7. Large-scale RCT
8. Regulatory approval
9. **Post approval. The discovery of unintended consequences, drug resistance, long-term effects AND sub-group differences, compliance problems, treatment heterogeneity.**
10. The cycle continues P_1 > TS > EE > P_2

A summary (in systems language)

- DRUG RCTs attempt to manufacture a closed system in which a known agent produces an expected outcome in predetermined circumstances.
- BUT ... That closed system may not correspond particularly well to the more open, real-world system in which the drug is prescribed.
- MOREOVER ... The 'closed system' as created in the RCT is only ever partially closed and will generally display residual heterogeneity.
- THUS ... Phase III trails should be understood not so much as 'final arbiters' but as 'way stations' representing current distillations of knowledge.
- CONCLUSION ... The growth of knowledge is achieved by a perpetual programme of identifying , testing and refining system components. An individual RCT will provide some answers, but open other questions, which may be partially explored by further inquiry and further RCTs.
- And so on.

EBP: Pathway to social programme RCT



EBP stage 1: common sense theory

Evaluation research has long recognised its roots in everyday thinking:

> 'Literacy programs for adults, for example, should lead to measurable improvements in reading skills. Lowering speed limits on interstate highways should significantly reduce the number of automobile fatalities and save gasoline. Increasing the price of electricity during the middle of the day should visibly reduce consumption during 'peak-load' hours. Efforts to educate sexually active individuals about 'safe-sex' should ***plainly*** reduce the spread of AIDS'. *Berk and Rossi* (1979)

- **The key word in here is the adjective 'plainly'. Interventions are endorsed initially because the underlying logic appears sound or indeed incontestable.**

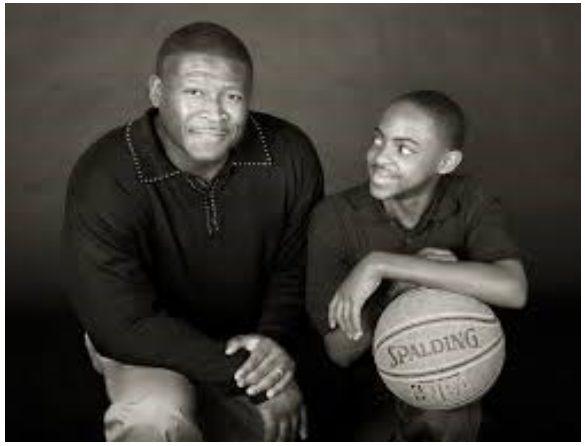
EBP stage 2: management & administration

Interventions do not arrive on the ground directly from the instincts of the policy makers. Their ideas are further articulated, negotiated, funded, staffed and implemented by programme managers and practitioners.

- Some of these preliminaries are informal and rely on 'experience', 'nous', 'custom and practice', 'borrowing', etc,
- Some are more formal and extensive. They might involve *developmental* exercises such as 'scenario-building', 'brainstorming', 'team-building', 'de-snagging', 'double-loop learning', 'quality circles', 'continuous improvement' and so on.

Programme implementation is inherently adaptive. The practitioner's role is not to standardise / reproduce interventions but rather to make them 'work on the day'. In order to conduct the RCT the trialist attempts to *freeze* the underlying processes ... FOR EXAMPLE >>>>

The Big Brother/ Big Sister Mentoring Program



"Taken together the results show that having a big brother or a big sister offers tangible benefits for youth. At the conclusion of the 18 month study we found that Little Brothers and Little Sisters were less likely to have started using drugs or taking alcohol, felt more competent about schoolwork, attended school more, got better grades, and have better relationships with their peers than they would have had if they had not participated in the program".

Grossman and Tierney, 1998

BB/BS: Waiting list 'controls'?

Most social programmes subjects enter voluntarily. A comparison of volunteers and non-volunteers would involve *selection bias*.

SOLUTION: Occasionally programmes are oversubscribed allowing an improvised randomised comparison of 'those who are served' and 'those who wait'.

The unanswered question is whether this methodological device influences behaviour of the controls. What is their frame of mind after volunteering for a programmes and then being told that it is not currently available? *Could significant outcome differences arise from the wait being detrimental rather than the intervention being beneficial?* Do the 'untreated' volunteers stoke up further resentment at the system and begin to go off the rails? Or, do they continue to 'keep their noses clean' in the hope of eventual entry to a cherished scheme?

The answer, alas, is that we do not know.

'Despite the number of studies employing this design, few have analysed intervention trials from the perspective of the waiting list controls rather than the experimental group' (Elliot 2002)

Eligibility and Implementation Fidelity in BB/BS

The BB/BS trial is arguably one of the best policed in these terms:

- certified training for mentors
- mentees are screened for minimal levels of social skills
- mentees required to attend and qualify via induction sessions
- family/guardians produce written agreement on attendance and behaviour
- matching of mentor/mentee pairing in terms of race and gender
- guidelines on number, duration and location of meetings
- ‘age and residential’ restrictions
- exclusion for those with ‘physical and learning difficulties’
- exclusion for those on ‘special programmes’ within the BBBS portfolio.
- exclusion for those under the care of other (statutory) child protection agencies/services

BUT NOTE

- Exclusions are extremely difficult to reproduce precisely because they take their meaning in local, historically-situated provision. It remains far from clear just WHO has been excluded.
- Implementation guidelines concerns the apparatus rather than the mechanisms of mentoring. It cannot prescribe what happens in the eighteen months when mentor and mentee are eyeball to eyeball. Mentoring is a process of human interaction and genuine human interaction is spontaneous and adaptive.
- **EXPERIMENTAL DESIGN IS IMPROVISED, INCOMPLETE AND ARBITRARY**

Conclusion: Fixed and Shifting Foundations

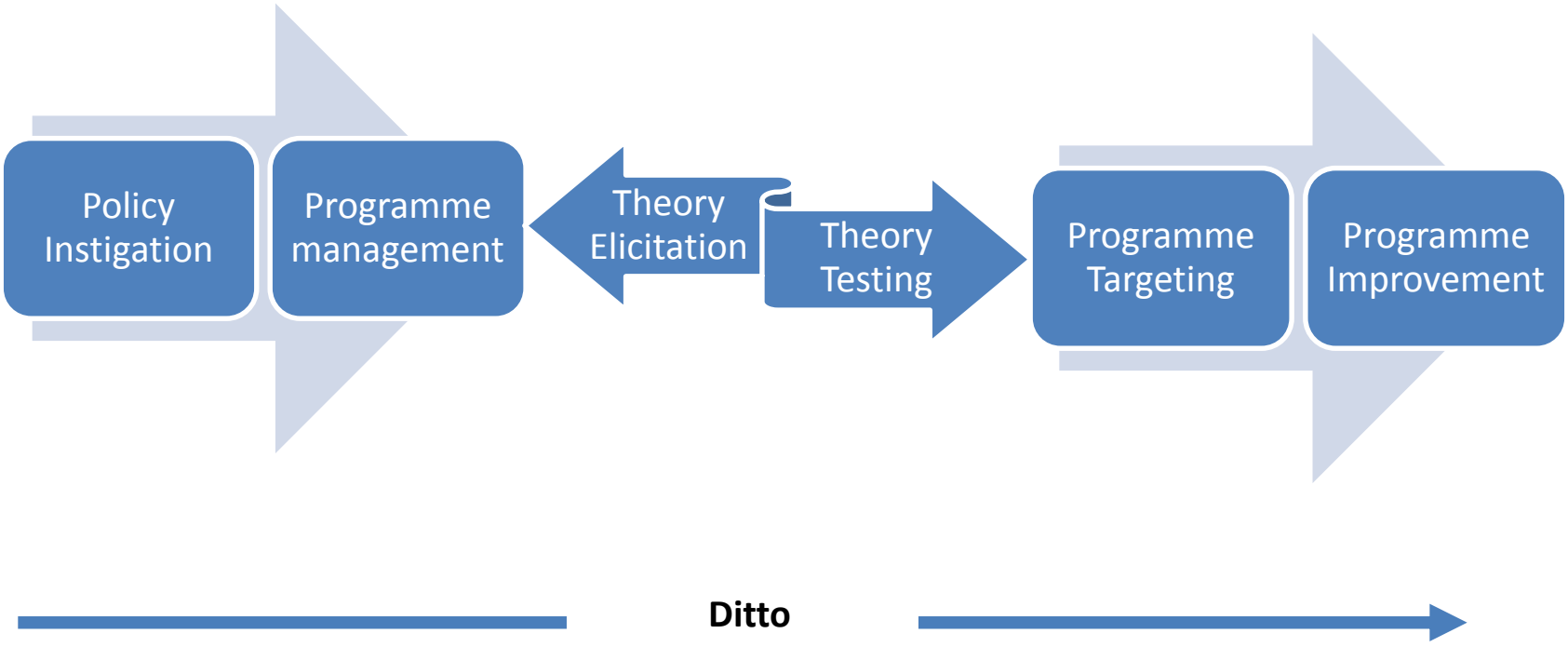
All of the design features of drug RCTs are interrogated and thus fixed in prior research.

The net effects reported in drug RCTs speak to ideal conditions but they are reproducible ideals.

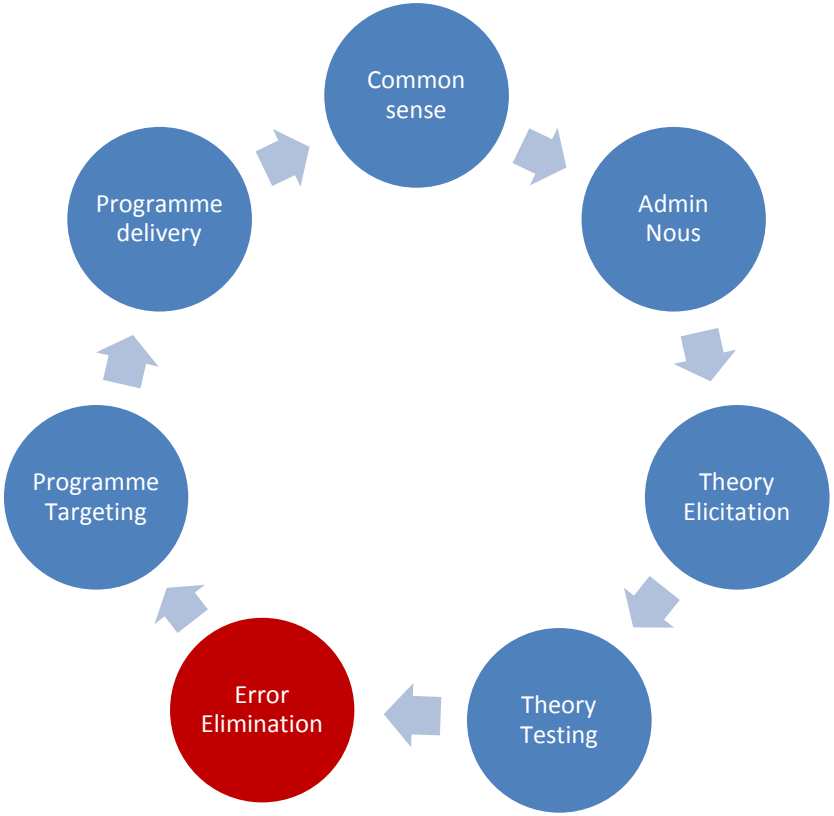
All the design features of a social programme RCTs have to be improvised to capture the current thinking of local administrators.

The net effects reported in social programme RCTs are ad hoc, partial artefacts.

EBP: Realist evaluation research pathway



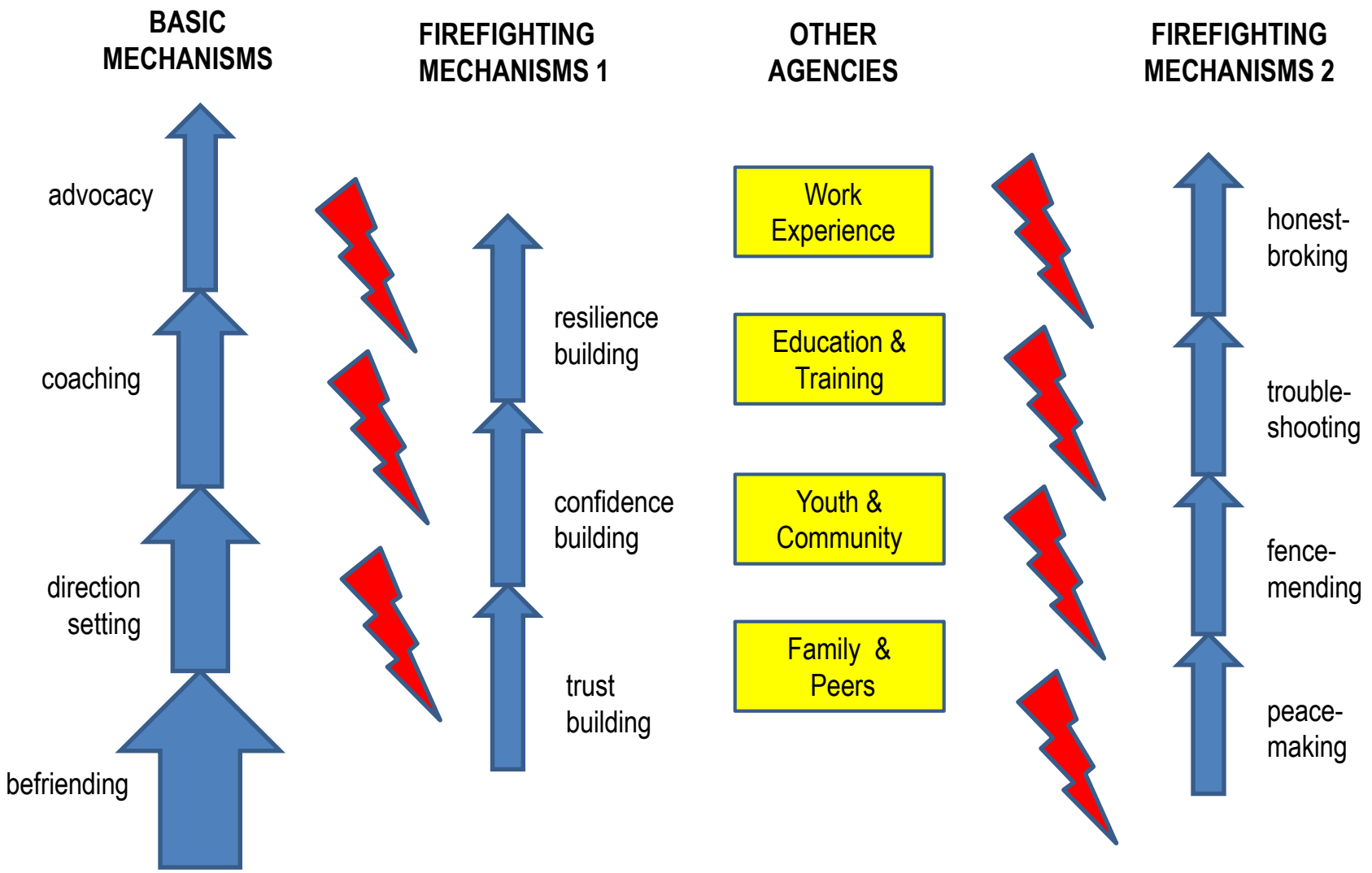
Learning from the real gold standard: the cumulation of inquiry



- Instead of one-off evaluations think of continuing cycles of evaluations. Common sense is (frankly) repetitious. The same policy ideas are endlessly recycled.
- Each evaluation will confirm some expectations about the how, for whom and in what circumstances a programme works but it will also discover unintended consequences, adaptation, heterogeneity, sub-group differences, etc.
- Use the next evaluation to explore these contingencies.

1. Build 'mechanism libraries'. Explore how the programme mechanism is adopted and adapted in different circumstances.
2. Build accounts (pictures) of 'system dynamics' **>>**

Mentoring – Some System Dynamics



Happy Ending - The Unity of Method

Conclusion 2: Evidence-based Physics!

‘Determining the 'true' treatment effect of a given therapy is a bit like determining the 'true' weight of a liter of water. Those who answer that a liter of water weighs a kilogram are either assuming an implicit 'on planet earth, at sea level, at four degrees Celsius' or confusing the intrinsic property of mass with the extrinsic property of weight. *Like weight, treatment effect is an extrinsic property, emerging only through an interaction between the intervention, the patient, and the circumstances in which it is being measured.* Adjust the context and a different effect emerges – just as a liter of water weighs a little over a third of a kilogram on Mars’.

[Kent and Kitsios, 2009]