

The problem of evaluating complex interventions - a recent example

Prepared as an appetizer for the workshop on evaluation of complex interventions of the Danish Research Network for Patient Safety and Quality in Healthcare (FPKS, Forskningsnetværk for patientsikkerhed og kvalitet i sundhedsvæsenet), 21.09.2015.

A recent paper by Baines et al (1, open access), evaluating the effect of patient safety initiatives in Dutch hospitals, serves as a very pertinent illustration of the problems for both policy makers and entrepreneurs of complex interventions in healthcare. Conducting a study of this magnitude and meticulousness has required an enormous effort, and yet the conclusion is equivocal: in a series of three measurements, involving a significant proportion of the Dutch hospitals, a not statistically significant decrease of 30% in preventable adverse events ($p=0.10$) was seen, with no decrease in the total incidence of preventable and non-preventable adverse events (see Figure 1).

Two editorials accompany the paper; they question the use of preventable adverse events as a measure for patient safety (2), or indeed the idea that patient safety can be captured in one measure (3).

Add to this that by design the explanatory power of the study is weak: it is not a controlled study; there are examples to show that a control group can change the conclusion of a time series based study drastically (4). And it is also not based on any kind of program theory that could be corroborated or refuted (5).

Even if these flaws could be fixed, it would still not be straightforward to translate the experiences from the experiment to one's own setting. Would we ever replicate the intervention, exactly as carried out in the study, if that can be defined at all? And would there likely be factors in our context that could affect the effect of the intervention importantly (6)?

Patient safety initiatives are not alone in this predicament. Accreditation or strategies based on setting national goals are other examples of interventions aimed at grand improvement, such as "improving patient safety", "improving healthcare quality", or "improving population health", and they are facing the same challenges:

- It is hard to assess the outcome - what is a valid measure of patient safety or quality of a healthcare system?
- It is challenging to conduct a study that can attribute the outcome to any intervention
- It is hard to translate experiences from one setting to another

What can policy makers do? Obviously, the answer is not "there is no evidence to guide us, so we must continue to do, what we happen to be doing at the moment".

In my opinion, the way forward includes

- accepting that grand improvements may not be measurable by entirely the same methods as specific improvements

- stopping to look for “proof that it works” and turning to studying “which works, how, why, for whom, to what extent and in what context” (6), with an added attention to costs and side effects, weighing likely costs and harm against expected benefits to arrive at a prudent decision
- accepting that the design of a grand improvement intervention does not stop, once it is launched. Evaluation to inform adjustments should be an integral part of any such enterprise.

(1) Baines R, Langelaan M, de Bruijne M et al. How effective are patient safety initiatives? A retrospective patient record review study of changes to patient safety over time. *BMJ Qual Saf* 2015; 24: 561–571.

(2) Shojania KG, Marang-van de Mheen PJ. Temporal trends in patient safety in the Netherlands: reductions in preventable adverse events or the end of adverse events as a useful metric? *BMJ Qual Saf* 2015; 24: 541-544.

(3) Vincent C, Amalberti R. Safety in healthcare is a moving target. *BMJ Qual Saf* 2015; 24: 539-540.

(4) Benning A, Dixon-Woods M, Nwulu U et al. Multiple component patient safety intervention in English hospitals: controlled evaluation of second phase. *BMJ* 2011; 342: d199.

(5) Davidoff F, Dixon-Woods M, Leviton L, et al. Demystifying theory and its use in improvement. *BMJ Qual Saf* 2015; 24: 228-238.

(6) Wears RL. Improvement and evaluation. *BMJ Qual Saf* 2015; 24: 92-94.

Carsten Engel