

The implementation of care.data has been disastrous (*Significance*, April 2014), with the Department of Health failing to learn the lessons of the Summary Care Record. However, the value of care.data is also far from clear. The suggestion that it will deliver “accurate” data is dubious: those who work with healthcare records know they are often erroneous. Challenges to using routinely collected data have long been discussed. [1]

A core idea in statistics, one public and politicians can fail to grasp, is the value of an unbiased sample that is relatively small in proportion to the size of the population. Good data on a sample can be more useful than having poor data on most of the population. Yet policymakers are wedded to techno-utopian dreams of big computer systems and would rather spend money there than on staff costs hiring researchers.

To praise the idea of care.data but criticise its implementation is to miss the possibility that flaws in implementation arise from the same viewpoint that unquestioningly accepts the supposed benefits. [2] As statisticians, we need a research base to determine the utility of routinely collected data versus its costs.

1: Byar DP (1980). Why data bases should not replace randomized clinical trials. *Biometrics*, 36, 337-42.

2: Greenhalgh T, Potts HWW, Wong G, Bark P, Swinglehurst D (2009). Tensions and paradoxes in electronic patient record research: A systematic literature review using the meta-narrative method. *Milbank Quarterly*, 87(4), 729-88.