Evaluating a complex system-wide intervention using the difference in differences method: the Delivering Choice Programme

Jeff Round,[1] Robyn Drake,[1] Edward Kendall,[2] Rachael Addicott,[3] Nicky Agelopoulos,[4] Louise Jones[1]

[1]Marie Curie Palliative Care Research Unit, University College London, London, UK
[2]Health Care Evaluation Group, University College London, London, UK
[3]The King's Fund, London, UK
[4]Marie Curie Cancer Care, London, UK

Correspondence to
Jeff Round,
Marie Curie Palliative Care Research Unit, UCL Mental Health Science Unit, Room 121. b, Charles Bell House, 67–73 Riding House Street, London W1W 7EJ, UK; j.round@ucl.ac.uk

CrossMark

ABSTRACT

Objectives We report the use of difference in differences (DiD) methodology to evaluate a complex, system-wide healthcare intervention. We use the worked example of evaluating the Marie Curie Delivering Choice Programme (DCP) for advanced illness in a large urban healthcare economy.

Methods DiD was selected because a randomised controlled trial was not feasible. The method allows for before and after comparison of changes that occur in an intervention site with a matched control site. This enables analysts to control for the effect of the intervention in the absence of a local control. Any policy, seasonal or other confounding effects over the test period are assumed to have occurred in a balanced way at both sites. Data were obtained from primary care trusts. Outcomes were place of death, inpatient admissions, length of stay and costs.

Results Small changes were identified between pre- and post-DCP outputs in the intervention site. The proportion of home deaths and median cost increased slightly, while the number of admissions per patient and the average length of stay per admission decreased slightly. None of these changes was statistically significant.

Conclusions Effects estimates were limited by small numbers accessing new services and selection bias in sample population and comparator site. In evaluating the effect of a complex healthcare intervention, the choice of analysis method and output measures is crucial. Alternatives to randomised controlled trials may be required for evaluating large scale complex interventions and the DiD approach is suitable, subject to careful selection of measured outputs and control population.

INTRODUCTION

In evaluating the effects of complex healthcare interventions such as large system service reorganisation, the gold standard randomised controlled trial is not always feasible. We report on the use of difference in differences (DiD) analysis to assess the effects of the Marie Curie Delivering Choice Programme (DCP), a service redesign initiative which aimed to enable people to achieve their preferences for end-of-life care.[1 2] The challenge for researchers is how to examine rigorously changes in outputs that arise following such an intervention. Simple estimates of resource use do not improve understanding of change over time and the relation to causality, be it complexity of illness, individual need or preference, or the choices that are available through new models of care. In addition, in the UK there have been a number of recent government-led initiatives that influence ongoing change in end-of-life care such as the Department of Health National Cancer Plan,[3] national service framework strategies for non-cancer conditions,[4 5] the Department of Health End of Life Care Strategy[6] and strategic changes in end-of-life care investment patterns. Against this complex and changing background, simple before and after analysis of resource use by an end-of-life population is not informative. In addition, the numbers of patients participating in and affected by local service redesign initiatives are often small and may be subject to a range of selection biases.

We chose to explore the use of DiD analysis. This method is used to investigate the effects of an intervention over time on two groups, one subject directly to the intervention and one control group at a suitably matched comparator site.[7] This allows researchers to control

for the possibility that both groups have changed over time for reasons unrelated to the intervention. In effect, the analysis isolates the impact of the intervention by removing other known or unknown factors which may have affected either or both groups during the study period. Such changes might include alterations to national policy or demographics—such as an ageing population—that would affect the intervention and comparator populations. If there were no control group then any changes seen in the intervention group might be ascribed to the intervention itself without accounting for the possible effects of other factors.

Our primary aim was to assess whether DiD is a suitable and robust approach for identifying significant changes in measured outputs following the introduction of a system-wide complex intervention, in this case the DCP. We compared four measures of output at intervention and control sites: (1) place of death; and over the last 8 weeks of life, (2) number of hospital admissions, (3) length of each admission and (4) the total cost per patient of providing inpatient care. These measures of outcome were selected according to the UK policy aims of reducing unnecessary inpatient admissions at the end of life and supporting people to be cared for and die in the place of their choice.

The DCP

This service redesign initiative[1][2] aims to improve the care of people who are dying from all advanced illnesses and increase the opportunities for them to be cared for and die in the place of their choice. It is a complex intervention in which a number of components in local health systems are adapted to deliver change. Details of two local service redesigns are published elsewhere.[2] Examples include introduction of discharge liaison services, facilitating transfer to home from hospital using specialised ambulance provision and rapid response teams to deal with emergencies arising in the home. The intervention is developed following theoretical and practical pilot work to assess the current state of service delivery in that locality, followed by consultation with local stakeholders to devise and refine a menu of new services. These processes conform to the guidance proposed by the Medical Research Council (MRC) on the development of complex healthcare interventions[8] and use action research to achieve new models of care. The MRC guidance goes further to suggest ways to assess the effects of an intervention in practice. During DCP, much process data is collected. Knowledge of effects on patient flows and pathways is supplemented by qualitative data collected from patients and carers to increase understanding of their experience of the new service models. Use of DiD may provide a rigorous quantitative method to assess change in measurable outputs over time.

Ethical approval for this analysis was obtained from Oxfordshire Research Ethics Committee B for the release of anonymised hospital activity data for the participating primary care trust (PCT) and from a comparator PCT within a similar site in another city matched on age, ethnicity and socioeconomic factors. In this paper, the intervention and comparator site remain anonymised.

METHODS

Population

All individuals living within the PCT in which the intervention was delivered were able to access the intervention. There was no a priori selection of patients, randomly or otherwise, as the service redesign was applied across the entire local health economy. Data on all individuals who died between April 2006 and March 2008 were obtained from PCTs at the test site and comparator site, and included the corresponding hospital admission data. The control site was chosen to match as closely as possible the test site to minimise the risk of any bias arising in the results owing to differences between populations. Factors that influenced selection of the control site were the population size and characteristics, and the Trust's structural design. Both sites provided anonymous data on admissions and deaths, including place of death.

We considered hospital use for the last 8 weeks of life. This time period was chosen to represent the period during which patients were most likely to require the end-of-life care services provided through the DCP. Data on the last 8 weeks of life were extracted from the data set and costs assigned to patients on the basis of the Healthcare Resource Group (HRG) codes.

Difference in differences

The DiD approach was selected as an approach that allowed for comparisons over time between non-random populations and[7][9] by comparing the treatment group before and after an intervention with a control group from a suitably matched comparator control site that did not receive the intervention.[10] The DiD analysis is a quasi-experimental method used to measure the effects of a treatment or intervention over time. Analysis was undertaken for four outputs (1) place of death, (2) number of admissions, (3) length of stay and (4) costs. Data were obtained from the intervention and control PCTs for all patients who died 1 year prior to introduction of the intervention (April 2006–March 2007) and for 1 year afterwards (April 2007–March 2008).

The analysis is a fixed-effect multiple regression model. We tested a DiD effect between the intervention and control PCTs for patients aged over 18 years registered with a general practitioner (GP) within the PCT between the period April 2006 to March 2008

using inpatient attendance data supplied by the relevant PCTs and national reference cost data. Patients without full records available were excluded from the analysis. Variables (1–3) were calculated for hospital admissions over the last 8 weeks of life. We considered hospital activity only within the last 8 weeks for periods of care which commenced before but overlapped the 8 weeks period and adjusted for age, sex and admission method (elective or non-elective).

We adjusted for time trends and focused at the level of the month of any observation. This required the model to account for correlation between months. For all outputs, fixed effects models with random intercepts were fitted, controlling for autoregression but not heteroskedacity. This model type can be used with time-dependent observations and enables any change due to the intervention to be isolated from changes that occur over time due to other trends in the data (eg, an increase in hospital admissions during colder winter months). A paired Z test was used to test for statistical significance.

Baseline differences between the populations are not normally tested using DiD methods. The DiD approach seeks to measure the effect of a change while controlling for many unknown factors (unobserved heterogeneity). Therefore the comparator is picked based on a priori expectations that it exhibits similar unknown factors as the intervention. We measure the effect of the DCP in the intervention site. The non-intervention site then acts as a proxy controlling for the unobserved heterogeneity.

Analysis at the level of the PCT, HRG, age and sex groups, separately by preintervention and postintervention time period and month allowed for the longterm growth in costs and attendance numbers, but assessed the institutional intervention at the level of the institution at which it occurred. The observational unit was the averaged output variable per month per

age category (18–108 in categories 5 years wide) per PCT. For the analysis of bed days and admissions, HRG was treated as a fixed effect, essentially fitting a separate model to each HRG within each age and sex group and PCT. This is a type of case-mix adjustment, and allowed for the possibility of different intensities of resource use in different HRGs. Cost data were summed across HRGs and analysed at the level of age group, sex and admission method (elective/non-elective) within PCTs. Costs for the entire study were estimated using National Health Service (NHS) national tariffs for 2010–2011.[11] No discounting was required as all costs were estimated using the same base year.

Serial dependence was fitted as a simple autoregressive model, since there were insufficient time points to test seasonal effects. A separate seasonal (winter) term was included in the model to adjust for excessive attendances in winter. The starting model was fitted with all possible interactions. Non-significant terms were removed using backwards stepwise removal, but the key terms of interest (intervention group, study period and the interaction of these two terms) were retained regardless of significance. The interaction of study period and intervention group constitutes the DiD effect, adjusted in this case for the growth over time of attendances within the PCT, and any differences between significant age and sex groups, and assessed separately within each HRG.

RESULTS

Descriptive statistics of intervention and control sites for the four output measures are shown in table 1. Primary DiD grouped analysis results are presented in figures 1–4. We present the difference between the control site and the intervention site at base year, change in both sites over the study period and the difference between the two sites at the end of the study

Table 1 Descriptive statistics of intervention and control sites

Time period	Intervention site		Comparator site		
	Preintervention (April 2006 to March 2007)	Intervention (April 2007 to March 2008)	Preintervention (April '06 to March 2007)	Intervention (April '07 to March '08)	Difference in difference estimator (95% CI)
Total number of deaths	6715	6291	2265	2032	–
Deaths at home (proportion)	1245 (0.185)	1176 (0.187)	368 (0.162)	333 (0.164)	−0.012 (−0.063 to 0.039)
Mean number of admissions (SD)	0.84 (1.03)	0.82 (0.96)	1.16 (1.22)	1.17 (1.16)	0.085 (−0.061 to 0.231)
Mean length of stay per patient in days (SD)	9.16 (12.62)	8.66 (21.24)	11.49 (13.18)	11.92 (13.61)	0.176 (−1.301 to 1.654)
Median cost of acute care admissions (min, max*)	£2592 (£0, £34 595)	£2599 (£0, £34 595)	£2795 (£0, £18 394)	£2852 (£0, £18 304)	£215 (−£188 to £620)
Mean cost of acute care admission (SD)	£2120 (£1112)	£2125 (£1116)	£3285 (£1956)	£3102 (£1534)	–

*The estimated maximum cost is based on the maximum length of stay of 56 days of the analysis period for this study. Where stays exceeded 56 days actual costs will be higher, though they are not considered as they fall outside our study analysis period.
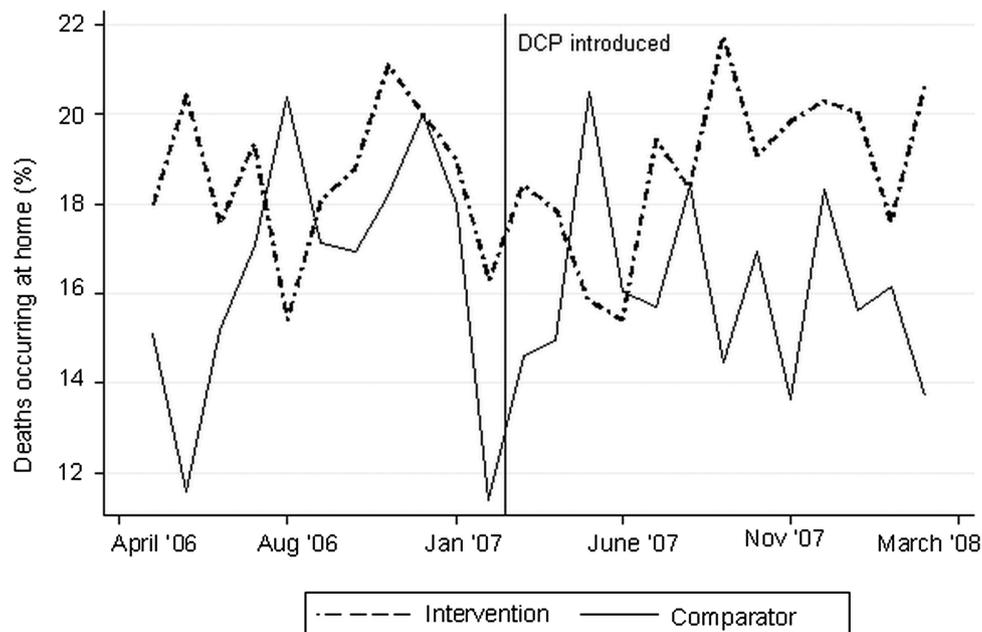
**Figure 1** Percentage of deaths that occur at home. DCP, Delivering Choice Programme.

period. There were small changes following the intervention period in the intervention site. The proportion of home deaths and median cost increased slightly, while the number of admissions per patient and the average length of stay per admission decreased slightly. None of the changes that occurred during the period of analysis were found to be statistically significant when the intervention site was compared with the control site.

Figures 1–4 show unadjusted results for the grouped analysis. Data in the figures are presented as changes over time, with the intervention and control sites plotted on the same graph. The time point where the intervention was introduced is marked by the vertical line.

Place of death
Due to restrictions in the data provided by the intervention group PCT, place of death is classified as either a home death or not a home death (figure 1). An aim of the DCP was to increase the numbers of people who were able to die in the place of their choice, working under the assumption that more people would prefer to die at home than currently do so.[12] There is wide
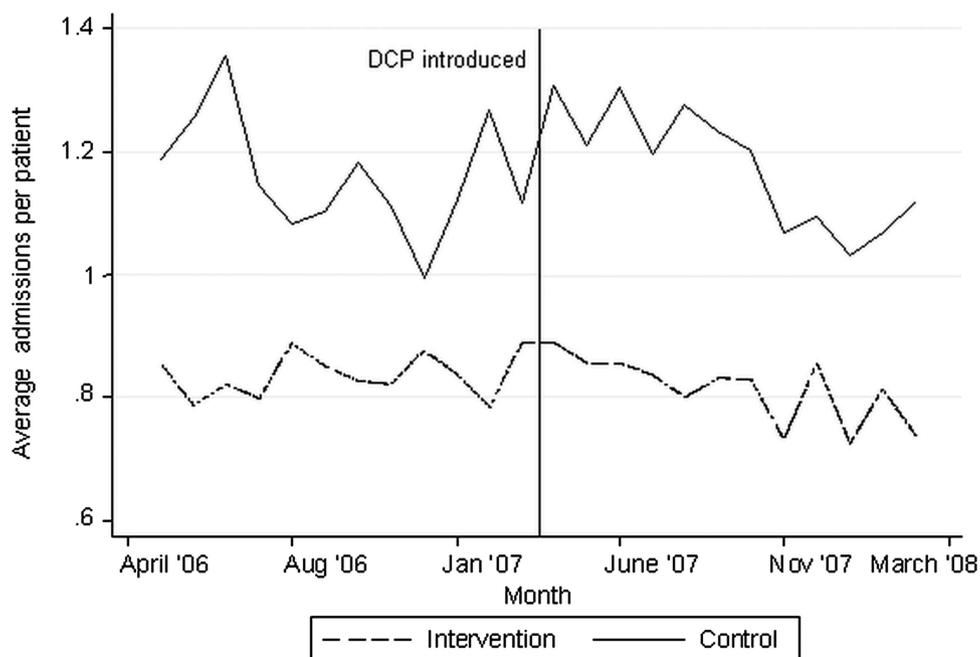
Figure 2 Average number of admissions per patient in the last 8 weeks of life. DCP, Delivering Choice Programme.
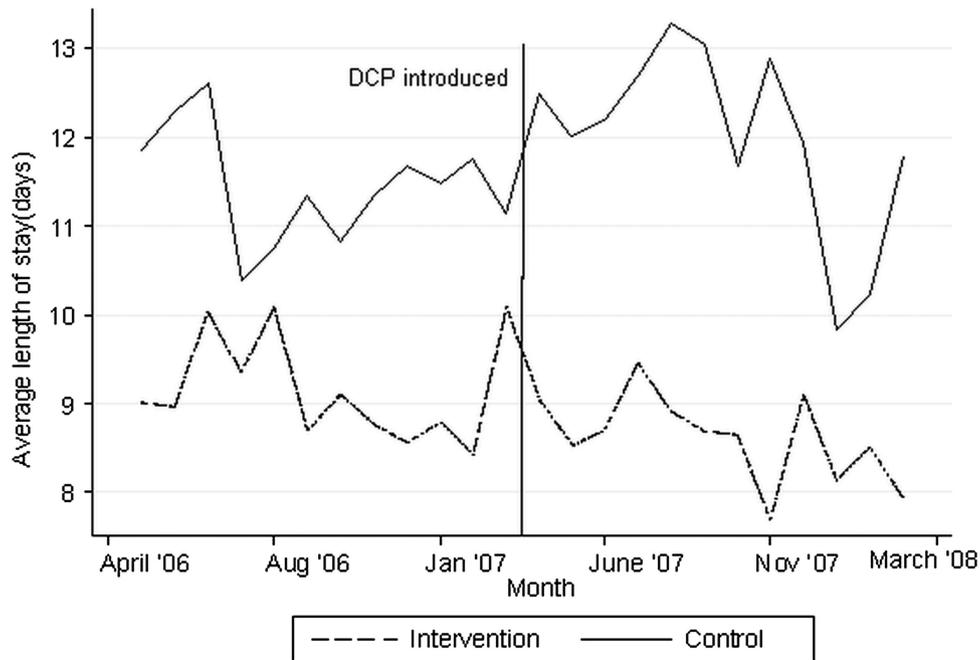
Figure 3 Average length of stay in hospital per patient in the last 8 weeks of life. DCP, Delivering Choice Programme.

variation in the number of patients dying at home in any given month over the analysis time period and there is no clear trend in place of death over time. It cannot be concluded that the DCP increased the number of home deaths in the test site compared with the control.

Average number of admissions
Over time, there was a small but statistically significant reduction in the average number of admissions per

patient at the test site over the study period (figure 2). However, when set against the control site, this difference disappears. It therefore cannot be concluded that the DCP has reduced the number of admissions per patient in the last 8 weeks of their life.

Length of stay
A small but significant decrease in length of stay was observed in the intervention site relative to the
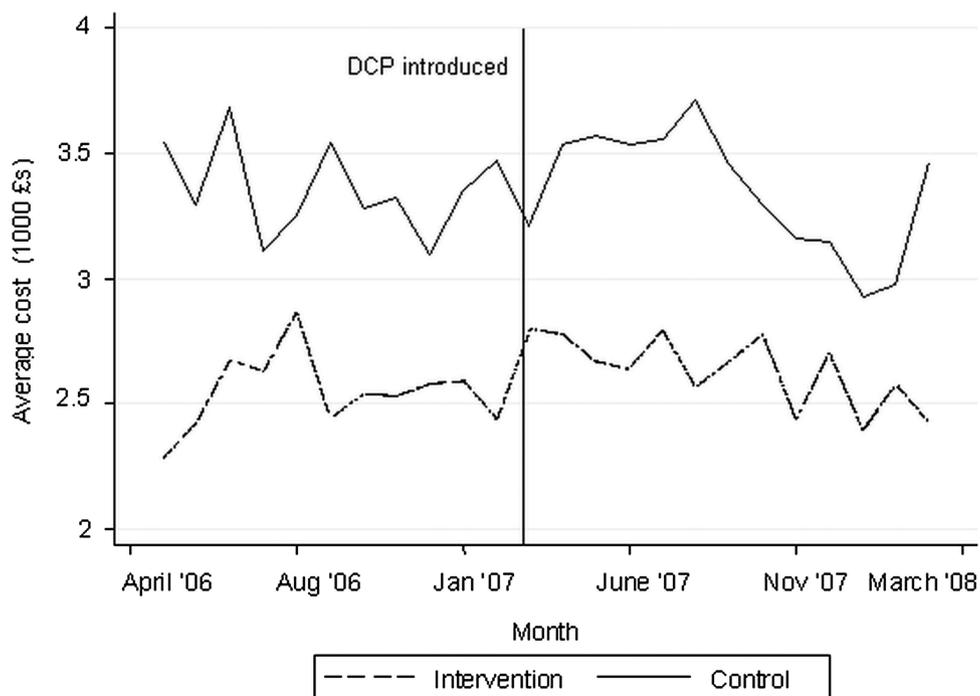
Figure 4 Average cost per admission for patients in the last 8 weeks of life. DCP, Delivering Choice Programme.

control site over the period of the evaluation (figure 3). No statistically significant reduction in length of stay was observed between the two sites. As with home deaths and number of admissions, wide monthly variations are clearly present. It cannot be concluded that the DCP led to a reduction in length of stay.

Costs

The DiD shows no statistically significant change between the sites over the study period (figure 4). There is a wide difference in mean cost per admission between the two sites, though the reasons for this are not clear, nor does such a difference have an impact on whether costs have changed as a result of the DCP. As with the other measures of output, it cannot be concluded that the DCP has led to any changes in the mean cost of care per admission.

DISCUSSION

The results we present from the DiD analysis do not suggest that there was any statistically significant difference between the test site and the comparator site for any of the measured outputs. That is to say, any changes observed in the measured outputs during the study period cannot necessarily be attributed to the DCP. This is not to say that the DCP has no value or that it has not improved the care of the dying in the test site. For example, an evaluation of qualitative evidence collected for two elements of the service redesign showed that quality of care and patient choice at the end of life did improve.[13]

However, the DiD methodology has been demonstrated as useful in previous studies[7][9] and it was an appropriate method for use in our evaluation. Designing a study using such a methodology for evaluation presents difficulties, and choices made during the study design phase have a significant impact on the reliability and accuracy of results. This paper highlights some of these difficulties. Alternative analytical approaches to DiD analysis may be possible. Change point detection methods were considered, as they are suitable for examining whether and when a change has occurred within a data set. Such an approach is often used with time series data.[14] However it was unsuitable in this context as we anticipated large seasonal effects. These seasonal effects would potentially mask broader changes as data for only a single year preintervention and postintervention were available. We are not aware of any additional accurate analytical approaches suitable for our data and study aims.

The choice of output measures is an important starting point for any evaluation. In this instance, our high level quantitative outputs act merely as proxy measures for the true objective of the DCP—to provide high quality care for the dying. These outputs should be considered in conjunction with the qualitative evidence collected to understand the full impact of the programme. Three of the outputs chosen for this analysis, namely place of death, number of admissions to hospital and length of stay in hospital, may closely reflect quality of care while the fourth, costs per hospital admission does not. While these measures are not a true reflection of patient outcomes, they do provide important information about healthcare processes. Although any number of alternative outcome measures may be considered, the three process outcomes measured are often accepted as proxies of clinical outcome in circumstances in palliative care research where patient reported outcomes are not available.[1][2] It is possible to assume a link between the negative aspects of processes and an expected improvement in health outcomes from changes to the frequency with which individuals experience each process. In the absence of more appropriate measures, these three outputs are adequate, though the strength of the relationship between changes in each process measure and outputs is unknown.

Place of death

While it may appear desirable to increase the numbers who die at home based on the belief that more people wish to die at home than so do at present,[15] place of death does not necessarily correlate with quality of death. People may die at home for a variety of reasons. This analysis found that approximately 19% of patients in the intervention site died at home; this compares with survey evidence that as many as 64% of the population express a preference for a home death.[15] By simply examining place of death, important information on the quality of the death may be missed, or worse, assumed.

Hospital admissions

A reduction in hospital admissions and length of stay may indicate a better quality death, but not necessarily better quality care. We have assumed that a reduction in these measures is desirable. However, it would need to be shown that patients who spent less time in hospital achieved better health outcomes and that patients who would be most appropriately cared for in hospital were not denied appropriate care. It is outwith the remit of our analysis to explore these issues; however we recommend that any future evaluation of the DCP or similar programmes should consider such factors when selecting outcome measures at the study design stage.

Costs

Although it is important to understand the cost implications, inpatient costs are not a good indicator of effect for the DCP as any change in mean or median costs cannot be interpreted independently as good or bad. An increase in mean costs may suggest that the programme is effective in moving easier to manage

cases from hospital to home, leaving more complicated, and therefore expensive, patients to remain in hospital. We do not know whether or not those patients left in hospital are there appropriately. Possibly those patients who are easier to care for are dying at home simply because it is easier to provide a home death for them. It might also be that costs are merely shifted from one service provider to another. If fewer people die in hospital, then more are likely to die in other settings and must be supported by appropriate resources, often provided by social services or informally by family and carers. If these costs are not captured in an evaluation, it is difficult to assess whether or not the intervention represents a good use of scarce resources. We must conclude that our evaluation of the DCP does not provide the necessary evidence to estimate whether it has provided value for money for the stakeholders involved.

Selecting outcome measures when studying complex system-wide interventions is difficult. By their nature, such interventions are likely to have an impact on a wide range of factors. Our analytical approach relied on the use of existing data provided by healthcare commissioning organisations and so we were limited by the types of outcomes or outputs we could measure. Prospective study designs that collect data directly from patients would be advantageous, as patient-reported outcomes, and a wide range of clinical data regarding primary, secondary, emergency and social care service usage and experience could be collected. The down side would be that fewer patients can typically be recruited given the costs of research and collecting data directly from people very near the end of life is challenging. A trade-off must be made between study power (through a large cohort) and the details on outcomes it is possible to collect.

Choice of control site

The choice of comparator group in DiD analysis is crucial. The principle is that when changes that occur in the control and intervention group are observed, it cannot be ruled out that they had the same cause—such as a change in government policy or a change in the sociodemographic make-up of the population. But neither can DiD conclusively rule out that the intervention had no effect—correlation is not causation—and if the changes are in the same direction, this may be the result of unrelated factors. What can be said is that compared with the control site, the DCP at the intervention site does not appear to have had any impact on any of the outputs measured. Such an interpretation of the result is corroborated by the descriptive statistics, which do not show any significant change between the preintervention and postintervention outputs.

Data quality

A key complication was the poor quality of the data provided by the PCTs. At the intervention site,

records of where a patient died were supplied separately from records on patient hospital admissions. These data sets were therefore merged using a unique identifier, in this case the NHS number. However, not all patient records could be matched in this way and a number were dropped from the analysis. The total number of records dropped was small (<1%) relative to the number of patient deaths that occurred during the years under analysis and could not have had a significant impact on the results. Even if all records were assumed to be home deaths, these would not be sufficient to show a significant difference in the results.

CONCLUSION

This worked example showed that DiD analysis can be useful in increasing our understanding of the effects of multicomponent, complex interventions when randomised trials are not suitable or feasible but is limited by the quality of data available for analysis. Intervention and comparator populations must be selected carefully to increase confidence that results are an effect of the interventions and do not arise from latent differences in the populations being studied. Careful attention must be paid to selection of meaningful outcome (output) measures. Researchers may wish to consider DiD as a methodologically robust analysis tool when designing evaluations of service redesign programmes. Additional learning may be gained from supplementing quantitative outcomes with qualitative data on patient and carer experiences of care in test and comparator sites.

REFERENCES

1 Agelopoulos N, Tate T. The Marie Curie Delivering Choice Programme. *Eur J Palliat Care* 2009;16:290–4.

2 Agelopoulos N. How improving services can hold the key to end of life care choices. *J Care Serv Manage* 2009;491:16–23.

3 Department of Health. *The NHS cancer plan.* London: Department of Health, 2000.

4 Department of Health. *Coronary heart disease: national service framework for coronary heart disease.* London: Department of Health, 2000.

5 Department of Health. *Living well with dementia: A National Dementia Strategy.* London: Department of Health, 2009.

6 Department of Health. *End of Life Care Strategy—promoting high quality care for all adults at the end of life*, 2008. bit.ly/14Y949U (accessed 8 Aug 2013).

7 Nolan A. Evaluating the impact of eligibility for free care on the use of general practitioner (GP) services: a difference-in-difference matching approach. *Soc Sci Med* 2008;67:1164–72.

8 Medical Research Council. Developing and evaluating complex interventions: new guidance; Medical Research Council, 2008. bit.ly/fEREl3 (accessed 8 Aug 2013).

9 Madden D, Nolan A, Nolan B. GP reimbursement and visiting behaviour in Ireland. *Health Econ* 2005;14:1047–60.

10 Card D, Krueger AB. Minimum wages and employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *Am Econ Rev* 1994;84:772–93.

11 Department of Health. NHS Payment by Results 2010–11 National Tariff Information, 2010. bit.ly/19Qwt2u (accessed 8 Aug 2013).

12 Higginson IJ, Astin P, Dolan S. Where do cancer patients die? Ten-year trends in the place of death of cancer patients in England. *Palliat Med* 1998;12:353–363.

13 Payne S, Seymour J, Ingleton C, *et al*. Independent evaluation of the Marie Curie Cancer Care Delivering Choice Programme, 2009. bit.ly/16Dgr9a (accessed 8 Aug 2013).

14 Kim JY, Amador RB. Detection of change in persistence of a linear timeseries. *J Econometrics* 2000;95:97–116.

15 Marie Curie Cancer Care. Preferences for place of death. http://www.mariecurie.orguk/en-gb/press-media/news-comment/marie-curie-welcomes-palliative-care-funding-review/