S1 Table: Supporting information for Harvest plots

All results as reported in the included papers and the decision process underpinning the Harvest plot

Where outcomes within a category were conflicting, the decision process attached priority as follows:

- Defined primary outcomes in an adequately powered study
- Outcomes that measured impact in the whole eligible population (typically using routine data rather than data from a sub-group who accepted/completed the intervention or were recruited for the evaluation)
- Outcomes which were measured with a validated instrument (as opposed to responses to non-validated questions)
- Outcomes that were clinically as well as statistically significant (e.g. achieved s defined minimum clinically important difference)

Finally, if there were any remaining doubt, the authors' interpretation was considered as providing the context for our decision.

Abbreviations used in this table

Study Design: ITS: Interrupted time series RM: Repeated measures Long: Longitudinal QE: Quasi experimental B&A: Before and after

PAAP: Personalised asthma action plan HbA1c testing: Glycated haemoglobin testing QOF: Quality Outcome Framework

Citation design, size and quality	**Reported outcomes** * indicates the primary outcome (if stated).	**Researcher's interpretation for the Harvest plot**
Beck 2004 QE 1 hospital, 16 paediatric patients who had an incident of DKA. Quality score = 15	**Organisational process outcomes** Programme participation • Participants greater telephone contact (16 crisis management calls vs 0; p=.001)	Organisational processes and disease control both improved. **Illustrated as positive effect**
	Disease control Hospital admissions • Decrease in hospital admissions from intervention group (1 emergency department visit or diabetic ketoacidosis episode vs 5 diabetic ketoacidosis hospitalisations; p=.039)	
	Individual behaviour outcomes	

	Not assessed	
Chien 2012 QE 118 practices, 5557 diabetes patients. Quality score = 13	**Organisational process outcomes** * Hba1C testing Intervention group • HbA1c testing : 2003 = 84% & 2004=85%, 2006 = 86% & 2007 = 91% Control Group • HbA1c testing : 2003 = 83% & 2004=85%, 2006 = 86% & 2007 = 87%	Diabetes care processes and outcomes did not improve significantly **Illustrated as no effect**
	Disease control HbA1c levels Intervention group • HbA1c <9b: 2003 = 36% & 2004 = 35%, 2006 = NA & 2007 = 32% Control group • HbA1c <9b: 2003 = 43% & 2004 = 38%, 2006 = NA & 2007 = 33% The coefficient on intervention*post (difference in difference) was reported as not significant in these results, no p value provided.)	
	Individual behaviour outcomes Not assessed	

Conrad 2013 QE 19 medical groups, 21,365 patients Quality score = 10	**Organisational process outcomes** Quality Incentive Programme • regression results : 2003-04= -0.001 & 2005-07 = -0.04 Quality scorecard • regression results: 2003-04 = -0.019 & 2005-07 = -0.004	no significant positive effect on general clinical quality QIP 05-07 statistically significant negative result showing a reduction in quality **Illustrated as negative effect**
	Disease control Not assessed	
	Individual behaviour outcomes Not assessed	
Fagan 2010 QE 20,943 65+ year old patients. Quality score = 16	**Organisational process outcomes** *HbA1c testing • Intervention Group – Odds ratio = 1.66; 95%CI (1.14, 2.43) • Comparison Group – Odds ratio = 3.76; 95%CI (3.42, 4.13) • Intervention relative to Comparison – Odds ratio = 0.44; 95%CI (0.30, 0.65)	**Illustrated as no effect**
	Disease control Not assessed	
	Individual behaviour outcomes Not assessed	
Gulliford 2007 Long,	**Organisational process outcomes** HbA1c testing • HbA1c recorded in year: 2000 = 60, 2001 = 72, 2002 = 80, 2003 = 78, 2005 = 95	▪Increase in tests performance (until 2002)

26 general practices, 2099 patients. Quality score = 17	**Disease control** HbA1c levels • HbA1c ≤7.4%: 2000 = 22, 2001 = 32, 2002 = 37, 2003 = 38, 2005 = 57 • HbA1c ≤10%: 2000 = 52, 2001 = 64, 2002 = 70, 2003 = 72, 2005 = 89 (No further statistics provided on these outcomes)	▪ increase in HbAc1 target of <7.4% & HbA1c <10% **Illustrated as positive effect**
	Individual behaviour outcomes Not assessed	
Kontopantelis 2012 ITS 148 practices, 23,920 patients. Quality score = 17	**Organisational process outcomes** HbA1c testing • HbA1c recorded in year (SD): 2000/1 = 71.1 (45.3), 2001/2 = 77.9 (41.5), 2002/3 = 82.8 (37.7), 2003/4 = 89.2 (31.1), 2004/5 = 93.0 (25.5), 2005/6 = 93.7 (24.3), 2006/7 = 93.5 (24.6)	▪Increase in tests performance (until 2005/6) ▪Increase in HbAc1 target of ≤7.4% ▪Increase in HbAc1 target of ≤10% (until 2004/5). **Illustrated as positive effect**
	Disease control HbA1c levels • HbA1c ≤7.4% (SD): 2000/1 = 45.5 (49.8), 2001/2 = 48.4 (50.0), 2002/3 = 50.2 (50.0), 2003/4 = 52.2 (50.0), 2004/5 = 55.6 (49.7), 2005/6 = 56.4 (49.6), 2006/7 = 59.3 (49.1) • HbA1c ≤10% (SD): 2000/1 = 88.5 (31.9), 2001/2 = 90.4 (29.4), 2002/3 = 90.8 (28.9), 2003/4 = 91.8 (27.4), 2004/5 = 92.6 (26.3), 2005/6 = 92.5 (26.3), 2006/7 = 92.7 (26.0)	
	Individual behaviour outcomes Not assessed	
LeBlanc 2017 Long 583 physicians, 83,580 adult patients	**Organisational Process outcomes** HbA1c testing • ≤2 HbA1c tests per year: univariate model OR = 1.16 (p<0.0001); 99%CI (1.11 1.20). Multivariate model OR = 1.23 (p<0.0001); 99%CI (1.18, 1.28)	**Illustrated as positive effect**

Quality score = 13		
	Disease control HbA1c levels • All patients: univariate model OR = 0.00; 99%CI (-0.03, 0.02). Multivariate model OR = -0.01; 99%CI (-0.03, 0.02 • HbA1C 6.5% to 7.0%: univariate model OR = -0.02 (p<0.0001); 99%CI (-0.04, 0.01). Multivariate model OR = -0.02 (p<0.0001); 99%CI (-0.04, 0.01). • HbA1C 7.1% to 8.9%: univariate model OR = 0.03; 99%CI (-0.01, 0.08). Multivariate model OR = 0.02; 99%CI (-0.02, 0.06). • HbA1C ≥9%: univariate model OR = 0.04; 99%CI (-0.06, 0.15). Multivariate model OR = 0.00; 99%CI (-0.10, 0.10)	▪No statistically significant changes in mean HbA1c levels **Illustrated as no effect**
	Individual behaviour outcomes Not assessed	
Mandel 2007 RM 44 paediatric practices 13 380 children. Quality score = 16	**Organisational process outcomes** Asthma action plan ownership. • 19 (70%) achieved the 80% threshold for the PAAP. • The cumulative percentage of the network all-payer asthma population receiving "perfect care" increased from 4% to 88%, with 18 of 44 practices (41%) achieving a perfect care percentage of 95% or greater (no statistics reported)	**Illustrated as positive effect**
	Disease control Not assessed	
	Individual behaviour outcomes Not assessed	
Pape 2015	**Organisational process outcomes**	

B&A 1 primary care trust, 6,142 patients. Quality score = 18	Not assessed	
	Disease control HbA1c levels HbA1c of ≤8% : • Exception reporting Baseline = 0.085, Secular trend effect = 0.001 (p = 0.910), QOF+ baseline = 0.060 (p=0.018) • Controlled Patients Baseline = 0.725, Secular trend effect = 0.015 (p=0.005), QOF+ baseline = 0.002 (p=0.968) HbA1c of ≤9% : • Exception reporting Baseline = 0.062, Secular trend effect = 0.001 (p = 0.891), QOF+ baseline = 0.043 (p=0.049) • Controlled Patients Baseline = 0.822, Secular trend effect = 0.015 (p=0.002), QOF+ baseline = 0.003 (p=0.934)	▪No statistically significant improvements in mean HbA1c levels ▪ Increase can be attributed to increase in exception reporting since intro of QOF+ **Illustrated as no effect**
	Individual behaviour outcomes Not assessed	
Rosenthal 2005 QE 205 physician groups, 1,174,294 patients. Quality score = 18	**Organisational process outcomes** HbA1c testing Intervention group • Pre Quality Incentive Programme - 62.0%, after QIP 64.1%, • Difference (Post-pre), 2.1% (SE 1.0) • P value .02 Control group • Pre Quality Incentive Programme - 62.0%, after QIP 64.1%, • Difference (Post-pre), 2.1% (SE 1.0) • P value .02	▪Slight improvement but not significantly different from comparison group **Illustrated as no effect**

	Disease control Not assessed	
	Individual behaviour outcomes Not assessed	
Vamos 2011 ITS, 422 general practices 154 945 patients. Quality score = 15	**Organisational process outcomes** • HbA1c measured (95% CI)- 1997, by quintile: 32.8 (31.8-33.7), 31.2 (30.2-32.0), 34.6 (33.7-35.6), 32.2 (31.2-33.0), 37.7 (36.7-38.7) • HbA1c measured (95% CI)- 2005, by quintile: 74.0 (73.4-74.6), 76.4 (75.8-76.9), 77.3 (76.7-77.8), 73.9 (73.3-74.5), 76.2 (75.6-76.8)	
	Disease control HbA1c mean levels • HbA1c mean (95% CI)- 1997, by quintile, 7.6 (7.5-7.7), 7.6 (7.5-7.7), 7.7 (7.6-7.8), 7.5 (7.4-7.6), 8.2 (8.1-8.3) • HbA1c mean (95% CI)- 2005, by quintile, 7.5 (7.5-7.5), 7.4 (7.4-7.4), 7.4 (7.4-7.4), 7.5 (7.4-7.5), 7.4 (7.4-7.5) • Baseline proportion of patients meeting HbA1c <7.0% in 1997: 35.3, 95% CI = 31.0-39.7, p<0.05 • Annual change before introduction of P4P: 2.0,95% CI = 1.3-2.7, p<0.05 • Annual change in the year P$P introduced: 0.8, 95% CI = -1.8-3.5, • Annual change after P4P was introduced: -2.2, 95% CI = -4.0- -0.4, p<0.01	▪ No significant additional improvement **Illustrated as no effect**
	Individual behaviour outcomes Not assessed	
Young 2007	**Organisational process outcomes** HbA1c testing	▪ No difference between post & pre-intervention trends.

ITS, 334 Primary care physicians, unknown number of patients. Quality score = 16	• Adherence rates: mean (SD) pre-intervention: 1999 = 0.56 (0.23), 2000 = 0.57 (0.19), 2001 = 0.59 (0.17) • Adherence rates: mean (SD) post-intervention: 2002 = 0.62 (0.17), 2003 = 0.61 (0.18), 2004 = 0.63 (0.18) • Change in adherence rate: 2000-2001 = 0.018; 2001-2002= 0.026, p<.05 • Difference in rate of change (2001-2000)(vs (2002-2004) = 0.009 (no p value given)	▪ Overall increase in performance result of secular trends **Illustrated as no effect**
	Disease control Not assessed	
	Individual behaviour outcomes Not assessed	