The TrueBlue model of collaborative care using practice nurses as case managers for depression alongside diabetes or heart disease

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-002171
Article Type:	Research
Date Submitted by the Author:	01-Oct-2012
Complete List of Authors:	Morgan, Mark; Flinders and Deakin Universities, Greater Green Triangle University Department of Rural Health Coates, Michael; Flinders and Deakin Universities, Greater Green Triangle University Department of Rural Health Dunbar, James; Flinders and Deakin Universities, Greater Green Triangle University Department of Rural Health Reddy, Prasuna; University of Newcastle, School of Medicine and Public Health Schlicht, Kate; Flinders and Deakin Universities, Greater Green Triangle University Department of Rural Health Fuller, Jeff; Flinders University, School of Nursing and Midwifery
Primary Subject Heading:	Public health
Secondary Subject Heading:	Diabetes and endocrinology
Keywords:	Coronary heart disease < CARDIOLOGY, General diabetes < DIABETES & ENDOCRINOLOGY, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, MENTAL HEALTH, PUBLIC HEALTH

Abstract:

Objectives: Determine the effectiveness of collaborative-care in reducing depression in primary-care patients with diabetes or heart disease using practice nurses as case managers.

Design: A two-arm open randomised-cluster trial with wait-list control for 6 months. The intervention was followed over 12 months.

Setting: Eleven Australian general practices, five randomly allocated to the intervention and six to the control.

Participants: 404 primary-care patients (211 intervention, 193 control) with depression and type 2 diabetes, coronary heart disease or both.

Intervention: The practice nurse acted as case manager identifying depression, reviewing pathology results, lifestyle risk factors, and patient goals and priorities. Usual care continued in the controls.

Main outcome measure: A five-point reduction in depression scores for patients with moderate to severe depression. Secondary outcome was improvements in physiological measures.

Results: Mean depression scores after six months of intervention for patients with moderate to severe depression decreased by 5.7±1.3 compared with 4.3±1.2 in the control, a significant (p=0.012) difference. (The plus-minus is the 95% confidence range.) Intervention practices demonstrated adherence to treatment guidelines and intensification of treatment for depression, where exercise increased by 19%, referrals to exercise programs by 16%, referrals to mental health workers (MHWs) by 7%, and visits to MHWs by 17%. Control-practice exercise did not change, referrals to exercise programs dropped by 5%, and visits to MHWs by 3%. Only referrals to MHW increased, by 12%. Intervention improvements were sustained over 12 months, with significant (p=0.015) decrease in 10-year cardiovascular-disease risk from 27.4±3.4% to 24.8±3.8%. A review of patients indicated the study's safety protocols were followed.

Conclusion: TrueBlue participants showed significantly improved depression and treatment intensification, sustained over 12-months of intervention, and reduced 10-year CVD risk. Collaborative care using practice nurses appears to be an effective primary-care intervention.

Trial registration: Australia and New Zealand Clinical Trials Registry ACTRN12609000333213.

Article summary

Article focus

- To determine the effectiveness of a collaborative-care model to reduce depression in primary-care patients with diabetes or heart disease.

- To determine the effectiveness of using practice nurses as case managers of patients with depression and diabetes, heart disease or both.

Key messages

- The TrueBlue model of collaborative care can be introduced within the general practice workforce with practice nurses taking on the role of case manager.

- Practice nurses can improve care of depression in patients with diabetes or heart disease, leading to better outcomes and reduced 10-year CVD risk.

- The care of patients using the TrueBlue model is closer to "best practice" guidelines, with substantially better levels of adherence to guideline-recommended checks than occur in usual care.

Strengths and limitations of this study

- The TrueBlue model of collaborative care overcomes many of the difficulties in implementing a guideline for the treatment of co-morbid depression.

- The study's purpose-designed care plan gives patients and their carers, allied health professionals, specialists and general practitioners ready access to patient details enabling them to see at glance where improve clinical care may be needed.

- Clinics were able to recover the costs of the collaborative care through Australian Medicare rebates.

- The study could only be run in practices that had a practice nurse on staff and had access to clinical software capable of generating a disease registry from which participants could be selected.

- Differences between TrueBlue- and control-practice outcomes may have been reduced by patients completing the PHQ-9 depression questionnaire and reading the project description, and by GPs being made aware of individual PHQ-9 results so that they could take action where warranted.

Introduction

Managing diabetes and heart disease has been highlighted as one of the global "grand challenges in chronic non-communicable diseases"[1] because the prevalence of these two preventable diseases is increasing[2]. Along with depression, they have been identified as health priority areas in many countries. A vicious cycle exists between depression and these chronic diseases, with each being a

risk factor for the other[3]. Higher mortality has been demonstrated for people with depression and type 2 diabetes (T2DM) or coronary heart disease (CHD) beyond that due to the separate diseases alone[4]. For patients with depression and T2DM or CHD or both, there are increased risks of adverse outcomes[5] but this co-morbid depression is often missed in primary care[6]. Consequently, the identification of depression has now been incorporated in many heart disease guidelines as one of the requirements for optimal management. Meeting these challenges will require an innovative use of the existing general practice workforce and such a re-orientation of resources has been identified as one of the grand challenges[1].

Collaborative care is a system that has been shown to be more effective for chronic disease management than standard care[7]. It includes a reorientation of the medical workforce through new or adjusted roles for team members, particularly using practice nurses as the identified case manager[8,9]. It also includes the use of evidence based guidelines, systematic screening and monitoring of risk factors, timetabled recall visits, information support for the clinician, enhanced patient self-management, a means of effective communication between all members of the care team and audit information for the practice. Since self-care for diabetes has been found to be suboptimal across a range of self-managed activities, particularly for patients with depression, a collaborative care model may be able to achieve better quality of care through the case manager monitoring patient progress[10,11].

Evaluation of a change in the way general practice clinics look after patients requires complex intervention methodology[12] beyond single interventions such as introduction of a guideline with financial incentives[13]. This methodology began with a search for potential models of care (step I), and led to adopting the University of Washington's successful IMPACT model of Collaborative Care for depression[14,15]. In the exploratory trial (step II), our pilot project[16] adapted IMPACT by training practice nurses as case managers. Practice nurses were trained to screen for depression using a patient self-report measure, the nine-item Patient Health Questionnaire (PHQ-9)[17], as part of comprehensive chronic disease management. They were also trained to use a protocol for care management based on the depression scores. The depression screening and management was embedded in routine visits for patients with diabetes or CHD. The pilot demonstrated that it was feasible to detect, monitor and treat depression in routine general practice alongside the usual biophysical measures, and identified moderate to severe depression in 34% of participants. The TrueBlue study was a randomised cluster trial (step III) that built on and extended the pilot. It investigated whether a collaborative care model (the intervention) is better than usual care (the control) for the management of patients with depression and T2DM, CHD or both in Australian

general practice. It was designed to fit into normal clinic operations, making use of practice nurses and medical software, and able to be funded by existing Australian Medicare rebates.

Methods/design

Study design

The design and methodology of the study have been described in detail elsewhere[18]. The study commenced in 2009 and was undertaken in two phases. The first phase was a cluster randomised intervention trial in which general practices were randomly allocated to either an intervention group in which nurse-led collaborative care was undertaken or to a wait-list control group in which usual care led by the general practitioner (GP) was continued. At six months, the TrueBlue training was provided to the control practices. The key aims of the first phase were to determine whether participants with moderate to severe depression in the intervention group showed at least a five-point reduction from the baseline depression scores after six months of intervention and whether this reduction was significantly better than in the control group. A five-point reduction reflects a clinically-relevant change in individuals receiving depression treatment[19]. The secondary outcome was to determine whether the intervention also led to improvements in the patients' physiological measures. The second phase followed the intervention group for an additional six months to determine how the collaborative care model affected health outcomes over a twelve-month period.

Sample size

The sample size calculation was based on detecting a 50% reduction in depression score at the 0.05 significance level with 80% power. Detecting a 50% reduction is more stringent than detecting a five-point reduction and provided some additional buffering. Using depression scores extracted from an earlier study[11], the calculation indicated that 237 patients would be required in each group. An intra-cluster correlation of 0.04 was used (S. K. Lo, pers. comm.), with a recruitment target of 50 patients per clinic. (Fifty patients were chosen so that clinics could budget for a nurse's time to carry out the intervention with four patients each week over the 3-month cycle of care.) To allow for a 50% dropout, the study required 450 patients from nine clinics in the intervention group and the same in the control group.

Practice recruitment

Practices selected from metropolitan, regional and rural areas were invited to participate in the study on the basis of having a practice nurse to provide the collaborative care and being able to identify eligible patients, those with CHD or T2DM or both, from their registries. The unit of randomisation was the clinic. Clinics agreeing to participate were allocated by a random number generator to either the intervention or control arm of the study. Five practices (3 regional, 2

metropolitan) in the intervention group and six (2 regional, 4 metropolitan) in the control group completed the study. Two regional intervention clinics withdrew whilst first-visit data were being collected when their TrueBlue-trained practice nurses left the clinic, but some (13) patients from one of these clinics did complete the study and data were collected for them.

Patient selection

Eligible patients were sent a postal survey that included a consent form and were asked to complete and return the enclosed PHQ-9 questionnaire, a self-report measure of depression.[17] The PHQ-9 has nine items, each scored from 0 (no problems) to 3 (problems nearly every day). The sum of the scores of the nine items will lie in one of five depression categories: none (0–4), mild (5–9), moderate (10–14), moderately severe (15–19) and severe (20–27). (While it is known that responses to some of the PHQ-9 items may overlap with diabetes symptoms[20], our pilot demonstrated that nurses and patients preferred using the PHQ-9 because the patient's response to each of its items became the basis for the problem solving and goal setting activities that were part of TrueBlue.) Patients with scores of five or above, indicating some form of depression, were invited to participate in the study. A maximum of 50 patients per practice were invited. Patients in residential care or under 18 years of age were not eligible. Figure 1 presents the CONSORT diagram of the patient-recruitment process.

Patient safety

Participation in the intervention included a series of patient visits to their practice nurse (PN) and usual GP every three months over a 12-month period. Patients in the control group continued with their "usual care". The control clinics were also provided with the PHQ-9 depression scores to ensure patient safety during the trial. The protocol required that practice nurses take action if severe depression was recorded in the returned PHQ-9 or if the patient had responded to the suicidal-ideation question (question 9) on the questionnaire. This action was to be taken *irrespective of whether the clinic was in the intervention or in the control group.*

Practice nurse training

The PN training included a two-day workshop to prepare them for their enhanced roles in nurse-led collaborative care. Topics in the workshop included identifying and monitoring depression using the PHQ-9 questionnaire, and quality of life responses using version 2 of the SF36 questionnaire[21]. Patient goal setting and problem solving were key components of the training with a particular emphasis on behavioural techniques to achieve improved mental health[22]. The training also prepared the PNs for their role as case managers including ensuring Diabetes Australia and Australian National Heart Foundation guidelines were being followed, and referrals were provided

to appropriate services, such as allied health and mental health professionals, through discussion with the GPs.

Data collection

The research team developed a protocol-driven care-plan template from which study data could be extracted automatically and sent to the research team. The template was designed to be a multi-purpose document in which the patient's past medical history, current medications, allergies, biophysical and psychosocial measures, lifestyle risks, personal goals and referrals were recorded. It was designed to comply with the requirements to claim Australian Medicare rebates for care planning and to provide a checklist for "gold-standard" care. A copy of the care plan was provided to the patient as a written record of their progress.

The care-plan template collected physical measures, including body-mass index, waist circumference, weight and blood pressure, and the latest pathology results, including lipid profile, glycaemic control (HbA1c) and renal function. Data also included lifestyle risk factors, such as smoking, alcohol consumption and level of physical activity, and depression score as measured by the PHQ-9 questionnaire. Referrals to and attendance at exercise programmes and with mental health workers were also recorded, along with the patient's own goals and possible barriers to achieving these goals. The care-plan template was used by the intervention-group clinics to acquire patient data at three-monthly intervals over a twelve-month period.

In the control group, the only complete dataset recorded using our comprehensive protocol-driven care-plan template was obtained after the six months of "usual care" when the TrueBlue training was offered to the control clinics. No baseline or three-month datasets were acquired since the study was deliberately designed to avoid changing the "usual care" that would have otherwise occurred by introducing our care-plan template. The study was designed in this way to be run pragmatically in the context of the clinics' normal activities. The only baseline measure obtained was the depression score. On completion of the study, we retrospectively collected what baseline data the control clinics routinely recorded in their electronic medical records.

TrueBlue collaborative care

As part of the TrueBlue model, patients were scheduled to visit the practice every three months for a 45-minute nurse consult followed by a 15-minute consult with their usual GP, in which stepped care (psycho- or pharmacotherapy) was offered if depression scores had not improved or had not dropped below a value of five. The PN used the care-plan template and obtained current physical measures and reviewed recent pathology results. PNs also reviewed lifestyle risk factors. They re-administered the PHQ-9 and worked with the patient to identify possible barriers to

achieving their goals and discussed ways to overcome the barriers. This information gathering phase of the consultation was an opportunity to assist the patient with self-management by discussing available educational resources, such as the library of fact sheets on aspects of self-management of depression, and setting personal goals for review at the next three-monthly visit.

Statistical analysis

Participants in this study were clustered under clinics by design. It is known that clinics are likely to be different from each other and that ignoring the nested nature of the data may lead to biased estimates of parameter standard errors. However, statistical techniques for correcting for the effects of clustering tend to be overly severe and conservative[23] when a small number of higher level units (the clusters) are used and, therefore, we tested whether the clinics were in fact significantly different from each other. ANCOVAs[24,25] were used to adjust for baseline values and test the significance of changes in depression scores between clinics after six months.

Of the five clinics in the intervention (clinics 4, 5, 13, 15 and 17), only clinics 4 and 17 were significantly different from each other ($F(1,76)=9.6$, $p<0.01$). No other comparisons were significant between intervention clinics. Of the six clinics in the control group (clinics 1, 2, 3, 6, 16 and 18), only clinics 6 and 18 were significantly different from other ($F(1,78)=14.5$, $p<0.01$). No other comparisons were significant between control clinics. Furthermore, the intra-correlation coefficient (ICC) of 0.058 for the primary outcome suggests that only 6% of variance could be attributed to the clinics' level. Given this lack of difference between the clinics in each arm coupled with the sample-size requirements for reliable multilevel modelling[26], we analysed our data at the patient level.

In order to compare the effectiveness of the TrueBlue care model to the usual-care control, ANCOVAs were used to adjust for baseline values and test the significance of changes in continuous variables between the two groups after six months. Mixed-effects logistic regression was used to test the significance of changes in the binary (categorical) variables between the two groups after six months. Within each group, changes between the baseline and six-month visits were tested using paired *t*-tests for the continuous variables and matched-case-control McNemar χ^2 tests for the binary variables.

The longer-term effects of the intervention were evaluated by assessing the three-monthly changes over the 12-month period using linear mixed models for the continuous variables and mixed-effects logistic regression for the binary variables. Note that the study design could not collect 12-month "usual care" data from the control clinics since TrueBlue training was provided to these clinics at six-months after which they ceased to be a control.

Patients from the clinics that withdrew before or during collection of first-visit data were excluded from the analyses. (Data for the thirteen patients from one of these clinics who did complete the study have been included.) Available clinics' characteristics were compared between early dropouts and participating clinics and addressed in terms of their possible impact on the generalisability of the results. Missing six-month data were replaced with their baseline values using the "no change" formulation of intention-to-treat by assuming that no change occurred between baseline and six months. The underlying assumptions of the statistical tests used were assessed. STATA version 11.2 was used for the statistical analyses.

Results

Demographics (Table 1)

Table 1 presents the characteristics of the patients in both the intervention and control groups who commenced the study, and shows that these characteristics were similar across both groups. There were no significant differences in patient characteristics between the intervention and control at baseline. Approximately 70% of patients in the intervention group completed all five visits over the 12-month study period.

Phase 1: Comparison of outcomes between control and intervention groups after 6 months (Table 2)

Table 2 presents baseline and 6-month data for markers used to monitor control of chronic disease for both the intervention and the control group. While the six-month depression scores for all patients were significantly lower than those at baseline in both the intervention group (10.7±0.7 to 7.1±0.8, $t(163)=8.38$, $p<0.001$) and the control group (11.6±0.9 to 9.0±0.9, $t(145)=6.01$, $p<0.001$), the ANCOVA adjusting for the baseline scores showed that the improvement was significantly better in the intervention group than in the control ($F(1,309)=6.40$, $p=0.012$). (The 95% confidence ranges are indicated by the plus-minus sign.)

Half of the patients had only mild-depression at baseline (PHQ-9 scores between five and nine). Because the reported score for many of these patients may be due to their diabetes rather than depression[20], the intervention is unlikely to be able to change these scores. This is one reason that Katon and colleagues used a score of ten or more as an inclusion criterion in their study[15]. Consequently, we examined the change to baseline PHQ-9 scores for patients with moderate to severe depression (PHQ-9 scores of 10 or more) at baseline. These patients showed significant improvement, with the mean depression score in the intervention group dropping by 5.7±1.3, from 14.4±1.1 to 8.7±1.3 ($t(80)=9.00$, $p<0.001$), a clinically-significant change[19]. The improvement in the intervention for these patients was significantly better than in the control group ($F(1,161)=4.02$,

p=0.047) where the depression score dropped by 4.3±1.2, from 15.1±1.1 to 10.8±1.4 ($t(82)=6.88$, $p<0.001$).

Except for the HDL measurements, there were no significant changes in biophysical measures after six months in either group. Smoking rates were low at baseline in the patients with established cardiovascular risk factors. Recording of alcohol was sub-optimal, although better than other Australian primary care surveys[27].

The intervention group also showed significantly greater number of patients exercising, referred to and attending an exercise program, and referred to and attending a mental health worker after six months of collaborative care. In the control group, there were no significant changes observed after six months, except that referrals to a mental health worker increased significantly ($p<0.001$) from 9% to 21%, consistent with the action being taken by the nurses as required by the protocol. Neither group showed any significant changes in the number of patients taking antidepressant medication.

Phase 2: Chronic disease outcomes over 12 months using TrueBlue collaborative care (Table 3)

Table 3 presents data at baseline and 12 months for the intervention group for markers used to monitor control of existing diabetes and CHD. The improvement in mental health observed after 6 months was maintained at 12 months, with the significant reduction in the mean depression score maintained (10.7±0.7 to 6.6±0.7, $t(163)=9.92$, $p<0.001$), and nearly 70% of patients having lower depression scores than baseline after one year. Patients with moderate to severe depression at baseline showed an even greater improvement after 12 months of collaborative care, with the mean depression score dropping by 6.4±1.2, from 14.4±0.8 to 8.0±1.2 ($t(80)=10.41$, $p<0.001$). The significant improvement in the mean SF36v2 composite mental-health and physical-health scores observed after 6 months was also maintained at 12 months.

Physiological measures showed a trend, although not significant, to improvement in weight, systolic blood pressure and HDL. Mean baseline lipids and HbA1c were close to guideline targets. The 10-year CVD risk calculated with the Framingham risk equations[28] suggests a small but significant (p=0.015) reduction in risk from 27.4% to 24.8% for the patients with only T2DM. (The Framingham risk equations cannot be used for those patients who have CHD.)

The most notable changes in lifestyle after 12-months of the intervention were a significant increase in the numbers of patients who reported taking regular exercise or being referred to an exercise program. Reported referrals and visits to a mental health worker and numbers taking antidepressant medication were also significantly greater at 12 months.

The TrueBlue protocol also included goal setting so that patients could become more pro-active in their own care. An analysis of participant goals revealed that two-thirds of visits resulted in at least one behavioural activation goal being set and, over the course of the study, 86% of patients identified a behavioural activation goal.

Adherence to guidelines (Figure 2)

Figure 2 shows the percentage of TrueBlue patients who had psychosocial and biophysical checks undertaken as recommended by the Australian National Heart Foundation and Diabetes Australia guidelines, with the corresponding percentages for usual care taken from a study of a large sample of Australian general practices[27].

Discussion

Outcomes of phase 1

Depression scores were significantly lower at six months for patients in the intervention practices compared with those in the control group, and the improvement was clinically significant for patients with moderate to severe depression[19], with patients moving one depression category. Patients experienced increased nurse contact time through the nurse consultations, were provided with information about mental health and their physical health through psycho-education resources, and had their treatment intensified when required. Modalities included behavioural activation, antidepressant medication, and referrals to mental health professionals and exercise programs. Similar improvements in depression scores and stepped-up care were observed in the collaborative care model of Katon and colleagues[15]. The reduction in depression scores observed in the control group could be explained, in part, by control practices being provided with each patient's entry-level depression score during the recruitment process as part of the study's safety protocol. Usual care could have been influenced by drawing attention to co-morbid depression[15] as the protocol required that practice nurses take action if severe depression was recorded or if the patient had responded to the suicidal-ideation question. Referrals to mental-health workers by the control clinics had increased significantly consistent with the clinics taking action where warranted. It is also known[29] that recruiting interested patients (those who wanted to participate) from interested clinics (those that agreed to join) can affect the representativeness of the study population. GPs with a particular interest in the study may be more likely to participate and may manage their patients more effectively, irrespective of whether they are in the control or intervention arms. Consequently, a reduction in depression scores in the control group was expected but the structured TrueBlue model did produce a significantly better reduction in depression. While the effect size may be small (Cohen's $f = 0.15$), it is important to note that TrueBlue was designed to be implemented easily within general practices, with running costs funded by existing Australian Medicare rebates, and to

make better use of their existing resources. These features mean that TrueBlue could be easily applied to patients across general practices at a population level, making the benefits clinically important.

Outcomes of phase 2

The key clinical outcomes over a 12-month period in the intervention group (Table 3) were a sustained improvement in mental health, demonstrated by symptom severity score (PHQ-9 total score) and by the patient's function and subjective evaluation of mental health (SF36 mental health composite score) and physical health (SF36 physical health composite score). Regular physical exercise has been shown to be important for reducing depression[30]. The self-reported exercise rates showed significant improvement over the 12 months of collaborative care intervention. The biophysical measures reported in Table 3 showed modest improvements after 12 months and the Framingham risk equations[28] suggest a small but significant reduction in the 10-year CVD risk for the T2DM patients. These improvements were achieved despite that fact that we did not specifically select patients whose physiological parameters exceeded guidelines. Rather, our recruitment process selected from the practice's disease registry on the basis of only the presence of depression and T2DM or CHD and, consequently, many patients were already being treated to target on measures such as cholesterol and HbA1c, leaving little room for improvement.

Limitations

We were able to run TrueBlue only in practices that had access to clinical software capable of generating a disease registry from which participants could be selected, and had a practice nurse on staff. Clinics that chose to take part in the study may not have been representative of wider general practice. Operational limitations further reduced the number of practices over the duration of the study. Patient response rates to the mail-out (28%) may reflect anxiety over the new model of care where the patient discloses depression and visits the PN first rather than only the GP. Usual care in the control clinics may have been changed by patients completing the PHQ-9 and reading the project description. GPs were made aware of individual PHQ-9 results, and took action where warranted. GP awareness of these biophysical and lifestyle risks may be expected to change clinical management. By design, TrueBlue practices needed to incorporate all research activities within the context of their busy clinics, and so only research data that could be extracted automatically were collected. The data dropout resulting from these two factors contributed to the observed small effect size. We were not able to obtain multiple data sets at three-monthly intervals over 12 months of 'usual care' because the act of inviting patients and measuring depression scores and biophysical measures would in itself change the nature of usual care. In addition, practices would not have been

willing to join the study if there was a chance of being randomly allocated to 12 months of being in such a control arm[29].

Collaborative Care

A recent UK study has shown the difficulties of disseminating a guideline without guidance on how to implement collaborative care. Organisational barriers included GPs finding the PHQ-9 awkward to use, nurses not feeling confident or competent due to lack of training, and no guidance on stepped care[13]. The TrueBlue model of collaborative care overcame many of these difficulties. Its successful components were:[9,31]

- *Use of evidence based guidelines.* National Heart Foundation and Diabetes Australia guidelines determined disease management targets and frequency of monitoring.
- *Systematic screening and monitoring of risk factors.* Patients attended three monthly visits in which a care plan with its checklist was completed. By providing a comprehensive collation of all necessary information, this document made clinical management by the patient's GP easier, quicker and more accurate.
- *Timetabled recall visits.* The date of the next appointment was set during each visit, PHQ-9 was readministered and if improvement was insufficient, *stepped care* was followed by initiating drug therapy or increasing the dose, or referral to a mental-health worker according to the guidelines.
- *New or adjusted roles for team members.* PNs took responsibility for organising and monitoring the outcome of referrals, goals and targets. They used a depression questionnaire (the PHQ-9) to open a discussion with patients about their depression symptoms.
- *Information support for the clinician.* GPs were provided with the care plan by the PNs.
- *Enhanced patient self-management.* Patients received their own copy of the care plan with personalised goals, current measurements, targets and safety advice. A component of each visit was to discuss and update their plan, and receive education material on depression.
- *Identified case manager.* PNs became case managers but the GP remained the key clinician.
- *Means of effective communication between all members of the care team.* The care plan was designed to provide relevant clinical information in a succinct format while still being comprehensible to patients.
- *Audit information for the practice.* De-identified data was provided automatically through the care plan.

Applicability of TrueBlue

TrueBlue used existing clinical software and improved the focus of the GP consultation by delegating some tasks to the PN. Higher levels of adherence to guideline-recommended checks were also reported for TrueBlue. Patients and their carers, allied health professionals, specialists and general practitioners gained ready access to patient details provided in TrueBlue's care plan enabling them to see at glance where improve clinical care may be needed. The study achieved improved outcomes with the potential for prevention of heart attack and stroke through reduced 10-year CVD risk. The care plan template also allowed the practice to collect high quality audit data without taking up clinical time. While it was not possible to obtain complete financial data from the clinics specifically relating to the TrueBlue visits, the data that are available suggest that clinics did indeed cover their costs in implementing TrueBlue through Australian Medicare rebates. The success of TrueBlue and TeamCare[15] demonstrates that collaborative care is feasible in routine general practice in Australia and the USA, and could lead to improved outcomes for patients with depression and other chronic diseases[7,32].

Acknowledgements

The authors wish to thank the patients, practice nurses, general practitioners and support staff of the participating clinics Evans Head Medical Centre, Lennox Head Medical Centre, Keen Street Clinic, Tintenbar Medical Centre, Health on Grange, Warradale Medical Centre, Seaton Medical and Specialist Centre, Woodcroft Medical Centre and Southcare Medical Centre. Professors Wayne Katon and Juergen Unützer of the University of Washington were unstinting in their advice on adaptation of the IMPACT model. We would also like to thank Bob Leahy for managing the project, Vince Versace for his statistical advice and Vicki Brown for her assistance during the course of the study.

Contributors

All authors have full access to the complete study dataset, contributed to the design, implemented the project, and co-wrote and approved the manuscript. MM, MC, JD and PR analysed the data. MM, PR and KS developed and ran the practice nurse training program. JD and PR conceived the TrueBlue model during a visit to the IMPACT team. JD is the guarantor.

Funding

Funding was provided by *beyondblue*, the National Depression Initiative in Australia, but it had no other involvement in any phase of the study. All researchers were independent of and not influenced by *beyondblue.*

Competing interests

All authors have completed the Unified Competing Interest form and declare that (1) funding was received from *beyondblue* to carry out the study; (2) they do not have any relationships with any companies that might have an interest in the submitted work in the previous 3 years; (3) their spouses, partners, or children do not have any financial relationships that may be relevant to the submitted work; and (4) they do not have any non-financial interests that may be relevant to the submitted work.

Ethical approval

Ethics approval was obtained from Flinders University's Social and Behavioural Research Ethics Committee. All patients gave informed consent to participate in the study.

Data sharing

The dataset used for the analysis and the computer codes used to produce the results are available from the corresponding author.

Licence

References

1. Daar AS, Singer PA, Persad DL, Pramming SK, Matthews DR, Beaglehole R, et al. Grand challenges in chronic non-communicable diseases. *Nature* 2007;450(7169):494-96.
2. World Health Organisation. Preventing Chronic Disease: A Vital Investment. Geneva, 2005.
3. Ali S, Stone MA, Peters JL, Davies MJ, Khunti K. The prevalence of co-morbid depression in adults with Type 2 diabetes: a systematic review and meta-analysis. *Diabet Med* 2006;23(11):1165-73.
4. Egede LE, Nietert PJ, Zheng D. Depression and all-cause and coronary heart disease mortality among adults with and without diabetes. *Diabetes Care* 2005;28(6):1339-45.
5. Clarke DM, Currie KC. Depression, anxiety and their relationship with chronic diseases: a review of the epidemiology, risk and treatment evidence. *Med J Aust* 2009;190(7 Suppl):S54-60.
6. Freeling P, Rao BM, Paykel ES, Sireling LI, Burton RH. Unrecognised depression in general practice. *Br Med J (Clin Res Ed)* 1985;290(6485):1880-83.
7. Gilbody S, Bower P, Fletcher J, Richards D, Sutton AJ. Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. *Arch Intern Med* 2006;166(21):2314-21.
8. Katon W, Von Korff M, Lin E, Walker E, Simon GE, Bush T, et al. Collaborative management to achieve treatment guidelines. Impact on depression in primary care. *JAMA* 1995;273(13):1026-31.
9. Hickie IB, McGorry PD. Increased access to evidence-based primary mental health care: will the implementation match the rhetoric? *Med J Aust* 2007;187(2):100-03.
10. Lin EH, Katon W, Von Korff M, Rutter C, Simon GE, Oliver M, et al. Relationship of depression and diabetes self-care, medication adherence, and preventive care. *Diabetes Care* 2004;27(9):2154-60.
11. Reddy P, Ford D, Dunbar JA. Improving the quality of diabetes care in general practice. *Aust J Rural Health* 2010;18(5):187-93.
12. Campbell MJ. Cluster randomized trials in general (family) practice research. *Stat Methods Med Res* 2000;9(2):81-94.
13. Mitchell C, Dwyer R, Hagan T, Mathers N. Impact of the QOF and the NICE guideline in the diagnosis and management of depression: a qualitative study. *Br J Gen Pract* 2011;61(586):e279-e89.
14. Unutzer J, Katon W, Williams JW, Jr., Callahan CM, Harpole L, Hunkeler EM, et al. Improving primary care for depression in late life: the design of a multicenter randomized trial. *Med Care* 2001;39(8):785-99.
15. Katon WJ, Lin EH, Von Korff M, Ciechanowski P, Ludman EJ, Young B, et al. Collaborative care for patients with depression and chronic illnesses. *N Engl J Med* 2010;363(27):2611-20.
16. Morgan MA, Dunbar J, Reddy P. Collaborative care - The role of practice nurses. *Aust Fam Physician* 2009;38(11):925-26.
17. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16(9):606-13.
18. Morgan M, Dunbar J, Reddy P, Coates M, Leahy R. The TrueBlue study: is practice nurse-led collaborative care effective in the management of depression for patients with heart disease or diabetes? *BMC Fam Pract* 2009;10:46.
19. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. *Medical care* 2004;42(12):1194-201.
20. Reddy P, Philpot B, Ford D, Dunbar JA. Identification of depression in diabetes: the efficacy of PHQ-9 and HADS-D. *Brit J Gen Prac* 2010;DOI: 10.3399/bjgp10X502128.
21. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30(6):473-83.

22. Dobson KS, Hollon SD, Dimidjian S, Schmaling KB, Kohlenberg RJ, Gallop RJ, et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the prevention of relapse and recurrence in major depression. *Journal of Consulting and Clinical Psychology* 2008;76(3):468-77.
23. Kauermann G, Carroll RJ. A Note on the Efficiency of Sandwich Covariance Matrix Estimation. *J Amer Statistical Assoc* 2001;96(456):1387-96.
24. Senn S. Change from baseline and analysis of covariance revisited. *Statistics in Medicine* 2006;25(24):4334-44.
25. Vickers A. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: A simulation study. *BMC Medical Research Methodology* 2001;1:1-4.
26. Hox JJ. *Applied multilevel analysis*. Second ed. Amsterdam: TT-Publikaties, 1995.
27. Wan Q, Harris MF, Jayasinghe UW, Flack J, Georgiou A, Penn DL, et al. Quality of diabetes care and coronary heart disease absolute risk in patients with type 2 diabetes mellitus in Australian general practice. *Qual Saf Health Care* 2006;15(2):131-35.
28. Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J* 1991;121(1 Pt 2):293-98.
29. Wilson S, Delaney BC, Roalfe A, Roberts L, Redman V, Wearn AM, et al. Randomised controlled trials in primary care: case study. *BMJ* 2000;321:24-27.
30. Wiles NJ, Haase AM, Gallacher J, Lawlor DA, Lewis G. Physical activity and common mental disorder: results from the Caerphilly study. *Am J Epidemiol* 2007;165(8):946-54.
31. Fuller J, Perkins D, Parker S, Holdsworth L, Kelly B, Roberts R, et al. Effectiveness of service linkages in primary mental health care: a narrative review part 1. . *BMC Health Services Research* 2011;11:71.
32. Richards DA, Lovell K, Gilbody S, Gask L, Torgerson D, Barkham M, et al. Collaborative care for depression in UK primary care: a randomized controlled trial. *Psychol Med* 2008;38(2):279-87.

Table 1: Patient characteristics at the baseline visits. There were no significant differences between the intervention and control at baseline.

Characteristic	Intervention group (*n*=170)	Control group (*n*=147)
Male (%) / Female (%)	51.8% / 48.2%	55.2% / 44.8%
Age (yr)	68.0 ± 11.7	67.6 ± 11.2
Aboriginal or Torres Strait Islander (%)	0.0%	0.7%
Diagnosis: Type-2 diabetes	37.6%	47.6%
CHD	45.3%	35.8%
Both	17.1%	16.6%
Body mass index (kg/m^2)	31.4 ± 6.0 (*n*=170)	30.8 ± 6.0 (*n* =103)
Systolic blood pressure (mmHg)	134.1 ± 19.0 (*n*=169)	133.5 ± 19.6 (*n* =112)
Total cholesterol (mmol/L)	4.21 ± 0.94 (*n*=165)	4.41 ± 1.06 (*n* =110)
Triglycerides (mmol/L)	1.73 ± 0.88 (*n*=165)	1.92 ± 1.37 (*n* =105)
LDL (mmol/L)	2.22 ± 0.74 (*n*=159)	2.37 ± 0.88 (*n* =89)
HDL (mmol/L)	1.23 ± 0.36 (*n*=159)	1.18 ± 0.33 (*n* =97)
HbA1c (mmol/L)	7.00 ± 1.21 (*n*=94)	7.19 ± 1.42 (*n* =69)
PHQ-9 score	10.7 ± 4.7 (*n*=164)	11.6 ± 5.5 (*n*=146)
PHQ-9 score range at baseline	5 to 24	5 to 27

Table 2: TrueBlue outcomes at six months in the intervention and control groups. (See Table 3 for a list of abbreviations.)

	Intervention				Control				Between groups
	n	Baseline	6-months	Within	*n*	Baseline	6-months	Within	
PHQ9 depression score	164	10.7±0.8	7.1±0.8	$p<0.001^{\dagger}$	146	11.6±0.9	9.0±0.9	$p<0.001^{\ddagger}$	$p=0.012$
SF36v2 mental-health score[§]	71	37.2±3.4	41.1±3.4	$p=0.034^{\dagger}$		Not recorded			NS
SF36v2 physical-health score[§]	71	39.9±2.2	42.5±2.6	$p=0.023^{\dagger}$		Not recorded			NS
Body mass index (kg/m^2)	162	31.3±1.0	31.2±1.0	NS	103	30.8±1.2	31.0±1.0	NS	NS
Waist (cm)	161	104.7±2.4	105.0±2.4	NS	80	104.2±4.0	105.8±3.2	NS	NS
Systolic blood pressure (mmHg)	161	134.2±3.0	132.4±2.8	NS	112	133.5±3.8	131.2±3.4	NS	NS
Total cholesterol (mmol/L)	158	4.21±0.16	4.22±0.14	NS	109	4.41±0.20	4.44±0.20	NS	NS
LDL (mmol/L)	154	2.23±0.12	2.17±0.14	NS	86	2.37±0.18	2.29±0.20	NS	NS
HDL (mmol/L)	154	1.23±0.06	1.29±0.06	$p=0.023^{\dagger}$	93	1.17±0.06	1.27±0.08	$p=0.011^{\ddagger}$	NS
Triglycerides (mmol/L)	158	1.72±0.14	1.66±0.12	NS	104	1.84±0.22	1.75±0.18	NS	NS
HbA1c (%)[¶]	89	6.97±0.24	6.90±0.26	NS	67	7.22±0.34	7.40±0.36	NS	$p=0.049$
Ten-year CVD risk[*]	61	26.9±3.2	26.1±3.2	NS	46	26.3±3.6	24.7±3.2	NS	NS
The 95% confidence ranges are indicated by the plus-minus (±) sign. Note that lower scores indicate improvement for all items except the SF36v2 and HDL results, where higher scores indicate improvement.									
Smoking	162	15 (9%)	13 (8%)	NS	110	13 (12%)	13 (12%)	NS	NS
Alcohol	104	47 (45%)	51 (49%)	NS	42	27 (64%)	27 (64%)	NS	NS
Exercises 30 mins/day, 5 d/wk	162	66 (41%)	97 (60%)	$p<0.001^{\dagger}$	75	22 (29%)	22 (29%)	NS	$p<0.001$
Referred to exercise program	162	32 (20%)	58 (36%)	$p<0.001^{\dagger}$	111	15 (14%)	10 (9%)	NS	$p<0.001$
Attends exercise program	162	12 (7%)	23 (14%)	$p=0.041^{\dagger}$	79	12 (15%)	9 (11%)	NS	NS
On antidepressant medication	162	27 (17%)	34 (21%)	NS	113	31 (27%)	36 (32%)	NS	$p=0.025$
Referred to mental health worker	162	47 (29%)	58 (36%)	$p=0.022^{\dagger}$	114	10 (9%)	24 (21%)	$p<0.001^{\ddagger}$	$p<0.001$
Attends mental health worker	162	10 (6%)	37 (23%)	$p<0.001^{\dagger}$	109	14 (13%)	11 (10%)	NS	$p=0.044$

The values in brackets are the percentages of the total *n*.

(NS) No significant difference. (†) Significant difference between baseline and six-months within the *intervention* clinics. (‡) Significant difference between baseline and six-months within the *control* clinics. (§) SF36v2 questionnaires were not collected by all clinics. (¶) HbA1c results were only available for patients with T2DM. (*) CVD risk could only be calculated for patients with T2DM.

Table 3: TrueBlue outcomes at 12 months within the intervention clinics only.

	Intervention			
	n	Baseline	12-months	
PHQ9 depression score	164	10.7±0.7	6.6±0.7	*p*<0.001
SF36v2 mental-health score[§]	70	36.0±3.2	41.3±2.8	*p*<0.001
SF36v2 physical-health score[§]	70	40.6±2.2	44.3±2.8	*p*<0.001
Body mass index (kg/m^2)	142	31.4±1.0	31.1±1.0	*p*=0.006
Waist (cm)	141	105.0±2.4	105.2±2.6	NS
Systolic blood pressure (mmHg)	141	135.2±3.2	130.2±3.0	*p*=0.016
Total cholesterol (mmol/L)	138	4.18±0.16	4.28±0.16	NS
LDL (mmol/L)	135	2.19±0.12	2.24±0.20	NS
HDL (mmol/L)	135	1.22±0.06	1.36±0.08	*p*<0.001
Triglycerides (mmol/L)	138	1.73±0.16	1.63±0.14	*p*=0.004
HbA1c (%)[¶]	79	7.01±0.26	7.04±0.28	NS
Ten-year CVD risk[*]	55	27.4±3.4	24.9±3.6	*p*=0.015

The 95% confidence ranges are indicated by the plus-minus (±) sign. Lower scores indicate improvement for all items except the SF36v2 and HDL results, where higher scores indicate improvement.

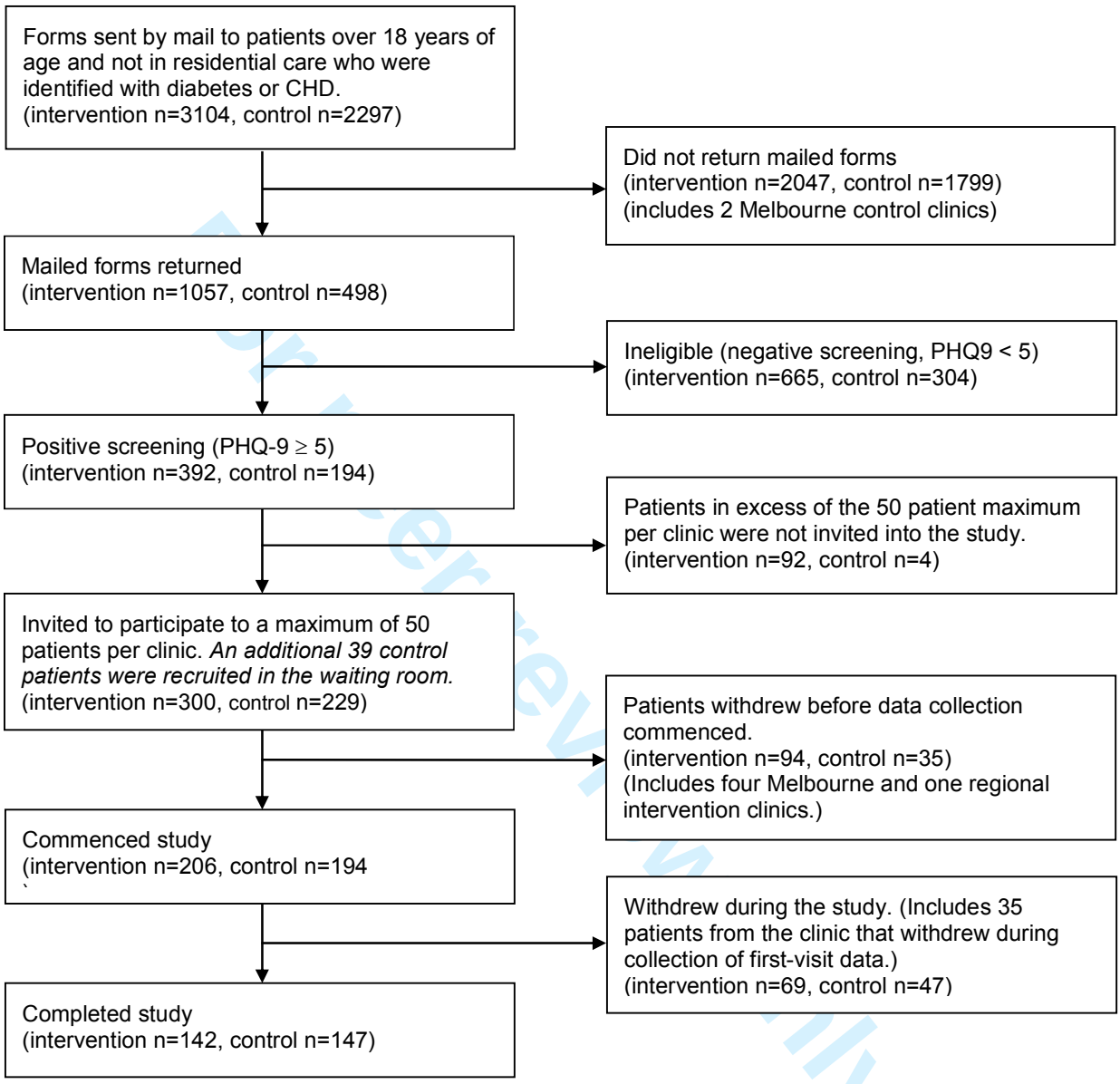
Smoking	142	15 (11%)	11 (8%)	NS
Alcohol	95	45 (47%)	47 (49%)	NS
Exercises 30 mins/day, 5 days/wk	142	57 (40%)	83 (58%)	*p*<0.001
Referred to exercise program	142	26 (18%)	53 (37%)	*p*<0.001
Attends exercise program	142	10 (7%)	17 (12%)	NS
On antidepressant medication	142	22 (15%)	33 (23%)	*p*=0.001
Referred to mental health worker	142	40 (28%)	59 (42%)	*p*<0.001
Attends mental health worker	142	8 (6%)	25 (18%)	*p*<0.001

The values in brackets are the percentages of the total *n*.

(NS) No significant difference. (§) SF36v2 questionnaires were not collected by all clinics. (¶) HbA1c results were only available for patients with T2DM. (*) CVD risk could only be calculated for patients with T2DM.

NOTE:
PHQ9 is the 9-question Patient Health Questionnaire.
SF36v2 is version 2 of the Short Form 36-Question health survey.
LDL is low-density lipoprotein.
HDL is high-density lipoprotein.
Glycated haemoglobin (HbA1c) was measured only for patients with diabetes.
Unit of alcohol is 10g of ethanol.

Figure 1. CONSORT flow diagram of the recruitment process.



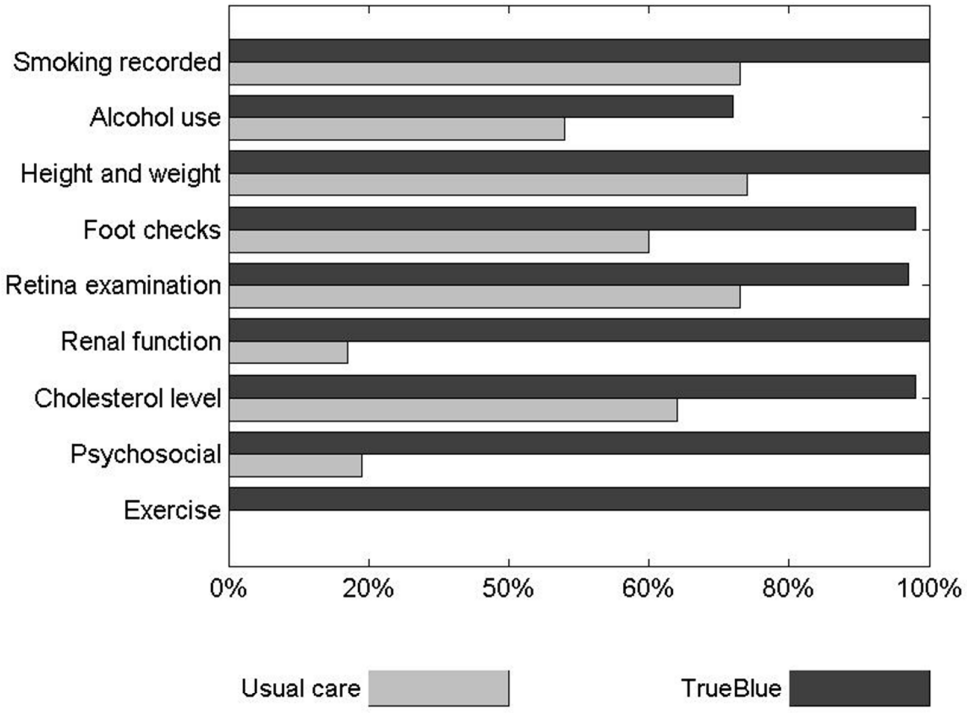


Figure 2: Recording of checks recommended by the National Heart Foundation and Diabetes Australia guidelines. Data for "usual care" were adapted from reference 27. No usual-care data were available for exercise.

Checklist of items to include when reporting a cluster randomised trial

* = addition to CONSORT *Modifications to checklist in italics*			
PAPER SECTION and topic	Item	Descriptor	Reported on Page No.
TITLE & ABSTRACT	1*	How participants were allocated to interventions (e.g., "random allocation", "randomised", or "randomly assigned"), *specifying that allocation was based on clusters*	1
INTRODUCTION Background	2*	Scientific background and explanation of rationale, *including the rationale for using a cluster design.*	2, 3
METHODS Participants	3*	Eligibility criteria for participants *and clusters* and the settings and locations where the data were collected.	3, 4
Interventions	4*	Precise details of the interventions intended for each group, *whether they pertain to the individual level, the cluster level or both,* and how and when they were actually administered.	Ref 17
Objectives	5*	Specific objectives and hypotheses, *and whether they pertain to the individual level, the cluster level or both*.	1, 3
Outcomes	6*	Report clearly defined primary and secondary outcome measures, *whether they pertain to the individual level, the cluster level or both*, and, when applicable, any methods used to enhance the quality of measurements (e.g., multiple observations, training of assessors).	4, 5
Sample size	7*	How *total* sample size was determined *(including method of calculation, number of clusters, cluster size, a coefficient of intracluster correlation (ICC or k), and an indication of its uncertainty*) and, when applicable, explanation of any interim analyses and stopping rules.	Ref 17
Randomisation.			
Sequence generation	8*	Method used to generate the random allocation sequence, including details of any restriction (e.g., blocking, stratification, *matching*).	
Allocation concealment	9*	Method used to implement the random allocation sequence, *specifying that allocation was based on clusters rather than individuals and* clarifying whether the sequence was concealed until interventions were assigned.	Ref 17
Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	
Blinding (Masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. If done, how the success of blinding was evaluated.	N/A
Statistical methods	12*	Statistical methods used to compare groups for primary outcome(s) *indicating how clustering was taken into account*; methods for additional analyses, such as subgroup analyses and adjusted analyses.	5
RESULTS Participant flow	13*	Flow of *clusters and* individual participants through each stage (a diagram is strongly recommended). Specifically, for each group report the numbers of *clusters and* participants randomly assigned, receiving intended treatment, completing the study protocol, and analyzed for the primary outcome. Describe protocol deviations from study as planned, together with reasons.	15 (Fig 1)
Recruitment	14	Dates defining the periods of recruitment and follow-up.	4
Baseline data	15*	Baseline information for each group *for the individual and cluster levels as applicable*	17 (Table 1)
Numbers analyzed	16*	Number of *clusters and* participants (denominator) in each group included in each analysis and whether the analysis was by "intention-to-treat". State the results in absolute numbers when feasible (e.g., 10/20, not 50%).	17, 18, 19 (Tables)
Outcomes and Estimation	17*	For each primary and secondary outcome, a summary of results for each group measures *for the individual or cluster level as applicable*, and the estimated effect size and its precision (e.g., 95% confidence interval) *and a coefficient of intracluster correlation (ICC or k) for each primary outcome.*	17, 18, 19 (Tables)
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those pre-	N/A

		specified and those exploratory.	
Adverse events	19	All important adverse events or side effects in each intervention group.	None
DISCUSSION Interpretation	20	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision and the dangers associated with multiplicity of analyses and outcomes.	7, 8
Generalisability	21*	Generalisability (external validity) *to individuals and/or clusters (as relevant)* of the trial findings.	9, 10
Overall evidence	22	General interpretation of the results in the context of current evidence.	9

The TrueBlue model of collaborative care using practice nurses as case managers for depression alongside diabetes or heart disease

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-002171.R1
Article Type:	Research
Date Submitted by the Author:	05-Dec-2012
Complete List of Authors:	Morgan, Mark; Flinders and Deakin Universities, Greater Green Triangle University Department of Rural Health Coates, Michael; Flinders and Deakin Universities, Greater Green Triangle University Department of Rural Health Dunbar, James; Flinders and Deakin Universities, Greater Green Triangle University Department of Rural Health Reddy, Prasuna; University of Newcastle, School of Medicine and Public Health Schlicht, Kate; Flinders and Deakin Universities, Greater Green Triangle University Department of Rural Health Fuller, Jeff; Flinders University, School of Nursing and Midwifery
Primary Subject Heading:	Public health
Secondary Subject Heading:	Diabetes and endocrinology
Keywords:	Coronary heart disease < CARDIOLOGY, General diabetes < DIABETES & ENDOCRINOLOGY, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, MENTAL HEALTH, PUBLIC HEALTH

The TrueBlue model of collaborative care using practice nurses as case managers for depression alongside diabetes or heart disease

MAJ Morgan[1], MJ Coates[1], JA Dunbar[1,*], P Reddy[2], K Schlicht[1] and J. Fuller[3]

1. Greater Green Triangle University Department of Rural Health, Flinders and Deakin Universities, PO Box 423, Warrnambool, Victoria, 3280, Australia
2. School of Medicine and Public Health, University of Newcastle, Callaghan, NSW, 2308, Australia
3. School of Nursing and Midwifery, Flinders University, GPO Box 2100, Adelaide, SA, 5001, Australia

* Corresponding author

Abstract

Objectives: Determine the effectiveness of collaborative-care in reducing depression in primary-care patients with diabetes or heart disease using practice nurses as case managers.

Design: A two-arm open randomised-cluster trial with wait-list control for 6 months. The intervention was followed over 12 months.

Setting: Eleven Australian general practices, five randomly allocated to the intervention and six to the control.

Participants: 400 primary-care patients (206 intervention, 194 control) with depression and type 2 diabetes, coronary heart disease or both.

Intervention: The practice nurse acted as case manager identifying depression, reviewing pathology results, lifestyle risk factors, and patient goals and priorities. Usual care continued in the controls.

Main outcome measure: A five-point reduction in depression scores for patients with moderate to severe depression. Secondary outcome was improvements in physiological measures.

Results: Mean depression scores after six months of intervention for patients with moderate to severe depression decreased by 5.7±1.3 compared with 4.3±1.2 in the control, a significant (p=0.012) difference. (The plus-minus is the 95% confidence range.) Intervention practices demonstrated adherence to treatment guidelines and intensification of treatment for depression, where exercise increased by 19%, referrals to exercise programs by 16%, referrals to mental health workers (MHWs) by 7%, and visits to MHWs by 17%. Control-practice exercise did not change, referrals to exercise programs dropped by 5%, and visits to MHWs by 3%. Only referrals to MHW increased, by 12%. Intervention improvements were sustained over 12 months, with significant

(p=0.015) decrease in 10-year cardiovascular-disease risk from 27.4±3.4% to 24.8±3.8%. A review of patients indicated the study's safety protocols were followed.

Conclusion: TrueBlue participants showed significantly improved depression and treatment intensification, sustained over 12-months of intervention, and reduced 10-year CVD risk. Collaborative care using practice nurses appears to be an effective primary-care intervention.

Trial registration: Australia and New Zealand Clinical Trials Registry ACTRN12609000333213.

Article summary

Article focus

- To determine the effectiveness of a collaborative-care model to reduce depression in primary-care patients with diabetes or heart disease.
- To determine the effectiveness of using practice nurses as case managers of patients with depression and diabetes, heart disease or both.

Key messages

- The TrueBlue model of collaborative care can be introduced within the general practice workforce with practice nurses taking on the role of case manager.
- Practice nurses can improve care of depression in patients with diabetes or heart disease, leading to better outcomes and reduced 10-year CVD risk.
- The care of patients using the TrueBlue model is closer to "best practice" guidelines, with substantially better levels of adherence to guideline-recommended checks than occur in usual care.

Strengths and limitations of this study

- The TrueBlue model of collaborative care overcomes many of the difficulties in implementing a guideline for the treatment of co-morbid depression.
- The study's purpose-designed care plan gives patients and their carers, allied health professionals, specialists and general practitioners ready access to patient details enabling them to see at glance where improve clinical care may be needed.
- Clinics were able to recover the costs of the collaborative care through Australian Medicare rebates.

- The study could only be run in practices that had a practice nurse on staff and had access to clinical software capable of generating a disease registry from which participants could be selected.

- Differences between TrueBlue- and control-practice outcomes may have been reduced by patients completing the PHQ-9 depression questionnaire and reading the project description, and by GPs being made aware of individual PHQ-9 results so that they could take action where warranted.

Introduction

Managing diabetes and heart disease has been highlighted as one of the global "grand challenges in chronic non-communicable diseases"[1] because the prevalence of these two preventable diseases is increasing[2]. Along with depression, they have been identified as health priority areas in many countries. A vicious cycle exists between depression and these chronic diseases, with each being a risk factor for the other[3]. Higher mortality has been demonstrated for people with depression and type 2 diabetes (T2DM) or coronary heart disease (CHD) beyond that due to the separate diseases alone[4]. For patients with depression and T2DM or CHD or both, there are increased risks of adverse outcomes[5] but this co-morbid depression is often missed in primary care[6]. Consequently, the identification of depression has now been incorporated in many heart disease guidelines as one of the requirements for optimal management. Meeting these challenges will require an innovative use of the existing general practice workforce and such a re-orientation of resources has been identified as one of the grand challenges[1].

Collaborative care is a system that has been shown to be more effective for chronic disease management than standard care[7]. It includes a reorientation of the medical workforce through new or adjusted roles for team members, particularly using practice nurses as the identified case manager[8,9]. It also includes the use of evidence based guidelines, systematic screening and monitoring of risk factors, timetabled recall visits, information support for the clinician, enhanced patient self-management, a means of effective communication between all members of the care team and audit information for the practice. Since self-care for diabetes has been found to be suboptimal across a range of self-managed activities, particularly for patients with depression, a collaborative care model may be able to achieve better quality of care through the case manager monitoring patient progress[10,11].

Evaluation of a change in the way general practice clinics look after patients requires complex intervention methodology[12] beyond single interventions such as introduction of a guideline with financial incentives[13]. This methodology began with a search for potential models of care (step I), and led to adopting the University of Washington's successful IMPACT model of Collaborative Care for depression[14,15]. In the exploratory trial (step II), our pilot project[16] adapted IMPACT by

training practice nurses as case managers. Practice nurses were trained to screen for depression using a patient self-report measure, the nine-item Patient Health Questionnaire (PHQ-9)[17], as part of comprehensive chronic disease management. They were also trained to use a protocol for care management based on the depression scores. The depression screening and management was embedded in routine visits for patients with diabetes or CHD. The pilot demonstrated that it was feasible to detect, monitor and treat depression in routine general practice alongside the usual biophysical measures, and identified moderate to severe depression in 34% of participants. The TrueBlue study was a randomised cluster trial (step III) that built on and extended the pilot. It investigated whether a collaborative care model (the intervention) is better than usual care (the control) for the management of patients with depression and T2DM, CHD or both in Australian general practice. It was designed to fit into normal clinic operations, making use of practice nurses and medical software, and able to be funded by existing Australian Medicare rebates.

Methods/design

Study design

The design and methodology of the study have been described in detail elsewhere[18]. The study commenced in 2009 and was undertaken in two phases. The first phase was a cluster randomised intervention trial in which general practices were randomly allocated to either an intervention group in which nurse-led collaborative care was undertaken or to a wait-list control group in which usual care led by the general practitioner (GP) was continued. At six months, the TrueBlue training was provided to the control practices. The key aims of the first phase were to determine whether participants with moderate to severe depression in the intervention group showed at least a five-point reduction from the baseline depression scores after six months of intervention and whether this reduction was significantly better than in the control group. A five-point reduction reflects a clinically-relevant change in individuals receiving depression treatment[19]. The secondary outcome was to determine whether the intervention also led to improvements in the patients' physiological measures. The second phase followed the intervention group for an additional six months to determine how the collaborative care model affected health outcomes over a twelve-month period.

Sample size

The sample size calculation was based on detecting a 50% reduction in depression score at the 0.05 significance level with 80% power and a two-tailed test. Detecting a 50% reduction is more stringent than detecting a five-point reduction and provided some additional buffering. Using depression scores from an earlier study (a mean of 5.5 and standard deviation of 6.1)[11], the calculation indicated that 237 patients would be required in each group. An intra-cluster correlation of 0.04 was used (S. K. Lo, pers. comm.), with a recruitment target of 50 patients per clinic. (Fifty

patients were chosen so that clinics could budget for a nurse's time to carry out the intervention with four patients each week over the 3-month cycle of care.) To allow for a 50% dropout, the study required 450 patients from nine clinics in the intervention group and the same in the control group.

Practice recruitment

Practices were identified in city and country areas on the basis of having a practice nurse to provide the collaborative care and being able to identify eligible patients, those with CHD or T2DM or both, from their registries; these were invited to participate in the study until the eighteen clinics required by the sample-size calculation were recruited. They were allocated by a random number generator to either the intervention or control arm of the study. The unit of randomisation was the clinic. Five practices (3 country, 2 city) in the intervention group and six (2 country, 4 city) in the control group completed the study. One country intervention clinic withdrew whilst first-visit data were being collected when its TrueBlue-trained practice nurse left the clinic, but some (n=13) patients from it did complete the study and data were collected from them. The study team was not able to determine why the other clinics withdrew.

Patient selection

Eligible patients were sent a postal survey that included a consent form and were asked to complete and return the enclosed PHQ-9 questionnaire, a self-report measure of depression.[17] The PHQ-9 has nine items, each scored from 0 (no problems) to 3 (problems nearly every day). The sum of the scores of the nine items will lie in one of five depression categories: none (0–4), mild (5–9), moderate (10–14), moderately severe (15–19) and severe (20–27). (While it is known that responses to some of the PHQ-9 items may overlap with diabetes symptoms[20], our pilot demonstrated that nurses and patients preferred using the PHQ-9 because the patient's response to each of its items became the basis for the problem solving and goal setting activities that were part of TrueBlue.) Patients with scores of five or above, indicating some form of depression, were invited to participate in the study. A maximum of 50 patients per practice were invited. Patients in residential care or under 18 years of age were not eligible. Figure 1 presents the CONSORT diagram of the patient-recruitment process.

Patient safety

Participation in the intervention included a series of patient visits to their practice nurse (PN) and usual GP every three months over a 12-month period. Patients in the control group continued with their "usual care". The control clinics were also provided with the PHQ-9 depression scores to ensure patient safety during the trial. The protocol required that practice nurses take action if severe depression was recorded in the returned PHQ-9 or if the patient had responded to the suicidal-

ideation question (question 9) on the questionnaire. This action was to be taken *irrespective of whether the clinic was in the intervention or in the control group*.

Practice nurse training

The PN training included a two-day workshop to prepare them for their enhanced roles in nurse-led collaborative care. Topics in the workshop included identifying and monitoring depression using the PHQ-9 questionnaire, and quality of life responses using version 2 of the SF36 questionnaire[21]. Patient goal setting and problem solving were key components of the training with a particular emphasis on behavioural techniques to achieve improved mental health[22]. The training also prepared the PNs for their role as case managers including ensuring Diabetes Australia and Australian National Heart Foundation guidelines were being followed, and referrals were provided to appropriate services, such as allied health and mental health professionals, through discussion with the GPs.

Data collection

The research team developed a protocol-driven care-plan template from which study data could be extracted automatically and sent to the research team. The template was designed to be a multi-purpose document in which the patient's past medical history, current medications, allergies, biophysical and psychosocial measures, lifestyle risks, personal goals and referrals were recorded. It was designed to comply with the requirements to claim Australian Medicare rebates for care planning and to provide a checklist for "gold-standard" care. A copy of the care plan was provided to the patient as a written record of their progress.

The care-plan template collected physical measures, including body-mass index, waist circumference, weight and blood pressure, and the latest pathology results, including lipid profile, glycaemic control (HbA1c) and renal function. Data also included lifestyle risk factors, such as smoking, alcohol consumption and level of physical activity, and depression score as measured by the PHQ-9 questionnaire. Referrals to and attendance at exercise programmes and with mental health workers were also recorded, along with the patient's own goals and possible barriers to achieving these goals. The care-plan template was used by the intervention-group clinics to acquire patient data at three-monthly intervals over a twelve-month period.

In the control group, the only complete dataset recorded using our comprehensive protocol-driven care-plan template was obtained after the six months of "usual care" when the TrueBlue training was offered to the control clinics. No baseline or three-month datasets were acquired since the study was deliberately designed to avoid changing the "usual care" that would have otherwise occurred by introducing our care-plan template. The study was designed in this way to be run

pragmatically in the context of the clinics' normal activities. The only baseline measure obtained was the depression score. On completion of the study, we retrospectively collected what baseline data the control clinics routinely recorded in their electronic medical records in order to have data for two time points, baseline and six-months.

TrueBlue collaborative care

As part of the TrueBlue model, patients were scheduled to visit the practice every three months for a 45-minute nurse consult followed by a 15-minute consult with their usual GP, in which stepped care (psycho- or pharmacotherapy) was offered if depression scores had not improved or had not dropped below a value of five. The PN used the care-plan template and obtained current physical measures and reviewed recent pathology results. PNs also reviewed lifestyle risk factors. They re-administered the PHQ-9 and worked with the patient to identify possible barriers to achieving their goals and discussed ways to overcome the barriers. This information gathering phase of the consultation was an opportunity to assist the patient with self-management by discussing available educational resources, such as the library of fact sheets on aspects of self-management of depression, and setting personal goals for review at the next three-monthly visit.

Statistical analysis

Participants in this study were clustered under clinics by design. It is known that clinics are likely to be different from each other and that ignoring the nested nature of the data may lead to biased estimates of parameter standard errors. However, statistical techniques for correcting for the effects of clustering tend to be overly severe and conservative[23] when a small number of higher level units (the clusters) are used and, therefore, we tested whether the clinics were in fact significantly different from each other. ANCOVAs[24,25] were used to adjust for baseline values and test the significance of changes in depression scores between clinics after six months, using STATA version 11.1 for the statistical analyses.

Of the five clinics in the intervention (clinics 4, 5, 13, 15 and 17), only clinics 4 and 17 were significantly different from each other ($F(1,76)=9.6$, $p<0.01$). No other comparisons were significant between intervention clinics. Of the six clinics in the control group (clinics 1, 2, 3, 6, 16 and 18), only clinics 6 and 18 were significantly different from other ($F(1,78)=14.5$, $p<0.01$). No other comparisons were significant between control clinics. Furthermore, the intra-correlation coefficient (ICC) of 0.058 for the primary outcome suggests that only 6% of variance could be attributed to the clinics' level. Given this lack of difference between the clinics in each arm coupled with the sample-size requirements for reliable multilevel modelling[26], we analysed our data at the patient level.

In order to compare the effectiveness of the TrueBlue care model to the usual-care control, ANCOVAs were used to adjust for baseline values and test the significance of changes in continuous variables between the two groups after six months. A multi-level mixed-effects logistic regression (STATA's xtmelogit) was used to test the significance of changes in the binary (categorical) variables between the two groups after six months, with time and group as the independent variables and with random effects at the patient level. (We used mixed-effects logistic-regression model since the pairs of observations over time are not independent, i.e. observations at six months would be expected to be related to the initial baseline observations.) Within each group, changes between the two time points (baseline and six-month visits) were tested using paired *t*-tests for the continuous variables and matched-case-control McNemar χ^2 tests for the binary variables.

The longer-term effects of the intervention were evaluated over the 12-month period using multi-level mixed-effects linear regression (STATA's xtmixed) for the continuous variables and multi-level mixed-effects logistic regression (xtmelogit) for the binary variables. All three-monthly data available in the intervention group over the twelve months were used. Note that the study design could not collect such "usual care" data from the control clinics since the data collection-protocol was part of the intervention. In addition, TrueBlue training was provided to these clinics at six-months after which they ceased to be a control.

Patients from the clinics that withdrew before or during collection of first-visit data were excluded from the analyses. (Data for the thirteen patients from one of these clinics who did complete the study have been included.) Available clinics' characteristics were compared between early dropouts and participating clinics and addressed in terms of their possible impact on the generalisability of the results. Missing six-month data were replaced with their baseline values using the "no change" formulation of intention-to-treat by assuming that no change occurred between baseline and six months. The underlying assumptions of the statistical tests used were assessed.

Results

Demographics (Table 1)

A total of 5401 invitations (3104 intervention and 2297 control; see figure 1) were posted to patients with either T2DM or CHD (or both) identified in the clinics' registers. Approximately 30% (1057 intervention and 537 control, including 39 additional patients invited in the waiting room) invitations were returned with completed constant forms and PHQ-9 questionnaires. This proportion is typical in studies of this type reported in the literature. Of these, 34% (300 intervention and 229 control) were eligible (a depression score or 5 or more) and were invited to participate. However,

25% of these (94 intervention and 36 control) did not commence when their clinics withdrew before data collection began.

Of the 206 patients in the intervention who commenced the study (figure 1), 17% (*n*=36) were forced to leave when their clinics withdraw the study. A further 14% (*n*=28) patients withdrew as the study progressed, with 4% leaving after 6 months, 5% after 9 months and 5% after the full year. Reasons included leaving the area, going into residential care or becoming too ill to continue, but no consistent pattern could be identified. (Exact numbers for each reason are not known.) In the control group, 24% (*n*=47) of the 194 patients who agreed to participate had forgotten about the study by the time the 6-month review was to be undertaken and did not want to proceed.

Table 1 presents the characteristics of the patients in both the intervention and control groups who commenced the study, and shows that these characteristics were similar across both groups. There were no significant differences in patient characteristics between the intervention and control at baseline.

Phase 1: Comparison of outcomes between control and intervention groups after 6 months (Table 2)

Table 2 presents baseline and 6-month data for markers used to monitor control of chronic disease for both the intervention and the control group. While the six-month depression scores for all 310 patients (164 intervention and 146 control) were significantly lower than those at baseline in both the intervention group (10.7±0.7 reducing to 7.1±0.8, $t(163)=8.38$, $p<0.001$) and the control group (11.6±0.9 reducing to 9.0±0.9, $t(145)=6.01$, $p<0.001$), the ANCOVA adjusting for the baseline scores showed that the improvement was significantly better in the intervention group than in the control ($F(1,309)=6.40$, $p=0.012$). (The 95% confidence ranges are indicated by the plus-minus sign.)

Half of the patients had only mild-depression at baseline (PHQ-9 scores between five and nine). Because the reported score for many of these patients may be due to their diabetes rather than depression[20], the intervention is unlikely to be able to change these scores. This is one reason that Katon and colleagues used a score of ten or more as an inclusion criterion in their study[15]. Consequently, we examined the change to baseline PHQ-9 scores for the 164 patients (81 intervention and 83 control) with moderate to severe depression (PHQ-9 scores of 10 or more) at baseline. These patients showed significant improvement, with the mean depression score in the intervention group dropping by 5.7±1.3, from 14.4±1.1 down to 8.7±1.3 ($t(80)=9.00$, $p<0.001$), a clinically-significant change[19]. The improvement in the intervention for these patients was significantly better than in the control group ($F(1,161)=4.02$, $p=0.047$) where the depression score dropped by 4.3±1.2, from 15.1±1.1 down to 10.8±1.4 ($t(82)=6.88$, $p<0.001$).

Except for the HDL measurements, there were no significant changes in biophysical measures after six months in either group. Smoking rates were low at baseline in the patients with established cardiovascular risk factors. Recording of alcohol was sub-optimal, although better than other Australian primary care surveys[27].

The intervention group also showed significantly greater number of patients exercising, referred to and attending an exercise program, and referred to and attending a mental health worker after six months of collaborative care. In the control group, there were no significant changes observed after six months, except that referrals to a mental health worker increased significantly (p<0.001) from 9% to 21%, consistent with the action being taken by the nurses as required by the protocol. Neither group showed any significant changes in the number of patients taking antidepressant medication.

Phase 2: Chronic disease outcomes over 12 months using TrueBlue collaborative care (Table 3)

Table 3 presents data at baseline and 12 months for the intervention group for markers used to monitor control of existing diabetes and CHD. The improvement in mental health observed after 6 months was maintained at 12 months, with the significant reduction in the mean depression score maintained (10.7±0.7 to 6.6±0.7, $t(163)=9.92$, p<0.001), and nearly 70% of patients having lower depression scores than baseline after one year. Patients with moderate to severe depression at baseline showed an even greater improvement after 12 months of collaborative care, with the mean depression score dropping by 6.4±1.2, from 14.4±0.8 to 8.0±1.2 ($t(80)=10.41$, p<0.001). The significant improvement in the mean SF36v2 composite mental-health and physical-health scores observed after 6 months was also maintained at 12 months.

Physiological measures showed a trend, although not significant, to improvement in weight, systolic blood pressure and HDL. Mean baseline lipids and HbA1c were close to guideline targets. The 10-year CVD risk calculated with the Framingham risk equations[28] suggests a small but significant (p=0.015) reduction in risk from 27.4% to 24.8% for the patients with only T2DM. (The Framingham risk equations cannot be used for those patients who have CHD.)

The most notable changes in lifestyle after 12-months of the intervention were a significant increase in the numbers of patients who reported taking regular exercise or being referred to an exercise program. Reported referrals and visits to a mental health worker and numbers taking antidepressant medication were also significantly greater at 12 months.

The TrueBlue protocol also included goal setting so that patients could become more pro-active in their own care. An analysis of participant goals revealed that two-thirds of visits resulted in at

least one behavioural activation goal being set and, over the course of the study, 86% of patients identified a behavioural activation goal.

Adherence to guidelines (Figure 2)

Figure 2 shows the percentage of TrueBlue patients who had psychosocial and biophysical checks undertaken as recommended by the Australian National Heart Foundation and Diabetes Australia guidelines, with the corresponding percentages for usual care taken from a study of a large sample of Australian general practices[27].

Discussion

Outcomes of phase 1

Depression scores were significantly lower at six months for patients in the intervention practices compared with those in the control group, and the improvement was clinically significant for patients with moderate to severe depression[19], with patients moving one depression category. Patients experienced increased nurse contact time through the nurse consultations, were provided with information about mental health and their physical health through psycho-education resources, and had their treatment intensified when required. Modalities included behavioural activation, antidepressant medication, and referrals to mental health professionals and exercise programs. Similar improvements in depression scores and stepped-up care were observed in the collaborative care model of Katon and colleagues[15]. The reduction in depression scores observed in the control group could be explained, in part, by control practices being provided with each patient's entry-level depression score during the recruitment process as part of the study's safety protocol. Usual care could have been influenced by drawing attention to co-morbid depression[15] as the protocol required that practice nurses take action if severe depression was recorded or if the patient had responded to the suicidal-ideation question. Referrals to mental-health workers by the control clinics had increased significantly consistent with the clinics taking action where warranted. It is also known[29] that recruiting interested patients (those who wanted to participate) from interested clinics (those that agreed to join) can affect the representativeness of the study population. GPs with a particular interest in the study may be more likely to participate and may manage their patients more effectively, irrespective of whether they are in the control or intervention arms. Consequently, a reduction in depression scores in the control group was expected but the structured TrueBlue model did produce a significantly better reduction in depression. While the effect size may be small (Cohen's $f = 0.15$), it is important to note that TrueBlue was designed to be implemented easily within general practices, with running costs funded by existing Australian Medicare rebates, and to make better use of their existing resources. These features mean that TrueBlue could be easily

applied to patients across general practices at a population level, making the benefits clinically important.

Outcomes of phase 2

The key clinical outcomes over a 12-month period in the intervention group (Table 3) were a sustained improvement in mental health, demonstrated by symptom severity score (PHQ-9 total score) and by the patient's function and subjective evaluation of mental health (SF36 mental health composite score) and physical health (SF36 physical health composite score). Regular physical exercise has been shown to be important for reducing depression[30]. The self-reported exercise rates showed significant improvement over the 12 months of collaborative care intervention. The biophysical measures reported in Table 3 showed modest improvements after 12 months and the Framingham risk equations[28] suggest a small but significant reduction in the 10-year CVD risk for the T2DM patients. These improvements were achieved despite that fact that we did not specifically select patients whose physiological parameters exceeded guidelines. Rather, our recruitment process selected from the practice's disease registry on the basis of only the presence of depression and T2DM or CHD and, consequently, many patients were already being treated to target on measures such as cholesterol and HbA1c, leaving little room for improvement.

Limitations

We were able to run TrueBlue only in practices that used clinical software, which we used to generate a disease registry from which participants could be selected, and had a practice nurse on staff. Clinics that chose to take part in the study may not have been representative of wider general practice. Operational limitations further reduced the number of practices over the duration of the study. Patient response rates to the mail-out (28%) may reflect anxiety over the new model of care where the patient discloses depression and visits the PN first rather than only the GP. Usual care in the control clinics may have been changed by patients completing the PHQ-9 and reading the project description. GPs were made aware of individual PHQ-9 results, and took action where warranted. GP awareness of these biophysical and lifestyle risks may be expected to change clinical management. By design, TrueBlue practices needed to incorporate all research activities within the context of their busy clinics, and so only research data that could be extracted automatically were collected. The data dropout resulting from these two factors contributed to the observed small effect size. We were not able to obtain multiple data sets at three-monthly intervals over 12 months of 'usual care' because the act of inviting patients and measuring depression scores and biophysical measures would in itself change the nature of usual care. In addition, practices would not have been willing to join the study if there was a chance of being randomly allocated to 12 months of being in such a control arm[29].

Collaborative Care

A recent UK study has shown the difficulties of disseminating a guideline without guidance on how to implement collaborative care. Organisational barriers included GPs finding the PHQ-9 awkward to use, nurses not feeling confident or competent due to lack of training, and no guidance on stepped care[13]. The TrueBlue model of collaborative care overcame many of these difficulties. Its successful components were:[9,31]

- *Use of evidence based guidelines.* National Heart Foundation and Diabetes Australia guidelines determined disease management targets and frequency of monitoring.
- *Systematic screening and monitoring of risk factors.* Patients attended three monthly visits in which a care plan with its checklist was completed. By providing a comprehensive collation of all necessary information, this document made clinical management by the patient's GP easier, quicker and more accurate.
- *Timetabled recall visits.* The date of the next appointment was set during each visit, PHQ-9 was readministered and if improvement was insufficient, *stepped care* was followed by initiating drug therapy or increasing the dose, or referral to a mental-health worker according to the guidelines.
- *New or adjusted roles for team members.* PNs took responsibility for organising and monitoring the outcome of referrals, goals and targets. They used a depression questionnaire (the PHQ-9) to open a discussion with patients about their depression symptoms.
- *Information support for the clinician.* GPs were provided with the care plan by the PNs.
- *Enhanced patient self-management.* Patients received their own copy of the care plan with personalised goals, current measurements, targets and safety advice. A component of each visit was to discuss and update their plan, and receive education material on depression.
- *Identified case manager.* PNs became case managers but the GP remained the key clinician.
- *Means of effective communication between all members of the care team.* The care plan was designed to provide relevant clinical information in a succinct format while still being comprehensible to patients.
- *Audit information for the practice.* De-identified data was provided automatically through the care plan.

Applicability of TrueBlue

TrueBlue used existing clinical software and improved the focus of the GP consultation by delegating some tasks to the PN. Higher levels of adherence to guideline-recommended checks were also reported for TrueBlue. Patients and their carers, allied health professionals, specialists and general practitioners gained ready access to patient details provided in TrueBlue's care plan

enabling them to see at glance where improve clinical care may be needed. The study achieved improved outcomes with the potential for prevention of heart attack and stroke through reduced 10-year CVD risk. The care plan template also allowed the practice to collect high quality audit data without taking up clinical time. While it was not possible to obtain complete financial data from the clinics specifically relating to the TrueBlue visits, the data that are available suggest that clinics did indeed cover their costs in implementing TrueBlue through Australian Medicare rebates. The success of TrueBlue and TeamCare[15] demonstrates that collaborative care is feasible in routine general practice in Australia and the USA, and could lead to improved outcomes for patients with depression and other chronic diseases[7,32].

Acknowledgements

The authors wish to thank the patients, practice nurses, general practitioners and support staff of the participating clinics Evans Head Medical Centre, Lennox Head Medical Centre, Keen Street Clinic, Tintenbar Medical Centre, Health on Grange, Warradale Medical Centre, Seaton Medical and Specialist Centre, Woodcroft Medical Centre and Southcare Medical Centre. Professors Wayne Katon and Juergen Unützer of the University of Washington were unstinting in their advice on adaptation of the IMPACT model. We would also like to thank Bob Leahy for managing the project, Vince Versace for his statistical advice and Vicki Brown for her assistance during the course of the study.

Contributors

All authors have full access to the complete study dataset, contributed to the design, implemented the project, and co-wrote and approved the manuscript. MM, MC, JD and PR analysed the data. MM, PR and KS developed and ran the practice nurse training program. JD and PR conceived the TrueBlue model during a visit to the IMPACT team. JD is the guarantor.

Funding

Funding was provided by *beyondblue*, the National Depression Initiative in Australia, but it had no other involvement in any phase of the study. All researchers were independent of and not influenced by *beyondblue.*

Competing interests

All authors have completed the Unified Competing Interest form and declare that (1) funding was received from *beyondblue* to carry out the study; (2) they do not have any relationships with any companies that might have an interest in the submitted work in the previous 3 years; (3) their

spouses, partners, or children do not have any financial relationships that may be relevant to the submitted work; and (4) they do not have any non-financial interests that may be relevant to the submitted work.

Ethical approval

Ethics approval was obtained from Flinders University's Social and Behavioural Research Ethics Committee. All patients gave informed consent to participate in the study.

Data sharing

The dataset used for the analysis and the computer codes used to produce the results are available from the corresponding author.

Licence

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in BMJ editions and any other BMJPGL products and sub-licences to exploit all subsidiary rights, as set out in our licence (http://resources.bmj.com/bmj/authors/checklists-forms/licence-for-publication).

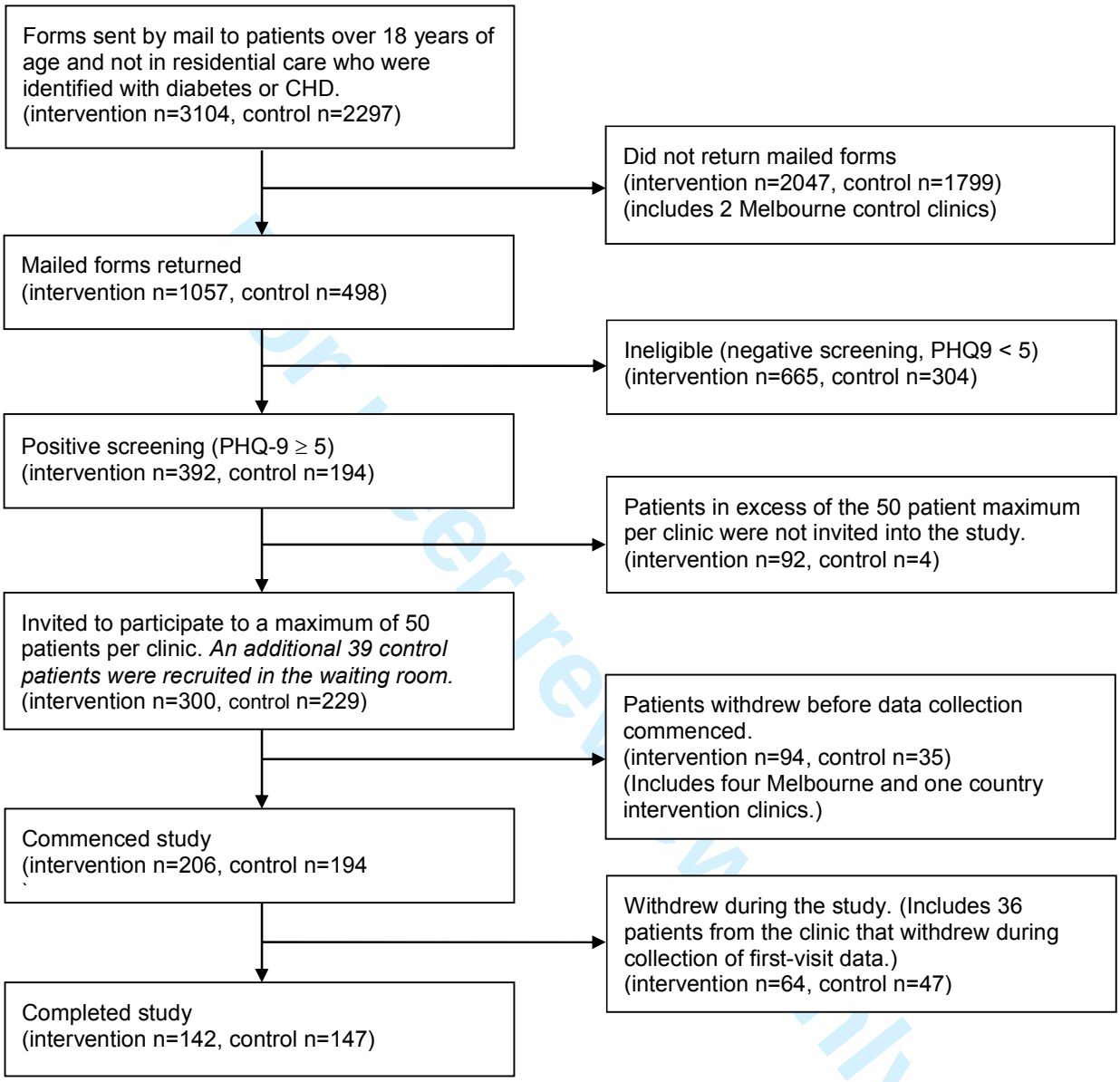
References

1. Daar AS, Singer PA, Persad DL, Pramming SK, Matthews DR, Beaglehole R, et al. Grand challenges in chronic non-communicable diseases. *Nature* 2007;450(7169):494-96.
2. World Health Organisation. Preventing Chronic Disease: A Vital Investment. Geneva, 2005.
3. Ali S, Stone MA, Peters JL, Davies MJ, Khunti K. The prevalence of co-morbid depression in adults with Type 2 diabetes: a systematic review and meta-analysis. *Diabet Med* 2006;23(11):1165-73.
4. Egede LE, Nietert PJ, Zheng D. Depression and all-cause and coronary heart disease mortality among adults with and without diabetes. *Diabetes Care* 2005;28(6):1339-45.
5. Clarke DM, Currie KC. Depression, anxiety and their relationship with chronic diseases: a review of the epidemiology, risk and treatment evidence. *Med J Aust* 2009;190(7 Suppl):S54-60.
6. Freeling P, Rao BM, Paykel ES, Sireling LI, Burton RH. Unrecognised depression in general practice. *Br Med J (Clin Res Ed)* 1985;290(6485):1880-83.
7. Gilbody S, Bower P, Fletcher J, Richards D, Sutton AJ. Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. *Arch Intern Med* 2006;166(21):2314-21.
8. Katon W, Von Korff M, Lin E, Walker E, Simon GE, Bush T, et al. Collaborative management to achieve treatment guidelines. Impact on depression in primary care. *JAMA* 1995;273(13):1026-31.
9. Hickie IB, McGorry PD. Increased access to evidence-based primary mental health care: will the implementation match the rhetoric? *Med J Aust* 2007;187(2):100-03.
10. Lin EH, Katon W, Von Korff M, Rutter C, Simon GE, Oliver M, et al. Relationship of depression and diabetes self-care, medication adherence, and preventive care. *Diabetes Care* 2004;27(9):2154-60.
11. Reddy P, Ford D, Dunbar JA. Improving the quality of diabetes care in general practice. *Aust J Rural Health* 2010;18(5):187-93.
12. Campbell MJ. Cluster randomized trials in general (family) practice research. *Stat Methods Med Res* 2000;9(2):81-94.
13. Mitchell C, Dwyer R, Hagan T, Mathers N. Impact of the QOF and the NICE guideline in the diagnosis and management of depression: a qualitative study. *Br J Gen Pract* 2011;61(586):e279-e89.
14. Unutzer J, Katon W, Williams JW, Jr., Callahan CM, Harpole L, Hunkeler EM, et al. Improving primary care for depression in late life: the design of a multicenter randomized trial. *Med Care* 2001;39(8):785-99.
15. Katon WJ, Lin EH, Von Korff M, Ciechanowski P, Ludman EJ, Young B, et al. Collaborative care for patients with depression and chronic illnesses. *N Engl J Med* 2010;363(27):2611-20.
16. Morgan MA, Dunbar J, Reddy P. Collaborative care - The role of practice nurses. *Aust Fam Physician* 2009;38(11):925-26.
17. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16(9):606-13.
18. Morgan M, Dunbar J, Reddy P, Coates M, Leahy R. The TrueBlue study: is practice nurse-led collaborative care effective in the management of depression for patients with heart disease or diabetes? *BMC Fam Pract* 2009;10:46.
19. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. *Medical care* 2004;42(12):1194-201.
20. Reddy P, Philpot B, Ford D, Dunbar JA. Identification of depression in diabetes: the efficacy of PHQ-9 and HADS-D. *Brit J Gen Prac* 2010;DOI: 10.3399/bjgp10X502128.
21. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30(6):473-83.
22. Dobson KS, Hollon SD, Dimidjian S, Schmaling KB, Kohlenberg RJ, Gallop RJ, et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication

in the prevention of relapse and recurrence in major depression. *Journal of Consulting and Clinical Psychology* 2008;76(3):468-77.

23. Kauermann G, Carroll RJ. A Note on the Efficiency of Sandwich Covariance Matrix Estimation. *J Amer Statistical Assoc* 2001;96(456):1387-96.
24. Senn S. Change from baseline and analysis of covariance revisited. *Statistics in Medicine* 2006;25(24):4334-44.
25. Vickers A. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: A simulation study. *BMC Medical Research Methodology* 2001;1:1-4.
26. Hox JJ. *Applied multilevel analysis*. Second ed. Amsterdam: TT-Publikaties, 1995.
27. Wan Q, Harris MF, Jayasinghe UW, Flack J, Georgiou A, Penn DL, et al. Quality of diabetes care and coronary heart disease absolute risk in patients with type 2 diabetes mellitus in Australian general practice. *Qual Saf Health Care* 2006;15(2):131-35.
28. Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J* 1991;121(1 Pt 2):293-98.
29. Wilson S, Delaney BC, Roalfe A, Roberts L, Redman V, Wearn AM, et al. Randomised controlled trials in primary care: case study. *BMJ* 2000;321:24-27.
30. Wiles NJ, Haase AM, Gallacher J, Lawlor DA, Lewis G. Physical activity and common mental disorder: results from the Caerphilly study. *Am J Epidemiol* 2007;165(8):946-54.
31. Fuller J, Perkins D, Parker S, Holdsworth L, Kelly B, Roberts R, et al. Effectiveness of service linkages in primary mental health care: a narrative review part 1. . *BMC Health Services Research* 2011;11:71.
32. Richards DA, Lovell K, Gilbody S, Gask L, Torgerson D, Barkham M, et al. Collaborative care for depression in UK primary care: a randomized controlled trial. *Psychol Med* 2008;38(2):279-87.

Figure 1. CONSORT flow diagram of the recruitment process.



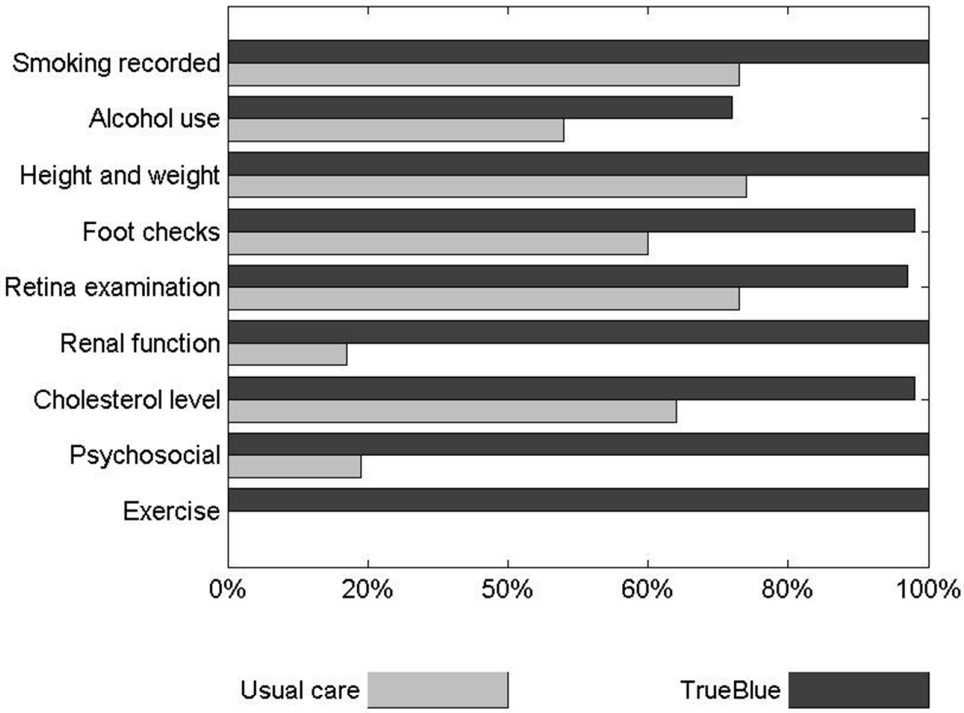


Figure 2: Recording of checks recommended by the National Heart Foundation and Diabetes Australia guidelines. Data for "usual care" were adapted from reference 27. No usual-care data were available for exercise.

Table 1: Patient characteristics at the baseline visits. There were no significant differences between the intervention and control at baseline.

Characteristic	Intervention group (*n*=170)	Control group (*n*=147)
Male (%) / Female (%)	51.8% / 48.2%	55.2% / 44.8%
Age (yr)	68.0 ± 11.7	67.6 ± 11.2
Aboriginal or Torres Strait Islander (%)	0.0%	0.7%
Diagnosis: Type-2 diabetes	37.6%	47.6%
CHD	45.3%	35.8%
Both	17.1%	16.6%
Body mass index (kg/m^2)	31.4 ± 6.0 (*n*=170)	30.8 ± 6.0 (*n* =103)
Systolic blood pressure (mmHg)	134.1 ± 19.0 (*n*=169)	133.5 ± 19.6 (*n* =112)
Total cholesterol (mmol/L)	4.21 ± 0.94 (*n*=165)	4.41 ± 1.06 (*n* =110)
Triglycerides (mmol/L)	1.73 ± 0.88 (*n*=165)	1.92 ± 1.37 (*n* =105)
LDL (mmol/L)	2.22 ± 0.74 (*n*=159)	2.37 ± 0.88 (*n* =89)
HDL (mmol/L)	1.23 ± 0.36 (*n*=159)	1.18 ± 0.33 (*n* =97)
HbA1c (mmol/L)	7.00 ± 1.21 (*n*=94)	7.19 ± 1.42 (*n* =69)
PHQ-9 score	10.7 ± 4.7 (*n*=164)	11.6 ± 5.5 (*n*=146)
PHQ-9 score range at baseline	5 to 24	5 to 27

Table 2: TrueBlue outcomes at six months in the intervention and control groups. (See Table 3 for a list of abbreviations.)

	Intervention				Control				Between groups
	n	Baseline	6-months	Within	*n*	Baseline	6-months	Within	
PHQ9 depression score	164	10.7±0.8	7.1±0.8	$p<0.001^{\dagger}$	146	11.6±0.9	9.0±0.9	$p<0.001^{\ddagger}$	$p=0.012$
SF36v2 mental-health score[§]	71	37.2±3.4	41.1±3.4	$p=0.034^{\dagger}$		Not recorded			NS
SF36v2 physical-health score[§]	71	39.9±2.2	42.5±2.6	$p=0.023^{\dagger}$		Not recorded			NS
Body mass index (kg/m^2)	162	31.3±1.0	31.2±1.0	NS	103	30.8±1.2	31.0±1.0	NS	NS
Waist (cm)	161	104.7±2.4	105.0±2.4	NS	80	104.2±4.0	105.8±3.2	NS	NS
Systolic blood pressure (mmHg)	161	134.2±3.0	132.4±2.8	NS	112	133.5±3.8	131.2±3.4	NS	NS
Total cholesterol (mmol/L)	158	4.21±0.16	4.22±0.14	NS	109	4.41±0.20	4.44±0.20	NS	NS
LDL (mmol/L)	154	2.23±0.12	2.17±0.14	NS	86	2.37±0.18	2.29±0.20	NS	NS
HDL (mmol/L)	154	1.23±0.06	1.29±0.06	$p=0.023^{\dagger}$	93	1.17±0.06	1.27±0.08	$p=0.011^{\ddagger}$	NS
Triglycerides (mmol/L)	158	1.72±0.14	1.66±0.12	NS	104	1.84±0.22	1.75±0.18	NS	NS
HbA1c (%)[¶]	89	6.97±0.24	6.90±0.26	NS	67	7.22±0.34	7.40±0.36	NS	$p=0.049$
Ten-year CVD risk[*]	61	26.9±3.2	26.1±3.2	NS	46	26.3±3.6	24.7±3.2	NS	NS
The 95% confidence ranges are indicated by the plus-minus (±) sign. Note that lower scores indicate improvement for all items except the SF36v2 and HDL results, where higher scores indicate improvement.									
Smoking	162	15 (9%)	13 (8%)	NS	110	13 (12%)	13 (12%)	NS	NS
Alcohol	104	47 (45%)	51 (49%)	NS	42	27 (64%)	27 (64%)	NS	NS
Exercises 30 mins/day, 5 d/wk	162	66 (41%)	97 (60%)	$p<0.001^{\dagger}$	75	22 (29%)	22 (29%)	NS	$p<0.001$
Referred to exercise program	162	32 (20%)	58 (36%)	$p<0.001^{\dagger}$	111	15 (14%)	10 (9%)	NS	$p<0.001$
Attends exercise program	162	12 (7%)	23 (14%)	$p=0.041^{\dagger}$	79	12 (15%)	9 (11%)	NS	NS
On antidepressant medication	162	27 (17%)	34 (21%)	NS	113	31 (27%)	36 (32%)	NS	$p=0.025$
Referred to mental health worker	162	47 (29%)	58 (36%)	$p=0.022^{\dagger}$	114	10 (9%)	24 (21%)	$p<0.001^{\ddagger}$	$p<0.001$
Attends mental health worker	162	10 (6%)	37 (23%)	$p<0.001^{\dagger}$	109	14 (13%)	11 (10%)	NS	$p=0.044$

The values in brackets are the percentages of the total *n*.

(NS) No significant difference. (†) Significant difference between baseline and six-months within the *intervention* clinics. (‡) Significant difference between baseline and six-months within the *control* clinics. (§) SF36v2 questionnaires were not collected by all clinics. (¶) HbA1c results were only available for patients with T2DM. (*) CVD risk could only be calculated for patients with T2DM.

Table 3: TrueBlue outcomes at 12 months within the intervention clinics only.

	Intervention			
	n	Baseline	12-months	
PHQ9 depression score	164	10.7±0.7	6.6±0.7	$p<0.001$
SF36v2 mental-health score[§]	70	36.0±3.2	41.3±2.8	$p<0.001$
SF36v2 physical-health score[§]	70	40.6±2.2	44.3±2.8	$p<0.001$
Body mass index (kg/m^2)	142	31.4±1.0	31.1±1.0	$p=0.006$
Waist (cm)	141	105.0±2.4	105.2±2.6	NS
Systolic blood pressure (mmHg)	141	135.2±3.2	130.2±3.0	$p=0.016$
Total cholesterol (mmol/L)	138	4.18±0.16	4.28±0.16	NS
LDL (mmol/L)	135	2.19±0.12	2.24±0.20	NS
HDL (mmol/L)	135	1.22±0.06	1.36±0.08	$p<0.001$
Triglycerides (mmol/L)	138	1.73±0.16	1.63±0.14	$p=0.004$
HbA1c (%)[¶]	79	7.01±0.26	7.04±0.28	NS
Ten-year CVD risk[*]	55	27.4±3.4	24.9±3.6	$p=0.015$

The 95% confidence ranges are indicated by the plus-minus (±) sign. Lower scores indicate improvement for all items except the SF36v2 and HDL results, where higher scores indicate improvement.

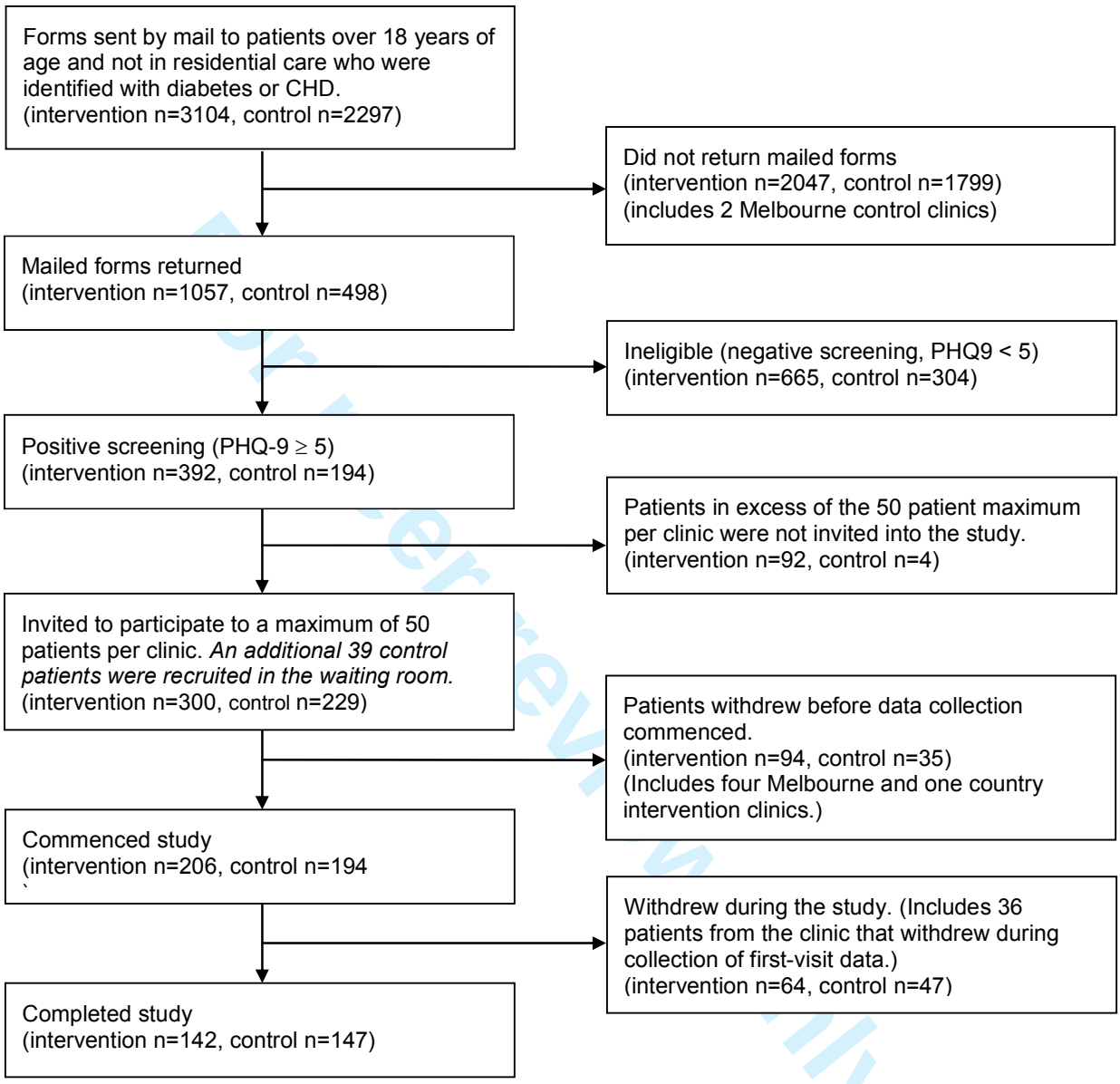
Smoking	142	15 (11%)	11 (8%)	NS
Alcohol	95	45 (47%)	47 (49%)	NS
Exercises 30 mins/day, 5 days/wk	142	57 (40%)	83 (58%)	$p<0.001$
Referred to exercise program	142	26 (18%)	53 (37%)	$p<0.001$
Attends exercise program	142	10 (7%)	17 (12%)	NS
On antidepressant medication	142	22 (15%)	33 (23%)	$p=0.001$
Referred to mental health worker	142	40 (28%)	59 (42%)	$p<0.001$
Attends mental health worker	142	8 (6%)	25 (18%)	$p<0.001$

The values in brackets are the percentages of the total *n*.

(NS) No significant difference. (§) SF36v2 questionnaires were not collected by all clinics. (¶) HbA1c results were only available for patients with T2DM. (*) CVD risk could only be calculated for patients with T2DM.

NOTE:
PHQ9 is the 9-question Patient Health Questionnaire.
SF36v2 is version 2 of the Short Form 36-Question health survey.
LDL is low-density lipoprotein.
HDL is high-density lipoprotein.
Glycated haemoglobin (HbA1c) was measured only for patients with diabetes.
Unit of alcohol is 10g of ethanol.

Figure 1. CONSORT flow diagram of the recruitment process.



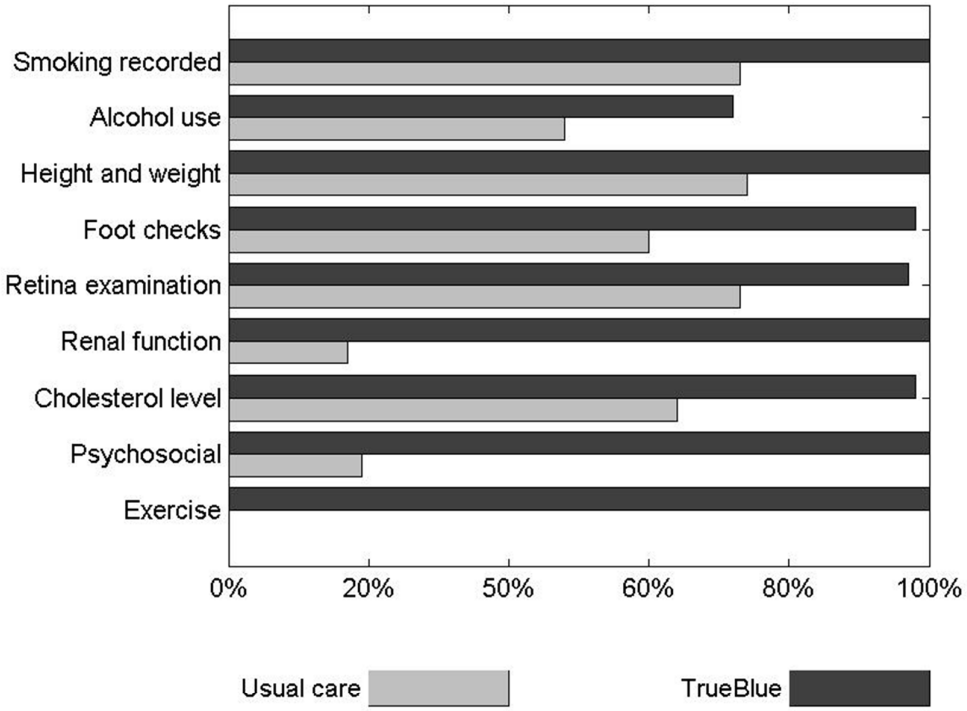


Figure 2: Recording of checks recommended by the National Heart Foundation and Diabetes Australia guidelines. Data for "usual care" were adapted from reference 27. No usual-care data were available for exercise.

The TrueBlue model of collaborative care using practice nurses as case managers for depression alongside diabetes or heart disease

MAJ Morgan[1], MJ Coates[1], JA Dunbar[1,*], P Reddy[2], K Schlicht[1] and J. Fuller[3]

1. Greater Green Triangle University Department of Rural Health, Flinders and Deakin Universities, PO Box 423, Warrnambool, Victoria, 3280, Australia
2. School of Medicine and Public Health, University of Newcastle, Callaghan, NSW, 2308, Australia
3. School of Nursing and Midwifery, Flinders University, GPO Box 2100, Adelaide, SA, 5001, Australia

* Corresponding author

Abstract

Objectives: Determine the effectiveness of collaborative-care in reducing depression in primary-care patients with diabetes or heart disease using practice nurses as case managers.

Design: A two-arm open randomised-cluster trial with wait-list control for 6 months. The intervention was followed over 12 months.

Setting: Eleven Australian general practices, five randomly allocated to the intervention and six to the control.

Participants: ~~404~~ 400 primary-care patients (~~211~~ 206 intervention, ~~193~~ 194 control) with depression and type 2 diabetes, coronary heart disease or both.

Intervention: The practice nurse acted as case manager identifying depression, reviewing pathology results, lifestyle risk factors, and patient goals and priorities. Usual care continued in the controls.

Main outcome measure: A five-point reduction in depression scores for patients with moderate to severe depression. Secondary outcome was improvements in physiological measures.

Results: Mean depression scores after six months of intervention for patients with moderate to severe depression decreased by 5.7±1.3 compared with 4.3±1.2 in the control, a significant (p=0.012) difference. (The plus-minus is the 95% confidence range.) Intervention practices demonstrated adherence to treatment guidelines and intensification of treatment for depression, where exercise increased by 19%, referrals to exercise programs by 16%, referrals to mental health workers (MHWs) by 7%, and visits to MHWs by 17%. Control-practice exercise did not change, referrals to exercise programs dropped by 5%, and visits to MHWs by 3%. Only referrals to MHW increased, by 12%. Intervention improvements were sustained over 12 months, with significant

(p=0.015) decrease in 10-year cardiovascular-disease risk from 27.4±3.4% to 24.8±3.8%. A review of patients indicated the study's safety protocols were followed.

Conclusion: TrueBlue participants showed significantly improved depression and treatment intensification, sustained over 12-months of intervention, and reduced 10-year CVD risk. Collaborative care using practice nurses appears to be an effective primary-care intervention.

Trial registration: Australia and New Zealand Clinical Trials Registry ACTRN12609000333213.

Article summary

Article focus

- To determine the effectiveness of a collaborative-care model to reduce depression in primary-care patients with diabetes or heart disease.
- To determine the effectiveness of using practice nurses as case managers of patients with depression and diabetes, heart disease or both.

Key messages

- The TrueBlue model of collaborative care can be introduced within the general practice workforce with practice nurses taking on the role of case manager.
- Practice nurses can improve care of depression in patients with diabetes or heart disease, leading to better outcomes and reduced 10-year CVD risk.
- The care of patients using the TrueBlue model is closer to "best practice" guidelines, with substantially better levels of adherence to guideline-recommended checks than occur in usual care.

Strengths and limitations of this study

- The TrueBlue model of collaborative care overcomes many of the difficulties in implementing a guideline for the treatment of co-morbid depression.
- The study's purpose-designed care plan gives patients and their carers, allied health professionals, specialists and general practitioners ready access to patient details enabling them to see at glance where improve clinical care may be needed.
- Clinics were able to recover the costs of the collaborative care through Australian Medicare rebates.

- The study could only be run in practices that had a practice nurse on staff and had access to clinical software capable of generating a disease registry from which participants could be selected.

- Differences between TrueBlue- and control-practice outcomes may have been reduced by patients completing the PHQ-9 depression questionnaire and reading the project description, and by GPs being made aware of individual PHQ-9 results so that they could take action where warranted.

Introduction

Managing diabetes and heart disease has been highlighted as one of the global "grand challenges in chronic non-communicable diseases"[1] because the prevalence of these two preventable diseases is increasing[2]. Along with depression, they have been identified as health priority areas in many countries. A vicious cycle exists between depression and these chronic diseases, with each being a risk factor for the other[3]. Higher mortality has been demonstrated for people with depression and type 2 diabetes (T2DM) or coronary heart disease (CHD) beyond that due to the separate diseases alone[4]. For patients with depression and T2DM or CHD or both, there are increased risks of adverse outcomes[5] but this co-morbid depression is often missed in primary care[6]. Consequently, the identification of depression has now been incorporated in many heart disease guidelines as one of the requirements for optimal management. Meeting these challenges will require an innovative use of the existing general practice workforce and such a re-orientation of resources has been identified as one of the grand challenges[1].

Collaborative care is a system that has been shown to be more effective for chronic disease management than standard care[7]. It includes a reorientation of the medical workforce through new or adjusted roles for team members, particularly using practice nurses as the identified case manager[8,9]. It also includes the use of evidence based guidelines, systematic screening and monitoring of risk factors, timetabled recall visits, information support for the clinician, enhanced patient self-management, a means of effective communication between all members of the care team and audit information for the practice. Since self-care for diabetes has been found to be suboptimal across a range of self-managed activities, particularly for patients with depression, a collaborative care model may be able to achieve better quality of care through the case manager monitoring patient progress[10,11].

Evaluation of a change in the way general practice clinics look after patients requires complex intervention methodology[12] beyond single interventions such as introduction of a guideline with financial incentives[13]. This methodology began with a search for potential models of care (step I), and led to adopting the University of Washington's successful IMPACT model of Collaborative Care for depression[14,15]. In the exploratory trial (step II), our pilot project[16] adapted IMPACT by

training practice nurses as case managers. Practice nurses were trained to screen for depression using a patient self-report measure, the nine-item Patient Health Questionnaire (PHQ-9)[17], as part of comprehensive chronic disease management. They were also trained to use a protocol for care management based on the depression scores. The depression screening and management was embedded in routine visits for patients with diabetes or CHD. The pilot demonstrated that it was feasible to detect, monitor and treat depression in routine general practice alongside the usual biophysical measures, and identified moderate to severe depression in 34% of participants. The TrueBlue study was a randomised cluster trial (step III) that built on and extended the pilot. It investigated whether a collaborative care model (the intervention) is better than usual care (the control) for the management of patients with depression and T2DM, CHD or both in Australian general practice. It was designed to fit into normal clinic operations, making use of practice nurses and medical software, and able to be funded by existing Australian Medicare rebates.

Methods/design

Study design

The design and methodology of the study have been described in detail elsewhere[18]. The study commenced in 2009 and was undertaken in two phases. The first phase was a cluster randomised intervention trial in which general practices were randomly allocated to either an intervention group in which nurse-led collaborative care was undertaken or to a wait-list control group in which usual care led by the general practitioner (GP) was continued. At six months, the TrueBlue training was provided to the control practices. The key aims of the first phase were to determine whether participants with moderate to severe depression in the intervention group showed at least a five-point reduction from the baseline depression scores after six months of intervention and whether this reduction was significantly better than in the control group. A five-point reduction reflects a clinically-relevant change in individuals receiving depression treatment[19]. The secondary outcome was to determine whether the intervention also led to improvements in the patients' physiological measures. The second phase followed the intervention group for an additional six months to determine how the collaborative care model affected health outcomes over a twelve-month period.

Sample size

The sample size calculation was based on detecting a 50% reduction in depression score at the 0.05 significance level with 80% power and a two-tailed test. Detecting a 50% reduction is more stringent than detecting a five-point reduction and provided some additional buffering. Using depression scores ~~extracted~~ from an earlier study (a mean of 5.5 and standard deviation of 6.1)[11], the calculation indicated that 237 patients would be required in each group. An intra-cluster correlation of 0.04 was used (S. K. Lo, pers. comm.), with a recruitment target of 50 patients per

clinic. (Fifty patients were chosen so that clinics could budget for a nurse's time to carry out the intervention with four patients each week over the 3-month cycle of care.) To allow for a 50% dropout, the study required 450 patients from nine clinics in the intervention group and the same in the control group.

Practice recruitment

Practices ~~selected from~~were identified in ~~metropolitan~~city~~, regional~~ and ~~rural~~ country areas ~~were invited to participate in the study~~ on the basis of having a practice nurse to provide the collaborative care and being able to identify eligible patients, those with CHD or T2DM or both, from their registries; these were invited to participate in the study until the eighteen clinics required by the sample-size calculation were recruited. ~~The unit of randomisation was the clinic.~~ ~~Clinics~~ They~~agreeing to participate~~ were allocated by a random number generator to either the intervention or control arm of the study. The unit of randomisation was the clinic. Five practices (3 ~~regional~~country, 2 ~~metropolitan~~city) in the intervention group and six (2 ~~regional~~country, 4 ~~metropolitan~~city) in the control group completed the study. ~~Two~~ One ~~regional~~ country intervention clinic~~s~~ withdrew whilst first-visit data were being collected when ~~their~~ its TrueBlue-trained practice nurse~~s~~ left the clinic, but some (*n*=13) patients from ~~one of these clinics~~it did complete the study and data were collected ~~for~~ from them. The study team was not able to determine why the other clinics withdrew.

Patient selection

Eligible patients were sent a postal survey that included a consent form and were asked to complete and return the enclosed PHQ-9 questionnaire, a self-report measure of depression.[17] The PHQ-9 has nine items, each scored from 0 (no problems) to 3 (problems nearly every day). The sum of the scores of the nine items will lie in one of five depression categories: none (0–4), mild (5–9), moderate (10–14), moderately severe (15–19) and severe (20–27). (While it is known that responses to some of the PHQ-9 items may overlap with diabetes symptoms[20], our pilot demonstrated that nurses and patients preferred using the PHQ-9 because the patient's response to each of its items became the basis for the problem solving and goal setting activities that were part of TrueBlue.) Patients with scores of five or above, indicating some form of depression, were invited to participate in the study. A maximum of 50 patients per practice were invited. Patients in residential care or under 18 years of age were not eligible. Figure 1 presents the CONSORT diagram of the patient-recruitment process.

Patient safety

Participation in the intervention included a series of patient visits to their practice nurse (PN) and usual GP every three months over a 12-month period. Patients in the control group continued with their "usual care". The control clinics were also provided with the PHQ-9 depression scores to ensure patient safety during the trial. The protocol required that practice nurses take action if severe depression was recorded in the returned PHQ-9 or if the patient had responded to the suicidal-ideation question (question 9) on the questionnaire. This action was to be taken *irrespective of whether the clinic was in the intervention or in the control group.*

Practice nurse training

The PN training included a two-day workshop to prepare them for their enhanced roles in nurse-led collaborative care. Topics in the workshop included identifying and monitoring depression using the PHQ-9 questionnaire, and quality of life responses using version 2 of the SF36 questionnaire[21]. Patient goal setting and problem solving were key components of the training with a particular emphasis on behavioural techniques to achieve improved mental health[22]. The training also prepared the PNs for their role as case managers including ensuring Diabetes Australia and Australian National Heart Foundation guidelines were being followed, and referrals were provided to appropriate services, such as allied health and mental health professionals, through discussion with the GPs.

Data collection

The research team developed a protocol-driven care-plan template from which study data could be extracted automatically and sent to the research team. The template was designed to be a multi-purpose document in which the patient's past medical history, current medications, allergies, biophysical and psychosocial measures, lifestyle risks, personal goals and referrals were recorded. It was designed to comply with the requirements to claim Australian Medicare rebates for care planning and to provide a checklist for "gold-standard" care. A copy of the care plan was provided to the patient as a written record of their progress.

The care-plan template collected physical measures, including body-mass index, waist circumference, weight and blood pressure, and the latest pathology results, including lipid profile, glycaemic control (HbA1c) and renal function. Data also included lifestyle risk factors, such as smoking, alcohol consumption and level of physical activity, and depression score as measured by the PHQ-9 questionnaire. Referrals to and attendance at exercise programmes and with mental health workers were also recorded, along with the patient's own goals and possible barriers to

achieving these goals. The care-plan template was used by the intervention-group clinics to acquire patient data at three-monthly intervals over a twelve-month period.

In the control group, the only complete dataset recorded using our comprehensive protocol-driven care-plan template was obtained after the six months of "usual care" when the TrueBlue training was offered to the control clinics. No baseline or three-month datasets were acquired since the study was deliberately designed to avoid changing the "usual care" that would have otherwise occurred by introducing our care-plan template. The study was designed in this way to be run pragmatically in the context of the clinics' normal activities. The only baseline measure obtained was the depression score. On completion of the study, we retrospectively collected what baseline data the control clinics routinely recorded in their electronic medical records in order to have data for two time points, baseline and six-months.

TrueBlue collaborative care

As part of the TrueBlue model, patients were scheduled to visit the practice every three months for a 45-minute nurse consult followed by a 15-minute consult with their usual GP, in which stepped care (psycho- or pharmacotherapy) was offered if depression scores had not improved or had not dropped below a value of five. The PN used the care-plan template and obtained current physical measures and reviewed recent pathology results. PNs also reviewed lifestyle risk factors. They re-administered the PHQ-9 and worked with the patient to identify possible barriers to achieving their goals and discussed ways to overcome the barriers. This information gathering phase of the consultation was an opportunity to assist the patient with self-management by discussing available educational resources, such as the library of fact sheets on aspects of self-management of depression, and setting personal goals for review at the next three-monthly visit.

Statistical analysis

Participants in this study were clustered under clinics by design. It is known that clinics are likely to be different from each other and that ignoring the nested nature of the data may lead to biased estimates of parameter standard errors. However, statistical techniques for correcting for the effects of clustering tend to be overly severe and conservative[23] when a small number of higher level units (the clusters) are used and, therefore, we tested whether the clinics were in fact significantly different from each other. ANCOVAs[24,25] were used to adjust for baseline values and test the significance of changes in depression scores between clinics after six months, using STATA version 11.21 was used for the statistical analyses.

Of the five clinics in the intervention (clinics 4, 5, 13, 15 and 17), only clinics 4 and 17 were significantly different from each other ($F(1,76)=9.6$, $p<0.01$). No other comparisons were

significant between intervention clinics. Of the six clinics in the control group (clinics 1, 2, 3, 6, 16 and 18), only clinics 6 and 18 were significantly different from other ($F(1,78)=14.5$, $p<0.01$). No other comparisons were significant between control clinics. Furthermore, the intra-correlation coefficient (ICC) of 0.058 for the primary outcome suggests that only 6% of variance could be attributed to the clinics' level. Given this lack of difference between the clinics in each arm coupled with the sample-size requirements for reliable multilevel modelling[26], we analysed our data at the patient level.

In order to compare the effectiveness of the TrueBlue care model to the usual-care control, ANCOVAs were used to adjust for baseline values and test the significance of changes in continuous variables between the two groups after six months. ~~M~~A multi-level mixed-effects logistic regression (STATA's xtmelogit) was used to test the significance of changes in the binary (categorical) variables between the two groups after six months, with ~~–~~time and group as the independent variables and with random effects at the patient level. (We used mixed-effects logistic-regression model since the pairs of observations over time are not independent, i.e. observations at six months would be expected to be related to the initial baseline observations.) Within each group, changes between the two time points (baseline and six-month visits) were tested using paired *t*-tests for the continuous variables and matched-case-control McNemar χ^2 tests for the binary variables.

The longer-term effects of the intervention were evaluated ~~by assessing the three monthly changes~~ over the 12-month period using multi-level ~~linear~~ mixed-effects linear regression ~~models~~ (STATA's xtmixed) for the continuous variables and multi-level mixed-effects logistic regression (xtmelogit) for the binary variables. All three-monthly data available in the intervention group over the twelve months were used. Note that the study design could not collect ~~12-months~~such "usual care" data from the control clinics since the data collection-protocol was part of the intervention. In addition, TrueBlue training was provided to these clinics at six-months after which they ceased to be a control.

Patients from the clinics that withdrew before or during collection of first-visit data were excluded from the analyses. (Data for the thirteen patients from one of these clinics who did complete the study have been included.) Available clinics' characteristics were compared between early dropouts and participating clinics and addressed in terms of their possible impact on the generalisability of the results. Missing six-month data were replaced with their baseline values using the "no change" formulation of intention-to-treat by assuming that no change occurred between baseline and six months. The underlying assumptions of the statistical tests used were assessed. ~~STATA version 11.2 was used for the statistical analyses.~~

Results

Demographics (Table 1)

A total of 5401 invitations (3104 intervention and 2297 control; see figure 1) were posted to patients with either T2DM or CHD (or both) identified in the clinics' registers. Approximately 30% (1057 intervention and 537 control, including 39 additional patients invited in the waiting room) invitations were returned with completed constant forms and PHQ-9 questionnaires. This proportion is typical in studies of this type reported in the literature. Of these, 34% (300 intervention and 229 control) were eligible (a depression score or 5 or more) and were invited to participate. However, 25% of these (94 intervention and 36 control) did not commence when their clinics withdrew before data collection began.

Of the 206 patients in the intervention who commenced the study (figure 1), 17% (*n*=36) were forced to leave when their clinics withdraw the study. A further 14% (*n*=28) patients withdrew as the study progressed, with 4% leaving after 6 months, 5% after 9 months and 5% after the full year. Reasons included leaving the area, going into residential care or becoming too ill to continue, but no consistent pattern could be identified. (Exact numbers for each reason are not known.) In the control group, 24% (*n*=47) of the 194 patients who agreed to participate had forgotten about the study by the time the 6-month review was to be undertaken and did not want to proceed.

Table 1 presents the characteristics of the patients in both the intervention and control groups who commenced the study, and shows that these characteristics were similar across both groups. There were no significant differences in patient characteristics between the intervention and control at baseline. ~~Approximately 70% of patients in the intervention group completed all five visits over the 12-month study period.~~

Phase 1: Comparison of outcomes between control and intervention groups after 6 months (Table 2)

Table 2 presents baseline and 6-month data for markers used to monitor control of chronic disease for both the intervention and the control group. While the six-month depression scores for all 310 patients (164 intervention and 146 control) were significantly lower than those at baseline in both the intervention group (10.7±0.7 reducing to 7.1±0.8, $t(163)=8.38$, $p<0.001$) and the control group (11.6±0.9 reducing to 9.0±0.9, $t(145)=6.01$, $p<0.001$), the ANCOVA adjusting for the baseline scores showed that the improvement was significantly better in the intervention group than in the control ($F(1,309)=6.40$, $p=0.012$). (The 95% confidence ranges are indicated by the plus-minus sign.)

Half of the patients had only mild-depression at baseline (PHQ-9 scores between five and nine). Because the reported score for many of these patients may be due to their diabetes rather than

depression[20], the intervention is unlikely to be able to change these scores. This is one reason that Katon and colleagues used a score of ten or more as an inclusion criterion in their study[15]. Consequently, we examined the change to baseline PHQ-9 scores for the 164 patients (81 intervention and 83 control) with moderate to severe depression (PHQ-9 scores of 10 or more) at baseline. These patients showed significant improvement, with the mean depression score in the intervention group dropping by 5.7±1.3, from 14.4±1.1 down to 8.7±1.3 ($t(80)=9.00$, $p<0.001$), a clinically-significant change[19]. The improvement in the intervention for these patients was significantly better than in the control group ($F(1,161)=4.02$, $p=0.047$) where the depression score dropped by 4.3±1.2, from 15.1±1.1 down to 10.8±1.4 ($t(82)=6.88$, $p<0.001$).

Except for the HDL measurements, there were no significant changes in biophysical measures after six months in either group. Smoking rates were low at baseline in the patients with established cardiovascular risk factors. Recording of alcohol was sub-optimal, although better than other Australian primary care surveys[27].

The intervention group also showed significantly greater number of patients exercising, referred to and attending an exercise program, and referred to and attending a mental health worker after six months of collaborative care. In the control group, there were no significant changes observed after six months, except that referrals to a mental health worker increased significantly ($p<0.001$) from 9% to 21%, consistent with the action being taken by the nurses as required by the protocol. Neither group showed any significant changes in the number of patients taking antidepressant medication.

Phase 2: Chronic disease outcomes over 12 months using TrueBlue collaborative care (Table 3)

Table 3 presents data at baseline and 12 months for the intervention group for markers used to monitor control of existing diabetes and CHD. The improvement in mental health observed after 6 months was maintained at 12 months, with the significant reduction in the mean depression score maintained (10.7±0.7 to 6.6±0.7, $t(163)=9.92$, $p<0.001$), and nearly 70% of patients having lower depression scores than baseline after one year. Patients with moderate to severe depression at baseline showed an even greater improvement after 12 months of collaborative care, with the mean depression score dropping by 6.4±1.2, from 14.4±0.8 to 8.0±1.2 ($t(80)=10.41$, $p<0.001$). The significant improvement in the mean SF36v2 composite mental-health and physical-health scores observed after 6 months was also maintained at 12 months.

Physiological measures showed a trend, although not significant, to improvement in weight, systolic blood pressure and HDL. Mean baseline lipids and HbA1c were close to guideline targets. The 10-year CVD risk calculated with the Framingham risk equations[28] suggests a small but

significant (p=0.015) reduction in risk from 27.4% to 24.8% for the patients with only T2DM. (The Framingham risk equations cannot be used for those patients who have CHD.)

The most notable changes in lifestyle after 12-months of the intervention were a significant increase in the numbers of patients who reported taking regular exercise or being referred to an exercise program. Reported referrals and visits to a mental health worker and numbers taking antidepressant medication were also significantly greater at 12 months.

The TrueBlue protocol also included goal setting so that patients could become more pro-active in their own care. An analysis of participant goals revealed that two-thirds of visits resulted in at least one behavioural activation goal being set and, over the course of the study, 86% of patients identified a behavioural activation goal.

Adherence to guidelines (Figure 2)

Figure 2 shows the percentage of TrueBlue patients who had psychosocial and biophysical checks undertaken as recommended by the Australian National Heart Foundation and Diabetes Australia guidelines, with the corresponding percentages for usual care taken from a study of a large sample of Australian general practices[27].

Discussion

Outcomes of phase 1

Depression scores were significantly lower at six months for patients in the intervention practices compared with those in the control group, and the improvement was clinically significant for patients with moderate to severe depression[19], with patients moving one depression category. Patients experienced increased nurse contact time through the nurse consultations, were provided with information about mental health and their physical health through psycho-education resources, and had their treatment intensified when required. Modalities included behavioural activation, antidepressant medication, and referrals to mental health professionals and exercise programs. Similar improvements in depression scores and stepped-up care were observed in the collaborative care model of Katon and colleagues[15]. The reduction in depression scores observed in the control group could be explained, in part, by control practices being provided with each patient's entry-level depression score during the recruitment process as part of the study's safety protocol. Usual care could have been influenced by drawing attention to co-morbid depression[15] as the protocol required that practice nurses take action if severe depression was recorded or if the patient had responded to the suicidal-ideation question. Referrals to mental-health workers by the control clinics had increased significantly consistent with the clinics taking action where warranted. It is also known[29] that recruiting interested patients (those who wanted to participate) from interested

clinics (those that agreed to join) can affect the representativeness of the study population. GPs with a particular interest in the study may be more likely to participate and may manage their patients more effectively, irrespective of whether they are in the control or intervention arms. Consequently, a reduction in depression scores in the control group was expected but the structured TrueBlue model did produce a significantly better reduction in depression. While the effect size may be small (Cohen's $f = 0.15$), it is important to note that TrueBlue was designed to be implemented easily within general practices, with running costs funded by existing Australian Medicare rebates, and to make better use of their existing resources. These features mean that TrueBlue could be easily applied to patients across general practices at a population level, making the benefits clinically important.

Outcomes of phase 2

The key clinical outcomes over a 12-month period in the intervention group (Table 3) were a sustained improvement in mental health, demonstrated by symptom severity score (PHQ-9 total score) and by the patient's function and subjective evaluation of mental health (SF36 mental health composite score) and physical health (SF36 physical health composite score). Regular physical exercise has been shown to be important for reducing depression[30]. The self-reported exercise rates showed significant improvement over the 12 months of collaborative care intervention. The biophysical measures reported in Table 3 showed modest improvements after 12 months and the Framingham risk equations[28] suggest a small but significant reduction in the 10-year CVD risk for the T2DM patients. These improvements were achieved despite that fact that we did not specifically select patients whose physiological parameters exceeded guidelines. Rather, our recruitment process selected from the practice's disease registry on the basis of only the presence of depression and T2DM or CHD and, consequently, many patients were already being treated to target on measures such as cholesterol and HbA1c, leaving little room for improvement.

Limitations

We were able to run TrueBlue only in practices that ~~had access to~~used clinical software, ~~capable of generating~~ which we used to generate a disease registry from which participants could be selected, and had a practice nurse on staff. Clinics that chose to take part in the study may not have been representative of wider general practice. Operational limitations further reduced the number of practices over the duration of the study. Patient response rates to the mail-out (28%) may reflect anxiety over the new model of care where the patient discloses depression and visits the PN first rather than only the GP. Usual care in the control clinics may have been changed by patients completing the PHQ-9 and reading the project description. GPs were made aware of individual PHQ-9 results, and took action where warranted. GP awareness of these biophysical and lifestyle

risks may be expected to change clinical management. By design, TrueBlue practices needed to incorporate all research activities within the context of their busy clinics, and so only research data that could be extracted automatically were collected. The data dropout resulting from these two factors contributed to the observed small effect size. We were not able to obtain multiple data sets at three-monthly intervals over 12 months of 'usual care' because the act of inviting patients and measuring depression scores and biophysical measures would in itself change the nature of usual care. In addition, practices would not have been willing to join the study if there was a chance of being randomly allocated to 12 months of being in such a control arm[29].

Collaborative Care

A recent UK study has shown the difficulties of disseminating a guideline without guidance on how to implement collaborative care. Organisational barriers included GPs finding the PHQ-9 awkward to use, nurses not feeling confident or competent due to lack of training, and no guidance on stepped care[13]. The TrueBlue model of collaborative care overcame many of these difficulties. Its successful components were:[9,31]

- *Use of evidence based guidelines.* National Heart Foundation and Diabetes Australia guidelines determined disease management targets and frequency of monitoring.
- *Systematic screening and monitoring of risk factors.* Patients attended three monthly visits in which a care plan with its checklist was completed. By providing a comprehensive collation of all necessary information, this document made clinical management by the patient's GP easier, quicker and more accurate.
- *Timetabled recall visits.* The date of the next appointment was set during each visit, PHQ-9 was readministered and if improvement was insufficient, *stepped care* was followed by initiating drug therapy or increasing the dose, or referral to a mental-health worker according to the guidelines.
- *New or adjusted roles for team members.* PNs took responsibility for organising and monitoring the outcome of referrals, goals and targets. They used a depression questionnaire (the PHQ-9) to open a discussion with patients about their depression symptoms.
- *Information support for the clinician.* GPs were provided with the care plan by the PNs.
- *Enhanced patient self-management.* Patients received their own copy of the care plan with personalised goals, current measurements, targets and safety advice. A component of each visit was to discuss and update their plan, and receive education material on depression.
- *Identified case manager.* PNs became case managers but the GP remained the key clinician.

- *Means of effective communication between all members of the care team.* The care plan was designed to provide relevant clinical information in a succinct format while still being comprehensible to patients.
- *Audit information for the practice.* De-identified data was provided automatically through the care plan.

Applicability of TrueBlue

TrueBlue used existing clinical software and improved the focus of the GP consultation by delegating some tasks to the PN. Higher levels of adherence to guideline-recommended checks were also reported for TrueBlue. Patients and their carers, allied health professionals, specialists and general practitioners gained ready access to patient details provided in TrueBlue's care plan enabling them to see at glance where improve clinical care may be needed. The study achieved improved outcomes with the potential for prevention of heart attack and stroke through reduced 10-year CVD risk. The care plan template also allowed the practice to collect high quality audit data without taking up clinical time. While it was not possible to obtain complete financial data from the clinics specifically relating to the TrueBlue visits, the data that are available suggest that clinics did indeed cover their costs in implementing TrueBlue through Australian Medicare rebates. The success of TrueBlue and TeamCare[15] demonstrates that collaborative care is feasible in routine general practice in Australia and the USA, and could lead to improved outcomes for patients with depression and other chronic diseases[7,32].

Acknowledgements

The authors wish to thank the patients, practice nurses, general practitioners and support staff of the participating clinics Evans Head Medical Centre, Lennox Head Medical Centre, Keen Street Clinic, Tintenbar Medical Centre, Health on Grange, Warradale Medical Centre, Seaton Medical and Specialist Centre, Woodcroft Medical Centre and Southcare Medical Centre. Professors Wayne Katon and Juergen Unützer of the University of Washington were unstinting in their advice on adaptation of the IMPACT model. We would also like to thank Bob Leahy for managing the project, Vince Versace for his statistical advice and Vicki Brown for her assistance during the course of the study.

Contributors

All authors have full access to the complete study dataset, contributed to the design, implemented the project, and co-wrote and approved the manuscript. MM, MC, JD and PR analysed the data.

MM, PR and KS developed and ran the practice nurse training program. JD and PR conceived the TrueBlue model during a visit to the IMPACT team. JD is the guarantor.

Funding

Funding was provided by *beyondblue*, the National Depression Initiative in Australia, but it had no other involvement in any phase of the study. All researchers were independent of and not influenced by *beyondblue.*

Competing interests

All authors have completed the Unified Competing Interest form and declare that (1) funding was received from *beyondblue* to carry out the study; (2) they do not have any relationships with any companies that might have an interest in the submitted work in the previous 3 years; (3) their spouses, partners, or children do not have any financial relationships that may be relevant to the submitted work; and (4) they do not have any non-financial interests that may be relevant to the submitted work.

Ethical approval

Ethics approval was obtained from Flinders University's Social and Behavioural Research Ethics Committee. All patients gave informed consent to participate in the study.

Data sharing

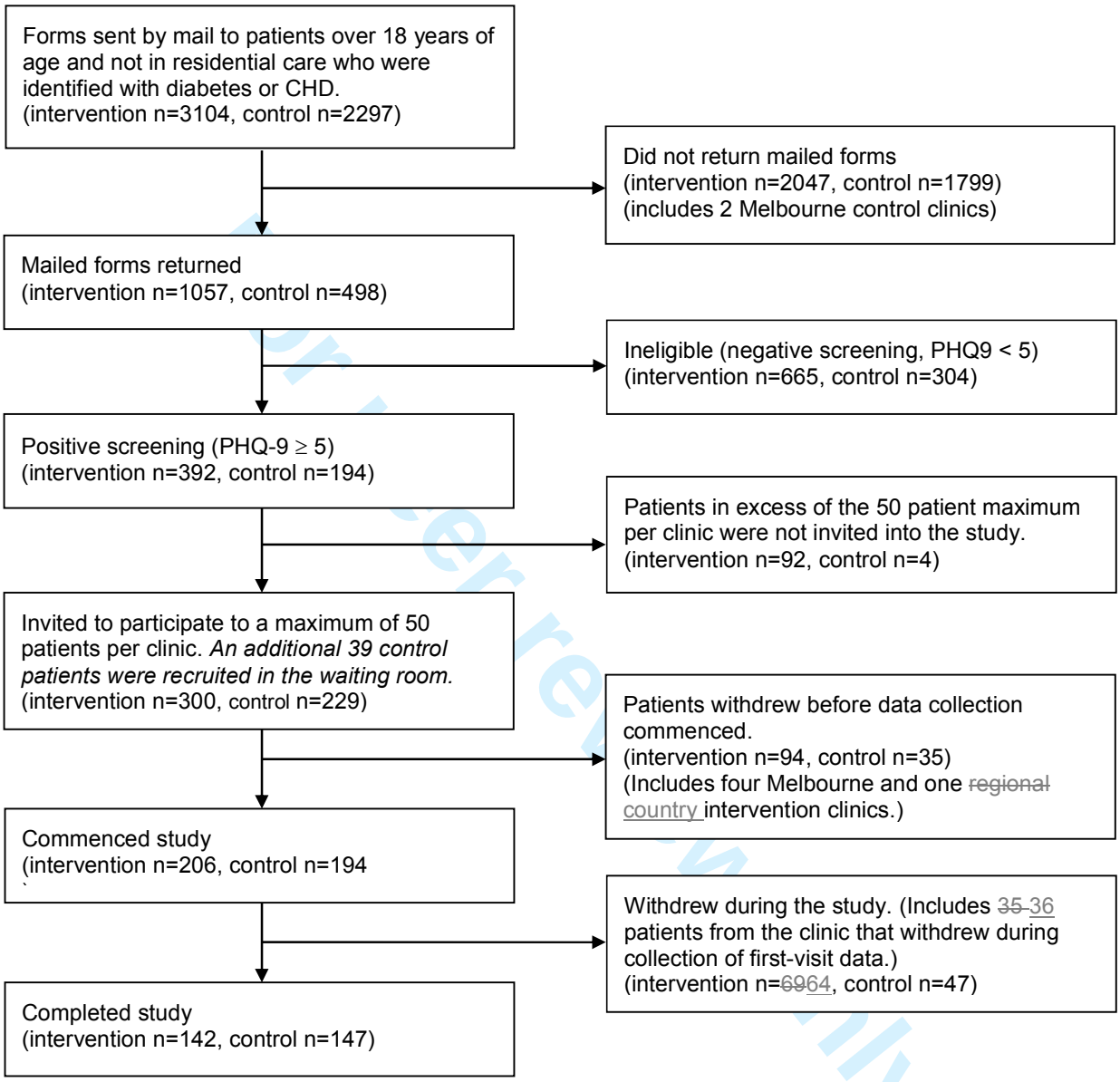
The dataset used for the analysis and the computer codes used to produce the results are available from the corresponding author.

References

1. Daar AS, Singer PA, Persad DL, Pramming SK, Matthews DR, Beaglehole R, et al. Grand challenges in chronic non-communicable diseases. *Nature* 2007;450(7169):494-96.
2. World Health Organisation. Preventing Chronic Disease: A Vital Investment. Geneva, 2005.
3. Ali S, Stone MA, Peters JL, Davies MJ, Khunti K. The prevalence of co-morbid depression in adults with Type 2 diabetes: a systematic review and meta-analysis. *Diabet Med* 2006;23(11):1165-73.
4. Egede LE, Nietert PJ, Zheng D. Depression and all-cause and coronary heart disease mortality among adults with and without diabetes. *Diabetes Care* 2005;28(6):1339-45.
5. Clarke DM, Currie KC. Depression, anxiety and their relationship with chronic diseases: a review of the epidemiology, risk and treatment evidence. *Med J Aust* 2009;190(7 Suppl):S54-60.
6. Freeling P, Rao BM, Paykel ES, Sireling LI, Burton RH. Unrecognised depression in general practice. *Br Med J (Clin Res Ed)* 1985;290(6485):1880-83.
7. Gilbody S, Bower P, Fletcher J, Richards D, Sutton AJ. Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. *Arch Intern Med* 2006;166(21):2314-21.
8. Katon W, Von Korff M, Lin E, Walker E, Simon GE, Bush T, et al. Collaborative management to achieve treatment guidelines. Impact on depression in primary care. *JAMA* 1995;273(13):1026-31.
9. Hickie IB, McGorry PD. Increased access to evidence-based primary mental health care: will the implementation match the rhetoric? *Med J Aust* 2007;187(2):100-03.
10. Lin EH, Katon W, Von Korff M, Rutter C, Simon GE, Oliver M, et al. Relationship of depression and diabetes self-care, medication adherence, and preventive care. *Diabetes Care* 2004;27(9):2154-60.
11. Reddy P, Ford D, Dunbar JA. Improving the quality of diabetes care in general practice. *Aust J Rural Health* 2010;18(5):187-93.
12. Campbell MJ. Cluster randomized trials in general (family) practice research. *Stat Methods Med Res* 2000;9(2):81-94.
13. Mitchell C, Dwyer R, Hagan T, Mathers N. Impact of the QOF and the NICE guideline in the diagnosis and management of depression: a qualitative study. *Br J Gen Pract* 2011;61(586):e279-e89.
14. Unutzer J, Katon W, Williams JW, Jr., Callahan CM, Harpole L, Hunkeler EM, et al. Improving primary care for depression in late life: the design of a multicenter randomized trial. *Med Care* 2001;39(8):785-99.
15. Katon WJ, Lin EH, Von Korff M, Ciechanowski P, Ludman EJ, Young B, et al. Collaborative care for patients with depression and chronic illnesses. *N Engl J Med* 2010;363(27):2611-20.
16. Morgan MA, Dunbar J, Reddy P. Collaborative care - The role of practice nurses. *Aust Fam Physician* 2009;38(11):925-26.
17. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16(9):606-13.
18. Morgan M, Dunbar J, Reddy P, Coates M, Leahy R. The TrueBlue study: is practice nurse-led collaborative care effective in the management of depression for patients with heart disease or diabetes? *BMC Fam Pract* 2009;10:46.
19. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. *Medical care* 2004;42(12):1194-201.
20. Reddy P, Philpot B, Ford D, Dunbar JA. Identification of depression in diabetes: the efficacy of PHQ-9 and HADS-D. *Brit J Gen Prac* 2010;DOI: 10.3399/bjgp10X502128.
21. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30(6):473-83.
22. Dobson KS, Hollon SD, Dimidjian S, Schmaling KB, Kohlenberg RJ, Gallop RJ, et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication

in the prevention of relapse and recurrence in major depression. *Journal of Consulting and Clinical Psychology* 2008;76(3):468-77.
23. Kauermann G, Carroll RJ. A Note on the Efficiency of Sandwich Covariance Matrix Estimation. *J Amer Statistical Assoc* 2001;96(456):1387-96.
24. Senn S. Change from baseline and analysis of covariance revisited. *Statistics in Medicine* 2006;25(24):4334-44.
25. Vickers A. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: A simulation study. *BMC Medical Research Methodology* 2001;1:1-4.
26. Hox JJ. *Applied multilevel analysis*. Second ed. Amsterdam: TT-Publikaties, 1995.
27. Wan Q, Harris MF, Jayasinghe UW, Flack J, Georgiou A, Penn DL, et al. Quality of diabetes care and coronary heart disease absolute risk in patients with type 2 diabetes mellitus in Australian general practice. *Qual Saf Health Care* 2006;15(2):131-35.
28. Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J* 1991;121(1 Pt 2):293-98.
29. Wilson S, Delaney BC, Roalfe A, Roberts L, Redman V, Wearn AM, et al. Randomised controlled trials in primary care: case study. *BMJ* 2000;321:24-27.
30. Wiles NJ, Haase AM, Gallacher J, Lawlor DA, Lewis G. Physical activity and common mental disorder: results from the Caerphilly study. *Am J Epidemiol* 2007;165(8):946-54.
31. Fuller J, Perkins D, Parker S, Holdsworth L, Kelly B, Roberts R, et al. Effectiveness of service linkages in primary mental health care: a narrative review part 1. . *BMC Health Services Research* 2011;11:71.
32. Richards DA, Lovell K, Gilbody S, Gask L, Torgerson D, Barkham M, et al. Collaborative care for depression in UK primary care: a randomized controlled trial. *Psychol Med* 2008;38(2):279-87.

Figure 1. CONSORT flow diagram of the recruitment process.



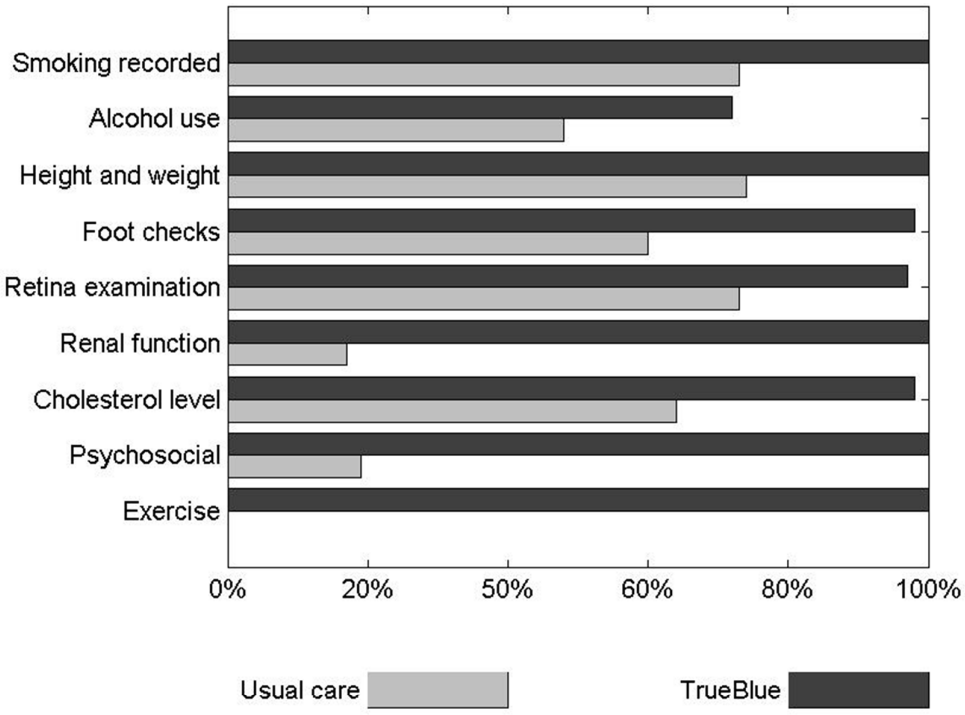


Figure 2: Recording of checks recommended by the National Heart Foundation and Diabetes Australia guidelines. Data for "usual care" were adapted from reference 27. No usual-care data were available for exercise.

Table 1: Patient characteristics at the baseline visits. There were no significant differences between the intervention and control at baseline.

Characteristic	Intervention group (*n*=170)	Control group (*n*=147)
Male (%) / Female (%)	51.8% / 48.2%	55.2% / 44.8%
Age (yr)	68.0 ± 11.7	67.6 ± 11.2
Aboriginal or Torres Strait Islander (%)	0.0%	0.7%
Diagnosis: Type-2 diabetes	37.6%	47.6%
CHD	45.3%	35.8%
Both	17.1%	16.6%
Body mass index (kg/m^2)	31.4 ± 6.0 (*n*=170)	30.8 ± 6.0 (*n* =103)
Systolic blood pressure (mmHg)	134.1 ± 19.0 (*n*=169)	133.5 ± 19.6 (*n* =112)
Total cholesterol (mmol/L)	4.21 ± 0.94 (*n*=165)	4.41 ± 1.06 (*n* =110)
Triglycerides (mmol/L)	1.73 ± 0.88 (*n*=165)	1.92 ± 1.37 (*n* =105)
LDL (mmol/L)	2.22 ± 0.74 (*n*=159)	2.37 ± 0.88 (*n* =89)
HDL (mmol/L)	1.23 ± 0.36 (*n*=159)	1.18 ± 0.33 (*n* =97)
HbA1c (mmol/L)	7.00 ± 1.21 (*n*=94)	7.19 ± 1.42 (*n* =69)
PHQ-9 score	10.7 ± 4.7 (*n*=164)	11.6 ± 5.5 (*n*=146)
PHQ-9 score range at baseline	5 to 24	5 to 27

Table 2: TrueBlue outcomes at six months in the intervention and control groups. (See Table 3 for a list of abbreviations.)

	Intervention				Control				Between groups
	n	Baseline	6-months	Within	*n*	Baseline	6-months	Within	
PHQ9 depression score	164	10.7±0.8	7.1±0.8	$p<0.001^{\dagger}$	146	11.6±0.9	9.0±0.9	$p<0.001^{\ddagger}$	$p=0.012$
SF36v2 mental-health score[§]	71	37.2±3.4	41.1±3.4	$p=0.034^{\dagger}$		Not recorded			NS
SF36v2 physical-health score[§]	71	39.9±2.2	42.5±2.6	$p=0.023^{\dagger}$		Not recorded			NS
Body mass index (kg/m^2)	162	31.3±1.0	31.2±1.0	NS	103	30.8±1.2	31.0±1.0	NS	NS
Waist (cm)	161	104.7±2.4	105.0±2.4	NS	80	104.2±4.0	105.8±3.2	NS	NS
Systolic blood pressure (mmHg)	161	134.2±3.0	132.4±2.8	NS	112	133.5±3.8	131.2±3.4	NS	NS
Total cholesterol (mmol/L)	158	4.21±0.16	4.22±0.14	NS	109	4.41±0.20	4.44±0.20	NS	NS
LDL (mmol/L)	154	2.23±0.12	2.17±0.14	NS	86	2.37±0.18	2.29±0.20	NS	NS
HDL (mmol/L)	154	1.23±0.06	1.29±0.06	$p=0.023^{\dagger}$	93	1.17±0.06	1.27±0.08	$p=0.011^{\ddagger}$	NS
Triglycerides (mmol/L)	158	1.72±0.14	1.66±0.12	NS	104	1.84±0.22	1.75±0.18	NS	NS
HbA1c (%)[¶]	89	6.97±0.24	6.90±0.26	NS	67	7.22±0.34	7.40±0.36	NS	$p=0.049$
Ten-year CVD risk[*]	61	26.9±3.2	26.1±3.2	NS	46	26.3±3.6	24.7±3.2	NS	NS
The 95% confidence ranges are indicated by the plus-minus (±) sign. Note that lower scores indicate improvement for all items except the SF36v2 and HDL results, where higher scores indicate improvement.									
Smoking	162	15 (9%)	13 (8%)	NS	110	13 (12%)	13 (12%)	NS	NS
Alcohol	104	47 (45%)	51 (49%)	NS	42	27 (64%)	27 (64%)	NS	NS
Exercises 30 mins/day, 5 d/wk	162	66 (41%)	97 (60%)	$p<0.001^{\dagger}$	75	22 (29%)	22 (29%)	NS	$p<0.001$
Referred to exercise program	162	32 (20%)	58 (36%)	$p<0.001^{\dagger}$	111	15 (14%)	10 (9%)	NS	$p<0.001$
Attends exercise program	162	12 (7%)	23 (14%)	$p=0.041^{\dagger}$	79	12 (15%)	9 (11%)	NS	NS
On antidepressant medication	162	27 (17%)	34 (21%)	NS	113	31 (27%)	36 (32%)	NS	$p=0.025$
Referred to mental health worker	162	47 (29%)	58 (36%)	$p=0.022^{\dagger}$	114	10 (9%)	24 (21%)	$p<0.001^{\ddagger}$	$p<0.001$
Attends mental health worker	162	10 (6%)	37 (23%)	$p<0.001^{\dagger}$	109	14 (13%)	11 (10%)	NS	$p=0.044$

The values in brackets are the percentages of the total *n*.

(NS) No significant difference. (†) Significant difference between baseline and six-months within the *intervention* clinics. (‡) Significant difference between baseline and six-months within the *control* clinics. (§) SF36v2 questionnaires were not collected by all clinics. (¶) HbA1c results were only available for patients with T2DM. (*) CVD risk could only be calculated for patients with T2DM.

Table 3: TrueBlue outcomes at 12 months within the intervention clinics only.

	Intervention			
	n	Baseline	12-months	
PHQ9 depression score	164	10.7±0.7	6.6±0.7	$p<0.001$
SF36v2 mental-health score[§]	70	36.0±3.2	41.3±2.8	$p<0.001$
SF36v2 physical-health score[§]	70	40.6±2.2	44.3±2.8	$p<0.001$
Body mass index (kg/m^2)	142	31.4±1.0	31.1±1.0	$p=0.006$
Waist (cm)	141	105.0±2.4	105.2±2.6	NS
Systolic blood pressure (mmHg)	141	135.2±3.2	130.2±3.0	$p=0.016$
Total cholesterol (mmol/L)	138	4.18±0.16	4.28±0.16	NS
LDL (mmol/L)	135	2.19±0.12	2.24±0.20	NS
HDL (mmol/L)	135	1.22±0.06	1.36±0.08	$p<0.001$
Triglycerides (mmol/L)	138	1.73±0.16	1.63±0.14	$p=0.004$
HbA1c (%)[¶]	79	7.01±0.26	7.04±0.28	NS
Ten-year CVD risk[*]	55	27.4±3.4	24.9±3.6	$p=0.015$

The 95% confidence ranges are indicated by the plus-minus (±) sign. Lower scores indicate improvement for all items except the SF36v2 and HDL results, where higher scores indicate improvement.

Smoking	142	15 (11%)	11 (8%)	NS
Alcohol	95	45 (47%)	47 (49%)	NS
Exercises 30 mins/day, 5 days/wk	142	57 (40%)	83 (58%)	$p<0.001$
Referred to exercise program	142	26 (18%)	53 (37%)	$p<0.001$
Attends exercise program	142	10 (7%)	17 (12%)	NS
On antidepressant medication	142	22 (15%)	33 (23%)	$p=0.001$
Referred to mental health worker	142	40 (28%)	59 (42%)	$p<0.001$
Attends mental health worker	142	8 (6%)	25 (18%)	$p<0.001$

The values in brackets are the percentages of the total *n*.

(NS) No significant difference. (§) SF36v2 questionnaires were not collected by all clinics. (¶) HbA1c results were only available for patients with T2DM. (*) CVD risk could only be calculated for patients with T2DM.

NOTE:
PHQ9 is the 9-question Patient Health Questionnaire.
SF36v2 is version 2 of the Short Form 36-Question health survey.
LDL is low-density lipoprotein.
HDL is high-density lipoprotein.
Glycated haemoglobin (HbA1c) was measured only for patients with diabetes.
Unit of alcohol is 10g of ethanol.

Checklist of items to include when reporting a cluster randomised trial

* = addition to CONSORT	*Modifications to checklist in italics*		
PAPER SECTION and topic	Item	Descriptor	Reported on Page No.
TITLE & ABSTRACT	1*	How participants were allocated to interventions (e.g., "random allocation", "randomised", or "randomly assigned"), *specifying that allocation was based on clusters*	1
INTRODUCTION Background	2*	Scientific background and explanation of rationale, *including the rationale for using a cluster design.*	2, 3
METHODS Participants	3*	Eligibility criteria for participants *and clusters* and the settings and locations where the data were collected.	3, 4
Interventions	4*	Precise details of the interventions intended for each group, *whether they pertain to the individual level, the cluster level or both,* and how and when they were actually administered.	Ref 17
Objectives	5*	Specific objectives and hypotheses, *and whether they pertain to the individual level, the cluster level or both*.	1, 3
Outcomes	6*	Report clearly defined primary and secondary outcome measures, *whether they pertain to the individual level, the cluster level or both*, and, when applicable, any methods used to enhance the quality of measurements (e.g., multiple observations, training of assessors).	4, 5
Sample size	7*	How *total* sample size was determined *(including method of calculation, number of clusters, cluster size, a coefficient of intracluster correlation (ICC or k), and an indication of its uncertainty*) and, when applicable, explanation of any interim analyses and stopping rules.	Ref 17
Randomisation.			
Sequence generation	8*	Method used to generate the random allocation sequence, including details of any restriction (e.g., blocking, stratification, *matching*).	Ref 17
Allocation concealment	9*	Method used to implement the random allocation sequence, *specifying that allocation was based on clusters rather than individuals and* clarifying whether the sequence was concealed until interventions were assigned.	
Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	
Blinding (Masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. If done, how the success of blinding was evaluated.	N/A
Statistical methods	12*	Statistical methods used to compare groups for primary outcome(s) *indicating how clustering was taken into account*; methods for additional analyses, such as subgroup analyses and adjusted analyses.	5
RESULTS Participant flow	13*	Flow of *clusters and* individual participants through each stage (a diagram is strongly recommended). Specifically, for each group report the numbers of *clusters and* participants randomly assigned, receiving intended treatment, completing the study protocol, and analyzed for the primary outcome. Describe protocol deviations from study as planned, together with reasons.	15 (Fig 1)
Recruitment	14	Dates defining the periods of recruitment and follow-up.	4
Baseline data	15*	Baseline information for each group *for the individual and cluster levels as applicable*	17 (Table 1)
Numbers analyzed	16*	Number of *clusters and* participants (denominator) in each group included in each analysis and whether the analysis was by "intention-to-treat". State the results in absolute numbers when feasible (e.g., 10/20, not 50%).	17, 18, 19 (Tables)
Outcomes and Estimation	17*	For each primary and secondary outcome, a summary of results for each group measures *for the individual or cluster level as applicable*, and the estimated effect size and its precision (e.g., 95% confidence interval) *and a coefficient of intracluster correlation (ICC or k) for each primary outcome.*	17, 18, 19 (Tables)
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those pre-	N/A

		specified and those exploratory.	
Adverse events	19	All important adverse events or side effects in each intervention group.	None
DISCUSSION Interpretation	20	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision and the dangers associated with multiplicity of analyses and outcomes.	7, 8
Generalisability	21*	Generalisability (external validity) *to individuals and/or clusters (as relevant)* of the trial findings.	9, 10
Overall evidence	22	General interpretation of the results in the context of current evidence.	9

The TrueBlue model of collaborative care using practice nurses as case managers for depression alongside diabetes or heart disease

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-002171.R2
Article Type:	Research
Date Submitted by the Author:	17-Dec-2012
Complete List of Authors:	Morgan, Mark; Flinders and Deakin Universities, Greater Green Triangle University Department of Rural Health Coates, Michael; Flinders and Deakin Universities, Greater Green Triangle University Department of Rural Health Dunbar, James; Flinders and Deakin Universities, Greater Green Triangle University Department of Rural Health Reddy, Prasuna; University of Newcastle, School of Medicine and Public Health Schlicht, Kate; Flinders and Deakin Universities, Greater Green Triangle University Department of Rural Health Fuller, Jeff; Flinders University, School of Nursing and Midwifery
Primary Subject Heading:	Public health
Secondary Subject Heading:	Diabetes and endocrinology, Cardiovascular medicine, Mental health
Keywords:	Coronary heart disease < CARDIOLOGY, General diabetes < DIABETES & ENDOCRINOLOGY, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, MENTAL HEALTH, PUBLIC HEALTH

The TrueBlue model of collaborative care using practice nurses as case managers for depression alongside diabetes or heart disease: a randomised trial

MAJ Morgan[1], MJ Coates[1], JA Dunbar[1,*], P Reddy[2], K Schlicht[1] and J. Fuller[3]

1. Greater Green Triangle University Department of Rural Health, Flinders and Deakin Universities, PO Box 423, Warrnambool, Victoria, 3280, Australia
2. School of Medicine and Public Health, University of Newcastle, Callaghan, NSW, 2308, Australia
3. School of Nursing and Midwifery, Flinders University, GPO Box 2100, Adelaide, SA, 5001, Australia

* Corresponding author

Abstract

Objectives: Determine the effectiveness of collaborative-care in reducing depression in primary-care patients with diabetes or heart disease using practice nurses as case managers.

Design: A two-arm open randomised-cluster trial with wait-list control for 6 months. The intervention was followed over 12 months.

Setting: Eleven Australian general practices, five randomly allocated to the intervention and six to the control.

Participants: 400 primary-care patients (206 intervention, 194 control) with depression and type 2 diabetes, coronary heart disease or both.

Intervention: The practice nurse acted as case manager identifying depression, reviewing pathology results, lifestyle risk factors, and patient goals and priorities. Usual care continued in the controls.

Main outcome measure: A five-point reduction in depression scores for patients with moderate to severe depression. Secondary outcome was improvements in physiological measures.

Results: Mean depression scores after six months of intervention for patients with moderate to severe depression decreased by 5.7±1.3 compared with 4.3±1.2 in the control, a significant (p=0.012) difference. (The plus-minus is the 95% confidence range.) Intervention practices demonstrated adherence to treatment guidelines and intensification of treatment for depression, where exercise increased by 19%, referrals to exercise programs by 16%, referrals to mental health workers (MHWs) by 7%, and visits to MHWs by 17%. Control-practice exercise did not change, referrals to exercise programs dropped by 5%, and visits to MHWs by 3%. Only referrals to MHW increased, by 12%. Intervention improvements were sustained over 12 months, with significant

(p=0.015) decrease in 10-year cardiovascular-disease risk from 27.4±3.4% to 24.8±3.8%. A review of patients indicated the study's safety protocols were followed.

Conclusion: TrueBlue participants showed significantly improved depression and treatment intensification, sustained over 12-months of intervention, and reduced 10-year CVD risk. Collaborative care using practice nurses appears to be an effective primary-care intervention.

Trial registration: ACTRN12609000333213 (Australia and New Zealand Clinical Trials Registry).

Article summary

Article focus

- To determine the effectiveness of a collaborative-care model to reduce depression in primary-care patients with diabetes or heart disease.
- To determine the effectiveness of using practice nurses as case managers of patients with depression and diabetes, heart disease or both.

Key messages

- The TrueBlue model of collaborative care can be introduced within the general practice workforce with practice nurses taking on the role of case manager.
- Practice nurses can improve care of depression in patients with diabetes or heart disease, leading to better outcomes and reduced 10-year CVD risk.
- The care of patients using the TrueBlue model is closer to "best practice" guidelines, with substantially better levels of adherence to guideline-recommended checks than occur in usual care.

Strengths and limitations of this study

- The TrueBlue model of collaborative care overcomes many of the difficulties in implementing a guideline for the treatment of co-morbid depression.
- The study's purpose-designed care plan gives patients and their carers, allied health professionals, specialists and general practitioners ready access to patient details enabling them to see at glance where improve clinical care may be needed.
- Clinics were able to recover the costs of the collaborative care through Australian Medicare rebates.

- The study could only be run in practices that had a practice nurse on staff to carry out the intervention and had access to clinical software capable of generating a disease registry from which patients could be selected to participate in the trial.

- Differences between TrueBlue- and control-practice outcomes may have been reduced by patients completing the PHQ-9 depression questionnaire and reading the project description, and by GPs being made aware of individual PHQ-9 results so that they could take action where warranted.

Introduction

Managing diabetes and heart disease has been highlighted as one of the global "grand challenges in chronic non-communicable diseases"[1] because the prevalence of these two preventable diseases is increasing[2]. Along with depression, they have been identified as health priority areas in many countries. A vicious cycle exists between depression and these chronic diseases, with each being a risk factor for the other[3]. Higher mortality has been demonstrated for people with depression and type 2 diabetes (T2DM) or coronary heart disease (CHD) beyond that due to the separate diseases alone[4]. For patients with depression and T2DM or CHD or both, there are increased risks of adverse outcomes[5] but this co-morbid depression is often missed in primary care[6]. Consequently, the identification of depression has now been incorporated in many heart disease guidelines as one of the requirements for optimal management. Meeting these challenges will require an innovative use of the existing general practice workforce and such a re-orientation of resources has been identified as one of the grand challenges[1].

Collaborative care is a system that has been shown to be more effective for chronic disease management than standard care[7]. It includes a reorientation of the medical workforce through new or adjusted roles for team members, particularly using practice nurses as the identified case manager to undertake the care of the patients[8,9]. It also includes the use of evidence based guidelines, systematic screening and monitoring of risk factors, timetabled recall visits, information support for the clinician, enhanced patient self-management, a means of effective communication between all members of the care team and audit information for the practice. Since self-care for diabetes has been found to be suboptimal across a range of self-managed activities, particularly for patients with depression, a collaborative care model may be able to achieve better quality of care through the case manager monitoring patient progress[10,11].

Evaluation of a change in the way general practice clinics look after patients requires complex intervention methodology[12] beyond single interventions such as introduction of a guideline with financial incentives[13]. This methodology began with a search for potential models of care (step I), and led to adopting the University of Washington's successful IMPACT model of Collaborative

Care for depression[14,15]. In the exploratory trial (step II), our pilot project[16] adapted IMPACT by training practice nurses as case managers. Practice nurses were trained to screen for depression using a patient self-report measure, the nine-item Patient Health Questionnaire (PHQ-9)[17], as part of comprehensive chronic disease management. They were also trained to use a protocol for care management based on the depression scores. The depression screening and management was embedded in routine visits for patients with diabetes or CHD. The pilot demonstrated that it was feasible to detect, monitor and treat depression in routine general practice alongside the usual biophysical measures, and identified moderate to severe depression in 34% of participants. The TrueBlue study was a randomised cluster trial (step III) that built on and extended the pilot. It investigated whether a collaborative care model (the intervention) is better than usual care (the control) for the management of patients with depression and T2DM, CHD or both in Australian general practice. It was designed to fit into normal clinic operations, making use of practice nurses and medical software, and able to be funded by existing Australian Medicare rebates.

Methods/design

Study design

The design and methodology of the study have been described in detail elsewhere[18]. The study commenced in 2009 and was undertaken in two phases. The first phase was a cluster randomised intervention trial in which general practices were randomly allocated to either an intervention group in which nurse-led collaborative care was undertaken or to a wait-list control group in which usual care led by the general practitioner (GP) was continued. At six months, the TrueBlue training was provided to the control practices. The key aims of the first phase were to determine whether participants with moderate to severe depression in the intervention group showed at least a five-point reduction from the baseline depression scores after six months of intervention and whether this reduction was significantly better than in the control group. A five-point reduction reflects a clinically-relevant change in individuals receiving depression treatment[19]. The secondary outcome was to determine whether the intervention also led to improvements in the patients' physiological measures. The second phase followed the intervention group for an additional six months to determine how the collaborative care model affected health outcomes over a twelve-month period.

Sample size

The sample size calculation was based on detecting a 50% reduction in depression score at the 0.05 significance level with 80% power and a two-tailed test. Detecting a 50% reduction is more stringent than detecting a five-point reduction and provided some additional buffering. Using depression scores from an earlier study (a mean of 5.5 and standard deviation of 6.1)[11], the calculation indicated that 237 patients would be required in each group. An intra-cluster correlation

of 0.04 was used (S. K. Lo, pers. comm.), with a recruitment target of 50 patients per clinic. (Fifty patients were chosen so that clinics could budget for a nurse's time to carry out the intervention with four patients each week over the 3-month cycle of care.) To allow for difficulties in recruitment, a 50% dropout was used. Based on these, the study required 450 patients from nine clinics in the intervention group and the same in the control group.

Practice recruitment

Practices were identified in city and country areas on the basis of having a practice nurse to provide the collaborative care and being able to identify eligible patients, those with CHD or T2DM or both, from their registries; these were invited to participate in the study until the eighteen clinics required by the sample-size calculation were recruited. They were allocated by a random number generator to either the intervention or control arm of the study. The unit of randomisation was the clinic. Five practices (3 country, 2 city) in the intervention group and six (2 country, 4 city) in the control group completed the study. One country intervention clinic withdrew whilst first-visit data were being collected when its TrueBlue-trained practice nurse left the clinic, but some (n=13) patients from it did complete the study and data were collected from them. The study team was not able to determine why the other clinics withdrew.

Patient selection

Eligible patients were sent a postal survey that included a consent form and were asked to complete and return the enclosed PHQ-9 questionnaire, a self-report measure of depression.[17] The PHQ-9 has nine items, each scored from 0 (no problems) to 3 (problems nearly every day). The sum of the scores of the nine items will lie in one of five depression categories: none (0–4), mild (5–9), moderate (10–14), moderately severe (15–19) and severe (20–27). (While it is known that responses to some of the PHQ-9 items may overlap with diabetes symptoms[20], our pilot demonstrated that nurses and patients preferred using the PHQ-9 because the patient's response to each of its items became the basis for the problem solving and goal setting activities that were part of TrueBlue.) Patients with scores of five or above, indicating some form of depression, were invited to participate in the study. A maximum of 50 patients per practice were invited. Patients in residential care or under 18 years of age were not eligible. Figure 1 presents the CONSORT diagram of the patient-recruitment process.

Patient safety

Participation in the intervention included a series of patient visits to their practice nurse (PN) and usual GP every three months over a 12-month period. Patients in the control group continued with their "usual care". The control clinics were also provided with the PHQ-9 depression scores to

ensure patient safety during the trial. The protocol required that practice nurses take action if severe depression was recorded in the returned PHQ-9 or if the patient had responded to the suicidal-ideation question (question 9) on the questionnaire. This action was to be taken *irrespective of whether the clinic was in the intervention or in the control group*.

Practice nurse training

The PN training included a two-day workshop to prepare them for their enhanced roles in nurse-led collaborative care. Topics in the workshop included identifying and monitoring depression using the PHQ-9 questionnaire, and quality of life responses using version 2 of the SF36 questionnaire[21]. Patient goal setting and problem solving were key components of the training with a particular emphasis on behavioural techniques to achieve improved mental health[22]. The training also prepared the PNs for their role as case managers including ensuring Diabetes Australia and Australian National Heart Foundation guidelines were being followed, and referrals were provided to appropriate services, such as allied health and mental health professionals, through discussion with the GPs.

Data collection

The research team developed a protocol-driven care-plan template from which study data could be extracted automatically and sent to the research team. The template was designed to be a multi-purpose document in which the patient's past medical history, current medications, allergies, biophysical and psychosocial measures, lifestyle risks, personal goals and referrals were recorded. It was designed to comply with the requirements to claim Australian Medicare rebates for care planning and to provide a checklist for "gold-standard" care. A copy of the care plan was provided to the patient as a written record of their progress.

The care-plan template collected physical measures, including body-mass index, waist circumference, weight and blood pressure, and the latest pathology results, including lipid profile, glycaemic control (HbA1c) and renal function. Data also included lifestyle risk factors, such as smoking, alcohol consumption and level of physical activity, and depression score as measured by the PHQ-9 questionnaire. Referrals to and attendance at exercise programmes and with mental health workers were also recorded, along with the patient's own goals and possible barriers to achieving these goals. The care-plan template was used by the intervention-group clinics to acquire patient data at three-monthly intervals over a twelve-month period.

In the control group, the only complete dataset recorded using our comprehensive protocol-driven care-plan template was obtained after the six months of "usual care" when the TrueBlue training was offered to the control clinics. No baseline or three-month datasets were acquired since

the study was deliberately designed to avoid changing the "usual care" that would have otherwise occurred by introducing our care-plan template. The study was designed in this way to be run pragmatically in the context of the clinics' normal activities. The only baseline measure obtained was the depression score. On completion of the study, we retrospectively collected what baseline data the control clinics routinely recorded in their electronic medical records in order to have data for two time points, baseline and six-months.

TrueBlue collaborative care

As part of the TrueBlue model, patients were scheduled to visit the practice every three months for a 45-minute nurse consult followed by a 15-minute consult with their usual GP, in which stepped care (psycho- or pharmacotherapy) was offered if depression scores had not improved or had not dropped below a value of five. The PN used the care-plan template and obtained current physical measures and reviewed recent pathology results. PNs also reviewed lifestyle risk factors. They re-administered the PHQ-9 and worked with the patient to identify possible barriers to achieving their goals and discussed ways to overcome the barriers. This information gathering phase of the consultation was an opportunity to assist the patient with self-management by discussing available educational resources, such as the library of fact sheets on aspects of self-management of depression, and setting personal goals for review at the next three-monthly visit.

Statistical analysis

Participants in this study were clustered under clinics by design. It is known that clinics are likely to be different from each other and that ignoring the nested nature of the data may lead to biased estimates of parameter standard errors. However, statistical techniques for correcting for the effects of clustering tend to be overly severe and conservative[23] when a small number of higher level units (the clusters) are used and, therefore, we tested whether the clinics were in fact significantly different from each other. ANCOVAs[24,25] were used to adjust for baseline values and test the significance of changes in depression scores between clinics after six months, using STATA version 11.1 for the statistical analyses.

Of the five clinics in the intervention (clinics 4, 5, 13, 15 and 17), only clinics 4 and 17 were significantly different from each other ($F(1,76)=9.6$, $p<0.001$). No other comparisons were significant between intervention clinics. Of the six clinics in the control group (clinics 1, 2, 3, 6, 16 and 18), only clinics 6 and 18 were significantly different from each other ($F(1,78)=14.5$, $p<0.001$). No other comparisons were significant between control clinics. Furthermore, the intra-correlation coefficient (ICC) of 0.058 for the primary outcome suggests that only 6% of variance could be attributed to the clinics' level. Given this lack of difference between the clinics in each arm coupled

with the sample-size requirements for reliable multilevel modelling[26], we analysed our data at the patient level.

In order to compare the effectiveness of the TrueBlue care model to the usual-care control, ANCOVAs were used to adjust for baseline values and test the significance of changes in continuous variables between the two groups after six months. A multi-level mixed-effects logistic regression (STATA's xtmelogit) was used to test the significance of changes in the binary (categorical) variables between the two groups after six months, with time and group as the independent variables and with random effects at the patient level. (We used mixed-effects logistic-regression model since the pairs of observations over time are not independent, i.e. observations at six months would be expected to be related to the initial baseline observations.) Within each group, changes between the two time points (baseline and six-month visits) were tested using paired *t*-tests for the continuous variables and matched-case-control McNemar χ^2 tests for the binary variables.

The longer-term effects of the intervention were evaluated over the 12-month period using multi-level mixed-effects linear regression (STATA's xtmixed) for the continuous variables and multi-level mixed-effects logistic regression (xtmelogit) for the binary variables. All three-monthly data available in the intervention group over the twelve months were used. Note that the study design could not collect such "usual care" data from the control clinics since the data collection-protocol was part of the intervention. In addition, TrueBlue training was provided to these clinics at six-months after which they ceased to be a control.

Patients from the clinics that withdrew before or during collection of first-visit data were excluded from the analyses. (Data for the thirteen patients from one of these clinics who did complete the study have been included.) Available clinics' characteristics were compared between early dropouts and participating clinics and addressed in terms of their possible impact on the generalisability of the results. Missing six-month data were replaced with their baseline values using the "no change" formulation of intention-to-treat by assuming that no change occurred between baseline and six months. The underlying assumptions of the statistical tests used were assessed.

Results

Demographics (Table 1)

A total of 5401 invitations (3104 intervention and 2297 control; see figure 1) were posted to patients with either T2DM or CHD (or both) identified in the clinics' registers. Approximately 30% (1057 intervention and 537 control, including 39 additional patients invited in the waiting room)

invitations were returned with completed constant forms and PHQ-9 questionnaires. This proportion is typical in studies of this type reported in the literature. Of these, 34% (300 intervention and 229 control) were eligible (a depression score or 5 or more) and were invited to participate. However, 25% of these (94 intervention and 36 control) did not commence when their clinics withdrew before data collection began.

Of the 206 patients in the intervention who commenced the study (figure 1), 17% (n=36) were forced to leave when their clinics withdraw the study. A further 14% (n=28) patients withdrew as the study progressed, with 4% leaving after 6 months, 5% after 9 months and 5% after the full year. Reasons included leaving the area, going into residential care or becoming too ill to continue, but no consistent pattern could be identified. (Exact numbers for each reason are not known.) In the control group, 24% (n=47) of the 194 patients who agreed to participate had forgotten about the study by the time that the 6-month review was to be undertaken and did not want to proceed.

Table 1 presents the characteristics of the patients in both the intervention and control groups who commenced the study, and shows that these characteristics were similar across both groups. There were no significant differences in patient characteristics between the intervention and control at baseline.

Phase 1: Comparison of outcomes between control and intervention groups after 6 months (Table 2)

Table 2 presents baseline and 6-month data for markers used to monitor control of chronic disease for both the intervention and the control group. While the six-month depression scores for all 310 patients (164 intervention and 146 control) were significantly lower than those at baseline in both the intervention group (10.7±0.7 reducing to 7.1±0.8, $t(163)=8.38$, $p<0.001$) and the control group (11.6±0.9 reducing to 9.0±0.9, $t(145)=6.01$, $p<0.001$), the ANCOVA adjusting for the baseline scores showed that the improvement was significantly better in the intervention group than in the control ($F(1,309)=6.40$, $p=0.012$). (The 95% confidence ranges are indicated by the plus-minus sign.)

Half of the patients had only mild-depression at baseline (PHQ-9 scores between five and nine). Because the reported score for many of these patients may be due to their diabetes rather than depression[20], the intervention is unlikely to be able to change these scores. This is one reason that Katon and colleagues used a score of ten or more as an inclusion criterion in their study[15]. Consequently, we examined the change to baseline PHQ-9 scores for the 164 patients (81 intervention and 83 control) with moderate to severe depression (PHQ-9 scores of 10 or more) at baseline. These patients showed significant improvement, with the mean depression score in the intervention group dropping by 5.7±1.3, from 14.4±1.1 down to 8.7±1.3 ($t(80)=9.00$, $p<0.001$), a

clinically-significant change[19]. The improvement in the intervention for these patients was significantly better than in the control group ($F(1,161)=4.02$, $p=0.047$) where the depression score dropped by 4.3±1.2, from 15.1±1.1 down to 10.8±1.4 ($t(82)=6.88$, $p<0.001$).

Except for the HDL measurements, there were no significant changes in biophysical measures after six months in either group. Smoking rates were low at baseline in the patients with established cardiovascular risk factors. Recording of alcohol was sub-optimal, although better than other Australian primary care surveys[27].

The intervention group also showed significantly greater number of patients exercising, referred to and attending an exercise program, and referred to and attending a mental health worker after six months of collaborative care. In the control group, there were no significant changes observed after six months, except that referrals to a mental health worker increased significantly ($p<0.001$) from 9% to 21%, consistent with the action being taken by the nurses as required by the protocol. Neither group showed any significant changes in the number of patients taking antidepressant medication.

Phase 2: Chronic disease outcomes over 12 months using TrueBlue collaborative care (Table 3)

Table 3 presents data at baseline and 12 months for the intervention group for markers used to monitor control of existing diabetes and CHD. The improvement in mental health observed after 6 months was maintained at 12 months, with the significant reduction in the mean depression score maintained (10.7±0.7 to 6.6±0.7, $t(163)=9.92$, $p<0.001$), and nearly 70% of patients having lower depression scores than baseline after one year. Patients with moderate to severe depression at baseline showed an even greater improvement after 12 months of collaborative care, with the mean depression score dropping by 6.4±1.2, from 14.4±0.8 to 8.0±1.2 ($t(80)=10.41$, $p<0.001$). The significant improvement in the mean SF36v2 composite mental-health and physical-health scores observed after 6 months was also maintained at 12 months.

Physiological measures showed a trend, although not significant, to improvement in weight, systolic blood pressure and HDL. Mean baseline lipids and HbA1c were close to guideline targets. The 10-year CVD risk calculated with the Framingham risk equations[28] suggests a small but significant ($p=0.015$) reduction in risk from 27.4% to 24.8% for the patients with only T2DM. (The Framingham risk equations cannot be used for those patients who have CHD.)

The most notable changes in lifestyle after 12-months of the intervention were a significant increase in the numbers of patients who reported taking regular exercise or being referred to an exercise program. Reported referrals and visits to a mental health worker and numbers taking antidepressant medication were also significantly greater at 12 months.

The TrueBlue protocol also included goal setting so that patients could become more pro-active in their own care. An analysis of participant goals revealed that two-thirds of visits resulted in at least one behavioural activation goal being set and, over the course of the study, 86% of patients identified a behavioural activation goal.

Adherence to guidelines (Figure 2)

Figure 2 shows the percentage of TrueBlue patients who had psychosocial and biophysical checks undertaken as recommended by the Australian National Heart Foundation and Diabetes Australia guidelines, with the corresponding percentages for usual care taken from a study of a large sample of Australian general practices[27].

Discussion

Outcomes of phase 1

Depression scores were significantly lower at six months for patients in the intervention practices compared with those in the control group, and the improvement was clinically significant for patients with moderate to severe depression[19], with patients moving one depression category. Patients experienced increased nurse contact time through the nurse consultations, were provided with information about mental health and their physical health through psycho-education resources, and had their treatment intensified when required. Modalities included behavioural activation, antidepressant medication, and referrals to mental health professionals and exercise programs. Similar improvements in depression scores and stepped-up care were observed in the collaborative care model of Katon and colleagues[15]. The reduction in depression scores observed in the control group could be explained, in part, by control practices being provided with each patient's entry-level depression score during the recruitment process as part of the study's safety protocol. Usual care could have been influenced by drawing attention to co-morbid depression[15] as the protocol required that practice nurses take action if severe depression was recorded or if the patient had responded to the suicidal-ideation question. Referrals to mental-health workers by the control clinics had increased significantly consistent with the clinics taking action where warranted. It is also known[29] that recruiting interested patients (those who wanted to participate) from interested clinics (those that agreed to join) can affect the representativeness of the study population. GPs with a particular interest in the study may be more likely to participate and may manage their patients more effectively, irrespective of whether they are in the control or intervention arms. Consequently, a reduction in depression scores in the control group was expected but the structured TrueBlue model did produce a significantly better reduction in depression. While the effect size may be small (Cohen's $f = 0.15$), it is important to note that TrueBlue was designed to be implemented easily within general practices, with running costs funded by existing Australian Medicare rebates, and to

make better use of their existing resources. These features mean that TrueBlue could be easily applied to patients across general practices at a population level, making the benefits clinically important.

Outcomes of phase 2

The key clinical outcomes over a 12-month period in the intervention group (Table 3) were a sustained improvement in mental health, demonstrated by symptom severity score (PHQ-9 total score) and by the patient's function and subjective evaluation of mental health (SF36 mental health composite score) and physical health (SF36 physical health composite score). Regular physical exercise has been shown to be important for reducing depression[30]. The self-reported exercise rates showed significant improvement over the 12 months of collaborative care intervention. The biophysical measures reported in Table 3 showed modest improvements after 12 months and the Framingham risk equations[28] suggest a small but significant reduction in the 10-year CVD risk for the T2DM patients. These improvements were achieved despite that fact that we did not specifically select patients whose physiological parameters exceeded guidelines. Rather, our recruitment process selected from the practice's disease registry on the basis of only the presence of depression and T2DM or CHD and, consequently, many patients were already being treated to target on measures such as cholesterol and HbA1c, leaving little room for improvement.

Limitations

We were able to run TrueBlue only in practices that used clinical software, which we used to generate a disease registry from which participants could be selected, and had a practice nurse on staff. Clinics that chose to take part in the study may not have been representative of wider general practice. Operational limitations further reduced the number of practices over the duration of the study. Patient response rates to the mail-out (28%) may reflect anxiety over the new model of care where the patient discloses depression and visits the PN first rather than only the GP. Usual care in the control clinics may have been changed by patients completing the PHQ-9 and reading the project description. GPs were made aware of individual PHQ-9 results, and took action where warranted. GP awareness of these biophysical and lifestyle risks may be expected to change clinical management. By design, TrueBlue practices needed to incorporate all research activities within the context of their busy clinics, and so only research data that could be extracted automatically were collected. The data dropout resulting from these two factors contributed to the observed small effect size. We were not able to obtain multiple data sets at three-monthly intervals over 12 months of 'usual care' because the act of inviting patients and measuring depression scores and biophysical measures would in itself change the nature of usual care. In addition, practices would not have been

willing to join the study if there was a chance of being randomly allocated to 12 months of being in such a control arm[29].

Collaborative Care

A recent UK study has shown the difficulties of disseminating a guideline without guidance on how to implement collaborative care. Organisational barriers included GPs finding the PHQ-9 awkward to use, nurses not feeling confident or competent due to lack of training, and no guidance on stepped care[13]. The TrueBlue model of collaborative care overcame many of these difficulties. Its successful components were:[9,31]

- *Use of evidence based guidelines.* National Heart Foundation and Diabetes Australia guidelines determined disease management targets and frequency of monitoring.
- *Systematic screening and monitoring of risk factors.* Patients attended three monthly visits in which a care plan with its checklist was completed. By providing a comprehensive collation of all necessary information, this document made clinical management by the patient's GP easier, quicker and more accurate.
- *Timetabled recall visits.* The date of the next appointment was set during each visit, PHQ-9 was readministered and if improvement was insufficient, *stepped care* was followed by initiating drug therapy or increasing the dose, or referral to a mental-health worker according to the guidelines.
- *New or adjusted roles for team members.* PNs took responsibility for organising and monitoring the outcome of referrals, goals and targets. They used a depression questionnaire (the PHQ-9) to open a discussion with patients about their depression symptoms.
- *Information support for the clinician.* GPs were provided with the care plan by the PNs.
- *Enhanced patient self-management.* Patients received their own copy of the care plan with personalised goals, current measurements, targets and safety advice. A component of each visit was to discuss and update their plan, and receive education material on depression.
- *Identified case manager.* PNs became case managers but the GP remained the key clinician.
- *Means of effective communication between all members of the care team.* The care plan was designed to provide relevant clinical information in a succinct format while still being comprehensible to patients.
- *Audit information for the practice.* De-identified data was provided automatically through the care plan.

Applicability of TrueBlue

TrueBlue used existing clinical software and improved the focus of the GP consultation by delegating some tasks to the PN. Higher levels of adherence to guideline-recommended checks were also reported for TrueBlue. Patients and their carers, allied health professionals, specialists and general practitioners gained ready access to patient details provided in TrueBlue's care plan enabling them to see at glance where improve clinical care may be needed. The study achieved improved outcomes with the potential for prevention of heart attack and stroke through reduced 10-year CVD risk. The care plan template also allowed the practice to collect high quality audit data without taking up clinical time. While it was not possible to obtain complete financial data from the clinics specifically relating to the TrueBlue visits, the data that are available suggest that clinics did indeed cover their costs in implementing TrueBlue through Australian Medicare rebates. The success of TrueBlue and TeamCare[15] demonstrates that collaborative care is feasible in routine general practice in Australia and the USA, and could lead to improved outcomes for patients with depression and other chronic diseases[7,32].

Acknowledgements

The authors wish to thank the patients, practice nurses, general practitioners and support staff of the participating clinics Evans Head Medical Centre, Lennox Head Medical Centre, Keen Street Clinic, Tintenbar Medical Centre, Health on Grange, Warradale Medical Centre, Seaton Medical and Specialist Centre, Woodcroft Medical Centre and Southcare Medical Centre. Professors Wayne Katon and Juergen Unützer of the University of Washington were unstinting in their advice on adaptation of the IMPACT model. We would also like to thank Bob Leahy for managing the project, Vince Versace for his statistical advice and Vicki Brown for her assistance during the course of the study.

Contributors

All authors have full access to the complete study dataset, contributed to the design, implemented the project, and co-wrote and approved the manuscript. MM, MC, JD and PR analysed the data. MM, PR and KS developed and ran the practice nurse training program. JD and PR conceived the TrueBlue model during a visit to the IMPACT team. JD is the guarantor.

Funding

Funding was provided by *beyondblue*, the National Depression Initiative in Australia, but it had no other involvement in any phase of the study. All researchers were independent of and not influenced by *beyondblue.*

Competing interests

All authors have completed the Unified Competing Interest form and declare that (1) funding was received from *beyondblue* to carry out the study; (2) they do not have any relationships with any companies that might have an interest in the submitted work in the previous 3 years; (3) their spouses, partners, or children do not have any financial relationships that may be relevant to the submitted work; and (4) they do not have any non-financial interests that may be relevant to the submitted work.

Ethical approval

Ethics approval was obtained from Flinders University's Social and Behavioural Research Ethics Committee. All patients gave informed consent to participate in the study.

Data sharing

The dataset used for the analysis and the computer codes used to produce the results are available from the corresponding author at director@greaterhealth.org.

Licence

References

1. Daar AS, Singer PA, Persad DL, et al. Grand challenges in chronic non-communicable diseases. *Nature* 2007;450(7169):494-96.
2. World Health Organisation. Preventing Chronic Disease: A Vital Investment. Geneva, 2005.
3. Ali S, Stone MA, Peters JL, et al. The prevalence of co-morbid depression in adults with Type 2 diabetes: a systematic review and meta-analysis. *Diabet Med* 2006;23(11):1165-73.
4. Egede LE, Nietert PJ, Zheng D. Depression and all-cause and coronary heart disease mortality among adults with and without diabetes. *Diabetes Care* 2005;28(6):1339-45.
5. Clarke DM, Currie KC. Depression, anxiety and their relationship with chronic diseases: a review of the epidemiology, risk and treatment evidence. *Med J Aust* 2009;190(7 Suppl):S54-60.
6. Freeling P, Rao BM, Paykel ES, et al. Unrecognised depression in general practice. *Br Med J (Clin Res Ed)* 1985;290(6485):1880-83.
7. Gilbody S, Bower P, Fletcher J, et al. Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. *Arch Intern Med* 2006;166(21):2314-21.
8. Katon W, Von Korff M, Lin E, et al. Collaborative management to achieve treatment guidelines. Impact on depression in primary care. *JAMA* 1995;273(13):1026-31.
9. Hickie IB, McGorry PD. Increased access to evidence-based primary mental health care: will the implementation match the rhetoric? *Med J Aust* 2007;187(2):100-03.
10. Lin EH, Katon W, Von Korff M, et al. Relationship of depression and diabetes self-care, medication adherence, and preventive care. *Diabetes Care* 2004;27(9):2154-60.
11. Reddy P, Ford D, Dunbar JA. Improving the quality of diabetes care in general practice. *Aust J Rural Health* 2010;18(5):187-93.
12. Campbell MJ. Cluster randomized trials in general (family) practice research. *Stat Methods Med Res* 2000;9(2):81-94.
13. Mitchell C, Dwyer R, Hagan T, et al. Impact of the QOF and the NICE guideline in the diagnosis and management of depression: a qualitative study. *Br J Gen Pract* 2011;61(586):e279-e89.
14. Unutzer J, Katon W, Williams JW, Jr., et al. Improving primary care for depression in late life: the design of a multicenter randomized trial. *Med Care* 2001;39(8):785-99.
15. Katon WJ, Lin EH, Von Korff M, et al. Collaborative care for patients with depression and chronic illnesses. *N Engl J Med* 2010;363(27):2611-20.
16. Morgan MA, Dunbar J, Reddy P. Collaborative care - The role of practice nurses. *Aust Fam Physician* 2009;38(11):925-26.
17. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16(9):606-13.
18. Morgan M, Dunbar J, Reddy P, et al. The TrueBlue study: is practice nurse-led collaborative care effective in the management of depression for patients with heart disease or diabetes? *BMC Fam Pract* 2009;10:46.
19. Löwe B, Unützer J, Callahan CM, et al. Monitoring depression treatment outcomes with the patient health questionnaire-9. *Medical care* 2004;42(12):1194-201.
20. Reddy P, Philpot B, Ford D, et al. Identification of depression in diabetes: the efficacy of PHQ-9 and HADS-D. *Brit J Gen Prac* 2010;DOI: 10.3399/bjgp10X502128.
21. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30(6):473-83.
22. Dobson KS, Hollon SD, Dimidjian S, et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the prevention of relapse and recurrence in major depression. *Journal of Consulting and Clinical Psychology* 2008;76(3):468-77.
23. Kauermann G, Carroll RJ. A Note on the Efficiency of Sandwich Covariance Matrix Estimation. *J Amer Statistical Assoc* 2001;96(456):1387-96.
24. Senn S. Change from baseline and analysis of covariance revisited. *Statistics in Medicine* 2006;25(24):4334-44.

25. Vickers A. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: A simulation study. *BMC Medical Research Methodology* 2001;1:1-4.
26. Hox JJ. *Applied multilevel analysis*. Second ed. Amsterdam: TT-Publikaties, 1995.
27. Wan Q, Harris MF, Jayasinghe UW, et al. Quality of diabetes care and coronary heart disease absolute risk in patients with type 2 diabetes mellitus in Australian general practice. *Qual Saf Health Care* 2006;15(2):131-35.
28. Anderson KM, Odell PM, Wilson PW, et al. Cardiovascular disease risk profiles. *Am Heart J* 1991;121(1 Pt 2):293-98.
29. Wilson S, Delaney BC, Roalfe A, et al. Randomised controlled trials in primary care: case study. *BMJ* 2000;321:24-27.
30. Wiles NJ, Haase AM, Gallacher J, et al. Physical activity and common mental disorder: results from the Caerphilly study. *Am J Epidemiol* 2007;165(8):946-54.
31. Fuller J, Perkins D, Parker S, et al. Effectiveness of service linkages in primary mental health care: a narrative review part 1. . *BMC Health Services Research* 2011;11:71.
32. Richards DA, Lovell K, Gilbody S, et al. Collaborative care for depression in UK primary care: a randomized controlled trial. *Psychol Med* 2008;38(2):279-87.

Table 1: Patient characteristics at the baseline visits. There were no significant differences between the intervention and control at baseline.

Characteristic	Intervention group (*n*=170)	Control group (*n*=147)
Male (%) / Female (%)	51.8% / 48.2%	55.2% / 44.8%
Age (yr)	68.0 ± 11.7	67.6 ± 11.2
Aboriginal or Torres Strait Islander (%)	0.0%	0.7%
Diagnosis: Type-2 diabetes	37.6%	47.6%
CHD	45.3%	35.8%
Both	17.1%	16.6%
Body mass index (kg/m^2)	31.4 ± 6.0 (*n*=170)	30.8 ± 6.0 (*n* =103)
Systolic blood pressure (mmHg)	134.1 ± 19.0 (*n*=169)	133.5 ± 19.6 (*n* =112)
Total cholesterol (mmol/L)	4.21 ± 0.94 (*n*=165)	4.41 ± 1.06 (*n* =110)
Triglycerides (mmol/L)	1.73 ± 0.88 (*n*=165)	1.92 ± 1.37 (*n* =105)
LDL (mmol/L)	2.22 ± 0.74 (*n*=159)	2.37 ± 0.88 (*n* =89)
HDL (mmol/L)	1.23 ± 0.36 (*n*=159)	1.18 ± 0.33 (*n* =97)
HbA1c (mmol/L)	7.00 ± 1.21 (*n*=94)	7.19 ± 1.42 (*n* =69)
PHQ-9 score	10.7 ± 4.7 (*n*=164)	11.6 ± 5.5 (*n*=146)
PHQ-9 score range at baseline	5 to 24	5 to 27

Table 2: TrueBlue outcomes at six months in the intervention and control groups. (See Table 3 for the list of abbreviations.)

	Intervention				Control				
	n	Baseline	6-months	Within group†	*n*	Baseline	6-months	Within group‡	Between groups
PHQ9 depression score	164	10.7±0.8	7.1±0.8	$p<0.001$	146	11.6±0.9	9.0±0.9	$p<0.001$	$p=0.012$
SF36v2 mental-health score§	71	37.2±3.4	41.1±3.4	$p=0.034$		Not recorded			NS
SF36v2 physical-health score§	71	39.9±2.2	42.5±2.6	$p=0.023$		Not recorded			NS
Body mass index (kg/m^2)	162	31.3±1.0	31.2±1.0	NS	103	30.8±1.2	31.0±1.0	NS	NS
Waist (cm)	161	104.7±2.4	105.0±2.4	NS	80	104.2±4.0	105.8±3.2	NS	NS
Systolic blood pressure (mmHg)	161	134.2±3.0	132.4±2.8	NS	112	133.5±3.8	131.2±3.4	NS	NS
Total cholesterol (mmol/L)	158	4.21±0.16	4.22±0.14	NS	109	4.41±0.20	4.44±0.20	NS	NS
LDL (mmol/L)	154	2.23±0.12	2.17±0.14	NS	86	2.37±0.18	2.29±0.20	NS	NS
HDL (mmol/L)	154	1.23±0.06	1.29±0.06	$p=0.023$	93	1.17±0.06	1.27±0.08	$p=0.011$	NS
Triglycerides (mmol/L)	158	1.72±0.14	1.66±0.12	NS	104	1.84±0.22	1.75±0.18	NS	NS
HbA1c (%)¶	89	6.97±0.24	6.90±0.26	NS	67	7.22±0.34	7.40±0.36	NS	$p=0.049$
Ten-year CVD risk*	61	26.9±3.2	26.1±3.2	NS	46	26.3±3.6	24.7±3.2	NS	NS
The 95% confidence ranges are indicated by the plus-minus (±) sign. Note that lower scores indicate improvement for all items except the SF36v2 and HDL results, where higher scores indicate improvement.									
Smoking	162	15 (9%)	13 (8%)	NS	110	13 (12%)	13 (12%)	NS	NS
Alcohol	104	47 (45%)	51 (49%)	NS	42	27 (64%)	27 (64%)	NS	NS
Exercises 30 mins/day, 5 d/wk	162	66 (41%)	97 (60%)	$p<0.001$	75	22 (29%)	22 (29%)	NS	$p<0.001$
Referred to exercise program	162	32 (20%)	58 (36%)	$p<0.001$	111	15 (14%)	10 (9%)	NS	$p<0.001$
Attends exercise program	162	12 (7%)	23 (14%)	$p=0.041$	79	12 (15%)	9 (11%)	NS	NS
On antidepressant medication	162	27 (17%)	34 (21%)	NS	113	31 (27%)	36 (32%)	NS	$p=0.025$
Referred to mental health worker	162	47 (29%)	58 (36%)	$p=0.022$	114	10 (9%)	24 (21%)	$p<0.001$	$p<0.001$
Attends mental health worker	162	10 (6%)	37 (23%)	$p<0.001$	109	14 (13%)	11 (10%)	NS	$p=0.044$

The values in brackets are the percentages of the total *n*.

(NS) No significant difference. (†) Significant difference between baseline and six-month values within the *intervention* clinics. (‡) Significant difference between baseline and six-months within the *control* clinics. (§) SF36v2 questionnaires were not collected by all clinics. (¶) HbA1c results were only available for patients with T2DM. (*) CVD risk could only be calculated for patients with T2DM.

Table 3: TrueBlue outcomes at 12 months within the intervention clinics only.

	Intervention			
	n	Baseline	12-months	Within group†
PHQ9 depression score	164	10.7±0.7	6.6±0.7	$p<0.001$
SF36v2 mental-health score§	70	36.0±3.2	41.3±2.8	$p<0.001$
SF36v2 physical-health score§	70	40.6±2.2	44.3±2.8	$p<0.001$
Body mass index (kg/m^2)	142	31.4±1.0	31.1±1.0	$p=0.006$
Waist (cm)	141	105.0±2.4	105.2±2.6	NS
Systolic blood pressure (mmHg)	141	135.2±3.2	130.2±3.0	$p=0.016$
Total cholesterol (mmol/L)	138	4.18±0.16	4.28±0.16	NS
LDL (mmol/L)	135	2.19±0.12	2.24±0.20	NS
HDL (mmol/L)	135	1.22±0.06	1.36±0.08	$p<0.001$
Triglycerides (mmol/L)	138	1.73±0.16	1.63±0.14	$p=0.004$
HbA1c (%)¶	79	7.01±0.26	7.04±0.28	NS
Ten-year CVD risk*	55	27.4±3.4	24.9±3.6	$p=0.015$

The 95% confidence ranges are indicated by the plus-minus (±) sign. Lower scores indicate improvement for all items except the SF36v2 and HDL results, where higher scores indicate improvement.

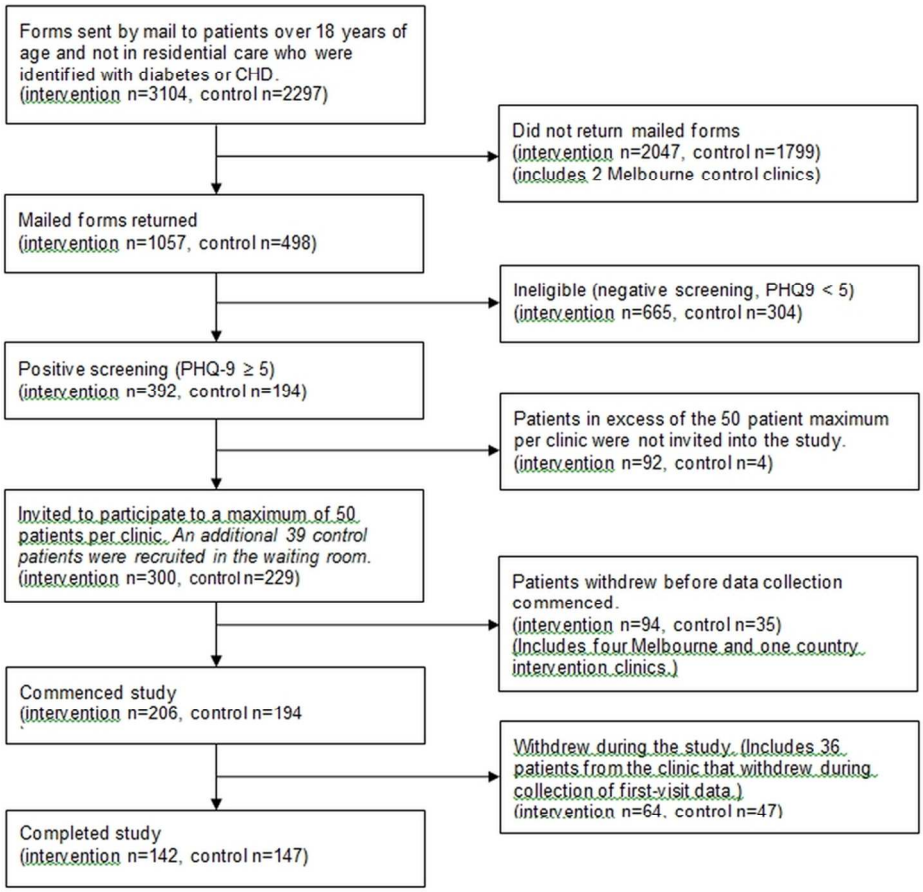
Smoking	142	15 (11%)	11 (8%)	NS
Alcohol	95	45 (47%)	47 (49%)	NS
Exercises 30 mins/day, 5 days/wk	142	57 (40%)	83 (58%)	$p<0.001$
Referred to exercise program	142	26 (18%)	53 (37%)	$p<0.001$
Attends exercise program	142	10 (7%)	17 (12%)	NS
On antidepressant medication	142	22 (15%)	33 (23%)	$p=0.001$
Referred to mental health worker	142	40 (28%)	59 (42%)	$p<0.001$
Attends mental health worker	142	8 (6%)	25 (18%)	$p<0.001$

The values in brackets are the percentages of the total *n*.

(NS) No significant difference. (†) Significant difference between baseline and twelve-month values. (§) SF36v2 questionnaires were not collected by all clinics. (¶) HbA1c results were only available for patients with T2DM. (*) CVD risk could only be calculated for patients with T2DM.

NOTE:
PHQ9 is the 9-question Patient Health Questionnaire.
SF36v2 is version 2 of the Short Form 36-Question health survey.
LDL is low-density lipoprotein.
HDL is high-density lipoprotein.
Glycated haemoglobin (HbA1c) was measured only for patients with diabetes.
Unit of alcohol is 10g of ethanol.

Figure 1. CONSORT flow diagram of the recruitment process.



90x94mm (300 x 300 DPI)

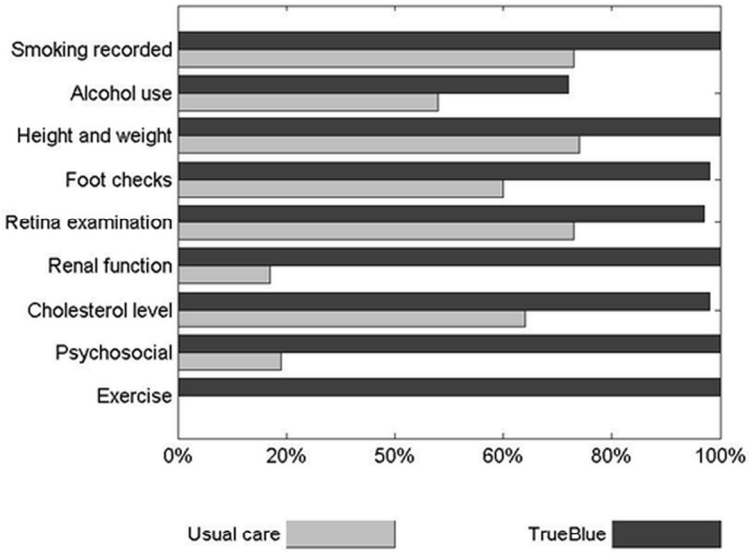


Figure 2: Recording of checks recommended by the National Heart Foundation and Diabetes Australia guidelines. Data for "usual care" were adapted from reference 27. No usual-care data were available for exercise.

95x90mm (300 x 300 DPI)

The TrueBlue model of collaborative care using practice nurses as case managers for depression alongside diabetes or heart disease: a randomised trial

MAJ Morgan[1], MJ Coates[1], JA Dunbar[1,*], P Reddy[2], K Schlicht[1] and J. Fuller[3]

1. Greater Green Triangle University Department of Rural Health, Flinders and Deakin Universities, PO Box 423, Warrnambool, Victoria, 3280, Australia
2. School of Medicine and Public Health, University of Newcastle, Callaghan, NSW, 2308, Australia
3. School of Nursing and Midwifery, Flinders University, GPO Box 2100, Adelaide, SA, 5001, Australia

* Corresponding author

Abstract

Objectives: Determine the effectiveness of collaborative-care in reducing depression in primary-care patients with diabetes or heart disease using practice nurses as case managers.

Design: A two-arm open randomised-cluster trial with wait-list control for 6 months. The intervention was followed over 12 months.

Setting: Eleven Australian general practices, five randomly allocated to the intervention and six to the control.

Participants: 400 primary-care patients (206 intervention, 194 control) with depression and type 2 diabetes, coronary heart disease or both.

Intervention: The practice nurse acted as case manager identifying depression, reviewing pathology results, lifestyle risk factors, and patient goals and priorities. Usual care continued in the controls.

Main outcome measure: A five-point reduction in depression scores for patients with moderate to severe depression. Secondary outcome was improvements in physiological measures.

Results: Mean depression scores after six months of intervention for patients with moderate to severe depression decreased by 5.7±1.3 compared with 4.3±1.2 in the control, a significant (p=0.012) difference. (The plus-minus is the 95% confidence range.) Intervention practices demonstrated adherence to treatment guidelines and intensification of treatment for depression, where exercise increased by 19%, referrals to exercise programs by 16%, referrals to mental health workers (MHWs) by 7%, and visits to MHWs by 17%. Control-practice exercise did not change, referrals to exercise programs dropped by 5%, and visits to MHWs by 3%. Only referrals to MHW increased, by 12%. Intervention improvements were sustained over 12 months, with significant

(p=0.015) decrease in 10-year cardiovascular-disease risk from 27.4±3.4% to 24.8±3.8%. A review of patients indicated the study's safety protocols were followed.

Conclusion: TrueBlue participants showed significantly improved depression and treatment intensification, sustained over 12-months of intervention, and reduced 10-year CVD risk. Collaborative care using practice nurses appears to be an effective primary-care intervention.

Trial registration: ACTRN12609000333213 (Australia and New Zealand Clinical Trials Registry) ~~ACTRN12609000333213~~.

Article summary

Article focus

- To determine the effectiveness of a collaborative-care model to reduce depression in primary-care patients with diabetes or heart disease.
- To determine the effectiveness of using practice nurses as case managers of patients with depression and diabetes, heart disease or both.

Key messages

- The TrueBlue model of collaborative care can be introduced within the general practice workforce with practice nurses taking on the role of case manager.
- Practice nurses can improve care of depression in patients with diabetes or heart disease, leading to better outcomes and reduced 10-year CVD risk.
- The care of patients using the TrueBlue model is closer to "best practice" guidelines, with substantially better levels of adherence to guideline-recommended checks than occur in usual care.

Strengths and limitations of this study

- The TrueBlue model of collaborative care overcomes many of the difficulties in implementing a guideline for the treatment of co-morbid depression.
- The study's purpose-designed care plan gives patients and their carers, allied health professionals, specialists and general practitioners ready access to patient details enabling them to see at glance where improve clinical care may be needed.
- Clinics were able to recover the costs of the collaborative care through Australian Medicare rebates.

- The study could only be run in practices that had a practice nurse on staff to carry out the intervention and had access to clinical software capable of generating a disease registry from which ~~participants~~ patients could be selected to participate in the trial.

- Differences between TrueBlue- and control-practice outcomes may have been reduced by patients completing the PHQ-9 depression questionnaire and reading the project description, and by GPs being made aware of individual PHQ-9 results so that they could take action where warranted.

Introduction

Managing diabetes and heart disease has been highlighted as one of the global "grand challenges in chronic non-communicable diseases"[1] because the prevalence of these two preventable diseases is increasing[2]. Along with depression, they have been identified as health priority areas in many countries. A vicious cycle exists between depression and these chronic diseases, with each being a risk factor for the other[3]. Higher mortality has been demonstrated for people with depression and type 2 diabetes (T2DM) or coronary heart disease (CHD) beyond that due to the separate diseases alone[4]. For patients with depression and T2DM or CHD or both, there are increased risks of adverse outcomes[5] but this co-morbid depression is often missed in primary care[6]. Consequently, the identification of depression has now been incorporated in many heart disease guidelines as one of the requirements for optimal management. Meeting these challenges will require an innovative use of the existing general practice workforce and such a re-orientation of resources has been identified as one of the grand challenges[1].

Collaborative care is a system that has been shown to be more effective for chronic disease management than standard care[7]. It includes a reorientation of the medical workforce through new or adjusted roles for team members, particularly using practice nurses as the identified case manager to undertake the care of the patients[8,9]. It also includes the use of evidence based guidelines, systematic screening and monitoring of risk factors, timetabled recall visits, information support for the clinician, enhanced patient self-management, a means of effective communication between all members of the care team and audit information for the practice. Since self-care for diabetes has been found to be suboptimal across a range of self-managed activities, particularly for patients with depression, a collaborative care model may be able to achieve better quality of care through the case manager monitoring patient progress[10,11].

Evaluation of a change in the way general practice clinics look after patients requires complex intervention methodology[12] beyond single interventions such as introduction of a guideline with financial incentives[13]. This methodology began with a search for potential models of care (step I), and led to adopting the University of Washington's successful IMPACT model of Collaborative

Care for depression[14,15]. In the exploratory trial (step II), our pilot project[16] adapted IMPACT by training practice nurses as case managers. Practice nurses were trained to screen for depression using a patient self-report measure, the nine-item Patient Health Questionnaire (PHQ-9)[17], as part of comprehensive chronic disease management. They were also trained to use a protocol for care management based on the depression scores. The depression screening and management was embedded in routine visits for patients with diabetes or CHD. The pilot demonstrated that it was feasible to detect, monitor and treat depression in routine general practice alongside the usual biophysical measures, and identified moderate to severe depression in 34% of participants. The TrueBlue study was a randomised cluster trial (step III) that built on and extended the pilot. It investigated whether a collaborative care model (the intervention) is better than usual care (the control) for the management of patients with depression and T2DM, CHD or both in Australian general practice. It was designed to fit into normal clinic operations, making use of practice nurses and medical software, and able to be funded by existing Australian Medicare rebates.

Methods/design

Study design

The design and methodology of the study have been described in detail elsewhere[18]. The study commenced in 2009 and was undertaken in two phases. The first phase was a cluster randomised intervention trial in which general practices were randomly allocated to either an intervention group in which nurse-led collaborative care was undertaken or to a wait-list control group in which usual care led by the general practitioner (GP) was continued. At six months, the TrueBlue training was provided to the control practices. The key aims of the first phase were to determine whether participants with moderate to severe depression in the intervention group showed at least a five-point reduction from the baseline depression scores after six months of intervention and whether this reduction was significantly better than in the control group. A five-point reduction reflects a clinically-relevant change in individuals receiving depression treatment[19]. The secondary outcome was to determine whether the intervention also led to improvements in the patients' physiological measures. The second phase followed the intervention group for an additional six months to determine how the collaborative care model affected health outcomes over a twelve-month period.

Sample size

The sample size calculation was based on detecting a 50% reduction in depression score at the 0.05 significance level with 80% power and a two-tailed test. Detecting a 50% reduction is more stringent than detecting a five-point reduction and provided some additional buffering. Using depression scores from an earlier study (a mean of 5.5 and standard deviation of 6.1)[11], the calculation indicated that 237 patients would be required in each group. An intra-cluster correlation

of 0.04 was used (S. K. Lo, pers. comm.), with a recruitment target of 50 patients per clinic. (Fifty patients were chosen so that clinics could budget for a nurse's time to carry out the intervention with four patients each week over the 3-month cycle of care.) To allow for difficulties in recruitment, ~~for~~ a 50% dropout was used. Based on these, the study required 450 patients from nine clinics in the intervention group and the same in the control group.

Practice recruitment

Practices were identified in city and country areas on the basis of having a practice nurse to provide the collaborative care and being able to identify eligible patients, those with CHD or T2DM or both, from their registries; these were invited to participate in the study until the eighteen clinics required by the sample-size calculation were recruited. They were allocated by a random number generator to either the intervention or control arm of the study. The unit of randomisation was the clinic. Five practices (3 country, 2 city) in the intervention group and six (2 country, 4 city) in the control group completed the study. One country intervention clinic withdrew whilst first-visit data were being collected when its TrueBlue-trained practice nurse left the clinic, but some (n=13) patients from it did complete the study and data were collected from them. The study team was not able to determine why the other clinics withdrew.

Patient selection

Eligible patients were sent a postal survey that included a consent form and were asked to complete and return the enclosed PHQ-9 questionnaire, a self-report measure of depression.[17] The PHQ-9 has nine items, each scored from 0 (no problems) to 3 (problems nearly every day). The sum of the scores of the nine items will lie in one of five depression categories: none (0–4), mild (5–9), moderate (10–14), moderately severe (15–19) and severe (20–27). (While it is known that responses to some of the PHQ-9 items may overlap with diabetes symptoms[20], our pilot demonstrated that nurses and patients preferred using the PHQ-9 because the patient's response to each of its items became the basis for the problem solving and goal setting activities that were part of TrueBlue.) Patients with scores of five or above, indicating some form of depression, were invited to participate in the study. A maximum of 50 patients per practice were invited. Patients in residential care or under 18 years of age were not eligible. Figure 1 presents the CONSORT diagram of the patient-recruitment process.

Patient safety

Participation in the intervention included a series of patient visits to their practice nurse (PN) and usual GP every three months over a 12-month period. Patients in the control group continued with their "usual care". The control clinics were also provided with the PHQ-9 depression scores to

ensure patient safety during the trial. The protocol required that practice nurses take action if severe depression was recorded in the returned PHQ-9 or if the patient had responded to the suicidal-ideation question (question 9) on the questionnaire. This action was to be taken *irrespective of whether the clinic was in the intervention or in the control group*.

Practice nurse training

The PN training included a two-day workshop to prepare them for their enhanced roles in nurse-led collaborative care. Topics in the workshop included identifying and monitoring depression using the PHQ-9 questionnaire, and quality of life responses using version 2 of the SF36 questionnaire[21]. Patient goal setting and problem solving were key components of the training with a particular emphasis on behavioural techniques to achieve improved mental health[22]. The training also prepared the PNs for their role as case managers including ensuring Diabetes Australia and Australian National Heart Foundation guidelines were being followed, and referrals were provided to appropriate services, such as allied health and mental health professionals, through discussion with the GPs.

Data collection

The research team developed a protocol-driven care-plan template from which study data could be extracted automatically and sent to the research team. The template was designed to be a multi-purpose document in which the patient's past medical history, current medications, allergies, biophysical and psychosocial measures, lifestyle risks, personal goals and referrals were recorded. It was designed to comply with the requirements to claim Australian Medicare rebates for care planning and to provide a checklist for "gold-standard" care. A copy of the care plan was provided to the patient as a written record of their progress.

The care-plan template collected physical measures, including body-mass index, waist circumference, weight and blood pressure, and the latest pathology results, including lipid profile, glycaemic control (HbA1c) and renal function. Data also included lifestyle risk factors, such as smoking, alcohol consumption and level of physical activity, and depression score as measured by the PHQ-9 questionnaire. Referrals to and attendance at exercise programmes and with mental health workers were also recorded, along with the patient's own goals and possible barriers to achieving these goals. The care-plan template was used by the intervention-group clinics to acquire patient data at three-monthly intervals over a twelve-month period.

In the control group, the only complete dataset recorded using our comprehensive protocol-driven care-plan template was obtained after the six months of "usual care" when the TrueBlue training was offered to the control clinics. No baseline or three-month datasets were acquired since

the study was deliberately designed to avoid changing the "usual care" that would have otherwise occurred by introducing our care-plan template. The study was designed in this way to be run pragmatically in the context of the clinics' normal activities. The only baseline measure obtained was the depression score. On completion of the study, we retrospectively collected what baseline data the control clinics routinely recorded in their electronic medical records in order to have data for two time points, baseline and six-months.

TrueBlue collaborative care

As part of the TrueBlue model, patients were scheduled to visit the practice every three months for a 45-minute nurse consult followed by a 15-minute consult with their usual GP, in which stepped care (psycho- or pharmacotherapy) was offered if depression scores had not improved or had not dropped below a value of five. The PN used the care-plan template and obtained current physical measures and reviewed recent pathology results. PNs also reviewed lifestyle risk factors. They re-administered the PHQ-9 and worked with the patient to identify possible barriers to achieving their goals and discussed ways to overcome the barriers. This information gathering phase of the consultation was an opportunity to assist the patient with self-management by discussing available educational resources, such as the library of fact sheets on aspects of self-management of depression, and setting personal goals for review at the next three-monthly visit.

Statistical analysis

Participants in this study were clustered under clinics by design. It is known that clinics are likely to be different from each other and that ignoring the nested nature of the data may lead to biased estimates of parameter standard errors. However, statistical techniques for correcting for the effects of clustering tend to be overly severe and conservative[23] when a small number of higher level units (the clusters) are used and, therefore, we tested whether the clinics were in fact significantly different from each other. ANCOVAs[24,25] were used to adjust for baseline values and test the significance of changes in depression scores between clinics after six months, using STATA version 11.1 for the statistical analyses.

Of the five clinics in the intervention (clinics 4, 5, 13, 15 and 17), only clinics 4 and 17 were significantly different from each other ($F(1,76)=9.6$, $p<0.001$). No other comparisons were significant between intervention clinics. Of the six clinics in the control group (clinics 1, 2, 3, 6, 16 and 18), only clinics 6 and 18 were significantly different from each other ($F(1,78)=14.5$, $p<0.001$). No other comparisons were significant between control clinics. Furthermore, the intra-correlation coefficient (ICC) of 0.058 for the primary outcome suggests that only 6% of variance could be attributed to the clinics' level. Given this lack of difference between the clinics in each arm coupled

with the sample-size requirements for reliable multilevel modelling[26], we analysed our data at the patient level.

In order to compare the effectiveness of the TrueBlue care model to the usual-care control, ANCOVAs were used to adjust for baseline values and test the significance of changes in continuous variables between the two groups after six months. A multi-level mixed-effects logistic regression (STATA's xtmelogit) was used to test the significance of changes in the binary (categorical) variables between the two groups after six months, with time and group as the independent variables and with random effects at the patient level. (We used mixed-effects logistic-regression model since the pairs of observations over time are not independent, i.e. observations at six months would be expected to be related to the initial baseline observations.) Within each group, changes between the two time points (baseline and six-month visits) were tested using paired *t*-tests for the continuous variables and matched-case-control McNemar χ^2 tests for the binary variables.

The longer-term effects of the intervention were evaluated over the 12-month period using multi-level mixed-effects linear regression (STATA's xtmixed) for the continuous variables and multi-level mixed-effects logistic regression (xtmelogit) for the binary variables. All three-monthly data available in the intervention group over the twelve months were used. Note that the study design could not collect such "usual care" data from the control clinics since the data collection-protocol was part of the intervention. In addition, TrueBlue training was provided to these clinics at six-months after which they ceased to be a control.

Patients from the clinics that withdrew before or during collection of first-visit data were excluded from the analyses. (Data for the thirteen patients from one of these clinics who did complete the study have been included.) Available clinics' characteristics were compared between early dropouts and participating clinics and addressed in terms of their possible impact on the generalisability of the results. Missing six-month data were replaced with their baseline values using the "no change" formulation of intention-to-treat by assuming that no change occurred between baseline and six months. The underlying assumptions of the statistical tests used were assessed.

Results

Demographics (Table 1)

A total of 5401 invitations (3104 intervention and 2297 control; see figure 1) were posted to patients with either T2DM or CHD (or both) identified in the clinics' registers. Approximately 30% (1057 intervention and 537 control, including 39 additional patients invited in the waiting room)

invitations were returned with completed constant forms and PHQ-9 questionnaires. This proportion is typical in studies of this type reported in the literature. Of these, 34% (300 intervention and 229 control) were eligible (a depression score or 5 or more) and were invited to participate. However, 25% of these (94 intervention and 36 control) did not commence when their clinics withdrew before data collection began.

Of the 206 patients in the intervention who commenced the study (figure 1), 17% (*n*=36) were forced to leave when their clinics withdraw the study. A further 14% (*n*=28) patients withdrew as the study progressed, with 4% leaving after 6 months, 5% after 9 months and 5% after the full year. Reasons included leaving the area, going into residential care or becoming too ill to continue, but no consistent pattern could be identified. (Exact numbers for each reason are not known.) In the control group, 24% (*n*=47) of the 194 patients who agreed to participate had forgotten about the study by the time that the 6-month review was to be undertaken and did not want to proceed.

Table 1 presents the characteristics of the patients in both the intervention and control groups who commenced the study, and shows that these characteristics were similar across both groups. There were no significant differences in patient characteristics between the intervention and control at baseline.

Phase 1: Comparison of outcomes between control and intervention groups after 6 months (Table 2)

Table 2 presents baseline and 6-month data for markers used to monitor control of chronic disease for both the intervention and the control group. While the six-month depression scores for all 310 patients (164 intervention and 146 control) were significantly lower than those at baseline in both the intervention group (10.7±0.7 reducing to 7.1±0.8, $t(163)=8.38$, $p<0.001$) and the control group (11.6±0.9 reducing to 9.0±0.9, $t(145)=6.01$, $p<0.001$), the ANCOVA adjusting for the baseline scores showed that the improvement was significantly better in the intervention group than in the control ($F(1,309)=6.40$, $p=0.012$). (The 95% confidence ranges are indicated by the plus-minus sign.)

Half of the patients had only mild-depression at baseline (PHQ-9 scores between five and nine). Because the reported score for many of these patients may be due to their diabetes rather than depression[20], the intervention is unlikely to be able to change these scores. This is one reason that Katon and colleagues used a score of ten or more as an inclusion criterion in their study[15]. Consequently, we examined the change to baseline PHQ-9 scores for the 164 patients (81 intervention and 83 control) with moderate to severe depression (PHQ-9 scores of 10 or more) at baseline. These patients showed significant improvement, with the mean depression score in the intervention group dropping by 5.7±1.3, from 14.4±1.1 down to 8.7±1.3 ($t(80)=9.00$, $p<0.001$), a

clinically-significant change[19]. The improvement in the intervention for these patients was significantly better than in the control group ($F(1,161)=4.02$, $p=0.047$) where the depression score dropped by 4.3±1.2, from 15.1±1.1 down to 10.8±1.4 ($t(82)=6.88$, $p<0.001$).

Except for the HDL measurements, there were no significant changes in biophysical measures after six months in either group. Smoking rates were low at baseline in the patients with established cardiovascular risk factors. Recording of alcohol was sub-optimal, although better than other Australian primary care surveys[27].

The intervention group also showed significantly greater number of patients exercising, referred to and attending an exercise program, and referred to and attending a mental health worker after six months of collaborative care. In the control group, there were no significant changes observed after six months, except that referrals to a mental health worker increased significantly ($p<0.001$) from 9% to 21%, consistent with the action being taken by the nurses as required by the protocol. Neither group showed any significant changes in the number of patients taking antidepressant medication.

Phase 2: Chronic disease outcomes over 12 months using TrueBlue collaborative care (Table 3)

Table 3 presents data at baseline and 12 months for the intervention group for markers used to monitor control of existing diabetes and CHD. The improvement in mental health observed after 6 months was maintained at 12 months, with the significant reduction in the mean depression score maintained (10.7±0.7 to 6.6±0.7, $t(163)=9.92$, $p<0.001$), and nearly 70% of patients having lower depression scores than baseline after one year. Patients with moderate to severe depression at baseline showed an even greater improvement after 12 months of collaborative care, with the mean depression score dropping by 6.4±1.2, from 14.4±0.8 to 8.0±1.2 ($t(80)=10.41$, $p<0.001$). The significant improvement in the mean SF36v2 composite mental-health and physical-health scores observed after 6 months was also maintained at 12 months.

Physiological measures showed a trend, although not significant, to improvement in weight, systolic blood pressure and HDL. Mean baseline lipids and HbA1c were close to guideline targets. The 10-year CVD risk calculated with the Framingham risk equations[28] suggests a small but significant ($p=0.015$) reduction in risk from 27.4% to 24.8% for the patients with only T2DM. (The Framingham risk equations cannot be used for those patients who have CHD.)

The most notable changes in lifestyle after 12-months of the intervention were a significant increase in the numbers of patients who reported taking regular exercise or being referred to an exercise program. Reported referrals and visits to a mental health worker and numbers taking antidepressant medication were also significantly greater at 12 months.

The TrueBlue protocol also included goal setting so that patients could become more pro-active in their own care. An analysis of participant goals revealed that two-thirds of visits resulted in at least one behavioural activation goal being set and, over the course of the study, 86% of patients identified a behavioural activation goal.

Adherence to guidelines (Figure 2)

Figure 2 shows the percentage of TrueBlue patients who had psychosocial and biophysical checks undertaken as recommended by the Australian National Heart Foundation and Diabetes Australia guidelines, with the corresponding percentages for usual care taken from a study of a large sample of Australian general practices[27].

Discussion

Outcomes of phase 1

Depression scores were significantly lower at six months for patients in the intervention practices compared with those in the control group, and the improvement was clinically significant for patients with moderate to severe depression[19], with patients moving one depression category. Patients experienced increased nurse contact time through the nurse consultations, were provided with information about mental health and their physical health through psycho-education resources, and had their treatment intensified when required. Modalities included behavioural activation, antidepressant medication, and referrals to mental health professionals and exercise programs. Similar improvements in depression scores and stepped-up care were observed in the collaborative care model of Katon and colleagues[15]. The reduction in depression scores observed in the control group could be explained, in part, by control practices being provided with each patient's entry-level depression score during the recruitment process as part of the study's safety protocol. Usual care could have been influenced by drawing attention to co-morbid depression[15] as the protocol required that practice nurses take action if severe depression was recorded or if the patient had responded to the suicidal-ideation question. Referrals to mental-health workers by the control clinics had increased significantly consistent with the clinics taking action where warranted. It is also known[29] that recruiting interested patients (those who wanted to participate) from interested clinics (those that agreed to join) can affect the representativeness of the study population. GPs with a particular interest in the study may be more likely to participate and may manage their patients more effectively, irrespective of whether they are in the control or intervention arms. Consequently, a reduction in depression scores in the control group was expected but the structured TrueBlue model did produce a significantly better reduction in depression. While the effect size may be small (Cohen's $f = 0.15$), it is important to note that TrueBlue was designed to be implemented easily within general practices, with running costs funded by existing Australian Medicare rebates, and to

make better use of their existing resources. These features mean that TrueBlue could be easily applied to patients across general practices at a population level, making the benefits clinically important.

Outcomes of phase 2

The key clinical outcomes over a 12-month period in the intervention group (Table 3) were a sustained improvement in mental health, demonstrated by symptom severity score (PHQ-9 total score) and by the patient's function and subjective evaluation of mental health (SF36 mental health composite score) and physical health (SF36 physical health composite score). Regular physical exercise has been shown to be important for reducing depression[30]. The self-reported exercise rates showed significant improvement over the 12 months of collaborative care intervention. The biophysical measures reported in Table 3 showed modest improvements after 12 months and the Framingham risk equations[28] suggest a small but significant reduction in the 10-year CVD risk for the T2DM patients. These improvements were achieved despite that fact that we did not specifically select patients whose physiological parameters exceeded guidelines. Rather, our recruitment process selected from the practice's disease registry on the basis of only the presence of depression and T2DM or CHD and, consequently, many patients were already being treated to target on measures such as cholesterol and HbA1c, leaving little room for improvement.

Limitations

We were able to run TrueBlue only in practices that used clinical software, which we used to generate a disease registry from which participants could be selected, and had a practice nurse on staff. Clinics that chose to take part in the study may not have been representative of wider general practice. Operational limitations further reduced the number of practices over the duration of the study. Patient response rates to the mail-out (28%) may reflect anxiety over the new model of care where the patient discloses depression and visits the PN first rather than only the GP. Usual care in the control clinics may have been changed by patients completing the PHQ-9 and reading the project description. GPs were made aware of individual PHQ-9 results, and took action where warranted. GP awareness of these biophysical and lifestyle risks may be expected to change clinical management. By design, TrueBlue practices needed to incorporate all research activities within the context of their busy clinics, and so only research data that could be extracted automatically were collected. The data dropout resulting from these two factors contributed to the observed small effect size. We were not able to obtain multiple data sets at three-monthly intervals over 12 months of 'usual care' because the act of inviting patients and measuring depression scores and biophysical measures would in itself change the nature of usual care. In addition, practices would not have been

willing to join the study if there was a chance of being randomly allocated to 12 months of being in such a control arm[29].

Collaborative Care

A recent UK study has shown the difficulties of disseminating a guideline without guidance on how to implement collaborative care. Organisational barriers included GPs finding the PHQ-9 awkward to use, nurses not feeling confident or competent due to lack of training, and no guidance on stepped care[13]. The TrueBlue model of collaborative care overcame many of these difficulties. Its successful components were:[9,31]

- *Use of evidence based guidelines.* National Heart Foundation and Diabetes Australia guidelines determined disease management targets and frequency of monitoring.
- *Systematic screening and monitoring of risk factors.* Patients attended three monthly visits in which a care plan with its checklist was completed. By providing a comprehensive collation of all necessary information, this document made clinical management by the patient's GP easier, quicker and more accurate.
- *Timetabled recall visits.* The date of the next appointment was set during each visit, PHQ-9 was readministered and if improvement was insufficient, *stepped care* was followed by initiating drug therapy or increasing the dose, or referral to a mental-health worker according to the guidelines.
- *New or adjusted roles for team members.* PNs took responsibility for organising and monitoring the outcome of referrals, goals and targets. They used a depression questionnaire (the PHQ-9) to open a discussion with patients about their depression symptoms.
- *Information support for the clinician.* GPs were provided with the care plan by the PNs.
- *Enhanced patient self-management.* Patients received their own copy of the care plan with personalised goals, current measurements, targets and safety advice. A component of each visit was to discuss and update their plan, and receive education material on depression.
- *Identified case manager.* PNs became case managers but the GP remained the key clinician.
- *Means of effective communication between all members of the care team.* The care plan was designed to provide relevant clinical information in a succinct format while still being comprehensible to patients.
- *Audit information for the practice.* De-identified data was provided automatically through the care plan.

Applicability of TrueBlue

TrueBlue used existing clinical software and improved the focus of the GP consultation by delegating some tasks to the PN. Higher levels of adherence to guideline-recommended checks were also reported for TrueBlue. Patients and their carers, allied health professionals, specialists and general practitioners gained ready access to patient details provided in TrueBlue's care plan enabling them to see at glance where improve clinical care may be needed. The study achieved improved outcomes with the potential for prevention of heart attack and stroke through reduced 10-year CVD risk. The care plan template also allowed the practice to collect high quality audit data without taking up clinical time. While it was not possible to obtain complete financial data from the clinics specifically relating to the TrueBlue visits, the data that are available suggest that clinics did indeed cover their costs in implementing TrueBlue through Australian Medicare rebates. The success of TrueBlue and TeamCare[15] demonstrates that collaborative care is feasible in routine general practice in Australia and the USA, and could lead to improved outcomes for patients with depression and other chronic diseases[7,32].

Acknowledgements

The authors wish to thank the patients, practice nurses, general practitioners and support staff of the participating clinics Evans Head Medical Centre, Lennox Head Medical Centre, Keen Street Clinic, Tintenbar Medical Centre, Health on Grange, Warradale Medical Centre, Seaton Medical and Specialist Centre, Woodcroft Medical Centre and Southcare Medical Centre. Professors Wayne Katon and Juergen Unützer of the University of Washington were unstinting in their advice on adaptation of the IMPACT model. We would also like to thank Bob Leahy for managing the project, Vince Versace for his statistical advice and Vicki Brown for her assistance during the course of the study.

Contributors

All authors have full access to the complete study dataset, contributed to the design, implemented the project, and co-wrote and approved the manuscript. MM, MC, JD and PR analysed the data. MM, PR and KS developed and ran the practice nurse training program. JD and PR conceived the TrueBlue model during a visit to the IMPACT team. JD is the guarantor.

Funding

Funding was provided by *beyondblue*, the National Depression Initiative in Australia, but it had no other involvement in any phase of the study. All researchers were independent of and not influenced by *beyondblue.*

Competing interests

All authors have completed the Unified Competing Interest form and declare that (1) funding was received from *beyondblue* to carry out the study; (2) they do not have any relationships with any companies that might have an interest in the submitted work in the previous 3 years; (3) their spouses, partners, or children do not have any financial relationships that may be relevant to the submitted work; and (4) they do not have any non-financial interests that may be relevant to the submitted work.

Ethical approval

Ethics approval was obtained from Flinders University's Social and Behavioural Research Ethics Committee. All patients gave informed consent to participate in the study.

Data sharing

The dataset used for the analysis and the computer codes used to produce the results are available from the corresponding author.

Licence

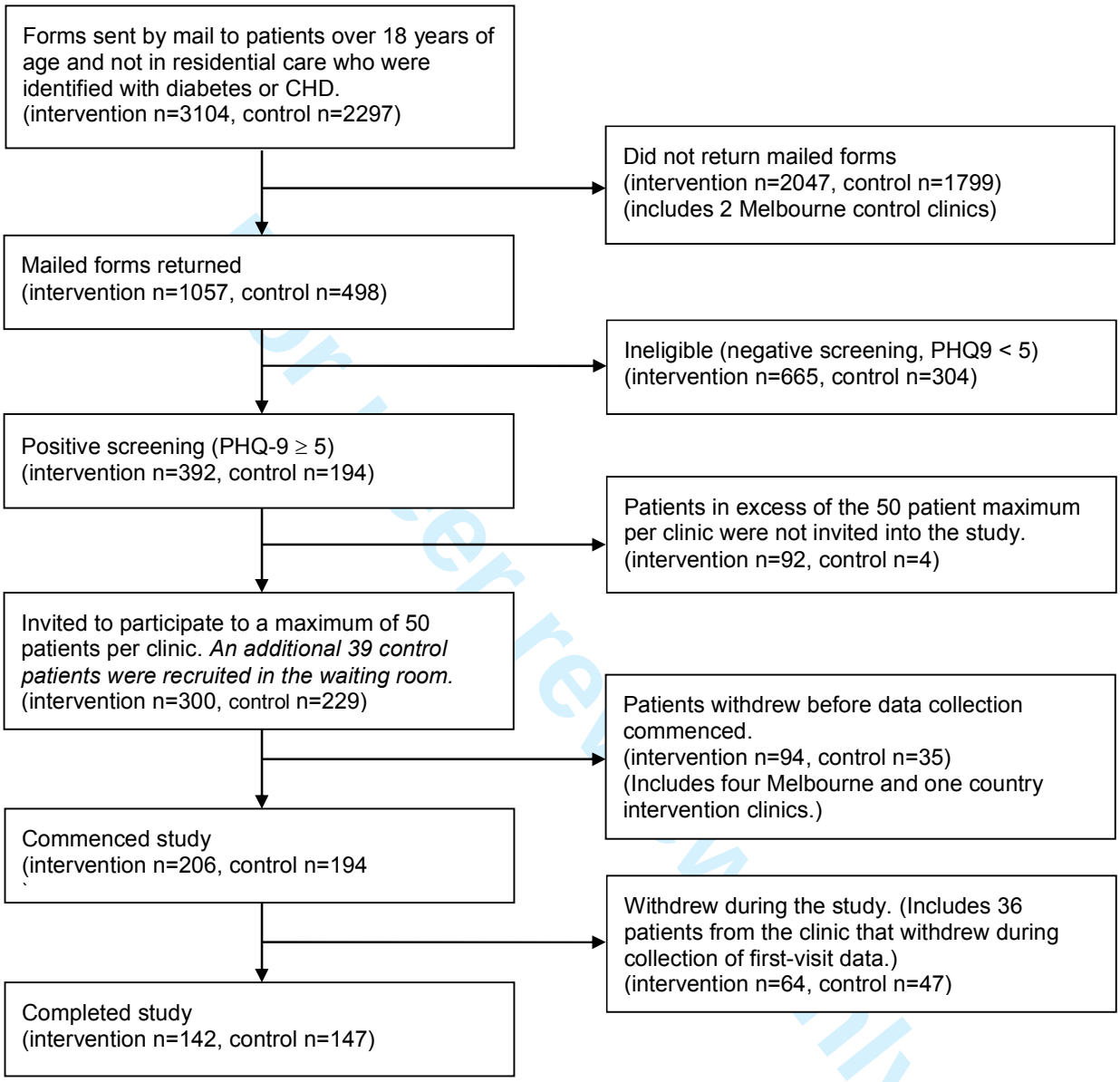
References

1. Daar AS, Singer PA, Persad DL, Pramming SK, Matthews DR, Beaglehole R, et al. Grand challenges in chronic non-communicable diseases. *Nature* 2007;450(7169):494-96.
2. World Health Organisation. Preventing Chronic Disease: A Vital Investment. Geneva, 2005.
3. Ali S, Stone MA, Peters JL, Davies MJ, Khunti K. The prevalence of co-morbid depression in adults with Type 2 diabetes: a systematic review and meta-analysis. *Diabet Med* 2006;23(11):1165-73.
4. Egede LE, Nietert PJ, Zheng D. Depression and all-cause and coronary heart disease mortality among adults with and without diabetes. *Diabetes Care* 2005;28(6):1339-45.
5. Clarke DM, Currie KC. Depression, anxiety and their relationship with chronic diseases: a review of the epidemiology, risk and treatment evidence. *Med J Aust* 2009;190(7 Suppl):S54-60.
6. Freeling P, Rao BM, Paykel ES, Sireling LI, Burton RH. Unrecognised depression in general practice. *Br Med J (Clin Res Ed)* 1985;290(6485):1880-83.
7. Gilbody S, Bower P, Fletcher J, Richards D, Sutton AJ. Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. *Arch Intern Med* 2006;166(21):2314-21.
8. Katon W, Von Korff M, Lin E, Walker E, Simon GE, Bush T, et al. Collaborative management to achieve treatment guidelines. Impact on depression in primary care. *JAMA* 1995;273(13):1026-31.
9. Hickie IB, McGorry PD. Increased access to evidence-based primary mental health care: will the implementation match the rhetoric? *Med J Aust* 2007;187(2):100-03.
10. Lin EH, Katon W, Von Korff M, Rutter C, Simon GE, Oliver M, et al. Relationship of depression and diabetes self-care, medication adherence, and preventive care. *Diabetes Care* 2004;27(9):2154-60.
11. Reddy P, Ford D, Dunbar JA. Improving the quality of diabetes care in general practice. *Aust J Rural Health* 2010;18(5):187-93.
12. Campbell MJ. Cluster randomized trials in general (family) practice research. *Stat Methods Med Res* 2000;9(2):81-94.
13. Mitchell C, Dwyer R, Hagan T, Mathers N. Impact of the QOF and the NICE guideline in the diagnosis and management of depression: a qualitative study. *Br J Gen Pract* 2011;61(586):e279-e89.
14. Unutzer J, Katon W, Williams JW, Jr., Callahan CM, Harpole L, Hunkeler EM, et al. Improving primary care for depression in late life: the design of a multicenter randomized trial. *Med Care* 2001;39(8):785-99.
15. Katon WJ, Lin EH, Von Korff M, Ciechanowski P, Ludman EJ, Young B, et al. Collaborative care for patients with depression and chronic illnesses. *N Engl J Med* 2010;363(27):2611-20.
16. Morgan MA, Dunbar J, Reddy P. Collaborative care - The role of practice nurses. *Aust Fam Physician* 2009;38(11):925-26.
17. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16(9):606-13.
18. Morgan M, Dunbar J, Reddy P, Coates M, Leahy R. The TrueBlue study: is practice nurse-led collaborative care effective in the management of depression for patients with heart disease or diabetes? *BMC Fam Pract* 2009;10:46.
19. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. *Medical care* 2004;42(12):1194-201.
20. Reddy P, Philpot B, Ford D, Dunbar JA. Identification of depression in diabetes: the efficacy of PHQ-9 and HADS-D. *Brit J Gen Prac* 2010;DOI: 10.3399/bjgp10X502128.
21. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30(6):473-83.
22. Dobson KS, Hollon SD, Dimidjian S, Schmaling KB, Kohlenberg RJ, Gallop RJ, et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication

in the prevention of relapse and recurrence in major depression. *Journal of Consulting and Clinical Psychology* 2008;76(3):468-77.

23. Kauermann G, Carroll RJ. A Note on the Efficiency of Sandwich Covariance Matrix Estimation. *J Amer Statistical Assoc* 2001;96(456):1387-96.
24. Senn S. Change from baseline and analysis of covariance revisited. *Statistics in Medicine* 2006;25(24):4334-44.
25. Vickers A. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: A simulation study. *BMC Medical Research Methodology* 2001;1:1-4.
26. Hox JJ. *Applied multilevel analysis*. Second ed. Amsterdam: TT-Publikaties, 1995.
27. Wan Q, Harris MF, Jayasinghe UW, Flack J, Georgiou A, Penn DL, et al. Quality of diabetes care and coronary heart disease absolute risk in patients with type 2 diabetes mellitus in Australian general practice. *Qual Saf Health Care* 2006;15(2):131-35.
28. Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J* 1991;121(1 Pt 2):293-98.
29. Wilson S, Delaney BC, Roalfe A, Roberts L, Redman V, Wearn AM, et al. Randomised controlled trials in primary care: case study. *BMJ* 2000;321:24-27.
30. Wiles NJ, Haase AM, Gallacher J, Lawlor DA, Lewis G. Physical activity and common mental disorder: results from the Caerphilly study. *Am J Epidemiol* 2007;165(8):946-54.
31. Fuller J, Perkins D, Parker S, Holdsworth L, Kelly B, Roberts R, et al. Effectiveness of service linkages in primary mental health care: a narrative review part 1. . *BMC Health Services Research* 2011;11:71.
32. Richards DA, Lovell K, Gilbody S, Gask L, Torgerson D, Barkham M, et al. Collaborative care for depression in UK primary care: a randomized controlled trial. *Psychol Med* 2008;38(2):279-87.

Figure 1. CONSORT flow diagram of the recruitment process.



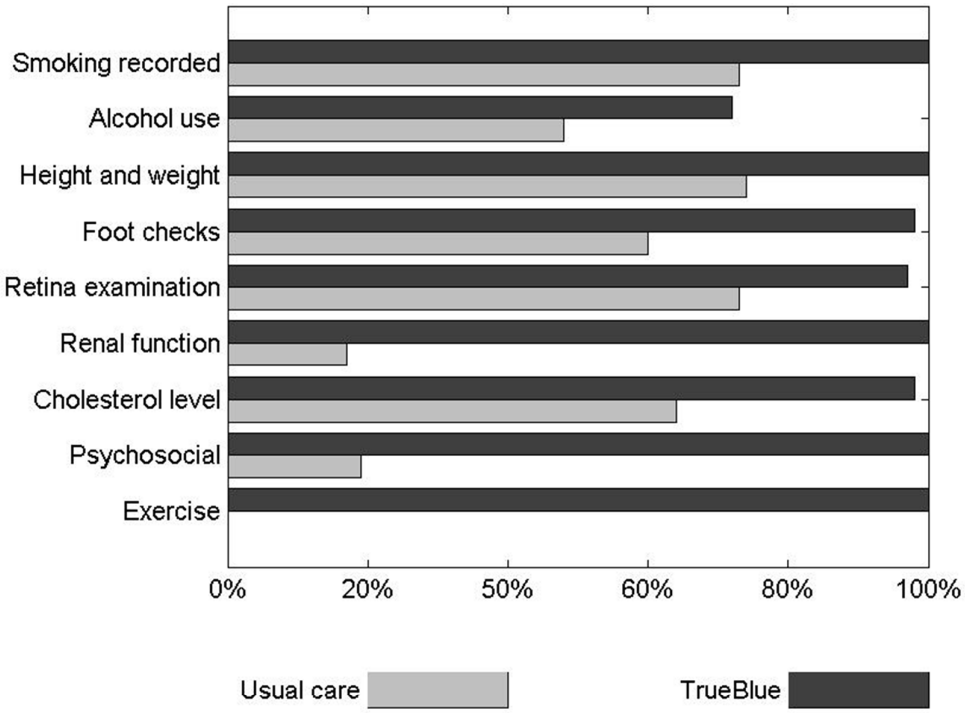


Figure 2: Recording of checks recommended by the National Heart Foundation and Diabetes Australia guidelines. Data for "usual care" were adapted from reference 27. No usual-care data were available for exercise.

Table 1: Patient characteristics at the baseline visits. There were no significant differences between the intervention and control at baseline.

Characteristic	Intervention group (*n*=170)	Control group (*n*=147)
Male (%) / Female (%)	51.8% / 48.2%	55.2% / 44.8%
Age (yr)	68.0 ± 11.7	67.6 ± 11.2
Aboriginal or Torres Strait Islander (%)	0.0%	0.7%
Diagnosis: Type-2 diabetes	37.6%	47.6%
CHD	45.3%	35.8%
Both	17.1%	16.6%
Body mass index (kg/m^2)	31.4 ± 6.0 (*n*=170)	30.8 ± 6.0 (*n* =103)
Systolic blood pressure (mmHg)	134.1 ± 19.0 (*n*=169)	133.5 ± 19.6 (*n* =112)
Total cholesterol (mmol/L)	4.21 ± 0.94 (*n*=165)	4.41 ± 1.06 (*n* =110)
Triglycerides (mmol/L)	1.73 ± 0.88 (*n*=165)	1.92 ± 1.37 (*n* =105)
LDL (mmol/L)	2.22 ± 0.74 (*n*=159)	2.37 ± 0.88 (*n* =89)
HDL (mmol/L)	1.23 ± 0.36 (*n*=159)	1.18 ± 0.33 (*n* =97)
HbA1c (mmol/L)	7.00 ± 1.21 (*n*=94)	7.19 ± 1.42 (*n* =69)
PHQ-9 score	10.7 ± 4.7 (*n*=164)	11.6 ± 5.5 (*n*=146)
PHQ-9 score range at baseline	5 to 24	5 to 27

Table 2: TrueBlue outcomes at six months in the intervention and control groups. (See Table 3 for ~~a~~ the list of abbreviations.)

	Intervention				Control				
	n	Baseline	6-months	Within group†	*n*	Baseline	6-months	Within group‡	Between groups
PHQ9 depression score	164	10.7±0.8	7.1±0.8	$p<0.001$‡	146	11.6±0.9	9.0±0.9	$p<0.001$‡	$p=0.012$
SF36v2 mental-health score§	71	37.2±3.4	41.1±3.4	$p=0.034$‡		Not recorded			NS
SF36v2 physical-health score§	71	39.9±2.2	42.5±2.6	$p=0.023$‡		Not recorded			NS
Body mass index (kg/m^2)	162	31.3±1.0	31.2±1.0	NS	103	30.8±1.2	31.0±1.0	NS	NS
Waist (cm)	161	104.7±2.4	105.0±2.4	NS	80	104.2±4.0	105.8±3.2	NS	NS
Systolic blood pressure (mmHg)	161	134.2±3.0	132.4±2.8	NS	112	133.5±3.8	131.2±3.4	NS	NS
Total cholesterol (mmol/L)	158	4.21±0.16	4.22±0.14	NS	109	4.41±0.20	4.44±0.20	NS	NS
LDL (mmol/L)	154	2.23±0.12	2.17±0.14	NS	86	2.37±0.18	2.29±0.20	NS	NS
HDL (mmol/L)	154	1.23±0.06	1.29±0.06	$p=0.023$‡	93	1.17±0.06	1.27±0.08	$p=0.011$‡	NS
Triglycerides (mmol/L)	158	1.72±0.14	1.66±0.12	NS	104	1.84±0.22	1.75±0.18	NS	NS
HbA1c (%)¶	89	6.97±0.24	6.90±0.26	NS	67	7.22±0.34	7.40±0.36	NS	$p=0.049$
Ten-year CVD risk*	61	26.9±3.2	26.1±3.2	NS	46	26.3±3.6	24.7±3.2	NS	NS
The 95% confidence ranges are indicated by the plus-minus (±) sign. Note that lower scores indicate improvement for all items except the SF36v2 and HDL results, where higher scores indicate improvement.									
Smoking	162	15 (9%)	13 (8%)	NS	110	13 (12%)	13 (12%)	NS	NS
Alcohol	104	47 (45%)	51 (49%)	NS	42	27 (64%)	27 (64%)	NS	NS
Exercises 30 mins/day, 5 d/wk	162	66 (41%)	97 (60%)	$p<0.001$‡	75	22 (29%)	22 (29%)	NS	$p<0.001$
Referred to exercise program	162	32 (20%)	58 (36%)	$p<0.001$‡	111	15 (14%)	10 (9%)	NS	$p<0.001$
Attends exercise program	162	12 (7%)	23 (14%)	$p=0.041$‡	79	12 (15%)	9 (11%)	NS	NS
On antidepressant medication	162	27 (17%)	34 (21%)	NS	113	31 (27%)	36 (32%)	NS	$p=0.025$
Referred to mental health worker	162	47 (29%)	58 (36%)	$p=0.022$‡	114	10 (9%)	24 (21%)	$p<0.001$‡	$p<0.001$
Attends mental health worker	162	10 (6%)	37 (23%)	$p<0.001$‡	109	14 (13%)	11 (10%)	NS	$p=0.044$

The values in brackets are the percentages of the total *n*.

(NS) No significant difference. (†) Significant difference between baseline and six-month values within the *intervention* clinics. (‡) Significant difference between baseline and six-months within the *control* clinics. (§) SF36v2 questionnaires were not collected by all clinics. (¶) HbA1c results were only available for patients with T2DM. (*) CVD risk could only be calculated for patients with T2DM.

Table 3: TrueBlue outcomes at 12 months within the intervention clinics only.

	Intervention			
	n	Baseline	12-months	Within group†
PHQ9 depression score	164	10.7±0.7	6.6±0.7	$p<0.001$
SF36v2 mental-health score§	70	36.0±3.2	41.3±2.8	$p<0.001$
SF36v2 physical-health score§	70	40.6±2.2	44.3±2.8	$p<0.001$
Body mass index (kg/m^2)	142	31.4±1.0	31.1±1.0	$p=0.006$
Waist (cm)	141	105.0±2.4	105.2±2.6	NS
Systolic blood pressure (mmHg)	141	135.2±3.2	130.2±3.0	$p=0.016$
Total cholesterol (mmol/L)	138	4.18±0.16	4.28±0.16	NS
LDL (mmol/L)	135	2.19±0.12	2.24±0.20	NS
HDL (mmol/L)	135	1.22±0.06	1.36±0.08	$p<0.001$
Triglycerides (mmol/L)	138	1.73±0.16	1.63±0.14	$p=0.004$
HbA1c (%)¶	79	7.01±0.26	7.04±0.28	NS
Ten-year CVD risk*	55	27.4±3.4	24.9±3.6	$p=0.015$

The 95% confidence ranges are indicated by the plus-minus (±) sign. Lower scores indicate improvement for all items except the SF36v2 and HDL results, where higher scores indicate improvement.

Smoking	142	15 (11%)	11 (8%)	NS
Alcohol	95	45 (47%)	47 (49%)	NS
Exercises 30 mins/day, 5 days/wk	142	57 (40%)	83 (58%)	$p<0.001$
Referred to exercise program	142	26 (18%)	53 (37%)	$p<0.001$
Attends exercise program	142	10 (7%)	17 (12%)	NS
On antidepressant medication	142	22 (15%)	33 (23%)	$p=0.001$
Referred to mental health worker	142	40 (28%)	59 (42%)	$p<0.001$
Attends mental health worker	142	8 (6%)	25 (18%)	$p<0.001$

The values in brackets are the percentages of the total *n*.

(NS) No significant difference. (†) Significant difference between baseline and twelve-month values. (§) SF36v2 questionnaires were not collected by all clinics. (¶) HbA1c results were only available for patients with T2DM. (*) CVD risk could only be calculated for patients with T2DM.

NOTE:
PHQ9 is the 9-question Patient Health Questionnaire.
SF36v2 is version 2 of the Short Form 36-Question health survey.
LDL is low-density lipoprotein.
HDL is high-density lipoprotein.
Glycated haemoglobin (HbA1c) was measured only for patients with diabetes.
Unit of alcohol is 10g of ethanol.

Checklist of items to include when reporting a cluster randomised trial

* = addition to CONSORT *Modifications to checklist in italics*			
PAPER SECTION and topic	Item	Descriptor	Reported on Page No.
TITLE & ABSTRACT	1*	How participants were allocated to interventions (e.g., "random allocation", "randomised", or "randomly assigned"), *specifying that allocation was based on clusters*	1
INTRODUCTION Background	2*	Scientific background and explanation of rationale, *including the rationale for using a cluster design.*	~~2, 3~~3, 4
METHODS Participants	3*	Eligibility criteria for participants *and clusters* and the settings and locations where the data were collected.	~~3, 4~~5,6
Interventions	4*	Precise details of the interventions intended for each group, *whether they pertain to the individual level, the cluster level or both,* and how and when they were actually administered.	3, 4, Ref ~~17~~18
Objectives	5*	Specific objectives and hypotheses, *and whether they pertain to the individual level, the cluster level or both.*	1, ~~3~~4
Outcomes	6*	Report clearly defined primary and secondary outcome measures, *whether they pertain to the individual level, the cluster level or both*, and, when applicable, any methods used to enhance the quality of measurements (e.g., multiple observations, training of assessors).	~~4, 5~~6, 7
Sample size	7*	How *total* sample size was determined *(including method of calculation, number of clusters, cluster size, a coefficient of intracluster correlation (ICC or k), and an indication of its uncertainty*) and, when applicable, explanation of any interim analyses and stopping rules.	~~Ref 17~~4,5
Randomisation. Sequence generation	8*	Method used to generate the random allocation sequence, including details of any restriction (e.g., blocking, stratification, *matching*).	
Allocation concealment	9*	Method used to implement the random allocation sequence, *specifying that allocation was based on clusters rather than individuals and* clarifying whether the sequence was concealed until interventions were assigned.	~~Ref 17~~5
Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	
Blinding (Masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. If done, how the success of blinding was evaluated.	N/A
Statistical methods	12*	Statistical methods used to compare groups for primary outcome(s) *indicating how clustering was taken into account*; methods for additional analyses, such as subgroup analyses and adjusted analyses.	~~5~~7,8
RESULTS Participant flow	13*	Flow of *clusters and* individual participants through each stage (a diagram is strongly recommended). Specifically, for each group report the numbers of *clusters and* participants randomly assigned, receiving intended treatment, completing the study protocol, and analyzed for the primary outcome. Describe protocol deviations from study as planned, together with reasons.	8, 9, ~~15 (Fig 1)~~Fig. 1
Recruitment	14	Dates defining the periods of recruitment and follow-up.	4
Baseline data	15*	Baseline information for each group *for the individual and cluster levels as applicable*	~~17 (~~Table 1~~)~~
Numbers analyzed	16*	Number of *clusters and* participants (denominator) in each group included in each analysis and whether the analysis was by "intention-to-treat". State the results in absolute numbers when feasible (e.g., 10/20, not 50%).	~~17, 18, 19 (Tables)~~Tables 1–3
Outcomes and Estimation	17*	For each primary and secondary outcome, a summary of results for each group measures *for the individual or cluster level as applicable*, and the estimated effect size and its precision (e.g., 95% confidence interval) *and a coefficient of intracluster correlation (ICC or k) for each primary outcome.*	Tables 1–3~~17, 18, 19 (Tables)~~
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those pre-	N/A

		specified and those exploratory.	
Adverse events	19	All important adverse events or side effects in each intervention group.	None
DISCUSSION Interpretation	20	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision and the dangers associated with multiplicity of analyses and outcomes.	~~7, 8~~11, 12
Generalisability	21*	Generalisability (external validity) *to individuals and/or clusters (as relevant)* of the trial findings.	~~9, 10~~14
Overall evidence	22	General interpretation of the results in the context of current evidence.	~~9~~13, 14