Research Report DFE-RR211

Randomised controlled trial of the 'Teens and Toddlers' programme

Ruth Maisey, Svetlana Speight, Peter Keogh and Ivonne Wollny
NatCen Social Research

Chris Bonell, Annik Sorhaindo and Kaye Wellings
London School of Hygiene and Tropical Medicine

Susan Purdon
Bryson Purdon Social Research

Department for Education

Content

Tables

Figures

Acknowledgements

We would like to thank all young women who took part in the study for their time, and all teachers and Teens and Toddlers facilitators who helped with the administration of the study and/or participated in the survey of teachers.

At the Department for Education, we are grateful to Sarah Butt, Richard White, Alison Hadley and Rajkiran Egan for their support of the study at its different stages and valuable feedback on the report.

Within the Teens and Toddlers programme, we would like to thank Diana Whitmore, Andrew McDowell, Kirsty Humphrey, Suzanne Wallis and Miranda Taggart for their help with the implementation of the study and helpful comments on the design of the study and the report.

We are also grateful to steering group members Meg Wiggins and Sally Hopewell for their many helpful comments at different stages of the study and on the report. We would also like to thank those who carried out peer reviews of the study's research protocol – Douglas Altman, Richard Hayes, Doug Kirby and Carole Torgerson – as well as those who carried out peer reviews of this report – Lyndal Bond and, again, Carole Torgerson.

Within the research team, we would like to thank Sara Scott for her very useful comments on the report, and Cathy O'Donnell and Alice Reeves for their help with the report production. We are also grateful to Tricia Jessiman who worked on the qualitative research part of the evaluation but provided us with useful advice on its RCT element. In addition, we would like to thank Sue Corbett in Project Computing and Emma Fenn and her colleagues in the Operations department, as well as all the interviewers who worked on the study.

Finally, we would like to thank the University of Oxford, where one of the lead authors of the report – Chris Bonell – moved in October 2011, for facilitating Chris's continued involvement with the study.

Executive summary

Introduction

This is a report of the main findings of a randomised controlled trial (RCT) conducted to assess the impact of the 'Teens and Toddlers' (T&T) youth development and teenage pregnancy prevention programme. This trial forms part of a wider evaluation that included a stage of formative qualitative work and a process evaluation (Jessiman et al. 2012). It was funded by the Department for Education and carried out by NatCen Social Research, the London School of Hygiene and Tropical Medicine, and Bryson Purdon Social Research.

The T&T programme aims to decrease teenage pregnancy by raising the aspirations and educational attainment of 13-17 year old teenagers at most risk of leaving education early, social exclusion and becoming pregnant. It seeks to achieve these aims through:

- Providing practical experience of working with children and developing an appreciation of the enormous privilege, responsibility and hard work involved in having a child.
- Raising awareness of the personal and financial responsibilities implied by an unexpected pregnancy through developing an understanding of the impact of a child on family life and personal lifestyle.
- Encouraging the development of alternative goals to being pregnant, such as satisfying work and relationships.
- Educating teens about the realities of pregnancy, the importance of sexual responsibility and the effects of harmful behaviours (such as drugs and violence) and health related factors - from foetus to neonate to child.
- Providing one-to-one counselling and mentoring/coaching to young people who are emotionally challenged and need support, to enhance the outcomes of T&T.
- Providing the knowledge and skills required for successful relationships and sexual responsibility, to prevent conception and protect sexual health.

T&T work with local authorities to implement the programme through secondary schools. Young women complete an 18 to 20 week programme during which they attend weekly three-hour sessions in a nursery setting. Each participant supports a child (for about an hour and a half per session), takes part in classroom-based group work, keeps a journal of their experience and learning, and has access to a trained counsellor. Ultimately, participation in the programme enables the young people to achieve a National Award in Interpersonal Skills, Level 1 (NCFE).

The RCT design

The T&T programme has previously been internally evaluated through analyses of retrospective data from annual surveys of young people aged under 20 who completed the programme at least six months ago (Humphrey 2011, Humphrey 2010, Humphrey 2009). However, the strength of these analyses is limited by the lack of a control group. Without a control group, it is impossible to assess the degree to which the outcomes observed are due to the intervention and not to other factors. A randomised controlled trial (RCT) overcomes this limitation since it allows the effects of the intervention to be disentangled from the effects of other factors and background noise (e.g. maturational, seasonal or other time-related factors that might affect outcomes). It is true that there are also disadvantages associated with RCTs, as there are with all methods of evaluating social interventions. For instance, participants might choose not to disclose sensitive information, which can reduce the study's power to detect impacts (a comprehensive discussion of the possible limitations to this study is included in section 5.2). Despite these limitations, however, RCTs provide the least biased estimates of intervention effects. This is because randomisation enables intervention and control groups to be comparable on both measured and unmeasured factors, which might otherwise confound estimates of intervention effects. It is for this reason that an RCT was employed on this study.

The design of the RCT included the following elements:

- At-risk young women were identified by their teachers using guidance provided by T&T.
- Individual girls were randomly allocated to an intervention or control group.
- Data for all participants in the study were collected by questionnaire at three points in time: prior to allocation (baseline), immediately post intervention (follow-up one) and a year after the intervention (follow-up two).
- Two cohorts of girls participated in the trial, one starting in September 2009 and one in January/February 2010. In total, 449 teenagers entered the trial (228 in the control arm and 221 in the intervention arm).
- The overall response rate of these baseline participants at follow-up one was 95% and at follow-up two 91%. The response rate was the same in the intervention and control groups.

Outcomes

The aim of the RCT was to evaluate the impacts of the T&T programme on participants in terms of the following specific outcomes at follow-up two (i.e. one year post-intervention):

Primary outcomes

- Did not use any contraception the last time they had sex (and had sex within the last three months);
- Has had more than one episode of not using contraception in the last three months;
- Expects teenage parenthood;

- Low youth development score.

Secondary outcomes

- Did not use a condom the last time they had sex (and had sex within the last three months);
- Has had more than one episode of not using a condom in the last three months;
- Believes that the best age to have sex for the first time is under 16 years of age;
- Is favourable to sometimes not using protection for sex;
- Low self-reflection;
- Low emotional vocabulary;
- Low self-esteem;
- Dislikes school;
- Lack of expectation regarding post-16 education, training or employment;
- Low sexual health knowledge;
- Difficulty in discussing sex with a boyfriend;
- Difficulty in discussing the pill in a clinic or with a doctor;
- Has become pregnant since baseline;
- Lacks awareness of the impact of parenthood on social life;
- Number of school days missed.

Characteristics of teenagers participating in the trial

Prior to randomisation, baseline data were collected on the teenagers taking part in the RCT. Their average age was 13.5 years. Many were from disadvantaged backgrounds, with 44% receiving free school meals and 32% living in workless households. Thirteen per cent had experienced heterosexual intercourse, 2% had been pregnant, and 22% expected to become parents while still in their teens. About a third disliked school, the same proportion missed school without permission, and about a fifth said they got drunk at least once a month. Overall, the prevalence of risk-taking behaviours and attitudes among participants at baseline suggests that teachers targeted an appropriate group of young women who were at risk of teenage pregnancy.

Analysis of the characteristics of the teenagers randomised to either the intervention or the control group showed that the groups were well balanced. The very few differences between them that were present by chance have been taken into account when analysing the effects of the intervention.

There were marked baseline differences between teenagers who dropped out of the programme and those who completed it. These differences indicated that those who dropped out were more likely to be from disadvantaged backgrounds and to engage in risk-taking behaviours (see Chapter 2). Attendance at the T&T programme in the trial was relatively high, with 73% completing the programme, although it was somewhat lower than the completion rates reported on T&T programmes most recently.

The overall response rates to the follow-up surveys were very high, with 95% of baseline respondents completing a questionnaire at follow-up one and 91% at follow-up two. The response rates did not vary between the intervention and control groups or by whether the teenagers completed the T&T programme or dropped out. There was some evidence of a non-response bias (i.e. systematic differences between those who did and did not respond) at follow-up one but not at follow-up two. There was no evidence to suggest that this negatively affected the comparability between the intervention and control groups.

Perceptions of the programme

The vast majority of teenagers enjoyed taking part in T&T (94% of those randomised to the intervention group). They particularly enjoyed the toddler time but were less keen on writing their journals. One-fifth of teenagers said that they found T&T difficult or challenging, with similar numbers specifically reporting that toddler time, group work, writing a journal or receiving counselling was often difficult or challenging for them.

The teenagers participating in the programme expressed overwhelmingly positive feelings about it. Sixty-five per cent of girls in the intervention group reported that participating in T&T made them feel responsible, 64% that it made them happy, 61% were more positive about things and 59% were positively interested in the programme. Substantially fewer teenagers reported negative emotions arising from their participation in T&T (e.g. 37% said that they had sometimes felt bored, 20% had been irritated or annoyed, 7% anxious, and 7% angry). These positive findings were reinforced by others showing that 91% felt proud of doing the T&T programme, while only 7% said that their participation made them feel embarrassed.

The majority of teenagers also felt they knew more about the different areas of the T&T curriculum as a result of the programme. In particular, 66% of teenagers in the intervention group felt they knew a lot more about child development, 62% said they knew a lot more about giving and receiving negative feedback, and 60% said that they knew more about communication, listening and conversation skills.

Teenagers also felt that T&T had changed their attitudes or behaviours: 78% of girls in the intervention group felt that it had helped them to communicate, 77% subsequently felt better about themselves, 76% had made new friendships, 75% said it had made them want to delay parenthood, and 53% said it had changed what they wanted to do when they finished school.

The programme requires participants to miss one afternoon of school every week for 18 to 20 weeks. The lessons teenagers most commonly missed were science (27%) and physical education (23%). Some reported that they had fallen behind with their school work as a result of attending the programme (31%).

There was limited evidence of stigma associated with T&T. Seventy-two per cent of teenagers in the intervention group reported that other people ‘said nice things’ about the fact that they were doing T&T, 87% felt that their parents were positive about their

involvement with the programme, and 47% said that other people were jealous that they were doing T&T. However, 9% of teenagers in the intervention group reported that they were teased as a result of T&T.

T&T participants are encouraged to share their experiences and learning with their friends. The majority of teenagers in the intervention group talked to their peers who did not do T&T about what they did or learnt on the programme, with only 18% saying that they never talked about T&T. When teenagers did talk about T&T, the topics they most commonly talked about were sex and relationships (45%), parenting and being a parent (40%) and child development (38%).

Impact analysis

The trial examined whether the Teens and Toddlers intervention was effective in improving the outcomes listed earlier in this Executive summary for those allocated to the programme. The outcome data were collected at two points in time – at the end of the programme (follow-up one) and one year later (follow-up two) – to enable analysis of short- and medium-term impacts.

The impact analysis was conducted on an intention-to-treat basis. This means that all teenagers who were originally randomised to the intervention or control groups as part of the trial were included in the analysis regardless of how many sessions of the T&T programme they attended in total. This approach is recommended in the Consolidated Standards of Reporting Trials (CONSORT) because individuals who drop out of interventions are generally (and in the T&T RCT in particular) different in profile, according to baseline data, from those who remain (CONSORT Statement 2010). This means that those individuals who remain with the programme are very different from those in the control group (since there is no similar process of drop-out in this group). Thus a comparison of outcome data from all those who were allocated to a programme versus those who were allocated to the control group (i.e. an 'intention-to-treat' analysis) is fairer than a comparison of outcome data from those who remain with the programme versus those in the control group (an 'in treatment' analysis).

The analysis showed that at follow-up one (immediately after the programme finished), there was no evidence that the intervention had been effective in changing any of the three primary outcomes. However, there was evidence of a positive impact of the programme on three of the 14 secondary outcomes. Teenagers in the intervention group were less likely to have low self-esteem (16%, compared with 28% in the control group); their knowledge of sexual health was less likely to be poor (73%, compared with 83%); and they were less likely to report difficulty in discussing the pill with a doctor or in a clinic (44%, compared with 56%).[1] These positive findings are consistent both with participants' perceptions of the effects of the intervention and with the perceptions of teachers and T&T

[1] All percentages relating to the impact analysis are based on regression models which controlled for relevant baseline differences between the intervention and control groups.

facilitators as discussed in Chapter 3 of this report and in the report on the integral process evaluation (Jessiman et al. 2012).

At follow-up two (one year later), there was no evidence that the programme had been effective in changing any of the four primary outcomes. However, the intervention was found to have had an effect in preventing low self-esteem. While low self-esteem was reported by 15% of teenagers randomised to the intervention group, the respective figure for the control group was 25%. Thus, the positive impact of the intervention on low self-esteem observed at follow-up one was sustained one year later. However, the positive impacts observed at follow-up one in regard to knowledge of sexual health and ability to discuss contraception with a doctor were no longer evident a year after the programme ended. No new impacts of the intervention were detected at follow-up two.

Evidence from other studies suggests that self-esteem may be protective against early sexual activity among girls and teenage pregnancy (Emler 2002; Spencer et al. 2002). Therefore improving teenagers' self-esteem may make an important contribution towards achieving the T&T intervention's objective to reduce teenage pregnancy.

Methodological limitations

There are a number of methodological limitations that might have affected the study's ability to detect possible impacts of the T&T programme (discussed in detail in Chapter 4).

One potential limitation relates to the intention-to-treat analytical approach. The concern here is that including in the analysis those who dropped out of the programme as part of the intervention group could dilute the impacts of the intervention. However, there is little evidence in the impact findings that dilution of positive impacts occurred. There is no overall pattern of statistically non-significant benefits to suggest that the intervention brought about a range of benefits which just failed to reach the level of statistical significance because of a dilution effect. Given that the drop-out rate was only 27%, if the intervention had had real effects on other outcomes, we would have expected to see this non-significant trend towards a range of benefits, which is not found in the data. What we find instead is that the non-significant associations are scattered either side of the threshold dividing positive from negative effects – with about half the differences suggesting potentially positive and the other half negative impacts.

Another possible methodological limitation of the study relates to the potential for contamination of the findings. If teenagers participating in the programme discussed the new knowledge they gained from participation with those who were in the control group and thus benefited the teenagers in the control group, this would have led to a certain amount of 'contamination' and therefore underestimation of any real intervention effects.

Although it is possible that some contamination occurred within this study, we do not believe that such effects are likely to have unduly affected the results. As with the above, if we had seen results that consistently tended in the direction of intervention benefit but

did not reach significance, then we might conclude from this that contamination could have resulted in our analysis having insufficient power to detect real intervention effects. However, this was not what we found. Furthermore, while we might anticipate a possible contamination effect on outcomes such as knowledge of sexual health, in which some benefits might plausibly be passed from intervention to control participants, we would not expect contamination to affect outcomes such as youth development or sexual behaviour.

Reluctance to disclose socially undesirable behaviours and attitudes is another methodological issue which could have affected the study's findings. If teenagers *as a whole* had under-reported certain behaviours, this would have made it more difficult to identify intervention effects but would not otherwise have biased the analysis. However, if teenagers *randomised to the intervention* were less or more likely to report risk-taking behaviours and attitudes than teenagers in the control group, this would have biased the findings in the direction of, respectively, over- or under-estimating any real benefits of the intervention.

Participants in the T&T programme are actively encouraged to be honest about their behaviours and attitudes. If teenagers in both the T&T and the control groups were under-reporting certain behaviours at the baseline but then those in the T&T group provided an honest picture post-intervention, this may have limited the study's potential to identify the intervention's positive effects. Another possibility is related to social desirability bias, which generates an over-reporting of benefits among intervention participants because they want to report what they perceive that the intervention providers and evaluators want to hear (e.g. Stuart and Grimes 2009).

If the teenagers who were recruited to the study were not sufficiently at risk (for example, due to the necessity to recruit a certain number or due to teenagers' reluctance to take part in the research), this too might have had a negative impact on the study's ability to detect the intervention's benefits. The process evaluation found that the T&T facilitators thought that the 'right' girls were being referred to the programme (Jessiman et al. 2012). In addition, comparisons of the prevalence of various sexual behaviours between the T&T RCT and the evaluation of the Young People's Development Programme (YPDP) (Wiggins et al. 2009) suggest that that the T&T RCT was successful in recruiting teenagers who were sufficiently at risk of teenage pregnancy.

Another possible limitation of the study is that it was designed to examine the short- and medium-term impacts of the programme and was not able to look at its long-term impacts, such as pregnancy incidents before age 20 or rates of being 'not in education, employment or training' (NEET).

The measures of youth development and self-esteem used in this study are not ones that have been previously validated (even though the self-esteem measure had been previously used in the Ripple study (Stephenson et al. 2004)). This means that these measures may be less sensitive to change than comprehensive validated measures would be.

The study's findings may have been affected by some of the limitations outlined above and it is possible that the ability to detect real intervention effects was thereby somewhat reduced. However, given that the overall pattern of differences between the intervention and control groups does not show a tendency towards positive (albeit statistically non-significant) impacts across a majority of outcomes, but is instead rather mixed, it is unlikely that there were a range of real intervention effects that the study failed to detect.

Implications of the findings of the study

While acknowledging the methodological limitations of the study, we do not believe that they explain the intervention's limited and mostly short-term impact on young women at risk of teenage pregnancy.

An alternative explanation is that T&T may not have provided sufficient sexual health education. There was evidence from the process evaluation that some facilitators felt that their main role was 'not teaching but facilitating', that is providing girls with an opportunity to discuss sexual health issues openly, rather than giving them accurate information. There was also variation in the facilitators' own knowledge of sexual health (Jessiman et al. 2012). While the RCT found evidence of the intervention's positive impact on sexual health knowledge at follow-up one, this effect was no longer evident a year after the programme ended.

Missing out on normal schooling to attend the programme may also have had a negative effect on T&T participants, with 31% reporting falling behind with schoolwork as a result of participation (see Chapter 3).

This intervention appears to have brought about benefits in terms of self-esteem – which evidence from other studies suggests may be protective against early sexual activity among girls and teenage pregnancy – but not other outcomes as measured in the RCT. Therefore, while we can conclude that it might have some potential for facilitating girls' personal development and possibly for reducing the risk of teenage pregnancy, we cannot conclude that its evidence base is at present strong. We would recommend further development and further evaluation of the intervention to address the current limitations suggested by our evaluation. The criteria for targeting the intervention need to be kept up to date with the most recent evidence on the risk factors for teenage pregnancy in the UK. Teachers need to be trained to use the criteria correctly. Participants and their parents should be fully informed about the aims of the intervention and why they have been targeted. The sex education aspects of the intervention should be significantly strengthened and facilitators should receive training so that they have consistent expertise in this area. It should be ensured that girls do not fall behind in their school work as a result of their participation in the programme, for example, through scheduling T&T sessions in such a way that the girls do not miss lessons in any key subjects.

While there was good fidelity in the intervention, it may be helpful to more closely follow the logic model framework developed for this evaluation (see Appendix B) and to develop it further (see Kirby 2004) on the basis of the above comments and their own expertise.

More generally, it is important for policy makers to appreciate that targeted interventions are unlikely on their own to have a significant impact on overall rates of teenage pregnancy in the population. This reflects the difficulty in predicting and then identifying in practice which girls are most at risk of teenage pregnancy; changes in what are the strongest risk factors for teenage pregnancy through time; and the fact that most teenage pregnancies actually arise among girls at low-to-moderate risk of teenage pregnancy simply because there are more girls in this group (Kneale et al. under review). Therefore, strategies to reduce teenage pregnancy should also emphasise the importance of universal interventions such as improving school-based sex and relationships education (DiCenso et al. 2002), expanding access to family planning and programmes to increase all girls' educational expectations (Harden et al. 2009).

1 Introduction

1.1 Overview

This report provides the main findings of a randomised controlled trial (RCT) that was conducted to assess the outcomes of the 'Teens and Toddlers' (T&T) youth development and teenage pregnancy prevention programme. This trial forms part of a wider evaluation that included a stage of formative qualitative work and an embedded process evaluation (Jessiman et al. 2012). It was funded by the Department for Education and carried out by NatCen Social Research, the London School of Hygiene and Tropical Medicine, and Bryson Purdon Social Research.

The T&T programme has previously been internally evaluated through analyses of retrospective data from annual surveys of young people aged under 20 who completed the programme at least six months ago (Humphrey 2011, Humphrey 2010, Humphrey 2009). However, the strength of these analyses is limited by the lack of a control group. Without a control group, it is impossible to assess the degree to which the outcomes observed are due to the intervention and not to other factors. An RCT overcomes this limitation since it allows the effects of the intervention to be disentangled from the effects of other factors and background noise (e.g. maturational, seasonal or other time-related factors that might affect outcomes). It is true that there are also disadvantages associated with RCTs, as there are with all methods of evaluating social interventions. For instance, participants might choose not to disclose sensitive information which can reduce the study's power to detect impacts (a comprehensive discussion of the possible limitations to this study is included in section 5.2). Despite these limitations, RCTs provide the least-biased estimates of intervention effects. This is because randomisation enables intervention and control groups to be comparable on both measured and unmeasured factors, which might otherwise confound estimates of intervention effects (Campbell and Russo 1999; Oakley 1990). It is for this reason that an RCT was employed on this study.

The aim of the RCT was to evaluate the outcomes of the T&T programme for participants in terms of the following measures.

Primary outcomes

- Did not use any contraception the last time they had sex (and had sex within the last three months);
- Has had more than one episode of not using contraception in the last three months;
- Expects teenage parenthood;
- Low youth development score.

Secondary outcomes

- Did not use a condom the last time they had sex (and had sex within the last three months);
- Has had more than one episode of not using a condom in the last three months;
- Believes that the best age to have sex for the first time is under 16 years of age;
- Is favourable to sometimes not using protection for sex;
- Low self-reflection;
- Low emotional vocabulary;
- Low self-esteem;
- Dislikes school;
- Lack of expectation regarding post-16 education, training or employment;
- Low sexual health knowledge;
- Difficulty in discussing sex with a boyfriend;
- Difficulty in discussing the pill with a doctor;
- Has become pregnant since baseline;
- Lack of awareness of the impact of parenthood on social life;
- Number of school days missed.

1.2 Policy and research background

Reducing the incidence of teenage pregnancies and parenthood continues to be a government priority in the UK, USA and other developed nations (Holgate and Evans 2006). In 2010, In England and Wales the conception rate for women aged under 20 was 54.6 conceptions per thousand women, which continued the overall downward trend over the last decade from 62.5 conceptions per thousand women in 2000 (ONS 2012).

In England, teenage pregnancy is most prevalent in the poorest communities and among the most vulnerable young people (Bailey 2005). In the 1992 waves of the British Household Panel Survey (BHPS) and the 1970 British Cohort Study (BCS70), low family social class and increased local area unemployment were associated with a higher risk of becoming a teenage mother (Ermisch and Pevalin 2003). The greater the level of childhood poverty, the more likely a young women is to become a parent in her teenage years. Conception rates are higher in deprived areas and the proportion of conceptions ending in abortion is higher in less deprived areas (Lee et. al. 2004; Teenage Pregnancy Unit, 2006). Births during teenage years are more prevalent among particular ethnic groups. Young people under the age of 20 from Bangladeshi backgrounds have an average fertility rate of 75 per 1,000 women in the relevant age range. The rate is lower for women from Pakistani or Caribbean origin (41 and 44 births per 1,000) and teenage birth rates in Britain are lowest for Indians and whites (17 and 29 births per 1,000) (Berthoud 2001).

Teenage parenthood is associated with adverse social, economic, and health outcomes (Ermisch 2003) which remain after adjusting for prior circumstances (Hobcraft and Kiernan 2001). The children of younger mothers in the BCS70 were more likely to be born

with low birthweight or preterm (Pevalin 2003), and at age five scored significantly lower on a standardised vocabulary test and behaviour scale. The children of young mothers are also more likely to be socially and economically disadvantaged than their peers. Pevalin (2003) measured outcomes at age five and ten years old and found that the children of mothers who gave birth under the age of 20 were more likely to have been separated from their mother for more than a month before they were five; more likely to not have their father living with them; and where their father was living with them, he was more likely to hold a manual occupation. Analyses of the National Child Development Study (NCDS) also found that having children before the age of 23 was associated with increased likelihood of being a lone parent, living in social housing and receiving benefits (Hobcraft and Kiernan 2001). These consequences of teenage parenthood often persist into later adult life. For example, when in their 30s and 40s, mothers who became parents under the age of 24 were: less likely to be employed and less likely to be living with a partner - and, if they did live with a partner, that partner was more likely to be unemployed or to be employed on a low wage (Ermisch 2003). Ermisch (2003) also found that young mothers experienced an overall standard of living throughout their 30s and 40s that was ten per cent lower than that of women who began having children at age 24 or older (Ermisch 2003). Further, despite the development of programmes and policies designed to keep young mothers in school, they are also less likely to have completed their education (Hofferth et. al. 2001) or to be in education, employment or training.

'Youth development' programmes aim to prevent teenage pregnancy by promoting personal development, self-esteem, positive career and other aspirations, and good relationships with adults among vulnerable young people through activities such as social and academic education, mentoring, arts, sports and volunteering (Kirby 2007). One such initiative, the 'Children's Aid Society-Carrera' programme was reported as reducing teenage pregnancies when implemented in New York City (Philliber et al. 2002) but replications elsewhere in the USA did not report similar benefits (Kirby, Rhodes and Campe 2005; Philliber, Kaye and Herrling 2001), and a non-randomised evaluation of a youth development intervention in the UK did not find any positive impacts on the participants (Wiggins et al. 2009). Recent reviews have called for refinement and further evaluation of youth development approaches (Harden et al. 2006).

One area identified for refinement was exemplified by the T&T programme, namely providing opportunities for young people to engage in voluntary service within youth development (see the next section for more details about the T&T programme). More generally, voluntary service has recently received significant attention from politicians in the UK and USA because of its potential benefits to volunteers, recipients and society (BBC online 2011; Fox News 2011). An experimental evaluation of the 'Teen Outreach Program', a youth development intervention that included voluntary service, reported reduced teenage pregnancies among participating girls; the authors hypothesised that this might reflect improved relationships with adults, as well as increased autonomy, competence and aspirations (Allen et al. 1997). Other evaluations have reported similar findings (Melchior 1998; O'Donnell et al. 1999).

However, in contrast to T&T, none of the other evaluated voluntary service programmes involved volunteering in child care. While 'virtual infant simulators' (dolls that mimic babies) are widely used to encourage young people to appreciate the responsibility involved in parenthood, despite evidence that they are ineffective in preventing teenage pregnancy (Barnett 2006; Chavaudra 2007), few programmes have exposed young people to looking after real children and none have been evaluated. The T&T programme is unique in combining youth development with experience of mentoring toddlers in nurseries. It has been delivered in the USA and UK for several years.

1.3 The Teens and Toddlers programme

T&T is the central project of Children: Our Ultimate Investment (COUI). It is a youth development and teenage pregnancy prevention programme that aims to decrease teenage pregnancy by raising the aspirations and educational attainment of 13-17 year old teenagers at most risk of leaving education early, social exclusion and becoming pregnant. It seeks to achieve these aims through:

- Providing practical experience of working with children and developing an appreciation of the enormous privilege, responsibility and hard work involved in having a child.
- Raising awareness of the personal and financial responsibilities of an unexpected pregnancy through developing an understanding of the impact on family life and personal lifestyle.
- Encouraging the development of alternative goals to being pregnant, such as satisfying work and relationships.
- Educating teens about the realities of pregnancy, the importance of sexual responsibility and the effects of harmful behaviours (e.g. drugs, violence) and health related factors - from foetus to neonate to child.
- Providing one-to-one counselling and mentoring/coaching to young people who are emotionally challenged and need support, to enhance the outcomes of T&T.
- Providing the knowledge and skills required for successful relationships and sexual responsibility, to prevent conception and protect sexual health.

T&T focuses on geographical areas with high rates of teenage pregnancy and works with local authorities to implement programmes through secondary schools. The programme is targeted at young people aged 13-17 considered to be at risk for teenage parenthood. The selection of teenagers for the programme is undertaken by teachers who are provided with guidance on the characteristics and attributes that T&T considers indicative of a risk for teenage parenthood (see Appendix D[2]). The young people selected are briefed by T&T operations staff on the nature of the programme. Those who are willing

[2] Appendix D shows both the original T&T guidance and the revised guidance document. The latter was developed specifically for the trial in order to improve the targeting of the recruitment to the programme. The process evaluation found that, in practice, very few schools used the revised guidance and most used the T&T original document, more for general guidance than as a formal selection tool (Jessiman et al. 2012).

and give informed consent to take part, and whose parents give informed consent to their participation, then enter the programme.

The programme involves working with young people for weekly three hour sessions over 18 to 20 weeks in a nursery setting. These sessions typically take place on one afternoon a week and participants are excused from school during that time in order to allow them to take part in the programme. Over the course of the programme, the young person supports a child, typically under the age of five and judged by nursery staff as potentially benefiting from special attention from a T&T young mentor, for about one and a half hours per session. They also have access to individual time with a trained counsellor and they spend around 90 minutes per session in a classroom working as a group with a facilitator trained in the pedagogy and curriculum of T&T. This classroom time focuses on child development, effective parenting skills, anger management, sexual health and relationships. Throughout the programme, participants engage in an ongoing journaling activity where they are encouraged to explore their thoughts and feelings about events occurring in their lives and to record their experiences working with the children in the nursery. Ultimately, participation in the programme enables the young people to achieve a National Award in Interpersonal Skills, Level 1 (NCFE). More information about the T&T curriculum can be found in Chapter 3 and in Jessiman et al. (2012).

Each T&T session begins with 'check-in' time when the teenagers tell the group how they are feeling that week so that the facilitators can identify any problems or issues the girls have been facing that could affect their behaviour or be addressed during the session. The nursery time and group work take place after check-in and the counselling sessions take place with individual participants during nursery time. These sessions usually last for around 45 minutes, meaning that two participants can receive counselling each week, missing half of their nursery time that week. The method for selecting participants for counselling each week varies between schools, and might be based on a rota or on individual needs (see Jessiman et al. 2012).

As mentioned earlier, the T&T programme had been self-evaluated before this RCT by analysing retrospective data from annual surveys of young people aged under 20 who completed the programme at least six months ago, to assess pregnancy rates and attitudes to teenage pregnancy (The Dream Mill 2004; McDowell 2004; Cater and McDowell 2007; Cater 2008; Humphrey 2011; Humphrey 2010; Humphrey 2009). The response rate to these annual surveys ranged from 47% to 71% and the findings showed low rates of pregnancy under the age of 18 in follow-up populations. However, as explained earlier, the strength of these analyses is limited by the lack of a control group.

1.4 Formative and process findings

The present evaluation of the T&T programme included a stage of formative qualitative research and a process evaluation (Jessiman et al. 2012) designed to support the RCT. This section summarises the findings from this qualitative work.

1.4.1 Formative evaluation

The aim of the formative evaluation was to examine empirically the extent to which the T&T programme was feasible, appropriately standardised and deemed acceptable by participants and other stakeholders, prior to the RCT. In addition it allowed the research team to map the intervention's logic model, which describes programme inputs and processes, and the causal mechanisms that were assumed to bring about programme outcomes (Kirby 2004). This was then used to inform the selection of outcomes and pathway variables for the RCT.

The formative work drew on:

- A review of documents from Children Our Ultimate Investment (COUI), the T&T programme, and the wider academic literature.
- Interviews with girls participating in routine delivery of the programme, i.e. outside this trial.
- Interviews with stakeholders including: officials within the Department for Education (DfE), teachers at the schools delivering the T&T programme, and COUI staff who deliver the programme or facilitate the sessions.
- A focus group of parents of young people who attended a school that delivered the T&T programme (but whose teenagers were not participating in the programme).

This formative research identified three central influences on the T&T logic model:

- 'Conscious conception' whereby potential parents are properly prepared to have a child emotionally, psychologically and physically before they conceive.
- 'Confluent education' whereby individuals learn through the confrontation, persistence and eventual resolution of a conflict. In the context of T&T, the relationship between the young people in the programme and the toddlers at the nursery is intended to create an opportunity for the young people to learn and grow by working through the challenging experience of caring for a young child.
- 'Youth development' that focuses on young people's existing assets and positive potential rather than merely their deficits and preventing risk.

A diagram illustrating the logic model can be found in Appendix B.

Recommendations

The formative evaluation also made a number of recommendations including that:

- The criteria for teachers' selection of young people for the programme should be objective, focused on factors that are evidence-based correlates of teenage pregnancy, and rely on information a teacher would have about every student.
- Both participants and parents should be provided with explicit information with regard to the aims of the programme in recruitment materials.

- There may be merit in considering the possible disadvantages that could arise for T&T participants as a result of them missing some school lessons to attend the programme. For instance, it might be appropriate for T&T to liaise with schools to develop ways to offer assistance to girls in catching up with missed lessons.
- Participants would consider individual random allocation to be acceptable and that girls could be randomly allocated either face-to-face or remotely.
- The potential for stigma should also be examined empirically in the evaluation.

As far as possible these recommendations were taken on board in the design of the RCT.

1.4.2 Integral process evaluation

The process evaluation built on the findings of the formative research to examine programme feasibility, fidelity and acceptability within the RCT. In other words, it investigated whether the programme was delivered *in vivo* as described both in the programme description and in the logic model, and how it was received by participants. It explored the perspectives of the key stakeholders involved in the programme (facilitators, teachers and young people) and the extent to which their understandings of the aims and purpose of the intervention tallied with those described in the logic model and programme description.

The process evaluation took place between February and July 2011 within schools that were also involved with the RCT. The research design incorporated the following main elements:

- A series of face-to-face in-depth interviews with school staff involved in the selection of pupils for participation in T&T in 16 of the 22 schools.
- Case studies of four schools, which involved following each school and associated nursery from the earliest stages of participation selection to the final awards ceremony at the end of the programme.
- 'Snapshot' observations of practice and interviews in a wider sample of 15 participating schools.

The key findings of this process evaluation are pertinent to the interpretation of the findings from the RCT and are therefore summarised below (for the complete findings from the integral process evaluation see Jessiman et al. 2012).

Selection process

The process evaluation found some variance in the process of selecting participants for the T&T programme. Those responsible for selection varied in their roles and consequently in their knowledge about individual young people. This was particularly the case with respect to risk factors such as sexual behaviour and family issues. The T&T selection tool was generally used to inform school staff's consideration of which young people may be suitable for the programme; ultimately staff members tended to apply their own subjective criteria when selecting young people. This is important because targeted interventions need to focus on those most at risk if they are to maximise their chances of

achieving reductions in the rate of teenage pregnancies. Having said that, the T&T facilitators typically felt that the 'right' girls were referred to the programme, and the prevalence of risk-taking attitudes and behaviours identified in the RCT suggests that the participants in the trial constituted an at-risk population as intended (for more detail, see Chapter 2).

Programme delivery

Overall, the programme was implemented with a high degree of fidelity. However, there were a few areas of variance. Some were an inevitable consequence of different facilitators working in different settings with different young people, but some could have been reduced by providing more specific guidance for facilitators and others involved in the programme.

Acceptability

Both adult stakeholders (i.e. T&T facilitators and teachers) and programme participants reported that the intervention was acceptable. However, it should be noted that some perspectives were out of scope for the evaluation and therefore not explored. These included: young women who refused to participate in the programme or who dropped out; schools who chose not to be involved in the programme; other teachers not involved in the implementation of T&T in their schools; and parents who withheld their consent for their child to be involved.

One issue highlighted by the process evaluation as meriting consideration by T&T and participating schools was how the aims of the programme and selection criteria should be communicated to young people. School staff generally did not inform young people about why they had been selected for the programme. Therefore, at the beginning of the intervention, young people themselves were typically unaware of the reasons for their selection, and when it became apparent later some participants expressed unhappiness about this aspect of the programme.

Perceived impacts

The final part of the process evaluation considered the perceptions of stakeholders and participants regarding the impacts of the programme. These are subjective perceptions rather than the kinds of impacts measured by the RCT, which are discussed in Chapter 4.

The impacts most commonly perceived by adult stakeholders and young people related to aspects of youth development such as self-esteem, self-confidence and interpersonal skills. These in turn were seen as having had an impact on participants' capacity to recognise the choices available to them, articulate those choices, and negotiate power more generally with those around them. Such personal and interpersonal skills may well be important elements of an overall approach to reducing the risk of early pregnancy since, for example, there is some evidence that higher self-esteem is associated with later sexual activity among girls (Spencer et al. 2002).

Adult stakeholders and young people were less certain about impacts on attitudes towards early pregnancy and motherhood. Adult stakeholders either thought that this impact would only become apparent in the longer term or believed that this intervention alone was unlikely to counteract all of the influences that lead to early pregnancy.

1.5 Randomised controlled trial

The effects of the T&T programme on primary and secondary outcomes were examined through a randomised controlled trial (RCT). This design provides the least-biased estimates of intervention effects because randomisation enables intervention and control groups to be comparable on both measured and unmeasured factors, which might otherwise confound estimates of intervention effects (Campbell and Russo 1999; Oakley 1990). In this instance the trial involved the random allocation of *individual* at-risk young women within schools to an intervention or control group (the latter engaging in school activities as usual). Other design options, such as cluster randomised trials, were considered at the design stage but not adopted: a short discussion of the decision process is included in Appendix A).

With participating schools at-risk young women were identified by their teachers. Guidance was provided on the characteristics and attributes considered indicative of a risk for teenage parenthood (see Appendix D for a copy of T&T's original guidance as well as a copy of the revised guidance which was designed for the trial to improve the selection of programme participants). In practice, teachers did not strictly adhere to the criteria set out in the guidance when selecting girls, also using their own judgement and subjective criteria to identify girls felt to be most at risk. The selected young women received preliminary information about the study prior to consenting to allocation, intervention and research. Data were collected by questionnaire at three points in time: prior to allocation (baseline), immediately post-intervention (follow-up one) and a year after intervention (follow-up two). Questionnaires collected quantitative data on the young women's socio-demographic characteristics as well as a range of attitudes and behaviours relevant to the programme outcomes. Girls who were randomised to the T&T programme completed an additional questionnaire that collected their views of the programme at follow-up one.

Two cohorts of young women participated in the trial, one starting in September 2009 and one starting in January/February 2010. The majority of the questionnaires were completed by girls in school though where necessary some girls completed the follow-up questionnaires at home. For the baseline survey with cohort one, questionnaires were administered using face-to-face Computer Assisted Personal Interviews (CAPI) in a private room at the school and the most sensitive parts of the questionnaire were self-administered via Computer Assisted Self Interview (CASI). The total interview length was 40 minutes. In contrast, the baseline survey with cohort two used a paper self-completion questionnaire with small groups of girls in a school classroom (the mode was changed between cohort one and cohort two to try to promote disclosures, see Appendix A for more details). Where possible these sessions were conducted under exam conditions and interviewers asked teachers either to absent themselves from the classroom or, if that was

not possible, to remain at their desk at the front of the classroom to ensure girls' privacy while completing the questionnaires. These sessions lasted between 40 minutes and 1 hour. Following the collection of these baseline data, girls were randomly allocated to the intervention and control groups (see Appendix A for more details on how the randomisation was conducted). For both follow-up stages, both cohorts of girls were asked to complete a paper self-completion questionnaire under the same conditions as described.

In determining the sample size for the RCT, it was necessary to balance the requirement for sufficient statistical power to detect an impact on the relevant outcome measures with the capacity of the T&T programme (see Appendix A for a detailed explanation of the power calculations). In total, 489 girls were randomised to the intervention and control arms. Eighty-five of these formed a reserve group to allow for drop-outs to be augmented (i.e. when girls dropped out of the programme they were retained within the trial but a reserve pair were added in order to ensure that the trial included a sufficient number of girls who completed the programme). Forty of these were not required, meaning that ultimately 449 girls entered the trial (228 in the control arm and 221 in the intervention arm). The overall response rate among these 449 girls at follow-up one was 95% and at follow-up two 91%. The response rate was the same in both arms of the trial and represents an excellent retention rate for a trial of this kind (for more details on sample size and response rates see Appendix A).

The main trial analysis has been conducted on an intention-to-treat basis. This means that all girls who were originally randomised to the intervention and control groups as part of the trial were included in the analysis regardless of how many sessions of the T&T programme they attended in total. This approach is recommended in the Consolidated Standards of Reporting Trials (CONSORT) because individuals who drop out of interventions are generally different in profile, according to baseline data, from those who remain (CONSORT Statement 2010). This means that those individuals who remain with the programme are very different from those in the control group (since there is no similar process of drop-out in this group). Thus a comparison between outcome data collected from all those who were allocated to a programme versus all those who were allocated to the control group (i.e. an 'intention-to-treat' analysis) is more robust than a comparison of outcome data from those who remain with the programme versus those in the control group (an 'in treatment' analysis). Full details of the analysis approach taken in this study are described in Chapter 4.

Key aspects of the RCT design underwent peer review to ensure that the approach taken represented best practice as far as possible. The trial was registered with http://clinicaltrials.gov/ and given the ID number NCT0131054. In addition it was approved by a research ethics committee at NatCen Social Research (ref P2922) and by a separate committee at LSHTM (ref 5932). Further details about the final approach and technical details relating to the implementation of the trial can be found in Appendix A.

1.6 Interpreting results in the report

Tables and figures

The tables in this report show the total number of cases in the whole sample for that wave of the study or in the particular sub-group being analysed, and the bases for different columns or rows (e.g. the two trial arms, 'completers' and 'drop-outs'). The total base figure includes all the eligible cases (i.e. all respondents or all respondents asked a particular question) minus cases with missing data (codes 'don't know' or 'not answered'). Thus, while the base description may be the same across several tables, the base sizes may differ slightly due to the exclusion of those coded 'don't know' or 'not answered'.

Due to rounding, percentage figures may not add up to exactly 100%. Furthermore, where the information in tables is based on multi-coded questions, the percentages in the table could add up to more than 100%.

For all graphs presenting data distributions, there are associated tables in Appendix C showing full details of the analysis. These are listed as sources underneath the graphs. Similarly, some tables in the main body of the report present key statistics from more detailed tables in Appendix C.

Some of the percentages in the impact analysis are based on regression models. Where this is the case, these percentages are labelled as 'adjusted' and full details of the adjustments are provided.

Statistical significance

Throughout the report, all comments on differences between sub-groups of the sample refer to differences tested for significance using STATA 10.1 and found to be statistically significant at the 95% confidence interval (CI) or above. Some exceptions, where the significance level was set at the 90% CI, are clearly identified.

2 Characteristics of teenagers participating in the trial

2.1 Introduction

This chapter examines the characteristics of teenagers who took part in the T&T trial and discusses whether the trial succeeded in recruiting young women who were sufficiently at risk of teenage pregnancy. Those who were randomised to the intervention group are compared with those allocated to the control group, and the chapter discusses whether the randomisation succeeded in creating balanced trial arms. The chapter also examines the programme attendance of the young women randomised to the intervention and whether there were differences between the characteristics of those who completed the programme and those who dropped out. Finally, the chapter examines whether there was evidence of a non-response bias (i.e. of systematic differences between those who did and did not respond to the surveys) and, if so, whether this affects the comparability of the intervention and control groups.

2.2 Characteristics of study participants at baseline

In this section, we describe the characteristics of the young women recruited to take part in the trial focusing on a number of key socio-demographic characteristics, behaviours and attitudes. These characteristics have been chosen from over 70 questions asked of teenagers at the baseline stage and reflect a range of topics covered in the questionnaire.

Results presented in Table 2.1 show the overall profile of the teenagers who took part in the baseline survey.

Table 2.1 Key characteristics of study participants at baseline		
Base: All at baseline		
	%	***Bases***
Socio-demographic characteristics		
Age	Mean=13.5	*446*
Non-owner housing tenure	76	*392*
Household worklessness	32	*447*
Non-white ethnicity	51	*446*
Family's main language not English	21	*447*
Receives free school meals	44	*446*
Sexual behaviours and attitudes		
Sexually active (had sex with a boy)	13	*446*
No contraception at last sex in last three months	2	*443*
Expects teenage parenthood	22	*443*
Has been pregnant	2	*449*
School behaviour and attitudes		
Dislike of school	32	*448*
Misses school without permission	35	*444*
Suspended or temporarily excluded from school in last six months	11	*447*
Youth development		
Youth development score	Mean=49.1	*439*
Low self-esteem	14	*445*
Other key characteristics		
Gets drunk (ever)	37	*445*
Gets drunk monthly or more	19	*447*

Note: The base for all questions is the whole baseline sample, including for sexually-related behaviours.

The results shown in Table 2.1 suggest that many teenagers participating in the study were from disadvantaged backgrounds: for example, 44% received free school meals (eligibility for which is based on family income), 76% lived in rented accommodation, and 32% were from families where no adult was in paid employment. With regard to sexual behaviours and attitudes, 13% had experienced heterosexual intercourse, 2% had been pregnant, and 22% expected to become a parent before reaching the age of 20. Just under a third disliked school (32%) and just over a third had missed school without permission (35%). Just over a third of teenagers said they had been drunk (37%) and just under a fifth (19%) said they got drunk at least monthly. These figures suggest a relatively high prevalence among the study sample of behaviours and attitudes known to be associated with increased risk of teenage pregnancy (Kneale 2008). Thus we conclude that the study was successful in recruiting young women who were sufficiently at risk.

2.3 Randomisation

As mentioned in Chapter 1, the young women in each school and cohort were randomised to an intervention or control group (see Appendix A for a detailed discussion of the randomisation process). This section examines the results of the randomisation, discusses whether it was successful in producing intervention and control groups with similar characteristics, and considers whether the impact analysis needs to be adjusted for any differences between the trial arms at baseline. It also examines whether there were any influences on the control group since the baseline which were similar to the T&T programme, and discusses implications of that for the comparability of the trial arms.

2.3.1 The trial arms at baseline

Table 2.2 shows the distributions of some key baseline characteristics among teenagers in the intervention and control arms of the trial. (See Table C2.1 in Appendix C for a more detailed version of this table.)

Table 2.2 Key baseline characteristics, by trial arm			
Base: All at baseline			
	Control group	**Intervention group**	**P value for difference**
	%	%	
Socio-demographic characteristics			
Age	Mean=13.5	Mean=13.5	0.436
Non-owner housing tenure	75	78	0.500
Household worklessness	31	34	0.503
Non-white ethnicity	50	51	0.846
Family's main language not English	21	21	0.955
Receives free school meals	42	46	0.423
Sexual behaviours and attitudes			
Sexually active (had sex with a boy)	13	13	0.964
No contraception at last sex in last three months	2	1	0.450
Expects teenage parenthood	18	26	**0.045**
Has been pregnant	2	3	0.539
School behaviour and attitudes			
Dislike of school	33	31	0.656
Misses school without permission	33	36	0.456
Suspended or temporarily excluded from school in last six months	11	10	0.735
Youth development			
Youth development score	Mean=49.5	Mean=48.8	0.379
Low self-esteem	15	14	0.785
Other key characteristics			
Gets drunk (ever)	36	38	0.634
Gets drunk monthly or more	18	20	0.688
Bases	*203-228*	*189-221*	*392-449*

Notes: The base for all questions is the whole baseline sample, including for sexually-related behaviours. P values refer to: chi-squared tests where the baseline characteristic is a categorical variable with an expected cell count of five or more, t tests where it is continuous, and Fisher's exact test (two-sided) where the baseline variable is categorical but at least one of the expected cell counts is under five. This table contains data extracted from Table C2.1 in Appendix C.

The results shown in Table 2.2 in this chapter and in Table C2.1 in Appendix C suggest that overall, the two trial arms were well-balanced. We compared around 70 characteristics of the two groups and there were only differences for five of them (see Figure 2.1), which is the number that would be expected to occur simply through chance.[3] The pattern of these differences was such that teenagers in the intervention group had a slightly riskier profile with regard to these characteristics than those in the control group.

[3] For this analysis, we took a conservative approach and treated as significant any differences with p value under 0.1. This was in order to minimise the risk of overlooking confounders important for the impact analysis.

For example, 26% of teenagers in the intervention group expected to become a parent by age 20, compared with 18% of those in the control group (see Table 2.2).

Next we checked whether the same profile of baseline differences applied to those teenagers who participated in follow-up one and follow-up two, as some teenagers did not respond to the follow-up surveys and this might have affected the differences in the data. (The results are shown in Table C2.2 and Table C2.3 in Appendix C.) From the point-of-view of the impact analysis, it is these baseline differences between the trial arms among follow-up one and follow-up two participants that matter, as the impact analysis only uses data for those teenagers who took part in the follow-up surveys.

The additional analysis of the baseline data revealed that there were indeed some variations in whether the two trial arms were different in terms of a particular characteristic depending on whether we included all baseline participants, only those who took part in follow-up one or only those who took part in follow-up two. Figure 2.1 summarises the findings of this analysis, and the full sets of results are shown in Table C2.2 and Table C2.3 in Appendix C.

Figure 2.1 Summary of baseline differences between the trial arms

	Differences between the trial arms*		
	All at baseline	Follow-up one participants	Follow-up two participants
Expects teenage parenthood	√		
Low self-reflection	√	√	√
Low sexual health knowledge	√	√	√
Worries a lot	√	√	√
Parents don't know most or all friends			√
Family member had a baby before age 20	√		
Bases	*449*	*425*	*408*

Source: Table C2.2 and Table C2.3 in Appendix C.
*at p<0.1

Where baseline differences between the trial arms are present, there is the potential that they could bias the analysis of impacts of the intervention. The way to prevent this from happening is to adjust the impact analysis for those differences which are relevant to individual outcomes.

On the basis of the results summarised in Figure 2.1, three baseline characteristics – namely low self-reflection, low sexual health knowledge and worrying a lot – were assessed as potential confounders in multivariate analyses of impact using follow-up one data. The same characteristics, plus whether parents don't know most or all friends, were assessed as potential confounders for analyses of impact using data from follow-up two. As discussed in Chapter 4, the actual decisions about which covariates to include in the

multivariate analyses of impact for individual outcomes were based on whether the potential confounders were associated with those individual outcomes, because to be a confounder, a variable must by definition differ by arm and be associated with the outcome in question. (For full details of which covariates are included in the regression models for which outcomes at follow-up one and follow-up two, see Table C4.2 and Table C4.4 in Appendix C.)

2.3.2 Influences on the control group since the baseline measurements

While the intervention group participated in the T&T programme, were there any influences on the control group that might have been similar in nature to the T&T intervention and thus might have affected the comparability between the trial arms at follow-ups? We examined two kinds of possible influences: (1) participation in other youth development programmes and (2) participation in sex and relationship education (SRE) lessons at school which were missed out by the intervention group.

At both follow-ups, teenagers in both the intervention and control groups were asked whether, since last survey, they had attended any personal development or education programmes for young people other than T&T that were not part of normal school. Twenty per cent of girls in the intervention group and 25% of those in the control group answered positively at least at one of the follow-ups (see Table C2.1 in Appendix C). The difference between these figures was not statistically significant, which suggests that other youth development programmes were unlikely to have affected comparability between the trial arms at follow-ups.

There was no specific question in the surveys about missing SRE lessons. However, the girls in the intervention group were asked which lessons they had missed because of participating in T&T, and 6% reported missing lessons in Personal, Social and Health Education (PSHE) (see Table C3.8 in Appendix C). It is possible that some of these lessons might have covered SRE topics and thus these girls missed out on education which was relevant to the focus of the intervention and which the control group participated in. However, given the low percentage reporting missing PSHE lessons and the fact that SRE topics were probably just part of the PSHE curriculum, we believe it is unlikely that missing these lessons by some girls in the intervention groups significantly affected the comparability between the trial arms at follow-ups.

2.4 Attendance at the programme

This section examines rates of attendance at the trial T&T programme by teenagers randomised to receive the intervention, and compares the characteristics of those who dropped out with those who completed the programme.

The programme consists of 18 to 20 weekly sessions. In order to receive a certificate as a result of attending the T&T programme, teenagers needed to have attended at least 50%

of the sessions. Those who attended 80%+ of the programme received an achievement certificate, and those who attended 50-79% of the programme received an attendance certificate.

For the trial, if a teenager dropped out of the programme within the first 8 weeks, where possible another teenager from the reserve sample was invited to join the programme. This was to ensure (1) that there was still a group of a sufficient size for the intervention's group sessions to be effective, and (2) that the number of teenagers completing the programme would be sufficiently large for an adequately powered analysis of impact. In analysis, the ex-reserve teenagers were treated as having *augmented* the original sample rather than having *replaced* those who dropped out, in order to maintain the 'intention-to-treat' design and avoid introducing a bias. In total, 45 out of 449 teenagers in the trial were ex-reserves (10%).

Table 2.3 shows rates of programme attendance among the 'starting' sample (that is, excluding reserves) and among all teenagers in the trial (i.e. those from the starting sample plus ex-reserves). They are very similar, as one would expect, given that the number of ex-reserves was relatively low. Overall, 73% of teenagers randomised to the intervention completed the programme. Accordingly, just over a quarter of teenagers randomised to the intervention (27%) dropped out of the programme. Of those completing the programme, most attended 80% or more of the sessions and received an achievement certificate.

Table 2.3 Attendance at the T&T programme

Base: Teenagers randomised to intervention		
	'Starting' intervention group (i.e. excluding reserves)	**Overall intervention group (including ex-reserves)**
Attendance at the programme	%	%
80%+ attendance (achievement certificate)	65	63
50-79% attendance (attendance certificate)	10	10
<50% attendance (no certificate): 'drop-outs'	25	27
Bases	*199*	*221*

Of those who dropped out, ten girls did not attend any sessions at all (17% of the drop-out group). Of those who attended at least one session, all but one girl dropped out within the first 8 weeks. The median number of sessions attended by those in the drop-out group

(including those who attended no sessions) was two sessions (results not shown in a table).[4]

The 73% completion rate for the trial T&T programme is relatively high but somewhat lower than the 80-88% completion reported on T&T programmes most recently.[5] This difference in completion rates is partially explained by differences between how this figure was calculated for the trial and is usually calculated by the T&T.[6] It is also probably explained, at least in part, by the effect of the research process on programme attendance (e.g. the individual-based randomisation of teenagers to the trial arms that would not allow friendship pairs to be kept together). However, the exact reasons are unknown as the evaluation did not collect information from teenagers who dropped out on why they stopped attending the programme sessions.

Were there any differences between teenagers who dropped out of the programme and those who completed it? Were T&T successful in retaining those girls who were the most at risk of teenage pregnancy? Table 2.4 compares these two groups and the control group on a number of key baseline characteristics.

[4] The median number of session attended by those who attended at least one session but did not complete the programme is also two sessions.
[5] The figures were provided by COUI and are taken from an email correspondence between COUI, DfE and NatCen Social Research on 13 December 2011. They refer to four T&T cohorts that took part in the programme over 2009 to 2011.
[6] The method of calculation used for the RCT included all teenagers randomised to the intervention in the base, while the method used by the T&T only includes those who attended at least one session. In the RCT, ten teenagers from those randomised to the intervention attended no programme sessions but were still included in the category of 'drop-outs'. When the RCT figure is recalculated while excluding these ten teenagers from the base and from the drop-outs, the attendance level increases from 73% to 76%.

Table 2.4 Key baseline characteristics, by attendance at T&T programme

Base: All at baseline

	Control	Interv.: completers	Interv.: drop-outs	*Bases*
	%	%	%	
Socio-demographic characteristics				
Non-owner housing tenure	75	76	85	*392*
Household worklessness	31	29	47	*447*
Non-white ethnicity	50	53	47	*446*
Family's main language not English	21	23	18	*447*
Receives free school meals	42	43	54	*446*
Sexual behaviours and attitudes				
Sexually active (had sex with a boy)	13	12	17	*446*
No contraception at last sex in last three months	2	1	0	*443*
Expects teenage parenthood	18	22	36	*443*
Has been pregnant	2	2	3	*449*
School behaviour and attitudes				
Dislike of school	33	29	37	*448*
Misses school without permission	33	32	49	*444*
Suspended or temporarily excluded from school in the last six months	11	7	20	*447*
Youth development				
Youth development score	Mean=49.5	Mean=49.2	Mean=47.5	*439*
Low self-esteem	15	14	14	*445*
Other key characteristics				
Gets drunk (ever)	36	33	53	*445*
Gets drunk monthly or more	18	17	27	*447*
Bases	*203-228*	*143-161*	*46-60*	*392-449*

Note: The base for all questions is the whole baseline sample, including for sexually-related behaviours. The category "Intervention: completers" includes teenagers who were randomised to intervention, completed at least 50% of the programme and received a certificate. The category "Intervention: drop-outs" includes teenagers who were randomised to intervention but dropped out of the programme and did not receive a certificate.

There were marked differences at the baseline stage between teenagers who dropped out of the programme and those who completed it, suggesting that those who dropped out were much more likely to be from disadvantaged backgrounds and to engage in risk-taking behaviours. For example, 47% of teenagers who dropped out lived in workless households (compared with 29% of those who completed the programme); 49% missed school without permission (compared with 32%); 53% had got drunk (compared with 33%); and 36% expected to become a parent before age 20 (compared with 22%, see Table 2.4).

2.5 Response to the research

This section is devoted to analysis of response to the study. The detailed response figures to follow-ups one and two are shown in Appendix A. Table 2.5 summarises response to the two follow-ups among the intervention and control groups.

Table 2.5 Response to follow-up one and follow-up two surveys among baseline participants, by trial arm and programme attendance

Base: All at baseline

	Completed follow-up one q-aire	Completed follow-up two q-aire	*Bases*
	%	%	
Control group	94	91	*228*
Intervention group: completed the programme	96	92	*161*
Intervention group: dropped out	92	88	*60*
All	95	91	*449*

While some modest differences in percentages are apparent in Table 2.5, there were no statistically significant differences in the rates of response, either at follow-up one or at follow-up two, between the intervention and control groups, or between those in the intervention group who completed the programme and those who dropped out.

Were there any differences between those who responded and did not respond to the study? Table 2.6 shows distributions of some key baseline characteristics by response to follow-up one. There is indeed some evidence that those who responded and did not respond were systematically different in some respects. For example, those who participated at baseline and then did not respond to follow-up one were more likely to expect teenage pregnancy than those who responded to both of these waves. They were also more likely to dislike school, to miss school without permission and to get drunk, and their average youth development score was lower. To sum up, those who did not respond to follow-up one were more likely to engage in risk-taking behaviours than those who completed a questionnaire at follow-up one. For most characteristics where we found these differences, this pattern of non-response was evident among both the intervention and control groups, although it was somewhat more pronounced among the intervention group (results not shown due to low absolute numbers of non-responders when analysed by trial arm).

Table 2.6 Key baseline characteristics, by response to follow-up one			
Base: All at baseline			
	Did not complete follow-up one questionnaire	**Completed follow-up one questionnaire**	**P value for difference**
	%	%	
Socio-demographic characteristics			
Non-owner housing tenure	82	76	0.529
Household worklessness	38	32	0.586
Non-white ethnicity	43	51	0.465
Family's main language not English	29	21	0.330
Receives free school meals	33	45	0.282
Sexual behaviours and attitudes			
Sexually active (had sex with a boy)	21	13	0.229
No contraception at last sex in last three months	4	1	0.325
Expects teenage parenthood	46	21	**0.006**
Has been pregnant	4	2	0.426
School behaviour and attitudes			
Dislike of school	58	31	**0.005**
Misses school without permission	58	33	**0.012**
Suspended or temporarily excluded from school in last six months	13	11	0.738
Youth development			
Youth development score	Mean=45.1	Mean=49.3	**0.021**
Low self-esteem	13	14	1.000
Other key characteristics			
Gets drunk (ever)	58	36	**0.027**
Gets drunk monthly or more	25	18	0.423
Bases	*22-24*	*370-425*	*392-449*

Notes: The base for all questions is the whole baseline sample, including for sexually-related behaviours.
P values refer to: chi-squared tests where the baseline characteristic is a categorical variable with an expected cell count of five or more, t tests where it is continuous, and Fisher's exact test (two-sided) where the baseline variable is categorical but at least one of the expected cell counts is under five.

Table 2.7 shows how some key baseline characteristics were distributed among those responding and not responding to follow-up two. In contrast to our findings with regard to follow-up one, there is no evidence that those who responded to follow-up two were different from those who did not respond. The only difference – marginally significant at $p<0.1$ – was with regard to household worklessness, with those who did not respond to follow-up two being marginally more likely to come from households where no-one was in paid work. Overall, even though the response to follow-up two was somewhat lower than that to follow-up one, the sample of respondents at follow-up two was more representative of baseline participants than the sample of respondents at follow-up one.

Table 2.7 Key baseline characteristics, by response to follow-up two			
Base: All at baseline			
	Did not complete follow-up two questionnaire	**Completed follow-up two questionnaire**	**P value for difference**
	%	%	
Socio-demographic characteristics			
Non-owner housing tenure	76	76	0.928
Household worklessness	45	31	0.075
Non-white ethnicity	55	50	0.586
Family's main language not English	24	21	0.606
Receives free school meals	48	44	0.635
Sexual behaviours and attitudes			
Sexually active (had sex with a boy)	10	14	0.491
No contraception at last sex in last three months	0	2	1.000
Expects teenage parenthood	21	22	0.827
Has been pregnant	2	2	1.000
School behaviour and attitudes			
Dislike of school	37	32	0.545
Misses school without permission	42	34	0.338
Suspended or temporarily excluded from school in the last six months	10	11	1.000
Youth development			
Youth development score	Mean=47.3	Mean=49.3	0.169
Low self-esteem	8	15	0.225
Other key characteristics			
Gets drunk (ever)	38	37	0.852
Gets drunk monthly or more	22	18	0.587
Bases	*37-41*	*355-408*	*392-449*

Notes: The base for all questions is the whole baseline sample, including for sexually-related behaviours.
P values refer to: chi-squared tests where the baseline characteristic is a categorical variable with an expected cell count of five or more, t tests where it is continuous, and Fisher's exact test (two-sided) where the baseline variable is categorical but at least one of the expected cell counts is under five.

To sum up, three main findings were discussed in this section:

1. Overall response rates were very high at both follow-ups.
2. Response rates did not vary by trial arm or according to whether the teenagers completed the T&T programme or dropped out.
3. While there was no evidence of a non-response bias for follow-up two participants, there were some systematic differences between those who responded and did not respond at follow-up one, with non-responders more likely to engage in risk-taking behaviours.

Does the differential response to follow-up one affect the robustness of the impact analysis of those data? We believe it does not, as even though there were differences between those who did and did not respond, the rate of non-response did not differ by arm and – at 5% – was very low. Thus the impact of the non-response bias on the representativeness of the productive sample is limited.

However, in order to make sure that any relevant implications of the non-response for the follow-ups could be adequately accounted for in the impact analysis, we checked whether the two trial arms were well-balanced at baseline for three samples: (1) all at baseline, (2) those at baseline who took part in follow-up one, and (3) those at baseline who took park in follow-up two (see section 2.3 for the results). Where baseline differences between the trial arms were found, those baseline characteristics have been checked as potential confounders for impacts at follow-ups one and two.

2.6 Two cohorts

As mentioned in Chapter 1, two cohorts of teenagers took part in the study: the first started in September 2009 and the second in January/February 2010. Once the baseline data were collected, it was apparent that there were substantial differences in the incidence of various behaviours and attitudes associated with increased risk of teenage pregnancy between the two cohorts. This section examines these differences and discusses possible explanations for their origins.

Two main factors need to be taken into consideration when discussing differences between cohort one and cohort two. The first factor relates to the schools in each cohort. There were ten schools in cohort one, all of which also took part in cohort two. In addition to these 'original' schools, 12 'supplementary' schools joined the trial at cohort two.

The second factor relates to a change in the mode of data collection. Teenagers in cohort one completed the baseline questionnaire via face-to-face interviews administered by NatCen interviewers using CAPI, with the most sensitive questions being self-administered via CASI. Those in cohort two completed paper self-completion questionnaires (for details about why this change of data collection mode took place, see Appendix A).

Table 2.8 compares the profile of teenagers in the following three groups:

- those in cohort one who were from the 'original' schools;
- those in cohort two who were from the 'original' schools;
- those in cohort two who were from the 'supplementary' schools.

Table 2.8 Key baseline characteristics, by cohort and whether schools took part in both cohorts or just cohort two

Base: All at baseline

	Cohort 1	Cohort 2 original schools	Cohort 2 suppl. schools'	*Bases*
	%	%	%	
Socio-demographic characteristics				
Non-owner housing tenure	86	81	64	*392*
Household worklessness	41	27	29	*447*
Non-white ethnicity	67	75	21	*446*
Family's main language not English	24	40	6	*447*
Receives free school meals	50	45	38	*446*
Sexual behaviours and attitudes				
Sexually active (had sex with a boy)	6	3	26	*446*
No contraception at last sex in last 3 months	1	0	4	*443*
Expects teenage parenthood	16	22	27	*443*
Has been pregnant	1	0	5	*449*
School behaviour and attitudes				
Dislike of school	20	31	45	*448*
Misses school without permission	25	39	41	*444*
Suspended or temporarily excluded from school in last 6 months	14	8	11	*447*
Youth development				
Youth development score	Mean=51.7	Mean=48.8	Mean=47.1	*439*
Low self-esteem	3	12	25	*445*
Other key characteristics				
Gets drunk (ever)	17	34	57	*445*
Gets drunk monthly or more	5	16	33	*447*
Bases	*148-153*	*93-120*	*151-176*	*392-449*

Note: The base for all questions is the whole baseline sample, including for sexually-related behaviours.

Results presented in Table 2.8 suggest that both factors discussed above contributed to the overall differences between teenagers in the two cohorts. Focusing first on differences between teenagers from the 'supplementary' schools in cohort two and those from the 'original' schools (in both cohorts), the evidence suggests that teenagers from the 'supplementary' schools were less likely to be from non-white ethnic backgrounds and to have English as a second language than those from the 'original' schools. Teenagers from the 'supplementary' schools were also much more likely be sexually active than those from the 'original' schools. In addition, teenagers from the 'supplementary' schools had a lower mean youth development score than those from the 'original' schools, and were more likely to have low self-esteem. Finally, teenagers from the 'supplementary' schools were more likely to get drunk and to dislike school.

Comparisons between teenagers from the same 'original' schools but taking part in cohort one and cohort two reveal that the prevalence of sexual behaviours and attitudes was similar between these two cohorts. Questions on these topics were self-administered in both cohorts: on the computer in cohort one and in a paper self-completion questionnaire in cohort two. However, with regard to other, less sensitive, questions, there are some notable differences between the two cohorts, which might be explained by the change of data collection mode from face-to-face interview to self-completion questionnaire. For example, teenagers in cohort two had a lower youth development score and lower self-esteem and were more likely to say that they disliked school. As questions on such topics are vulnerable to a social desirability effect, it is possible that the measurement of these phenomena was more accurate at cohort two than at cohort one due to the use of paper self-completion questionnaires for cohort two and of the interviewer administered CAPI for cohort one.

It is not clear why the profile of teenagers from the 'supplementary' schools was so different from those from the 'original' schools. It might be explained by the social make-up of their local areas, or it could be related to school-level factors. Geographical and school-level factors were not measured in this study.

It is our view that these differences do not bias the main impact analysis, as the randomisation of teenagers into the intervention and control groups was carried out within schools and cohorts, and therefore each school and cohort is equally represented in both arms of the trial.

2.7 Summary

The chapter examined the characteristics of teenagers who took part in the randomised controlled trial at baseline. Their average age was 13.5 years. Many were from disadvantaged backgrounds, with 44% receiving free school meals and 32% living in workless households. Thirteen per cent had experienced heterosexual intercourse, 2% had been pregnant, and 22% expected to become parents while still in their teens. About a third disliked school, the same proportion missed school without permission, and about a fifth said they got drunk at least once a month. Overall, the prevalence of risk-taking behaviours and attitudes among participants at baseline suggests that teachers targeted an appropriate group of young women who were at risk of teenage pregnancy.

Analysis of the results of the randomisation of teenagers to the intervention and control groups showed that the groups were well-balanced, only differing in regard to a few characteristics at baseline. In order to prevent these baseline differences from confounding the impact analysis, those characteristics where the trial arms were different at baseline are considered as potential confounders for effects on individual outcomes in the impact analysis of follow-up one and follow-up two data in Chapter 4.

Attendance at the T&T programme was relatively high with 73% completing the programme, although it was somewhat lower than the attendance reported on T&T

programmes most recently. There were marked baseline differences between teenagers who dropped out of the programme and those who completed it. These differences indicated that those who dropped out were more likely to be from disadvantaged backgrounds and to engage in risk-taking behaviours.

The overall response to the a follow-up surveys among baseline participants was very high, with 95% completing a questionnaire at follow-up one and 91% at follow-up two. Response rates did not vary by trial arm or by whether the teenagers completed the T&T programme or dropped out. There was some evidence of a non-response bias (i.e. systematic differences between those who did and did not respond) at follow-up one but not at follow-up two, and there was no evidence to suggest that this negatively affected the comparability between the trial arms.

The chapter also examined differences between the two cohorts involved in the trial and found that teenagers from schools which joined the trial at cohort two were more likely to engage in risk-taking activities than those from the schools which took part in both cohorts. The reasons for these differences are unclear and are most likely related to factors operating at geographical area level and school level, which were not measured in the study.

There were also some differences between the characteristics of the teenagers in cohort one and cohort two from the same 'original' schools, which might be explained by a change of data collection method at baseline from CAPI for cohort one to a paper self-completion questionnaire for cohort two. The self-completion method might have led to higher disclosure of behaviours and attitudes which are vulnerable to a social desirability effect. It is noteworthy that there were no differences between the two cohorts from the same schools with regard to the prevalence of sexual behaviours and attitudes, with cohort one completing that part of the questionnaire as CASI and cohort two filling in a paper questionnaire.

As teenagers from within the same schools and cohorts were randomly allocated to the intervention and control groups, these cohort differences would not bias our estimates of intervention effects.

3 Perceptions of the programme

3.1 Introduction

As discussed in Chapter 1, follow-up one was conducted shortly after teenagers had completed the T&T programme. After the teenagers had completed the main questionnaire at follow-up one, those who had been randomised to the intervention were asked to complete an additional questionnaire that collected their views of the programme and their perceptions of its impact. In total, 194 girls completed this additional questionnaire. This included 38 girls who dropped out of the programme before week eight. Sixteen members of the intervention group who took part in follow-up one chose not to complete the additional questionnaire. All 16 had dropped out of the programme before week eight and would not complete the additional questionnaire because they felt that it was not relevant to them.

This chapter presents the findings from the additional questionnaire and covers:

- The elements of T&T that teenagers enjoyed;
- The elements of T&T that teenagers found difficult or challenging;
- The topics teenagers felt they knew more about after attending T&T;
- The benefits of the T&T programme;
- The downsides of participating in the T&T programme;
- Teenagers' experiences of T&T, including how it made them feel and how other people reacted to their involvement with the programme;
- The extent to which teenagers talked about the programme to non-participants.

Where relevant the findings are shown separately for teenagers who began the programme with higher and lower youth development scores.[7]

3.2 Enjoyment and challenge of programme elements

Over the course of the T&T programme teenagers support a young child (typically a toddler judged by nursery staff as potentially benefiting from special attention) for about one and a half hours per session. They also have access to individual time with a trained counsellor and spend around 90 minutes per session in a classroom working with a facilitator trained in the pedagogy and curriculum of T&T. Throughout the programme they engage in ongoing journaling activity where they are encouraged to explore their thoughts and feelings about events occurring in their lives and record their experiences of working with the children in the nursery.

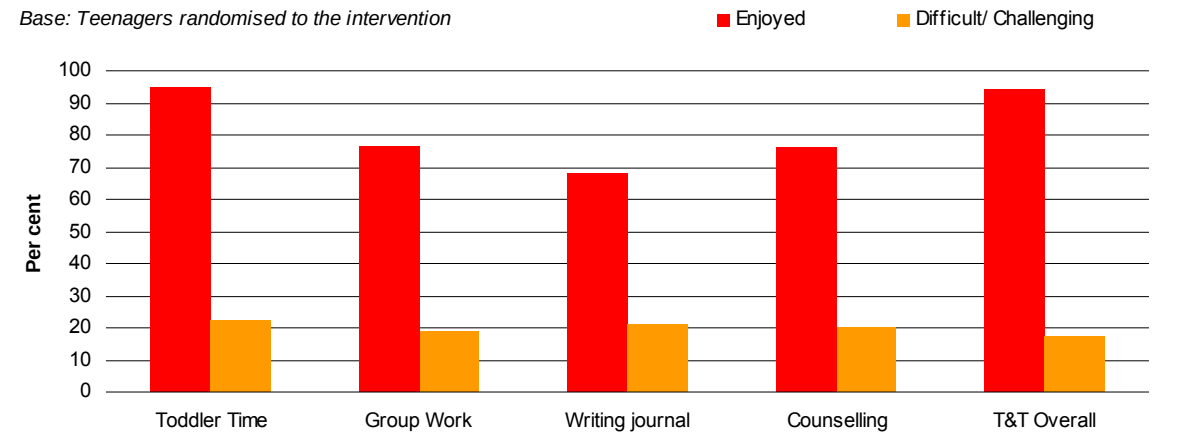
[7] We also looked at whether perceptions varied by ethnicity and take-up of free school meals. However, we did not find any consistent patterns of association and so have not reported these results.

As discussed in Chapter 1, the means of selecting participants for counselling each week varied across schools. As such, participants did not all receive the same number of counselling sessions. Eight per cent of teenagers in the intervention group said that they had never talked to the counsellor, 11% said they had spoken to the counsellor once, 37% had spoken to the counsellor two or three times, 24% four or five times, and 19% had spoken to the counsellor on more than five occasions (see Table C3.1).

Figure 3.1 shows how the teenagers felt about T&T overall and about each individual element of the programme. It shows that the vast majority of teenagers enjoyed taking part in T&T overall (94%, comprising 77% who said they enjoyed it a lot and 17% who said they enjoyed it a little, see Table C3.2). They enjoyed the toddler time in particular (95%) but were less keen on writing their journals (68%).

About one-fifth of teenagers said that taking part in T&T was difficult or challenging and this was equally the case for each different element of the programme.

Figure 3.1 Elements of the programme teenagers enjoyed or found difficult or challenging



Source: Table C3.2 and Table C3.3 in Appendix C.

Table C3.3 in Appendix C provides more details about the extent to which teenagers found each element of the programme difficult or challenging. Notably, teenagers were less likely to have found the counselling sessions difficult or challenging than other elements of the programme. Indeed, 56% of teenagers reported that they never found the counselling difficult or challenging compared with between 24% and 37% for other elements of the programme.

Since T&T is a youth development programme, we explored whether teenagers' perceptions of the programme varied depending on their baseline levels of youth development.[8] Table 3.1 shows that teenagers who scored lower on the youth

[8] This scale includes 8 items that assess various aspects of youth development such as confidence, self-esteem, conflict resolution, and communication. Each item was scored from 1 to 8

development scale enjoyed the programme a little less than those who scored higher on the youth development scale (89% compared with 99%). In particular they were less likely to enjoy the group work (69% compared with 84%) and the journal writing (58% compared with 76%).

Table 3.1 Elements of the programme that teenagers enjoyed, by youth development score

Base: Teenagers randomised to intervention

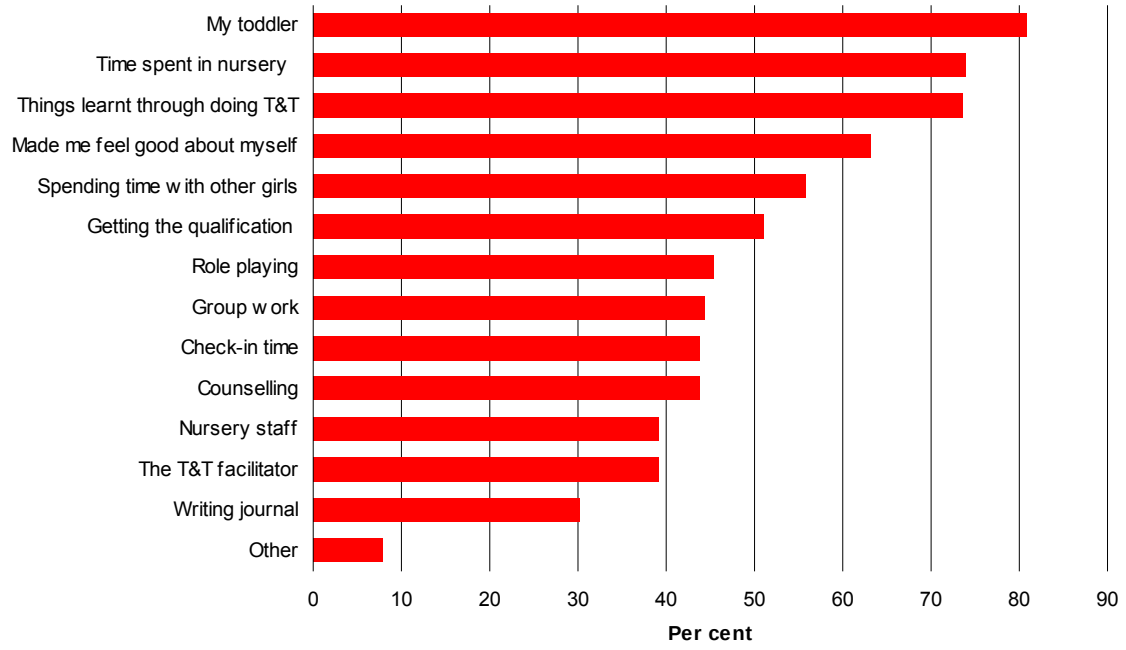
	Youth development score					
	Low		High		Total	
	%	*Base*	%	*Base*	%	*Base*
Toddler time	92	*93*	97	*95*	95	*190*
Group work	69	*93*	84	*95*	77	*190*
Writing journal	58	*86*	76	*92*	68	*180*
Counselling	77	*84*	74	*93*	76	*179*
Overall	89	*93*	99	*96*	94	*191*

Figure 3.2 looks at the elements of T&T in more detail, showing which aspects teenagers felt were the best. These findings correspond to those above, with the majority of teenagers reporting that their toddlers and time in the nursery were the best aspects of the programme (81% and 74% respectively). Other aspects identified by more than half the teenagers as best aspects of the programme were the things they learnt through doing T&T (73%); the fact that the programme made them feel good about themselves (63%); the time they spent with the other girls (56%); and getting the National Award in Interpersonal Skills qualification (51%).

and the youth development scale is a sum of the individual scores. Its values range from 8 to 64, with higher values indicating positive youth development. To assess whether teenagers' perceptions of the T&T programme varied depending on their baseline levels of youth development the scale was dichotomised in such a way that those who scored the average or lower (mean=49 for the 194 teenagers who completed the additional questionnaire) were coded as having a low level of youth development, and those who scored higher than the average were coded as having a high level of youth development.

Figure 3.2 Best things about T&T

Base: Teenagers randomised to the intervention

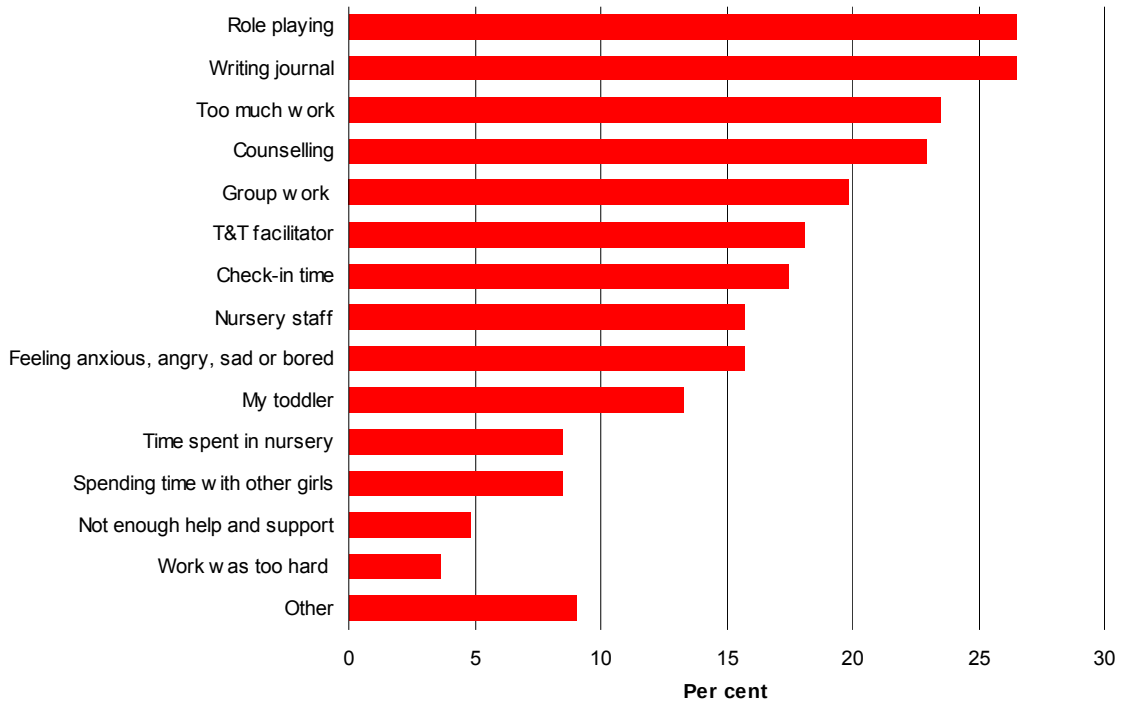


Source: Table C3.4 in Appendix C.

The aspects of the programme that teenagers felt were most difficult included role playing and journal writing (both 27%). In addition, 23% felt that the programme involved too much work and the same proportion found the counselling difficult or challenging (see Figure 3.3).

Figure 3.3 Difficult or challenging things about T&T

Base: Teenagers randomised to the intervention



Source: Table C3.5 in Appendix C.

3.3 Feelings about the programme

Teenagers expressed extremely positive feelings about the programme, as illustrated by Figure 3.4. Sixty-five per cent of teenagers reported that T&T had made them feel responsible, 64% said it had made them happy, 61% felt positive about things and 59% were positively interested in the programme.

Substantially fewer teenagers reported negative emotions arising from their participation in T&T. In response to a question about whether they had ever experienced negative emotions as a result of T&T, 37% said that they had sometimes felt bored, 20% had been irritated or annoyed, 16% had felt frustrated, 14% sad, 7% anxious, and 7% angry (see Figure 3.4)..

Figure 3.4 The ways teenagers felt when participating in T&T[9]

Source: Table C3.6 and Table C3.7 in Appendix C.

These positive feelings are reinforced by the fact that 91% of girls in the intervention group said that they felt proud of being involved with the T&T programme, while only 7% said that their participation made them feel embarrassed.

Teenagers who scored higher on the youth development scale were particularly likely to have experienced positive feelings about taking part in the T&T programme. In particular, they were more likely to say that it had made them feel positive about things (68% compared with 53%, see Table 3.2).

[9] Within this figure the size of the font for each emotion reflects the proportion of teenagers who felt that way about taking part in the T&T programme. The underlying data can be found in Table C3.6 and Table C3.7 in Appendix C.

Table 3.2 How teenagers felt when participating in T&T, by youth development score			
Base: Teenagers randomised to intervention			
	Youth development score		
	Low	High	Total
	%	%	%
Happy	61	66	64
Positive about things	53	68	61
Interested	55	62	59
Excited	43	48	46
Responsible	67	64	65
None of these	11	1	6
Bases	*93*	*98*	*193*

Furthermore, teenagers who scored higher on the youth development scale at baseline were more likely to agree that they felt proud of doing T&T (95% compared with 86%) and less likely to report feeling embarrassed about doing the programme (3% compared with 10%).

3.4 Perceived advantages of the programme

3.4.1 Knowledge

The T&T curriculum is comprised of the following elements:

- Ongoing journaling activity where participants are encouraged to explore their thoughts and feelings about events occurring in their lives and record their experiences working with the children in the nursery.
- Education in each session on an aspect of interpersonal skills, such as communication, listening, constructive conversation, receiving criticism, and giving and receiving negative feedback.
- Education on human development, particularly of toddlers and children.
- Education about healthy parenting skills including the importance and role of fathers in children's development.
- Discussions about teenage sex and the consequences of unplanned pregnancy.
- Information on sources of support for mental and physical health, including sexual health.

The additional questionnaire explored the extent to which teenagers thought that the T&T programme had improved their knowledge of these curriculum areas. The findings are presented in Table 3.3.

Table 3.3 Topics teenagers felt they knew more about					
Base: Teenagers randomised to intervention					
	Extent of improved knowledge				
	A lot	A little	Nothing	Not sure	*Bases*
	%	%	%	%	
Child development	66	24	7	3	*184*
Giving and receiving negative feedback	62	26	7	5	*186*
Communication, listening and conversation skills	60	27	8	5	*186*
Contraception	59	27	8	6	*182*
Managing anger	58	25	11	5	*185*
Sex and relationships	58	30	9	3	*182*
Sexually transmitted infections	57	34	6	3	*188*
Parenting and being a parent	49	34	11	6	*184*

The vast majority of teenagers in the intervention group felt that they knew more about these topics as a result of T&T. In particular, 66% of teenagers felt they knew a lot more about child development, 62% said they knew a lot more about giving and receiving negative feedback, and 60% said that they knew more about communication, listening and conversation skills.

The only significant difference between teenagers who had scored higher and lower on the youth development scale was that those who scored higher were more likely to say that they knew a lot more about giving and receiving negative feedback (68% compared with 48%; table not shown).

3.4.2 Attitudes and behaviours

When teenagers were asked whether T&T had changed their attitudes or behaviours:

- 79% said that it had taught them how to get advice about their health;
- 78% said that it had helped them to communicate;
- 77% felt better about themselves;
- 76% had made new friendships;
- 75% said it had made them want to delay parenthood;
- 68% said they got on better with people;
- 53% said it changed what they wanted to do when they finished school (see Table 3.4).

Table 3.4 Perceived effects of the programme, by youth development score						
Base: Teenagers randomised to intervention						
	Youth development score					
	Low		High		Total	
	%	*Bases*	%	*Bases*	%	*Bases*
Learnt how to get advice about my health	79	*92*	79	*95*	79	*189*
Communicate better with other people	78	*90*	79	*94*	78	*186*
Feels better about self	75	*92*	79	*94*	77	*188*
New friendships with other girls on programme	74	*92*	78	*95*	76	*189*
Want to delay parenthood	83	*92*	67	*94*	75	*188*
Get on with people better	70	*90*	65	*94*	68	*186*
Changed what wants to do when finish school	62	*91*	45	*94*	53	*187*

Table 3.4 shows a couple of differences between teenagers who scored higher and lower on the youth development scale. Those who scored lower were more likely to report that T&T had made them want to delay parenthood (83% compared with 67%). In addition, those who scored lower were more likely to report that T&T had changed what they want to do when they finish school (62% compared with 45%).

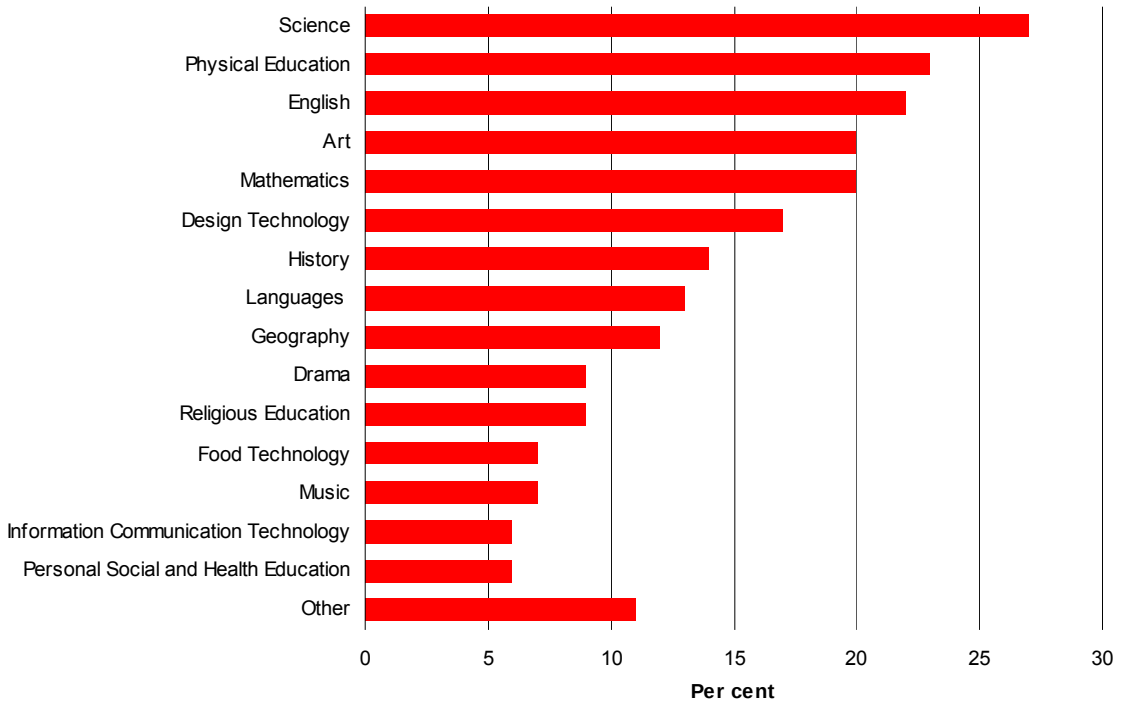
3.5 Potential disadvantages of the programme

As highlighted by Jessiman et al. (2012), one disadvantage of T&T is that teenagers who participate in the programme are required to miss one afternoon of school every week for 18 to 20 weeks.

Figure 3.5 shows which lessons were missed by the teenagers in the RCT's intervention group.

Figure 3.5 Lessons missed through doing T&T

Base: Teenagers randomised to the intervention



Source: Table C3.8 in Appendix C.

The proportion of teenagers missing particular lessons will to some extent reflect their prevalence in the school timetable. So, as might be expected, the lesson teenagers most frequently missed was science (27%). However, some schools specifically scheduled T&T so that it coincided with less academic subjects. This is illustrated by the fact that 23% of teenagers missed physical education in order to take part in T&T.

Since not all schools have a weekly timetable (e.g. some schools have a fortnightly timetable), teenagers participating in T&T do not necessarily miss the same lessons each week. However, teenagers' participation in T&T can still mean that they start to fall behind in the lessons that they miss. Indeed, of the teenagers in the RCT's intervention group, 31% felt that they had fallen behind with their schoolwork because of participating in T&T (see Table C3.9 in Appendix C). This did not differ between girls who had scored higher or lower on the youth development measure.

3.6 Other people and the programme

3.6.1 Other people's perceptions

As discussed in Chapter 1, the targeted nature of the intervention had some potential to create stigma regarding teenagers' participation in the programme. For that reason the additional questionnaire explored teenagers' perceptions of other people's responses to their involvement with T&T.

The findings showed limited evidence that there was any stigma associated with participation in T&T:

- 72% of teenagers reported that other people said nice things about the fact that they were doing T&T.
- 87% of teenagers felt that their parents were positive about their participation in the programme.
- 47% of teenagers said that other people were jealous that they were doing T&T.
- 9% of teenagers reported that they were teased as a result of doing T&T.

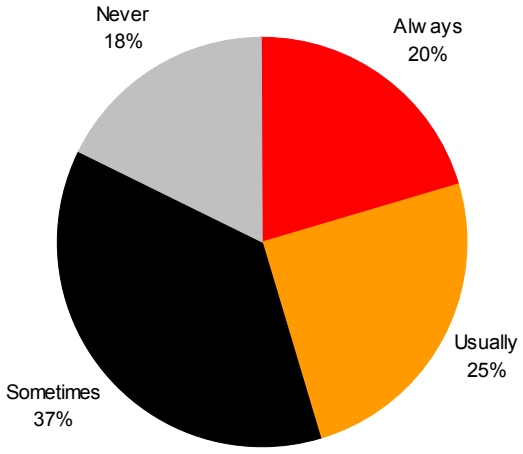
3.6.2 Discussions with other people

As part of the intervention, T&T participants are encouraged to share their experiences and learning with their friends. If such interactions took place and benefited girls who were in the control group then that might have led to contamination (see Chapter 4 for more detail). In order to be able to explore whether such contamination took place, the additional questionnaire collected information about whether teenagers talked to their peers about T&T.

Figure 3.6 shows how often teenagers talked to their peers who did not do T&T about what they did or learnt on the programme. Twenty per cent of teenagers said that they always talked about T&T to their peers, 25% that they usually did so, 37% sometimes did so, and 18% never talked about T&T to their peers.

Figure 3.6 How often teenagers talked to their peers about T&T

Base: Teenagers randomised to the intervention



Source: Table C3.10 in Appendix C.

Teenagers who scored higher on the youth development scale at baseline were particularly likely to talk about T&T to their peers (26% always talked about T&T compared with 12% of teenagers who scored lower on this scale).

The topics teenagers talked about to their peers are shown in Table 3.5. Sex and relationships was the most commonly discussed topic (45%), followed by parenting and being a parent (40%) and child development (38%).

Table 3.5 Topics teenagers talked about	
Base: Teenagers randomised to intervention	
Sex and relationships	45
Parenting and being a parent	40
Child development	38
Contraception	34
Communication, listening and conversation skills	31
Sexually transmitted infections	31
Managing anger	28
Giving and receiving negative feedback	22
Other	10
Nothing	18
Base	*186*

There was only one difference between girls who scored higher and lower on the youth development scale: those with higher scores were more likely to have talked to their peers about communication, listening and conversation skills (38% compared with 22%; table not shown).

To sum up, these results illustrate that there was a possibility of some contamination in the trial. The extent to which the impact findings may have been affected by such contamination (e.g. the extent to which the trial may have missed real intervention effects as a result of contamination) is discussed in Chapter 4.

3.7 Summary

This chapter explored T&T participants' views of the programme and their perceptions of its impact. The vast majority of teenagers enjoyed taking part in T&T (94% of those randomised to the intervention group). They particularly enjoyed the toddler time but were less keen on writing their journals. One-fifth of teenagers said that they found T&T difficult or challenging, with similar numbers specifically reporting that toddler time, group work, writing a journal or receiving counselling was often difficult or challenging for them.

The teenagers participating in the programme expressed overwhelmingly positive feelings about it. Sixty-five per cent of girls in the intervention group reported that participating in T&T made them feel responsible, 64% that it made them happy, 61% were more positive about things and 59% were positively interested in the programme. Substantially fewer

teenagers reported negative emotions arising from their participation in T&T (e.g. 37% said that they had sometimes felt bored, 20% had been irritated or annoyed, 7% anxious, and 7% angry). These positive findings were reinforced by others showing that 91% felt proud of doing the T&T programme, while only 7% said that their participation made them feel embarrassed

The majority of teenagers also felt they knew more about the different areas of the T&T curriculum as a result of the programme. In particular, 66% of teenagers in the intervention group felt they knew a lot more about child development, 62% said they knew a lot more about giving and receiving negative feedback, and 60% said that they knew more about communication, listening and conversation skills.

Teenagers also felt that T&T had changed their attitudes or behaviours: 78% of girls in the intervention group felt that it had helped them to communicate, 77% subsequently felt better about themselves, 76% had made new friendships, 75% said it had made them want to delay parenthood, and 53% said it had changed what they wanted to do when they finished school.

The programme requires participants to miss one afternoon of school every week for 18 to 20 weeks. The lessons teenagers most commonly missed were science (27%) and physical education (23%). Some reported that they had fallen behind with their school work as a result of attending the programme (31%).

There was limited evidence of stigma associated with T&T. Seventy-two per cent of teenagers in the intervention group reported that other people ‘said nice things’ about the fact that they were doing T&T, 87% felt that their parents were positive about their involvement with the programme, and 47% said that other people were jealous that they were doing T&T. However, 9% of teenagers in the intervention group reported that they were teased as a result of T&T.

T&T participants are encouraged to share their experiences and learning with their friends. The majority of teenagers in the intervention group talked to their peers who did not do T&T about what they did or learnt on the programme, with only 18% saying that they never talked about T&T. When teenagers did talk about T&T, the topics they most commonly talked about were sex and relationships (45%), parenting and being a parent (40%) and child development (38%).

4 Impact analysis

4.1 Introduction

This chapter examines the effects of the T&T intervention by comparing young women randomised to be offered the intervention with those randomised to the control group. The comparisons are made on the basis of a broad range of primary and secondary outcomes, reflecting the potentially broad benefits of participation in the T&T programme. The measures were all selected at the design stage in consultation with the study's steering group and informed by the logic model elucidated in the formative research. Data for all participants in the trial were collected by questionnaire at three points in time: prior to allocation (baseline), immediately post-intervention (follow-up one) and a year after the intervention (follow-up two) (for details, see Appendix A). While the primary focus of the analysis is on medium-term outcomes (that is, those measured at follow-up two), the chapter also examines short-term outcomes (at follow-up one). Both sets of findings are discussed together, and in the context of T&T facilitators', teachers' and teenagers' perceptions about the outcomes of the programme.

4.2 How impact is measured

The impact of the intervention is assessed using a broad range of primary and secondary outcomes selected to reflect the potentially broad benefits of participation in the T&T programme. The number of outcome measures selected was purposively kept low to reduce the possibility of false-positive results arising from multiple tests of statistical significance. The measures were all selected at the design stage in consultation with the steering group. All primary and a number of secondary outcomes were agreed in the research protocol at the beginning of the RCT, while the remaining secondary outcomes were agreed with the DfE and the study's steering group at a later stage but prior to the collection of final outcome data. The trial protocol was registered online (see Appendix A). Most measures had been previously used in other studies. See Figure 4.1 for a full list of outcomes used in the study and their origins, and Appendix A for details of how these measures were constructed from the original survey questions.

Figure 4.1 Primary and secondary outcomes	
Primary outcomes	***Secondary outcomes***
(1) No contraception use at last sex in last three months (Ripple trial) (2) >1 episodes of no contraception in the last three months (Ripple trial)[10] (3) Expectation of teenage parenthood (Ripple trial) (4) Youth development score (new measure informed by selected items from youth at risk version of the Life Effectiveness Questionnaire adapted to UK (Neill et al. 1997))	(1) Self-reported teenage pregnancy since baseline (Ripple trial) (2) No condom use at last sex in last three months (Ripple trial) (3) >1 episodes of no condom use in last three months (Ripple trial) (4) Low sexual health knowledge (Ripple trial) (5) Best age to have sex for the first time <16 (Ripple trial) (6) Unprotected sex regarded as okay (Ripple trial) (7) Hard to talk about sex with boyfriend (Ripple trial) (8) Hard to talk about contraception in clinic or with doctor (Ripple trial) (9) Lack of awareness of impact of parenthood on social life (new) (10) Low self-esteem (don't positively like self) (Ripple trial) (11) Low emotional self-reflection (not in touch with own feelings) (new) (12) Low emotional vocabulary (can't find words to say how feeling) (new) (13) Dislike of school (Ripple trial) (14) Lack of expectation of post-16 education, training or employment (Youth Cohort Study) (15) Number of school days missed (new)[11]

The analysis has been conducted on an intention-to-treat basis. This means that all teenagers who were originally randomised to the intervention or control group as part of the trial were included in the analysis regardless of how many sessions of the T&T programme they attended in total. This approach is recommended in the Consolidated Standards of Reporting Trials (CONSORT) because individuals who drop out of interventions are generally (and in the T&T RCT in particular) different in profile, according to baseline data, from those who remain (CONSORT Statement 2010). This means that those individuals who remain with the programme are very different from those in the control group (since there is no similar process of drop-out in this group). Thus a comparison of outcome data from all those who were allocated to a programme versus all those who were allocated to the control group (i.e. an 'intention-to-treat' analysis) is fairer than a comparison of outcome data from those who complete the programme versus

[10] This outcome is available at follow-up two only, as it was mistakenly omitted from the follow-up one questionnaire.

[11] This measure is available for follow-up two only as it covers a half-term period post intervention. The data came from the survey of teachers. Due to an error, this outcome was omitted from the online registration of the trial protocol.

those in the control group (an 'in treatment' analysis). However, the disadvantage of an intention-to-treat approach is that the size of an intervention's impact is diluted by the inclusion of girls who were randomised to the intervention group but dropped out of the programme (for more detail about the intention-to-treat approach, see Appendix A). We take account of this methodological limitation when discussing our findings regarding the impact of the programme.

Outcome data were collected at two points in time (in addition to baseline measurements): (1) at the end of the programme, and (2) one year later. Our analyses are focused primarily on outcomes at follow-up two. However, we also present findings for follow-up one and discuss which impacts were sustained from follow-up one to follow-up two and which dissipated.

It is important to note that, due to its time scale, the study did not aim to measure any longer-term outcomes for teenagers, such as their education, training or employment situation a few years after the intervention or whether they will have been pregnant before age 20.

The data analysis was carried out using logistic regression models for dichotomous outcomes and linear regression models for continuous outcomes. Only those individuals for whom we had complete data on all relevant variables were included in the analysis (on an outcome by outcome basis). We did not carry out any imputations for missing values.

As discussed in Chapter 2, while on the whole the intervention and control groups were well-balanced at baseline, there were a number of characteristics with respect to which these two groups were different (namely, three variables for follow-up one and four variables for follow-up two; see Figure 2.1 in Chapter 2).[12] In order to prevent these baseline differences from confounding the impact analysis, we undertook the following actions. For each of these baseline variables, we checked whether they were associated with each outcome (separately at follow-up one and follow-up two) in bivariate relationships.[13] If they were, then we adjusted for these baseline variables when modelling the effects of the intervention on these outcomes.

For example, in Chapter 2, we found that, at baseline, teenagers in the intervention group were more likely to worry a lot than those in the control group. So, in the analysis of impacts at follow-up two, we examined whether worrying a lot at baseline was associated with any of our outcomes. We found that there were associations present for six outcomes (namely, youth development score, no condom use at last sex, more than one episode of no condom use in the last three months, low emotional vocabulary, low self-esteem and low sexual health knowledge, see Table C4.4 in Appendix C). Therefore, worrying a lot at baseline was included as a covariate in regression models for these outcomes – but not for any other outcomes – at follow-up two. The details of which covariates were used in

[12] In this analysis of baseline differences, all those with $p<0.1$ were considered to be significant.
[13] In these bivariate analyses, all associations with $p<0.1$ were considered to be significant.

models for which outcomes at follow-up one are in Table C4.2 and similar details for follow-up two are in Table C4.4 in Appendix C.

The p values in tables with outcome data in this chapter are from logistic regression or linear regression models that included relevant confounders. A p value under 0.05 means that the difference in percentages (or in mean values) between the intervention and the control group is significant at the 5% significance level. The adjusted percentages for the control group in these tables have been calculated from odds ratios or linear regression coefficients in these models.[14]

4.3 Impact of the Teens and Toddlers intervention on teenagers in the intervention group

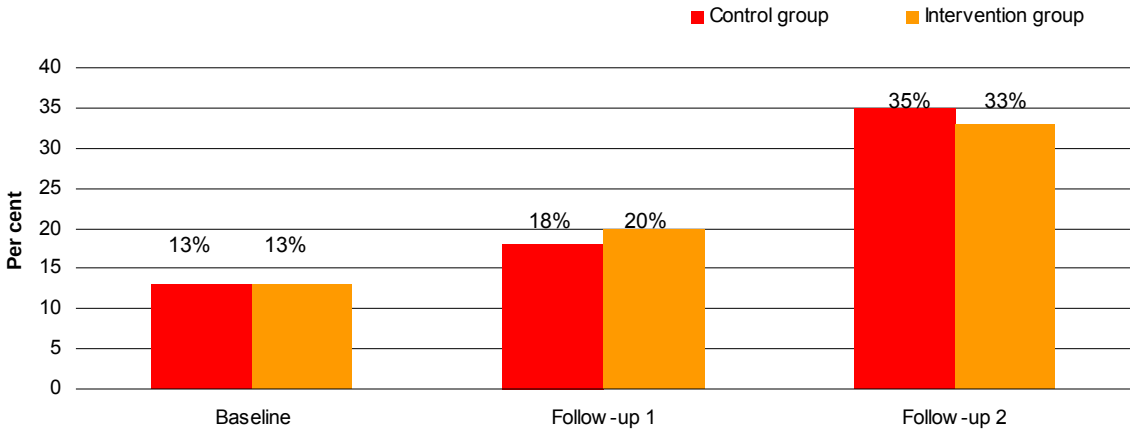
This section examines whether the T&T intervention was effective in influencing a number of outcomes for teenagers it targeted. The outcomes examined were measured at the end of the programme (follow-up one) and again one year later (follow-up two).

4.3.1 Context: experience of heterosexual sex

Before examining the findings for each outcome in detail, it is important to note that the denominator for all percentages discussed in the chapter is the overall sample, and this includes sexual-related outcomes (e.g. the proportion of girls who did not use a condom at last sex) even though not all girls in the trial were sexually active. This approach eliminates the risk of bias in impact findings, which could result from differences in rates of sexual activity among the intervention and control groups. However, the data presented in Figure 4.2 show that at all three points in time (that is, at baseline, follow-up one and follow-up two), the reported rates of heterosexual sex were similar in the intervention and control groups, and increased at a similar rate in both groups as the teenagers grew older.

[14] Adjusted percentages for the control group are calculated from the percentage for the intervention group in combination with the group membership odds ratio taken from the logistic regression. That is, the adjusted *odds* for the control group is calculated as $q = \frac{p}{1-p} * odds\,ratio$ where p is the proportion in the intervention group with a particular outcome. The adjusted *percentage* for the control group is then calculated as $\left(\frac{q}{1+q}\right) \times 100\%$. Adjusted means for the control group are calculated simply as the difference between the mean for the intervention group and the regression coefficient for the group membership binary variable.

Figure 4.2 Proportion of teenagers who had had sex with a boy (or man) at each wave, by trial arm



Source: Table C4.5

Note: Percentages for the control group have been adjusted for follow-up one and follow-up two for relevant baseline differences (see the source table for details).

4.3.2 Impact at follow-up one

Focusing first on findings at follow-up one (see Table 4.1 in this section and Table C4.1 in Appendix C), the analysis did not find evidence of the intervention's benefits with regard to any of the primary outcomes or most of the secondary outcomes. However, there was evidence of a positive impact for three out of 14 secondary outcomes. Specifically, teenagers in the intervention group were less likely to have low self-esteem (16%, compared with 28% in the control group)[15] or low sexual health knowledge (73%, compared with 83%), and were also less likely to report difficulty in discussing the (contraceptive) pill with a doctor or in a clinic (44%, compared with 56%).

[15] The quoted percentages that relate to impact have been adjusted for relevant baseline differences. (As the approach requires adjustments for only one arm of the trial, we have adjusted percentages for the control group, while those for the intervention group are from the unadjusted analysis.)

Table 4.1 Primary and secondary outcomes at follow-up one, by trial arm				
Base: All at follow-up one				
Outcomes	**Control group** Adjusted %	**Interv. group** %	**Percent. point difference** %	**P value for difference**
Primary outcomes				
No contraception at last sex in last 3 months	0	2	-2	0.132
Expects teenage pregnancy	27	30	-3	0.516
Youth development score	Mean=47.1	Mean=47.6	Difference between means -0.4	0.606
Secondary outcomes				
No condom use at last sex in last 3 months	4	4	-1	0.759
>1 episodes of no condom use in last 3 months	5	7	-2	0.385
Best age to have first sex under 16	18	19	-1	0.741
Favourable to not using protection sometimes in sex	8	10	-2	0.529
Low self-reflection	14	15	-1	0.843
Low emotional vocabulary	13	19	-6	0.098
Low self-esteem	28	16	11	**0.010**
Dislike of school	44	41	3	0.536
Lack of expectation of post-16 education, training or employment	1	3	-1	0.302
Low sexual health knowledge	83	73	10	**0.017**
Difficulty discussing sex with a boyfriend	39	43	-4	0.465
Difficulty discussing pill with a doctor	56	44	12	**0.016**
New pregnancy since baseline	2	2	0	0.973
Lack of awareness of impact of parenthood on social life	17	15	2	0.614
Bases	*215*	*210*		

Notes: The base for all questions is the whole sample including for sexually-related outcomes. Percentages for the comparison group have been adjusted, where necessary, for relevant baseline differences between the two arms; see Table C4.2 in Appendix C for details of which covariates were included in models for each outcome. P values refer to tests carried out using logistic regression models for dichotomous outcomes and linear regression models for continuous outcomes. For more complete detail about the tests and statistics for this analysis, see Table C4.1 in Appendix C.

The finding that the intervention had a positive effect with respect to the three outcomes of sexual health knowledge, low self-esteem and the ability to discuss contraception with a doctor is consistent with the perceptions of the teenagers in the intervention group and the T&T facilitators and teachers. As discussed in Chapter 3, most teenagers in the intervention group felt that their knowledge of contraception and of sexually transmitted

infections improved as a result of participating in the programme, many reported improvements to their communication skills, and many said they now felt better about themselves. The findings from the process evaluation also suggest that the intervention was perceived by facilitators, teachers and participants as leading to improvements in self-esteem, self-confidence and interpersonal skills (Jessiman et al. 2012).

4.3.3 Impact at follow-up two

This section examines whether the intervention was effective in influencing a number of outcomes for teenagers as measured at follow-up two, which was the primary focus of this study. In addition, we relate findings at follow-up two to those at follow-up one and discuss which of the follow-up one impacts were sustained by follow-up two and which dissipated.

Overview of main findings

Table 4.2 presents an overview of the results of the analysis of follow-up two data.

Table 4.2 Primary and secondary outcomes at follow-up two, by trial arm				
Base: All at follow-up two				
Outcomes	**Control group** Adjusted %	**Interv. group** %	**Percent. point difference** %	**P value for difference**
Primary outcomes				
No contraception at last sex in last 3 months	4	6	-1	0.606
>1 episodes of no contraception in last 3 months	12	13	-1	0.779
Expects teenage pregnancy	25	25	1	0.885
Youth development score	Mean=48.7	Mean=47.3	Difference between means 1.4	0.111
Secondary outcomes				
No condom use at last sex in last 3 months	11	12	-1	0.804
>1 episodes of no condom use in last 3 months	15	16	-1	0.786
Best age to have first sex under 16	18	16	1	0.717
Favourable to not using protection sometimes in sex	10	9	2	0.527
Low self-reflection	10	15	-5	0.169
Low emotional vocabulary	17	18	-1	0.806
Low self-esteem	25	15	9	**0.032**
Dislike of school	47	46	1	0.881
Lack of expectation of post-16 education, training or employment	1	1	0	0.976
Low sexual health knowledge	71	70	1	0.843
Difficulty discussing sex with a boyfriend	34	32	1	0.775
Difficulty discussing the pill with a doctor	42	44	-2	0.696
New pregnancy since baseline	6	4	1	0.548
Lack of awareness of impact of parenthood on social life	13	10	3	0.321
Number of school days missed	Mean=2.4	Mean=2.2	Difference between means 0.2	0.647
Bases	*207*	*201*		

Notes: The base for all questions is the whole sample including for sexually-related outcomes. Percentages for the comparison group have been adjusted, where necessary, for relevant baseline differences between the two arms; see Table C4.4 in Appendix C for details of which covariates were included in models for each outcome. P values refer to tests carried out using logistic regression models for dichotomous outcomes and linear regression models for continuous outcomes. For more complete detail about the tests and statistics for this analysis, see Table C4.3 in Appendix C.

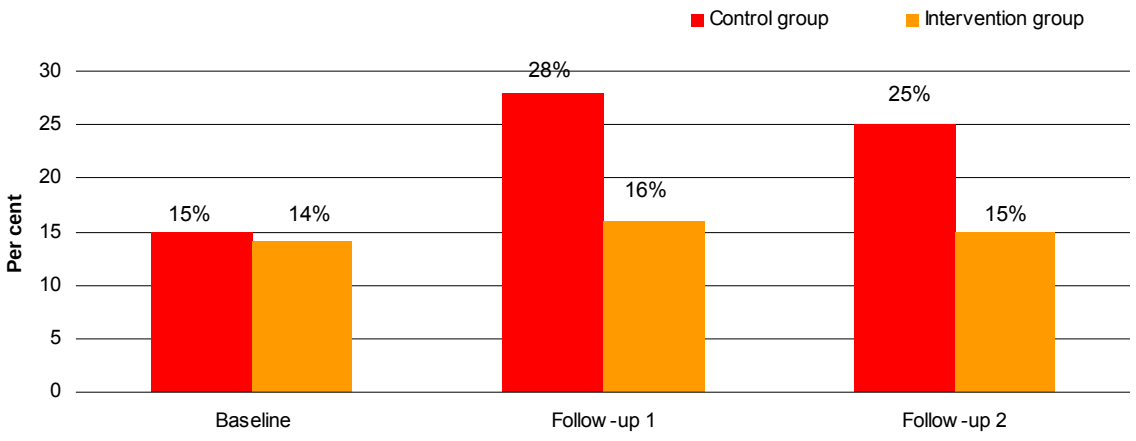
The data shows that, one year after completion of the programme, there were still differences between the intervention and control groups with regard to low self-esteem (see Table 4.2). However, there was no evidence that the intervention was effective in influencing any other outcomes for teenagers in the intervention group. The pattern of differences in prevalence of different outcomes among the intervention and control groups is such that some of these would suggest potentially positive impacts and about the same number potentially negative impacts.

Impact on self-esteem

As noted above, the impact analysis found that, at follow-up two, the intervention group reported lower rates of low self-esteem than the control group. Among teenagers randomised to the intervention, 15% had low self-esteem at follow-up two, while among those randomised to the control group the comparable figure was 25% (see Table 4.2).

Figure 4.3 illustrates this difference between the two trial arms at both follow-ups, as well as the baseline measurements for the prevalence of low self-esteem. A noteworthy aspect of the graph is that it shows that while, among the intervention group, the prevalence of low self-esteem remained more or less stable (at around 14-16%), there was a substantial increase in the proportion of teenagers with low self-esteem among the control group, from 15% at baseline to 28% and 25% at the respective follow-ups.

Figure 4.3 Proportion of teenagers who had low self-esteem at baseline and follow-ups, by trial arm



Source: Table 2.2, Table 4.1, and Table 4.2.
Note: Percentages for the control group have been adjusted for follow-up one and follow-up two for relevant baseline differences (see the source tables for details).

The measure of low self-esteem used in this study was based on a five-item scale. The distribution of answers to the self-esteem question on that original scale is presented in Table C4.6 in Appendix C (the percentages in that table are not adjusted for baseline differences unlike in Figure 4.3). The table shows that while there was a clear difference in the proportions of teenagers with low self-esteem in the intervention and control groups (that is, those who disagreed or disagreed strongly with the statement “I like myself”),

there was no mirroring difference in the proportions of teenagers with high self-esteem (those who agreed or agreed strongly with this statement – 51% in both groups at follow-up two).

We therefore carried out some additional analysis exploring whether relationships between low self-esteem and other variables in our data were consistent with evidence from other studies. This analysis revealed that, as expected, in the T&T RCT low self-esteem was strongly negatively associated with youth development score,[16] and also that low self-esteem at baseline was predictive of experience of heterosexual sex by follow-up two.[17]

To sum up, T&T did have a positive effect on self-esteem at both follow-up one and follow-up two, and this is consistent with the perceived impacts of the intervention as discussed in Chapter 3 of this report and in the integrated process evaluation report (Jessiman et al. 2012). Evidence from other studies suggests that self-esteem may be protective against early sexual activity among girls and teenage pregnancy (Emler 2002; Spencer et al. 2002). Therefore improving teenagers' self-esteem may make an important contribution towards achieving the T&T intervention's objective to reduce teenage pregnancy.

Other follow-up one impacts at follow-up two

Apart from low self-esteem, there were two other outcomes where we found the intervention to be effective at follow-up one: low sexual health knowledge and difficulty of discussing the pill in a clinic or with a doctor. Analysis of follow-up two data reveals that the impacts on these two outcomes dissipated by follow-up two, and the two trial arms were no longer different in terms of these characteristics. With regard to sexual health knowledge, at follow-up two, 70% of teenagers in the intervention group and 71% of those in the control group had low levels of knowledge. And with regard to the difficulty of discussing the pill with a doctor, at follow-up two, 44% of teenagers in the intervention group and 42% of those in the control group reported this (see Table 4.2).

Youth development

Youth development scores were somewhat lower among teenagers in the intervention group than among those in the control group. Before the adjustments for baseline differences, this difference was significant at the 5% level; however, the difference between the adjusted mean values (47.3 for the intervention group and 48.7 for the control group) is smaller and does not reach the normal threshold of 0.05 for statistical significance (p=0.111, see Table 4.2).

[16] At all three data collection points, there was a strong negative association between low self-esteem and youth development score. For example, at baseline, the mean score for those with low self-esteem was 43.8, and for the rest it was 50.0. The difference between these two figures was significant at p<0.001.

As it is somewhat surprising to find a positive impact of the intervention on self-esteem and at the same time an apparently negative impact in unadjusted analysis on youth development, we undertook some further analysis of the youth development measure. Table C4.7 in Appendix C shows values for each component of the youth development score among the intervention and control groups.[18] It appears that while the values for individual components of the score appear consistently lower among the intervention group than among the control group, there is only one component where the difference between the trial arms is statistically significant. This component relates to peaceful conflict resolution rather than to those components that are relevant to self-esteem. In both unadjusted and adjusted analyses, teenagers in the intervention group were less likely than those in the control group to report that they tended to sort out conflicts with other people peacefully.[19] However this finding should not be over-interpreted because this analysis was purely exploratory as explained above and conflict resolution was neither a primary nor a secondary outcome of the programme.

4.3.4 Summary of findings

To sum up, analysis revealed that the T&T intervention may have had a positive impact on teenagers' self-esteem at both follow-ups, and some short-term effects on their knowledge of sexual health and ability to discuss contraception in a clinical setting. However, these two short-term impacts appear to have dissipated within a year following the end of the programme, and no other medium-term impacts were found (see Figure 4.4).

Figure 4.4 Summary of findings in relation to impacts of the T&T intervention

Outcomes where a significant impact* was found at follow-up one or follow-up two	Follow-up one	Follow-up two
Low self-esteem	Positive impact	Positive impact
Low sexual health knowledge	Positive impact	No impact
Difficulty of discussing the pill with a doctor	Positive impact	No impact

* at p<0.05

[17] Of those teenagers who had low self-esteem at baseline, 58% had had sex with a boy (or man) by follow-up two. Of those who did not have low self-esteem at baseline, the comparable figure was 31%. The difference between these percentages was significant at p<0.001.
[18] The data in Table C4.7 relates to unadjusted analyses.
[19] Adjusted results for the components of the youth development score are not shown. For the component "I sort out my conflicts with other people peacefully", the adjusted mean for the control group was 5.5 (the same as the unadjusted mean), and the coefficient from the linear regression model was -0.5 (CI: -0.9 - -0.1), p=0.025.

4.3.5 Some illustrations of plausible effect sizes for those who completed the intervention

As not all teenagers randomised to intervention completed the T&T programme (only 73% did so), it is useful to know what the impacts of the intervention were on those who completed the programme. We explored the feasibility of a formal matched comparison 'on treatment' analysis but concluded that it could not be done reliably because of the presence of a very severe self-selection bias (i.e. teenagers staying on the programme were at baseline much less likely to have engaged in risk-taking behaviours and to report risk-taking attitudes than those who went on to drop out of the programme, as discussed in Chapter 2). This difference could not be adjusted for adequately when comparing the 'completers' in the intervention group with the comparison group. In the absence of a matched comparison analysis we have included some arithmetic illustrations of what maximum impacts on the 'treated' might be if we assume that the impact of the intervention on those who dropped out was zero.

The first illustration relates to low self-esteem, where we found a rate of 25% in the control group and 15% in the whole intervention group (i.e. an impact of nine percentage points).[20] Given the sample size of 201 in the intervention group for this outcome, this suggests that around 18 instances of low self-esteem might have been prevented by the intervention. On the assumption that all of these 18 must have occurred among those who completed the programme rather among those who dropped out (in keeping with the assumption of no impact on those who dropped out), it follows that 18 out of 147 'completers'[21] probably had better self-esteem because of the intervention. This represents a maximum 'on treatment' impact of around 12 percentage points (as compared with the nine percentage points impact on the overall intervention group).

Our second illustration relates to girls' lack of awareness of the potential impact of parenthood on social life, where we found a rate of 13% in the control group and 10% in the whole intervention group (i.e. a positive impact of three percentage points, which was not statistically significant). Given the sample size of 200 in the intervention group for this outcome, this suggests that six instances of lacking awareness might have been avoided among the intervention group. On the assumption that all of these instances occurred among those completing the programme, it follows that six out of 146 'completers' probably had better awareness of the impact of parenthood on social life because of the intervention. This represents a maximum 'on treatment impact' of around four percentage points (as compared with the three percentages points impact on the overall intervention group).

The next section discusses methodological issues which may have affected the impact results and appraises their significance.

[20] It is nine and not ten percentage points due to rounding of the prevalence figures for the two trial arms (24.5% and 15.4%).
[21] 147 represent 73% of 201 (the completion rate of the programme was 73%).

4.4 Methodological issues

This section discusses a number of methodological limitations that may have affected the study's ability to detect possible impacts of the programme and appraises their significance.

4.4.1 Intention-to-treat analytical approach and drop-outs

One limitation relates to the intention-to-treat analytical approach. While being the only robust method of analysing RCT data (see Appendix A), this approach does have a substantial limitation relating to the fact that all teenagers randomised to receive an intervention are treated as the intervention group, regardless of whether they completed the programme or even attended any sessions at all. As discussed in Chapter 2, in our study just under three-quarters (73%) of young women randomised to receive the intervention completed the programme. As such, one possible concern might be that including those who dropped out of the programme in the analysis (27% of all those randomised to the intervention) would have diluted the impacts.

However, there is little evidence in the impact findings that dilution of positive impacts occurred. There is no overall pattern of statistically non-significant benefits to suggest that the intervention brought about a range of benefits which fall just below the level of statistical significance because of a dilution effect. What we find instead is that for about half the outcomes at follow-up two, the differences between the two groups would suggest potentially positive impacts, while for the other half they would suggest potentially negative impacts. Where the differences are in the direction of positive impacts, it is possible that without dilution some of these differences would have been larger and reached statistically significance. However, where the differences in the intention-to-treat analysis are in the direction of negative impacts, the rate of attrition is simply not high enough for dilution to cause any putative real effects to become obscured.

It is also worth noting that while the drop-out rate from another youth-development intervention – the CAS Carrera programme as delivered in the New York City trial – was 21% per cent (that is, not much lower than the 27% rate in this study), the intention-to-treat analysis did reveal a number of positive impacts of that intervention including on teenage pregnancy (Philliber et al. 2002).

4.4.2 Contamination

Another possible methodological limitation of the study relates to the potential for contamination of the findings, which is associated with the individual-level randomisation of girls to the intervention or control group within each participating school. Although peer education is not a direct aim of the T&T intervention, T&T participants are encouraged to share their experiences and learning with their friends. As discussed in Chapter 3, most teenagers randomised to receive the intervention did indeed discuss the programme with their peers who were not involved with the programme (20% always talked about the programme to their peers, 25% talked about it usually, and 37% sometimes, see Figure 3.6 in Chapter 3), although we do not know whether these included girls in the control

group. It is possible that these interactions benefited teenagers who were in the control group and thus led to a certain amount of 'contamination'. It is worth noting that none of the specific topics covered by the programme were discussed by more than 45% of teenagers (see Table 3.5 in Chapter 3). Finally, it should be noted that while contamination can result in reduced power to identify intervention effects it does not otherwise bias the analysis.

Although it is possible that some contamination occurred within this study, we do not believe that such effects will have unduly affected the results. As with the above, if we had seen results that consistently tended in the direction of intervention benefit but did not reach significance, then we might conclude from this that contamination could have resulted in our analysis having insufficient power to detect real intervention effects. However, this was not what we found. Furthermore, while we might expect contamination to affect outcomes such as sexual health knowledge in which some (though probably not all or even most) benefits might plausibly be passed from intervention to control participants, we would not expect contamination to affect outcomes such as self-esteem or youth development (because these benefits are not easily transmittable between individuals) or sexual behaviour (because only some of the determinants of sexual behaviour such as knowledge are transmittable in this way whereas others such as self-esteem are not).

Finally, it is worth pointing out that other trials such as the New York City trial of the CAS Carrera programme also allocated *individuals*, rather than *clusters* of individuals, to intervention or control groups but nonetheless did find significant effects including on teenage pregnancy (Philliber at al. 2002).

4.4.3 Disclosure

The issue of whether study participants were disclosing socially undesirable behaviours and attitudes when completing questionnaires, and whether teenagers in the intervention and control groups were equally likely to disclose or not disclose sensitive information, are important methodological concerns to address when trying to understand our findings with regard to the effectiveness of the T&T intervention.

When conducting social research on topics as sensitive as sexual experiences and knowledge it will always be challenging to obtain disclosure from participants (Wellings et al. 1990), perhaps especially from the sorts of 'at-risk' young people taking part in this study. Within this evaluation we used standard trial methods to ensure confidentiality such as instructing that questionnaires should be completed under exam conditions with teachers unable to read responses, and assuring participants that data would be anonymised and not passed on to teachers or parents. However, it is still possible that participants in the research under-reported sensitive behaviours.

It is important to note here that if teenagers *as a whole* had under-reported sensitive behaviours this would have led to a reduced power to identify intervention effects but

would not otherwise bias the analysis.[22] However, if teenagers *randomised to the intervention* were less or more likely to report risky behaviours and attitudes than teenagers in the control group, this would have biased the findings in the direction of over- or under-estimating any benefits of the intervention.

Participants in the T&T programme are actively encouraged to be honest about their behaviours and attitudes. If teenagers in both the T&T and the control groups were under-reporting certain behaviours at the baseline but then those in the T&T group provided an honest picture post-intervention, this may have limited the study's potential to identify the intervention's positive effects. Another possibility is related to social desirability bias, which generates an over-reporting of benefits among intervention participants because they want to report what they perceive that the intervention providers and evaluators want to hear (e.g. Stuart and Grimes 2009).

4.4.4 Targeting the 'right' teenagers

According to its logic model, the success of the T&T intervention relies on its ability to target teenagers who are sufficiently at risk of teenage pregnancy. If the teenagers who were recruited to the study were not sufficiently at risk, this might have had a negative impact on the study's ability to detect the intervention's benefits.

It is possible that because of the requirement of the RCT to recruit about 16 teenagers to the study in each school,[23] the recruitment might have been broadened compared to how it would have been outside a trial. Also, it is possible that some teenagers who would normally have taken part in T&T might have chosen not to participate through reluctance to enter the trial. The integral process evaluation found that there was some variance in how teenagers were selected for participation in the programme, and those selecting them did not always follow the checklist provided by T&T. However, the T&T facilitators who were interviewed for the process evaluation thought that the 'right' girls were being referred to the programme (Jessiman et al. 2012).

In addition, a comparison between the T&T RCT and the evaluation of the YPDP (Wiggins et al. 2009) revealed that the prevalence of various sexual behaviours was similar in these two studies, which suggests that the T&T RCT was successful in recruiting teenagers who were sufficiently at risk of teenage pregnancy.

---

[22] As an example, if in reality 10% of an intervention group and 20% of a control group exhibit a particular behaviour which is socially undesirable the difference is ten percentage points. If in practice only half the people who exhibit that behaviour disclose this in a questionnaire, then the research would suggest that 5% of the intervention group and 10% of the control group exhibit that behaviour and the difference would be five percentage points. The smaller the difference between groups the bigger the sample needed to identify this difference as statistically significant (which is why under-reporting of sensitive behaviours would lead to reduced power to identify intervention effects). However, in both scenarios (reality and one where sensitive behaviour was under-reported) the control group are twice as likely to exhibit this risky behaviour as the intervention group, which illustrates that the analysis would not be biased.

[23] Six for the intervention group, six for the control group, and two reserves for each of the trial arms (see Appendix A for more detail).

4.4.5 Concerns around some outcome measures

Where possible, the study used validated outcome measures to assess the impact of the T&T intervention, and most of the measures used had been used on other studies. However, there are some methodological concerns relating to the measures of youth development and self-esteem.

Neither of these two measures had been validated in previous studies. In the case of youth development, this was because no existing validated score existed. For that reason we took an existing set of questions (Neill et al. 1997), which included but went beyond our notion of youth development, and used the most pertinent items (see Appendix A for information on the internal consistency of that scale). The lack of validation of this measure means that there is a risk that it may be less sensitive to change than a comprehensive validated measure would be.

We used a single-item measure of self-esteem that had previously been used in the Ripple study (Stephenson et al. 2004) instead of a more comprehensive validated measure such as the Rosenberg multi-item scale because self-esteem was not originally on our list of secondary outcomes. It was added to the list of outcomes in consultation with the DfE prior to our collection and analysis of follow-up data but after the questionnaires were finalised.

As discussed earlier in this chapter (see subsection “Impact on self-esteem” in section 4.3.3), there are some concerns about how well this single-item measure of low self-esteem performed in this study. However, some additional tests of the validity of this measure that we carried out, as well the fact that this impact finding was present at both follow-ups and was consistent with findings from the qualitative process evaluation (Jessiman et al. 2012) and from teenagers’ answers to the quantitative additional questionnaire about the evaluation at follow-up one (see Chapter 3 in this report), suggest that the impact of the T&T intervention on girls’ self-esteem was probably genuine rather than an artefact of using an unvalidated measure.

4.5 Summary

This chapter examined whether the T&T intervention was effective in improving a number of outcomes for the teenagers it targeted. The outcome data were collected at two points in time – at the end of the programme (follow-up one) and one year later (follow-up two) – to enable analysis of short- and medium-term impacts.

The impact analysis has been conducted on an intention-to-treat basis. This means that all teenagers who were originally randomised to the intervention or control group as part of the trial were included in the analysis regardless of how many sessions of the T&T programme they attended in total. This approach is recommended in the Consolidated Standards of Reporting Trials (CONSORT) as the preferred analysis strategy because it preserves the huge strengths of randomisation and allows strong inferences about cause

and effect that are not justified with other approaches to analysis (CONSORT Statement 2010).

The analysis showed that at follow-up one, there was no evidence that the intervention had been effective in changing any of the three primary outcomes. However, there was evidence of a positive impact of the programme on three of the 14 secondary outcomes. Teenagers in the intervention group were less likely to have low self-esteem (16%, compared with 28% in the control group), less likely to have low sexual health knowledge (73%, compared with 83%) and less likely to report difficulty with discussing the pill with a doctor or in a clinic (44%, compared with 56%). These positive findings are consistent with the perceptions of teenagers in the intervention group and of T&T facilitators and teachers regarding the effects of the intervention, as discussed in Chapter 3 and in the report on the integral process evaluation (Jessiman et al. 2012).

At follow-up two, there was no evidence that the programme had been effective in changing any of the four primary outcomes. However, the positive impact of the intervention on self-esteem was sustained one year later. While low self-esteem was reported by 15% of teenagers randomised to the intervention arm, the comparable figure for the control group was 25%. However, positive impacts on knowledge of sexual health and on teenagers' ability to discuss contraception with a doctor, which were present at follow-up one, dissipated within a year following the end of the programme. No new impacts of the intervention were detected at follow-up two.

There are a number of methodological issues that may have affected the study's ability to detect the true impacts of the programme. The chapter discussed the following concerns:

- The intention-to-treat analytical approach and the effect of drop-outs from the programme on its ability to detect the positive impacts of the intervention.
- The potential for contamination, where some benefits of the intervention might have been transmitted from teenagers in the intervention group to their peers in the control group.
- Disclosure of sensitive behaviours and attitudes among the study participants, and potential for bias if teenagers in one of the trial's arms were more likely to disclose these behaviours and attitudes than those in the other arm.
- Selection of teenagers to the study, and whether the study succeeded in recruiting teenagers who were sufficiently at risk of teenage pregnancy.
- The validity of some outcome measures, most notably the measures of self-esteem and of youth development.

The study's findings may have been affected by some of the limitations outlined above. In particular, it is possible that the ability to detect real intervention effects was somewhat reduced. However, given that the overall pattern of differences between the intervention and control groups does not show a tendency towards positive (albeit statistically non-

significant) impacts across a majority of outcomes, but is instead rather mixed, it is unlikely that there were a range of real intervention effects that the study failed to detect.

Chapter 5 discusses some other potential explanations for why the effectiveness of the intervention was so limited, together with implications of the findings for further policy development in this area.

5 Conclusions

In this concluding chapter, we focus on the main findings of the randomised controlled trial which was conducted to assess the impact of the Teens and Toddlers youth development and teenage pregnancy prevention programme.

5.1 Main findings

The study found no evidence that the T&T intervention had been effective in changing any of the primary outcomes, at either follow-up one or follow-up two, but it had a positive impact on teenagers' self-esteem, which was one of the secondary outcomes. One year after the completion of the programme (at follow-up two), those young women who were randomised to the intervention group were less likely to report low self-esteem than those who were randomised to the control group (15%, compared with 25%).[24] The evidence of the positive impact of the intervention on self-esteem was also present immediately after the end of the programme (at follow-up one). This finding is consistent with young people's perceptions of the impacts of the intervention (see Chapter 3 for the analysis of answers to the supplementary questionnaire). The integral process evaluation also reported that both adult stakeholders (that is, teachers and T&T facilitators) and teenagers randomised to the intervention group felt that the programme had a beneficial effect on the young women's self-esteem (Jessiman et al. 2012).

Evidence from other studies suggests that self-esteem may be protective against early sexual activity among girls and teenage pregnancy (Spencer et al. 2002; Stephenson et al. 2004). Therefore improving teenagers' self-esteem may make an important contribution towards achieving the T&T intervention's objective to reduce rates of teenage pregnancy.

The study did not find evidence that the T&T intervention had been effective in improving any of the other secondary outcomes measured one year after the end of the programme (at follow-up two). There was some evidence at follow-up one that the intervention had positive effects on knowledge of sexual health, and on teenagers' ability to discuss contraception in a clinic or with a doctor, but these impacts dissipated by follow-up two. Furthermore, the pattern of difference in the prevalence of different outcomes between the intervention and control groups at follow-up two did not show an unequivocal general non-significant trend towards benefit: some tended (non-significantly) towards beneficial outcomes, some towards adverse outcomes.

These findings contrast with the teenagers' own perceptions of the programme. As discussed in Chapter 3, the majority of teenagers in the intervention group reported

[24] All percentages relating to the impact analysis are based on regression models which controlled for relevant baseline differences between the trial arms.

various self-perceived benefits of participating in the programme such as improved knowledge of sexual health, better communication skills and an increased desire to delay parenthood.

The integrated process evaluation suggested that participating girls also considered the programme to have been beneficial in terms of various aspects of youth development, such as confidence, self-esteem and interpersonal skills. However, the process evaluation suggested fewer subjectively perceived impacts on sexual behaviours, teenagers' attitudes towards early pregnancy and parenthood, or their engagement or attainment in school (Jessiman at al. 2012).

5.2 Methodological limitations

There are a number of methodological limitations that might have affected the study's ability to detect possible impacts of the T&T programme (discussed in detail in Chapter 4). Here, we summarise these issues and appraise their significance.

One potential limitation relates to the intention-to-treat analytical approach. The concern here is that including in the analysis those who dropped out of the programme as part of the intervention group could dilute the impacts of the intervention. However, there is little evidence in the impact findings that dilution of positive impacts occurred. There is no overall pattern of statistically non-significant benefits to suggest that the intervention brought about a range of benefits which just failed to reach the level of statistical significance because of a dilution effect. Given that the drop-out rate was only 27%, if the intervention had had real effects on other outcomes, we would have expected to see this non-significant trend towards a range of benefits, which is not found in the data. What we find instead is that the non-significant associations are scattered either side of the threshold dividing positive from negative effects – with about half the differences suggesting potentially positive and the other half negative impacts.

Another possible methodological limitation of the study relates to the potential for contamination of the findings. If teenagers participating in the programme discussed the new knowledge they gained from participation with those who were in the control group and thus benefited the teenagers in the control group, this would have led to a certain amount of 'contamination' and therefore underestimation of any real intervention effects.

Although it is possible that some contamination occurred within this study, we do not believe that such effects are likely to have unduly affected the results. As with the above, if we had seen results that consistently tended in the direction of intervention benefit but did not reach significance, then we might conclude from this that contamination could have resulted in our analysis having insufficient power to detect real intervention effects. However, this was not what we found. Furthermore, while we might anticipate a possible contamination effect on outcomes such as knowledge of sexual health, in which some benefits might plausibly be passed from intervention to control participants, we would not expect contamination to affect outcomes such as youth development or sexual behaviour.

Reluctance to disclose socially undesirable behaviours and attitudes is another methodological issue which could have affected the study's findings. If teenagers *as a whole* had under-reported certain behaviours, this would have made it more difficult to identify intervention effects but would not otherwise have biased the analysis. However, if teenagers *randomised to the intervention* were less or more likely to report risk-taking behaviours and attitudes than teenagers in the control group, this would have biased the findings in the direction of, respectively, over- or under-estimating any real benefits of the intervention.

Participants in the T&T programme are actively encouraged to be honest about their behaviours and attitudes. If teenagers in both the T&T and the control groups were under-reporting certain behaviours at the baseline but then those in the T&T group provided an honest picture post-intervention, this may have limited the study's potential to identify the intervention's positive effects. Another possibility is related to social desirability bias, which generates an over-reporting of benefits among intervention participants because they want to report what they perceive that the intervention providers and evaluators want to hear (e.g. Stuart and Grimes 2009).

If the teenagers who were recruited to the study were not sufficiently at risk (for example, due to the necessity to recruit a certain number or due to teenagers' reluctance to take part in the research), this too might have had a negative impact on the study's ability to detect the intervention's benefits. The process evaluation found that the T&T facilitators thought that the 'right' girls were being referred to the programme (Jessiman et al. 2012). In addition, comparisons of the prevalence of various sexual behaviours between the T&T RCT and the evaluation of the Young People's Development Programme (YPDP) (Wiggins et al. 2009) suggest that that the T&T RCT was successful in recruiting teenagers who were sufficiently at risk of teenage pregnancy.

Another possible limitation of the study is that it was designed to examine the short- and medium-term impacts of the programme and was not able to look at its long-term impacts, such as pregnancy incidents before age 20 or rates of being 'not in education, employment or training' (NEET).

The measures of youth development and self-esteem used in this study are not ones that have been previously validated (even though the self-esteem measure had been previously used in the Ripple study (Stephenson et al. 2004)). This means that these measures may be less sensitive to change than comprehensive validated measures would be.

The study's findings may have been affected by some of the limitations outlined above and it is possible that the ability to detect real intervention effects was thereby somewhat reduced. However, given that the overall pattern of differences between the intervention and control groups does not show a tendency towards positive (albeit statistically non-

significant) impacts across a majority of outcomes, but is instead rather mixed, it is unlikely that there were a range of real intervention effects that the study failed to detect.

5.3 Implications of the findings of the study

While acknowledging the methodological limitations of the study, we do not believe that they explain the intervention's limited and mostly short-term impact on young women at risk of teenage pregnancy. This section discusses other possible explanations as well as implications for further policy development in this area.

T&T may not have provided sufficient sexual health education. There was evidence from the process evaluation that some facilitators felt that their main role was 'not teaching but facilitating', that is providing girls with an opportunity to discuss sexual health issues openly, rather than giving them accurate information. There was also variation in the facilitators' own knowledge of sexual health (Jessiman et al. 2012). While the RCT found evidence of the intervention's positive impact on sexual health knowledge at follow-up one, this effect was no longer evident a year after the programme ended.

Missing out on normal schooling to attend the programme may also have had a negative effect on T&T participants, with 31% reporting falling behind with schoolwork as a result of participation (see Chapter 3).

This intervention appears to have brought about benefits in terms of self-esteem – which evidence from other studies suggests may be protective against early sexual activity among girls and teenage pregnancy (Emler 2002; Spencer et al. 2002) – but not other outcomes as measured in the RCT. Therefore, while we can conclude that it might have some potential for facilitating girls' personal development and possibly for reducing the risk of teenage pregnancy, we cannot conclude that its evidence base is at present strong. We would recommend further development and further evaluation of the intervention to address the current limitations suggested by our evaluation. The criteria for targeting the intervention need to be kept up to date with the most recent evidence on the risk factors for teenage pregnancy in the UK. Teachers need to be trained to use the criteria correctly. Participants and their parents should be fully informed about the aims of the intervention and why they have been targeted. The sex education aspects of the intervention should be significantly strengthened and facilitators should receive training so that they have consistent expertise in this area. More thought is needed about how to ensure that girls do not fall behind in their school work as a result of their participation in the programme, for example, through scheduling T&T sessions in such a way that the girls do not miss lessons in any key subjects.

While there was good fidelity in the intervention, it may be helpful to more closely follow the logic model framework developed for this evaluation (see Appendix B) and to develop it further (see Kirby 2004) on the basis of the above comments and their own expertise.

More generally, it is important for policy makers to appreciate that targeted interventions are unlikely on their own to have a significant impact on overall rates of teenage pregnancy in the population. This reflects changes in the strong risk factors for teenage pregnancy through time; the difficulty in predicting and then identifying in practice which girls are most at risk of teenage pregnancy; and the fact that most teenage pregnancies actually arise among girls at low-to-moderate risk of teenage pregnancy simply because there are more girls in this group (Kneale et al. under review). Therefore, strategies to reduce teenage pregnancy should also increase emphasis on universal interventions such as improving school-based sex and relationships education (DiCenso et al. 2002), expanding access to family planning and programmes to increase all girls' educational expectations (Harden et al. 2009).

References

Allen, J. P., Philliber, S., Herrling, S. and Kuperminc, G. P. (1997) Preventing teen pregnancy and academic failure: experimental evaluation of a developmentally-based approach, *Child Development*, 64, 4, pp. 729-742.

Bailey, P. (2005) Teenage pregnancies: is the high rate of teenage pregnancy and parenthood in the UK a public health problem? *Journal of Family Planning and Reproductive Health Care*, 31, 4, pp. 315-319.

Barnett, J. E. (2006) Evaluating “baby think it over” infant simulators: a comparison group study, *Adolescence*, 41, 161, pp. 103-110.

BBC online. David Cameron launches Tories' 'big society' plan http://www.bbc.co.uk/news/uk-10680062. [accessed 8 April 2011].

Berthoud, R. (2001) Teenage births to ethnic minority women, *Population Trends,* 104, pp. 12-17.

Bland, J.M. and Kerry, S.M. (1997) Trials randomized in clusters, *British Medical Journal*, 315, p. 600.

Campbell, D. T. and Russo, J. J. (1999) *Social experimentation*. Thousand Oaks CA: Sage.

Cater, S. (2008) Pregnancy rates, attitudes and behaviour changes among graduates of Teens & Toddlers. Teens and Toddlers: Annual retrospective study 2008 (Unpublished report).

Cater, S. and McDowell, A. (2007) Pregnancy rates, attitudes and behaviour changes among graduates of Teens and Toddlers: Annual retrospective study 2007 (Unpublished report).

Chavaudra, N. (2007) The impact of virtual infant simulators in tackling under-18 conception rates in Rotherham, UK, *Journal of Family Planning and Reproductive Health Care*, 33, 1, p. 35.

Consolidated Standards of Reporting Trials (CONSORT) Statement (2010) http://www.consort-statement.org/consort-statement/further-explanations/box6_intention-to-treat-analysis/ [accessed 13 December 2011].

Coyle, K., Basen-Enqquist, K., Kirby, D., Parcel, G., Banspach, S., Harrist, R., Baumler, E. and Weil, M. (1999). ‘Short-term impact of safer choices: a multicomponent, school-based

HIV, other STD, and pregnancy prevention program', *Journal of School Health*, 69, 5, pp. 181-188.

DiCenso, A., Guyatt, G., Willan, A. and Griffith, L. (2002) 'Interventions to reduce unintended pregnancies among adolescents: systematic review of randomised controlled trials', *British Medical Journal*, 324, pp. 1426-34.

Emler, N. (2002) 'Costs and causes of low self-esteem', *Youth Studies Australia*, 21, 3, pp. 45-48.

Ermisch, J. (2003) Does a 'teen birth' have longer-term impacts on the mother? Suggestive evidence from the British Household Panel Study, Institute for Social and Economic Research Working Paper Series number 2003-32. Colchester: University of Essex.

Ermisch, J. and Pevalin, D. J. (2003) Who has a child as a teenager? Institute for Social and Economic Research Working Paper Series number 2003-30. Colchester: University of Essex.

Fox News. Barack Obama Proposes New and Expanded National Service Programs http://www.foxnews.com/story/0,2933,315361,00.html [accessed 8 April 2011].

Harden, A., Brunton, G., Fletcher, A., Oakley, A., Burchett, H. and Backhans, M. (2006) *Young people, pregnancy and social exclusion: a systematic synthesis of research evidence to identify effective, appropriate and promising approaches for prevention and support*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Harden, A., Brunton, G., Fletcher, A. and Oakley, A. (2009) Teenage pregnancy and social disadvantage: a systematic review integrating trials and qualitative studies. *British Medical Journal,* 2009;339:b4254.

Hobcraft, J. and Kiernan, K. (2001) Childhood poverty, early motherhood and adult exclusion, *British Journal of Sociology,* 52, pp. 495-517.

Hofferth, S. L., Reid, L. and Mott, F. (2001) The effects of early childbearing on schooling over time, *Family Planning Perspectives*, 33, 6, pp. 259-267.

Holgate, H.S. and Evans, R. (2006) Sexuality and young motherhood: discourses and definitions in Holgate, H. S., Evans, R. and Yuen, F. K. O. (eds) *Teenage pregnancy and parenthood: global perspectives, issues and interventions*, pp. 7-23. Abingdon/New York: Routledge.

Humphrey, K. (2009) Pregnancy rates, attitudes and behaviour changes among graduates of Teens and Toddlers: Annual retrospective study 2009 (Unpublished report).

Humphrey, K. (2010) A retrospective analysis exploring the efficacy of the Teens and Toddlers programme (Unpublished report).

Humphrey, K. (2011) A retrospective analysis exploring the efficacy of the Teens and Toddlers programme (Unpublished report) http://www.teensandtoddlers.org/images/stories/PDFs/Retrospective-Report-2011.pdf [accessed 7 March 2012].

Jessiman, T., Keogh, P., Scott, S., Wollny, I., Sorhaindo, A. and Bonell, C. (2012) *Teens and Toddlers integrated process evaluation*, DFE-RR212, DfE Research Report, London: DfE.

Kirby, D. (2007) *Emerging Answers 2007: research findings on programs to reduce teen pregnancy and sexually transmitted diseases*. Washington DC: The National Campaign to Prevent Teen and Unplanned Pregnancy.

Kirby, D.B., Rhodes, T. and Campe, S. (2005) *Implementation of multi-component youth programs to prevent teen pregnancy modelled after the Children's AID Society - Carrera Program*. Scotts Valley CA: ETR Associates.

Kirby, D. (2004) BDI Logic Models: a useful tool for designing, strengthening and evaluating programs to reduce adolescent sexual risk-taking, pregnancy, HIV and other STDs. recapp.etr.org/recapp/documents/**BDILOGICMODEL**20030924.pdf [accessed 13 March 2012]

Kneale, D. (2008) *Pathways to parenthood: exploring context as a predictor of time to first parenthood*. London: Centre for Longitudinal Studies, Institute of Education, University of London. PhD thesis.

Kneale, D., Bonell, C., Fletcher, A. and Wiggins, R. (under review) 'Distribution and determinants of risk of teenage motherhood in three British longitudinal studies: implications for targeted prevention interventions', *Journal of Epidemiology and Community Health*.

Lee, E., Clements, S., Ingham, R. and Stone, N. (2004) *A matter of choice? Explaining national variations in teenage abortion and motherhood*. York: Joseph Roundtree Foundation.

McDowell, A. (2004) A retrospective analysis of the efficacy of the Teen and Toddlers programme (Unpublished report).

Melchior, A. (1998) *National evaluation of learn and serve America school and community based programs*. Waltham MA: Brandeis University, Centre for Human Resources.

Neill, J. T., Marsh, H. W. And Richards, G. E. (1997) *Development and psychometrics of the Life Effectiveness Questionnaire.* Sydney: University of Western Sydney.

Oakley, A. (1990) Who's afraid of the randomized controlled trial? in Roberts, H. (ed.) *Women's health counts*, pp. 167-194. London: Routledge.

O'Donnell, L. Stueve, A. San Doval, A., Duran, R., Haber, D., Atnafou, R., Johnson, N., Grant, U., Murray, H., Juhn, G., Tang, J. and Piessens, P. (1999) The effectiveness of the Reach for Health Community Youth Service learning program in reducing early and unprotected sex among urban middle school students, American Journal of Public Health, 89, 2, pp. 176-81.

Office for National Statistics (2012) *Conceptions in England and Wales 2010. Statistical Bulletin.* http://www.ons.gov.uk/ons/dcp171778_258291.pdf [accessed 12 March 2012]

Pevalin, D. J. (2003) Outcomes in childhood and adulthood by mother's age at birth: evidence from the 1970 British Cohort Study, Institute of Social and Economic Research Working Papers Series number 2003-31. Colchester: University of Essex.

Philliber S., Kaye J.W., Herrling S. and West E. (2002) Preventing pregnancy and improving health care access among teenagers: an evaluation of the Children's Aid Society-Carrera Program. *Perspectives on Sexual and Reproductive Health,* 34, 5, pp. 244-251.

Philliber, S., Kaye, J.W. and Herrling, S. (2001) *The national evaluation of the Children's Aid Society Carrera Model Program to prevent teen pregnancy*. New York: Philliber Research Associations.

Spencer, J.M., Zimet, G.D., Aalsma, M.C. and Orr, D.P. (2002) 'Self-esteem as a predictor of initiation of coitus in early adolescents', *Pediatrics*, 109, 4, pp. 581-584.

Stephenson, J.M., Strange, V., Forrest, S., Oakley, A., Copas, A., Allen, E., Babiker, A., Black, S., Ali, M., Monteiro, H., Johnson, A. M., and the RIPPLE study team (2004) 'Pupil-led sex education in England (RIPPLE study): cluster-randomised intervention trial', *The Lancet*, 364, 9431, pp. 338-346.

Stuart, G.S. and Grimes, D.A. (2009) Social desirability bias in family planning studies: a neglected problem, *Contraception*, 80, pp. 108–112.

Teenage Pregnancy Unit (2006) Teenage pregnancy: working towards 2010 - good practice and self-assessment toolkit. UK Department for Education and Skills and UK Department of Health. http://webarchive.nationalarchives.gov.uk/20090618071133/dcsf.gov.uk/everychildmatters/resources-and-practice/ig00198/ [accessed 4 March 2012]

The Dream Mill (2004) Teens and Toddlers retrospective analysis focus groups, research debrief prepared for COUI UK (Unpublished report).

Torgerson, C. (2001) Contamination in trials: is cluster randomization the answer? *British Medical Journal*, 322, pp. 355-357.

Wellings, K., Field, J., Wadsworth, J., Johnson, A. M., Anderson, R. M. and Bradshaw, S. A. (1990) Sexual lifestyles under scrutiny, *Nature*, 348, 6299, pp. 276-8.

Wiggins, M., Bonell, C., Sawtell, M., Austerberry, H., Burchett, H., Allen, E. and Strange, V. (2009) Health outcomes of youth development programme in England: prospective matched comparison study, *British Medical Journal*, 2009;339:b2534.

Wiggins M., Bonell C., Burchett H., Sawtell M., Austerberry H., Allen E. and Strange V. (2008) *Young People's Development Programme final report*. London: Institute of Education.

Appendix A Technical details

The impact of the T&T programme was measured by means of a randomised controlled trial (RCT). This design was chosen because randomised trials are generally the strongest design when aiming to examine the effects of social interventions. They randomly allocate either individuals or clusters of individuals to receive an intervention or not, and follow participants from pre-intervention baseline measures to post-intervention outcome measures. Having a control group enables the effects of the intervention to be disentangled from the effects of background noise (e.g. maturational, seasonal or other time-related factors that might affect outcomes). If enough individuals are randomised this ensures that the intervention and control groups are similar for measured and unmeasured factors, which might themselves influence whether an outcome occurs and thus 'confound' an examination of intervention effects. Further, in the case of the T&T programme it would be difficult (if not impossible) to find a suitable control group in any other way because the girls who take part in the T&T programme are selected both by teachers and then by self-selection.

In this instance, the trial involved the random allocation of over 400 individual young women considered to be at risk of teenage pregnancy to an intervention or control group (the latter receiving 'standard care' in schools). At-risk young women were identified by their teachers using guidance on the characteristics and attributes that T&T consider to be indicative of a risk for teenage parenthood (a copy of the T&T guidance used during the trial, as well as a copy of the revised guidance, which was designed for the trial to improve the selection but was very rarely used, are included in Appendix D). The selected young women received preliminary information about the study prior to consenting to allocation, intervention and research. Data for all participants were collected by questionnaire at three points in time: prior to allocation (baseline), immediately post-intervention (follow-up one) and a year after intervention (follow-up two). The interviews collected quantitative data on key outcomes, and also explored participants' views of the programme.

This appendix contains details of the technical aspects of conducting the RCT.

Alternative designs to the individual-level RCT

At the design stage of the trial a number of alternative designs were considered. These included:

- A cluster randomised trial, with schools being the unit of randomisation;
- A cluster randomised trial, with friendship groups being the unit of randomisation;
- A quasi-experiment matched comparison design, with T&T girls being matched to similar girls from a small number of other schools.

The design was peer-reviewed, with two of the four reviewers suggesting that a cluster randomised trial (with schools as the unit of randomisation) may be preferable from the

point of view of avoiding contamination, and two in favour of the adopted design. There were two key reasons for our not using a cluster randomised trial of schools:

1. Sample size considerations. To run the study as a cluster randomised trial would have needed a very significant increase in the sample sizes: both in terms of the sample size of teenagers and the number of schools.
2. Identification of the relevant teenagers in control schools. For those schools assigned to the control group a means of identifying, and recruiting eligible girls would have been needed. As is detailed in Chapter 1, schools do not follow scripted and standardised protocols for the selection of girls for T&T, and putting in place procedures for control schools would have been difficult and potentially biasing.[25]

The second option, of running the RCT within schools but assigning girls to friendship 'clusters' prior to randomisation, was discussed at the evaluation steering group. The consensus was that it would add complication without any clear gains and it was not pursued.[26]

The issue of individual v. cluster randomisation is complex however and we include more discussion of the issue below.

The final design option, considered at the very start of the study, was a matched comparison design with girls participating in T&T being matched to 'similar' girls in other non-T&T schools. The problems here are similar to the problems with the school-cluster randomised trial, in that identifying and recruiting 'similar' girls from comparison schools is extremely difficult. Those difficulties would have been multiplied if a matched comparison group had been sought just for those girls fully participating in T&T (rather than all those girls who initially sign up): in Chapter 2 of this report we have shown that the girls fully participating in T&T have very different outcomes to the girls who enter the programme but leave early, and had we attempted to find a suitable matched comparison group for 'full participants' based on baseline questionnaires it is clear that we would almost certainly have failed.

Individual versus cluster allocation

Trials can randomly allocate either *individuals* or *clusters* of individuals (e.g. schools) to intervention and control groups. In choosing which is appropriate, the logic of an evaluation should follow the logic of an intervention (Bland & Kerry 1997). If an intervention principally aims to recruit targeted individuals and to intervene with these individuals in order to benefit them, then an individual-allocation trial is probably most appropriate. If an intervention recruits entire clusters, such as schools, and principally

[25] If there were any unobserved differences between the girls recruited in the T&T schools and the girls recruited in the control schools then this could generate bias. The alternative approach of using standardised recruitment procedures in both T&T and control schools was not discussed, but would have changed one important element of the T&T programme, namely the discretion exercised by school staff on who they put forward.

[26] This design would also have needed an increase in sample size.

aims to intervene with and/or achieve benefits for all or most individuals within that cluster, then a cluster-allocation trial is probably most appropriate.

Peer education is not an aim of the T&T intervention, but T&T participants are encouraged to share their experiences and learning with their friends. This interaction may benefit these friends and if these friends are in the control group then contamination could occur. However, we did not expect participants' conversations with friends in the control group to be so frequent and so powerful that a substantial portion of the benefit that might accrue to direct participants would be passed on to their friends.

Where contamination is unlikely to account for more than 30% of the intervention benefits being passed on to control participants, an individual allocation RCT tends to be more appropriate than a cluster RCT. This is largely because, in such circumstances, a cluster RCT would have less power to determine intervention effects than an individual allocation RCT because individuals within the same school resemble each other more closely than individuals in different schools, such that individual data are not statistically independent (Torgerson 2001). Moreover, our peer reviewers who scrutinised the design considered that one of the most notable risks to the evaluation was that it would be under-powered. However, increasing the sample size significantly to compensate for the reduced power of a cluster RCT would have placed too heavy a burden on T&T in terms of fundraising, liaison and delivering the programme to be practicable.

A further problem with using cluster trials to evaluate targeted interventions is that, as mentioned above, it would be extremely difficult to identify a comparable cohort of at-risk girls in any putative control schools, and therefore selection bias would be a considerable risk. The only means of avoiding this would be for the researchers to identify at-risk girls using standardised methods in all schools prior to randomisation, which would have been very difficult in this study because it would have represented a distortion of how the intervention is intended to work within schools.

Finally, evaluations of broadly similar interventions have used individual-allocation trials that have successfully measured benefits (e.g. Philliber et al. 2002), and experts regard individual-allocation trials as the strongest design for this particular form of individually targeted intervention (Harden et al. 2006; Kirby et al. 2007). So on balance it was concluded that the risk of a cluster design underestimating the benefits of the programme as a result of being under-powered and heavily subject to selection bias considerably outweighed the risk of an individual-allocation design under-estimating benefits as a result of contamination. Therefore the trial proceeded on the basis of an individual-allocation design. The process of randomisation is described later in this appendix.

Sample size and power calculations

The sample size for the trial needed to be a balance between the number required to give sufficient power to detect an impact on the relevant outcome measures and the number of young people that T&T had the capacity to include in the programme.

Our power calculation focused on our primary outcome, no contraception use at last sex (within last three months), which we predicted would be the least prevalent of our primary outcomes among the control group (approximately 25%, informed by Ripple, see Stephenson et al. 2004). Assuming this prevalence, in order to detect a halving of risk (in the range of effects reported by for example the Safer Choices and CAS-Carrera studies, see Coyle et al. 1999 and Kirby et al. 2005) we would require a sample of 180 per arm (assuming 80% power and 95% precision). A starting sample of 360 young women (180 in each arm) was judged achievable in terms of the capacity of T&T to deliver the intervention. The formative evaluation identified that up to 25% of girls allocated to receive T&T might drop out in the first eight weeks of the intervention, so the aim was to recruit a further 120 girls to hold on a reserve list (thus the overall target size for the baseline sample was 480 girls). Reserves would augment, rather than replace, drop-outs in the trial sample to avoid selection bias. Any girls who dropped out of T&T after eight weeks were not augmented.

The overall attrition rate within the trial's data collection was anticipated to be approximately 10% (i.e. the expectation was that 10% of teenagers would not complete the final follow-up questionnaire because they did not want to take part at that stage or had left school and changed their contact details, etc.) leaving an analytical sample of 430 girls. However the discussion of detectable effect sizes below uses 360 as the analytical sample size in order to err on the conservative side.

For information on the final sample size included in the trial, see the section on response later in this appendix.

Outcome measures: overview

When choosing the outcome measures for this study we tried to select a range of primary and secondary outcomes that reflected the potentially broad benefits of the T&T programme. The number of outcome measures selected was kept purposively low to reduce the possibility of false-positive results arising from multiple tests of statistical significance. The study steering group agreed that these outcome measures constituted a good representation of the proposed impacts of the intervention; reflected the logic model elucidated in the formative research; were sufficiently sensitive given the sample size; and were not susceptible to possible contamination. The outcome measures selected were:

Primary outcomes

- Did not use any contraception the last time they had sex (and had sex within the last three months);
- Has had more than one episode of not using contraception in the last three months;
- Expects teenage pregnancy;
- Low youth development score.

Secondary outcomes

- Did not use a condom the last time they had sex (and had sex within the last three months);
- Has had more than one episode of not using a condom in the last three months;
- Believes that the best age to have sex for the first time is under 16 years of age;*
- Is favourable to sometimes not using protection for sex;*
- Low self-reflection;
- Low emotional vocabulary;
- Low self-esteem;*
- Dislikes school;
- Lack of expectation regarding post-16 education, training or employment;
- Low sexual health knowledge;
- Difficulty in discussing sex with a boyfriend;*
- Difficulty in discussing the pill with a doctor;*
- Has become pregnant since baseline;
- Lack of awareness of the impact of parenthood on social life;
- Number of school days missed.* [27]

These measures were all selected at the design stage in consultation with the study steering group, with the exception of the six marked with an asterisk in the list above. Those marked with an asterisk were added in response to a request by the Department for Education (DfE) who believed that these were important measures that would complement those already chosen. They were added in November 2010, before the collection of the follow-up two data.

The detectable effect size for a trial of 360 participants with individual randomisation (180 in the intervention group and 180 in the control group) is presented below. These figures were based on an 80% power calculation and assume all statistical tests would be two-sided and use a 5% significance level.

Table A.1 Detectable effect sizes

% reporting outcome in control group	Effect size as odds ratio	Effect size as percentage point difference
25	0.47	11.5
50	0.55	14.5
30	0.49	12.5
60	0.56	14.5
40	0.54	13.7
65	0.55	14.5
20	0.42	10.5
6	0.11	5.3

[27] Unlike all other outcomes, the data for this outcome was from the survey of teachers. Due to an error, this outcome was omitted from the online registration of the trial protocol.

These figures indicate that the impact of T&T would need to be of the order shown in Table A.2 for an impact to be detected in this evaluation (this table only includes the measures that were selected at the design stage of the trial). Reference to similar evaluations indicated that it was not implausible to expect these (or even larger) effects on those outcomes. For example, the Safer Choices Evaluation found effect sizes for frequency of unprotected sex in the previous three months of OR=0.63 (p=0.02), condom use at last sex of OR=1.68 (p=0.01) and contraception use at last sex of OR=1.76 (p=0.07). The Carrera trial found effect sizes for female sexual debut of OR=0.5 p<0.05, female pregnancy of OR=0.3 p<0.05 and female contraception at last sex of OR=2.4 p<0.01.

Table A.2 Effect sizes required across primary and secondary outcomes

Outcome	Expected rate in control group	Rate necessary among T&T girls
Sexual/contraception behaviour		
1) No contraception use at last sex (within last 3 months)	25%	13.5%
2) >1 episodes of no contraception in last 3 months	50%	35.5%
Attitudes to teenage parenthood		
3) Expectation of teenage parenthood	30%	17.5%
Youth development		
4) Low score on youth development composite	30%	17.5%
Condom use		
5) No condom use at last sex (within last 3 months)	60%	45.5%
6) >1 episodes of no condom use in last 3 months	60%	45.5%
Emotional self-management		
7) Low self-reflection	40%	26.3%
8) Low emotional vocabulary	40%	26.3%
Educational attitudes and expectations		
9) Dislike of school	65%	50.5%
10) Lack of expectation of post-16 education, training or employment	20%	9.5%
Sexual health knowledge		
11) Low sexual health knowledge	50%	35.5%
Teenage pregnancy		
12) Teenage pregnancy	6%	0.7%
13) Lack of awareness of impact of parenthood on social life	30%	17.5%

Outcome measures: detailed specifications

Primary outcomes	
No contraception use at last sex in last three months	Measure based on two questions. Only those who said they had had sex with a boy (or man) were routed to these two questions. *How regularly have you had sex in the last three months with a boy (or a man)?* *- answer categories from 'at least once a week' to 'not at all'.* *Now thinking about the last time you had sex with a boy (or a man), did you use…* *A condom?* *The pill?* *The morning after pill?* *Any other contraception?* Issue positively recorded if: • Sexually active (has had sex with a boy/man) • Has had sex in the last three months (any regularity apart from 'none at all') • Did not use any contraception at last sex
More than one episodes of no contraception in last three months	This outcome is available for follow-up two but not for follow-up one (as the question about the number of episodes was omitted from baseline and follow-up one questionnaires by mistake). *And in the last three months, how many times would you say that you have had sex with a boy (or a man) without using contraception (e.g. a condom, the pill, the morning after pill, injections, or an implant)?* - *answer options from 'none' to 'more than ten times'.* Issue positively recorded if: - twice or more
Expectation of teenage parenthood	*How likely do you think it is that you will have a baby before you are 20, is it:* *Very likely, fairly likely, fairly unlikely or very unlikely?* Issue positively recorded if: - answered 'very likely' or 'fairly likely'
Youth development score	Based on eight questions. For each question, respondents were asked to give a score from 1 'false – not like you' to 8 'true – like

	you'. - *Overall most things I do turn out well* - *I know I have the ability to do anything I want to do* - *My own efforts and actions are what will shape my future* - *I work hard at solving my problems* - *Personal goals are important to me* - *I respect other people* - *I sort out my conflicts with other people peacefully* - *People understand me when I am talking* Youth development score is a sum of individual scores. Values range from 8 to 64, with higher values indicating positive youth development. The scale's internal reliability as measured by Cronbach's alpha was 0.78 at baseline, 0.77 at follow-up one, and 0.80 at follow-up two.
Secondary outcomes	
No condom use at last sex in last three months	Measure based on two questions. Only those who said they had had sex with a boy (or a man) were routed to these two questions. How regularly have you had sex in the last 3 months with a boy (or a man)? - answer categories from 'at least once a week' to 'not at all'. Now thinking about the last time you had sex with a boy (or a man), did you use…A condom? Issue positively recorded if: • Sexually active (has had sex with a boy/man) • Has had sex in the last 3 months (any regularity apart from 'none at all') • Did not use a condom at last sex
More than one episodes of no condom use in last three months	And in the last three months, how many times would you say that you have had sex with a boy (or a man) without using a condom? - answer options from 'none' to 'more than ten times'. Issue positively recorded if: - twice or more
Best age to have sex under 16	What do you think is a good age to have sex for the first time? - Answer options from 'Age 12 or younger' to 'Age 22 or older' Issue positively recorded if:

	- Answers from 'age 12 or younger' to 'age 15'
Favourable to not using protection sometimes in sex	Please tell me how much you agree or disagree with the following sentences… "Sex is about excitement and feeling good, so it's ok that people don't always think about using protection"? - Answer options 'agree strongly', 'agree', 'neither agree nor disagree', 'disagree', 'disagree strongly'. Issue positively recorded if: - Answers 'agree strongly' or 'agree'
Low self-reflection	Thinking over the last six months, how much do you agree or disagree with these sentences… "I am usually in touch with my own feelings"? • Answer options 'agree strongly', 'agree', 'neither agree nor disagree', 'disagree', 'disagree strongly'. Issue positively recorded if: • Answers 'disagree' or 'disagree strongly'
Low emotional vocabulary	Thinking over the last six months, how much do you agree or disagree with these sentences… "I can usually find the right words to say how I am feeling"? • Answer options 'agree strongly', 'agree', 'neither agree nor disagree', 'disagree', 'disagree strongly'. Issue positively recorded if: • Answers 'disagree' or 'disagree strongly'
Low self-esteem	Thinking over the last six months, how much do you agree or disagree with these sentences… "I like myself"? • Answer options 'agree strongly', 'agree', 'neither agree nor disagree', 'disagree', 'disagree strongly'. Issue positively recorded if: • Answers 'disagree' or 'disagree strongly'
Dislike of school	How much do you like school? • Answer options 'very much', 'quite a lot', 'not much' and 'not at all' Issue positively recorded if: • Answers 'not much' or 'not at all'
Lack of expectation of post-16 education, training	Based on two questions. When are you planning to leave school or sixth form? • As soon as possible

or employment	• Straight after your GCSEs • When you are 18 • After you are 18 And then what are you planning to do? • Continue in full-time education (uni., college) • Find a job • Training (e.g. skill-seekers, apprenticeship, youth training) • Go abroad for at least six months • Take some time off • Something else. Issues positively recorded if: • Planning to leave school as soon as possible or straight after GCSEs AND planning to go abroad, take some time off or do 'something else'.
Low sexual health knowledge	Based on four questions. As far as you know, are the following true or false… …People are less likely to catch infections passed on by sex if they use condoms? …People are less likely to catch infections passed on by sex if they use the pill? …People might have a sexually transmitted infection without realising it? • Answer options 'true', 'false' and 'not sure'. How many days after having sex do you think you can take the morning after pill (emergency contraceptive pill) to stop you getting pregnant? • Only within one day of having sex • Up to three days after having sex • Up to one week after having sex • Not sure. Issue positively recorded if: • one or more questions answered incorrectly or 'not sure'.
Difficulty of discussing sex with boyfriend	How easy or difficult would it be for you to... …talk openly about sex with a boyfriend? • 'Very easy', 'easy', 'difficult', 'very difficult', 'don't know'. Issue positively recorded if: • Answers 'difficult', 'very difficult' or 'don't know'.
Difficulty of discussing the pill with a doctor	How easy or difficult would it be for you to... …discuss going on the pill in a clinic or with a doctor? • 'Very easy', 'easy', 'difficult', 'very difficult', 'don't know'.

	Issue positively recorded if: • Answers 'difficult', 'very difficult' or 'don't know'.
New pregnancy since baseline	Based on one question but at baseline and follow-up two. Have you ever been pregnant (including any miscarriages or abortions (terminations))? • Yes/no. Issue positively recorded if: • at baseline the answer was 'no' (or the question was not asked as the teenager was not sexually active) and at follow-up two the answer was 'yes'. (The measure does not identify second pregnancies to those who reported previous pregnancies already at baseline)
Lack of awareness of impact of parenthood on social life	How much easier or harder would having a baby make a teenager's life…in terms of social life and time with friends? • Answer options 'a lot easier', 'a little easier', 'makes no difference', 'a little harder' and 'a lot harder'. Issue positively recorded if: • Answers 'a lot easier', 'a little easier' or 'makes no difference'.
Number of days of school missed[28]	From the survey of teachers. Calculated from the number of full days and half days missed during one half-term during the post-intervention year.

Schools and size of groups

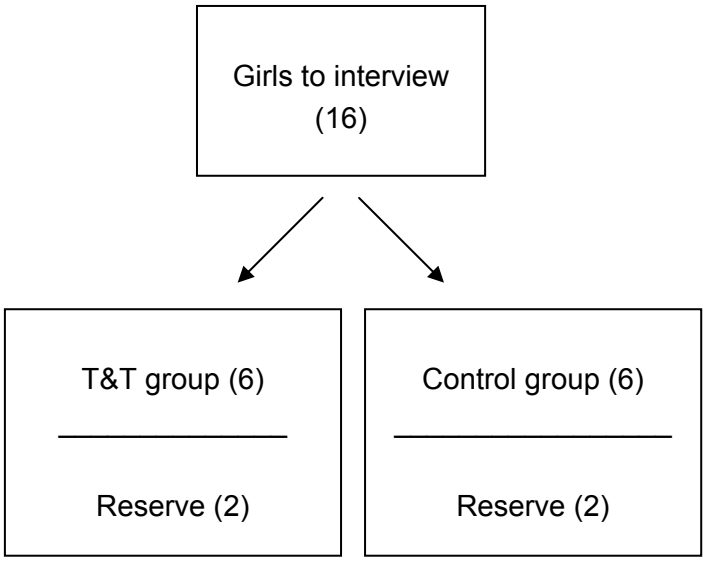
In total 22 schools took part in the RCT. These schools were selected by T&T and primarily consisted of schools with which they had established working relationships. The schools were located in various parts of the country but London was over-represented:

- 11 in London
- Three in Middlesex
- Three in Lincolnshire
- One in North Yorkshire
- Two in Lancashire
- Two in Greater Manchester.

The nurseries that support the delivery of the T&T programme were typically able to cater for groups of six girls. Since each group in the T&T programme required a control group

[28] Due to an error, this outcome was omitted from the online registration of the trial protocol.

and some reserve girls to augment drop-outs the average number of girls recruited to the trial in each school was typically 16 (see diagram below for details of how this number breaks down into reserves, girls offered the intervention, and control girls).



Randomisation procedure

The principle behind the randomisation was that a group of at risk girls would be recruited. The girls would be grouped into matched pairs within schools. Then one member of the pair would be randomly allocated to the T&T programme and one member of the pair would be randomly allocated to the control group. To cater for drop-out and maintain the number of girls experiencing the intervention, some pairs would be held in reserve and would enter the evaluation if one pair from their school drops out of the intervention within its first 8 weeks. Where this was necessary teachers and facilitators were advised to augment drop-outs with the reserve pairs using the order of the list that they were given by the research team.

In all likelihood the randomisation alone would produce groups that were naturally balanced across key characteristics. However, some differences inevitably arise by chance and so to reduce the risk of differences occurring across key characteristics we matched girls on a small number of characteristics before undertaking the randomisation. The details of our method are as follows:

- The most important factor to take into account when matching teenagers was considered to be the school the girls attended because different schools have different approaches to sex education and different social norms. Therefore girls were matched within their own schools.
- Since only 16 girls were typically available for matching within each school, we could only match on a small number of other criteria. In this instance, age and sexual experience were considered to be the most important.
- We started with age and sorted the girls into three groups representing the youngest third, the middle third and the oldest third. Then within these age groups

we sorted the girls in order of sexual experience.[29] This meant that the girls were sorted into a list whereby young inexperienced girls were next to other young inexperienced girls, whilst older experienced girls were next to other older experienced girls.

- These girls were subsequently paired[30] and each girl was given a random number using a computer-based random number generator. Within each pair the girl with the lowest number was selected for the T&T programme and the girl with the highest number was allocated to the control group.
- Finally, the reserve pairs were selected in a similar way, where each pair was given a random number, and the two pairs with the lowest random numbers were assigned to the reserve pool whilst the six pairs with the highest random numbers formed the main pairs for the evaluation.[31]

The randomisation was conducted in the offices of the research team after the collection of baseline data; schools had no influence on or knowledge of the randomisation process. The sorting and allocation of random numbers was implemented using the statistics software package SPSS 12.

The balance of characteristics between the intervention and control group was thoroughly explored in Chapter 2 of this report. This analysis shows that the groups were well-balanced on most characteristics. However to ensure that the differences that arose by chance did not affect the outcomes or our interpretation of them, where the differences were related to the outcome of interest, these differences were accounted for in the analysis and adjusted outcomes were calculated (see Chapter 4 for further details).

Questionnaire development

Cognitive pilot

A small scale cognitive pilot was conducted for this study in one London school participating in routine delivery of the programme outside this trial. Fieldwork was undertaken by two interviewers on two days of fieldwork in April 2009 and a total of eight cognitive interviews were conducted. The school that took part in the cognitive pilot had run T&T for a number of years and so was familiar with the eligibility criteria for T&T. They were asked to select a sample of 13-14 year old girls who, according to the usual T&T criteria, would be considered at risk of teenage pregnancy and therefore eligible for the T&T programme. These girls were invited to take part in the cognitive pilot and parental consent was sought for their participation. Those who did take part were given a £10 High Street voucher to thank them for their time.

The aims of the cognitive testing included exploring:

[29] Sexual experience was based on three questions about whether the girls had engaged in kissing, intimate touching, or sex. The answers to these questions were used to create a score where kissing scored 1 point, intimate touching scored 2 points, and sex scored 3 points. As such the score ranged from 0 (for a girl with no sexual experience) to 6 (for a girl who had relatively considerable sexual experience).

[30] Where an uneven number of girls had been interviewed in a particular school we formed a triplet in the oldest third of girls.

- How respondents understood the questions and their comprehension of key terms;
- Whether respondents felt willing and able to answer the questions;
- Whether respondents were able to recall accurately the information being sought;
- Whether the questions were considered to be too sensitive.

The questionnaire was initially designed to be administered through computer assisted interview. As such, the cognitive testing also aimed to assess whether the questions were appropriately split between: computer assisted personal interviewing (CAPI) and computer assisted self interviewing (CASI).

The topics covered in the cognitive questionnaire included participants':

- Experiences of bullying
- Behaviour at school (e.g. how often they truanted, answered back, swore at teachers)
- Attitudes to sexual relationships
- Sexual experience
- Knowledge of contraception and sexually transmitted infections
- Attitudes to parenting
- Relationship with parents.

Most of the cognitive questionnaire was administered via paper and pen interviewing with the interviewer exploring respondents' understanding through retrospective probing. However, the more sensitive topics (e.g. respondents' sexual experiences) were administered via a self-completion questionnaire. When interviewers probed these sections they offered respondents the option of writing their answers down and handing them over in a sealed envelope at the end of the interview (rather than providing their answers verbally). This was to ensure that we captured feedback from respondents about their understanding of the questions even when they felt too embarrassed to discuss their answers openly.

The feedback from the cognitive pilot was that overall the questions were not too sensitive or embarrassing. However, it seemed that the questions on bullying and truanting would probably be subject to social desirability bias and would benefit from being administered via self-completion (if possible, given the need to keep the self-completion section of the questionnaire to a reasonable length).

Respondents to the cognitive pilot emphasised the importance of the interviewer explaining that the data would be treated confidentially and also explaining what this meant in pragmatic terms (e.g. "we would never tell your teachers or your parents what you've said") in order to develop respondents' trust in this assurance.

[31] Given the relatively low reporting of sexual experience at baseline for cohort one, pairs of girls who reported a high level of sexual experience at the baseline for cohort 2 were automatically allocated to the trial rather than to the reserve group.

The other feedback from the pilot related to specific misunderstandings of the wording of particular questions and was used to make the questions more straightforward or to use terminology that was more familiar to respondents.

CAPI pilot

A small scale pilot of the CAPI programme and associated survey processes was conducted in one London school participating in routine delivery of the programme outside this trial. Fieldwork was undertaken between May and June 2009 by two interviewers and a total of 20 interviews were conducted. The school that took part in the CAPI pilot was not the same as the one that had taken part in the cognitive pilot. However they had also run T&T for a number of years and so were familiar with the eligibility criteria for T&T. As such they were able to select a sample of 13-14 year old girls who would be eligible for the T&T programme. These girls were invited to take part in the CAPI pilot and parental consent was sought for their participation.

The interviews lasted an average of 47 minutes and the feedback was extremely positive. The girls who participated enjoyed taking part and thought that the interview length was acceptable. There were no serious problems with the CAPI programme or the questionnaire. However the pilot did highlight some issues with fieldwork management and survey processes that were refined prior to the launch of the main fieldwork stage.

Firstly, with regard to fieldwork, the pilot highlighted difficulties in getting parental consent forms returned on time. Delays in receiving these forms jeopardised the success of the pilot, which needed to be completed within a defined time period. To speed up this process girls were ultimately offered a £10 High Street voucher to take part in the pilot (which did improve the speed at which the consent forms were returned). However, the delay in receiving consent forms was not expected to be as much of a problem at the main stage because the RCT design meant that girls needed to return their forms and complete the interview in order to enter the trial and have a chance of joining the T&T programme. This meant that, at the main stage, the chance of joining T&T acted as the incentive. Nevertheless, to ensure that the main stage operated as smoothly as possible we liaised with T&T operations staff to monitor the situation and ensure early identification of any problems. We also decided to offer incentives at both follow-up stages to ensure that they were motivated to turn up for their interview appointments.

In terms of developing the survey processes we:

- Added guidance to the interviewer instructions on how to answer respondents' questions about the reasons that they were selected for the T&T programme.
- Provided the interviewers with an additional showcard that included definitions of some of the terms that caused teenagers difficulties with comprehension (e.g. 'contraception', 'abortion', and 'diaphragm').
- Included a reference to 'sexual health' in the section of the teenager information sheet that explained what the questionnaire would be about, because some teenagers were surprised by the content of the questionnaire.

- Added guidance to the interviewer instructions on how they should best behave while respondents were completing the CASI section of the questionnaire (e.g. trying not to *watch* respondents since this might make them feel rushed).

Change of data collection mode

As discussed above, the questionnaire was initially designed to be administered via computer assisted personal interviewing (CAPI) with sections of computer assisted self interviewing (CASI) for sensitive topics. The baseline survey for respondents in cohort one was indeed administered via this mode.

However, some members of the study steering group felt that the baseline findings for cohort one showed a lower than expected prevalence of risky behaviours and attitudes. Three hypotheses were proposed to explain this:

1) The girls recruited by schools were not 'at risk';
2) The girls recruited by schools were not *currently* 'at risk' but might have become so;
3) The girls recruited by schools were 'at risk' but did not disclose this information at interview.

The baseline findings provided no evidence regarding the relative importance of each of these explanations. As such we decided to err on the side of caution and revise the data collection mode for cohort two, administering the survey as a paper self-completion questionnaire to small groups of young women. This was considered to be the more cautious approach because paper self-completion questionnaires have been successfully used in other UK projects to elicit disclosure from teenagers (e.g. Wiggins et al. 2008; Stephenson et al. 2004), whilst CAPI and CASI has not been tried and tested among respondents of this age range in the same way. The baseline survey for respondents in cohort two was therefore administered entirely via paper self-completion questionnaire.

The effect of this change in data collection mode was reviewed after completion of the baseline fieldwork for cohort two. The findings did show a higher prevalence of risky behaviours and attitudes among cohort two than cohort one. As an illustration, the percentage of girls at cohort one who reported that they had had sex was 6%, and the percentage of girls from cohort two who reported that they had had sex was 17%. On the surface, this suggests that the change in mode elicited greater disclosure.

However, as discussed in section 2.6 of Chapter 2, the difference in the data collection mode was not the only difference between the two cohorts, as there was also a difference between the schools participating in the trial. Cohort two contained a group of 'original' schools that had already participated in cohort one and some 'supplementary' schools that only started participating in the trial at cohort two. The analysis presented in Chapter 2 shows that the profile of the girls from the 'supplementary' schools was very different from those from the 'original' schools. To illustrate, among the cohort two girls who went to the 'original' schools, only 3% said that they had had sex which is very similar to the finding

from cohort one. In contrast, 26% of the girls from 'supplementary' schools reported that they had had sex. This suggests that the change of data collection mode for questions about sexual behaviours from CASI at cohort one to paper self-completion questionnaire at cohort two did not lead to a greater disclosure of these behaviours. However, where the data collection mode changed from CAPI at cohort one to paper self-completion questionnaire at cohort two (e.g. for questions around youth development, self-esteem, dislike of school), there was some evidence that the self-completion mode made it easier for the teenagers to disclose behaviours and attitudes that are sensitive to social desirability.

Fieldwork

Interviewers and briefings

Before the baseline survey was conducted with cohort one, all interviewers attended a full day briefing led by the NatCen Social Research research team. These covered an introduction to the study and its aims, an explanation of the sample, procedures for contacting schools and arranging fieldwork, the importance of explaining the confidential nature of the survey, and a dummy interview exercise designed to familiarise interviewers with the CAPI program.

When the survey mode was changed, it became simpler for interviewers to administer because they did not need to learn how to use the CAPI programme. For this reason, interviewers on all subsequent stages of the project self-briefed for the project by reading their written instructions and familiarising themselves with the various survey materials.

In total 21 interviewers worked on the baseline survey, 20 interviewers worked on follow-up one and 18 interviewers worked on follow-up two. Where possible, female interviewers worked on the project to help foster girls' confidence in answering the sensitive questions in the questionnaire (in fact, female interviewers conducted all but one of the data collection sessions).

Fieldwork arrangements in schools

The dates of the three stages of fieldwork for each of the two cohorts are shown in Table A.3.

Table A.3 Teens and Toddlers evaluation fieldwork dates

	Cohort 1	Cohort 2
Baseline	September 2009	January/February 2010
Follow-up one	March 2010	July/August 2010
Follow-up two	February/March 2011	June/August 2011

In total ten schools took part in cohort one. These ten schools also took part in cohort two and an additional 12 schools joined the trial solely for cohort two.

The majority of the interviews and data collection sessions were conducted in school. For the baseline survey with cohort one they were conducted face-to-face with girls in a private room at the school and the most sensitive parts of the questionnaire were administered as a self-completion section on the computer. The total interview length was 40 minutes. In contrast the baseline survey with cohort two was conducted as a paper self-completion questionnaire with small groups of girls in a school classroom. Where possible, these sessions were carried out under exam conditions and interviewers asked teachers either to absent themselves from the sessions or, if that was not possible, to avoid wandering around the classroom to ensure that the girls were able to complete the questionnaires privately. These sessions lasted between 40 minutes and one hour.

For follow-up one and follow-up two, girls were again asked to complete a paper self-completion questionnaire in small groups in a school classroom. These sessions also lasted between 40 minutes and one hour. In order to improve response rates to the follow-up surveys, where girls had left school or were otherwise absent from the group sessions, interviewers attempted to contact girls at home to see if they would be willing to take part. At both these follow-up stages girls who took part were given a £10 High Street voucher to thank them for their time.

Response to the surveys

As discussed in Chapter 1, T&T operations staff liaised with teachers to identify girls who were eligible for the programme. They were responsible for briefing girls on the nature of T&T and the RCT, and they also co-ordinated the process of gaining written consent to take part in the trial from the girls and their parents. Since these exercises were undertaken by T&T operations staff we do not know how many girls were identified as eligible for the programme, or how many of these girls agreed to take part and gained the necessary parental consent.

Baseline interviews were conducted with 493 girls who had agreed to take part in the trial. However, four of these girls were not ultimately entered into the trial because their school refused to allow them to miss the school lessons required to take part in the programme. Thus, 489 girls were randomised into the intervention and control arms. Eighty-five of these formed a reserve group to allow for drop-outs to be augmented (see the section on sample size earlier in this appendix for more details about the anticipated scale of drop-out). However, 40 of these were not required, meaning that ultimately 449 girls entered the trial. The overall response rate of these 449 girls at follow-up one was 95% and at follow-up two it was 91%, which represents an excellent retention rate for a trial of this kind (see Table A.4).[32]

As discussed in Chapter 1, participants who were randomised to the T&T intervention completed an additional questionnaire that collected their views of the programme. This

was administered after the main questionnaire and 194 of the T&T participants who took part in follow-up one also completed the additional questionnaire (this is equivalent to 92%).

The 16 teenagers who refused to complete the additional questionnaire had all dropped out of the programme, and chose not to take part because they felt that the additional questionnaire was not relevant to them.

Table A.4 Teens and Toddlers evaluation fieldwork outcomes

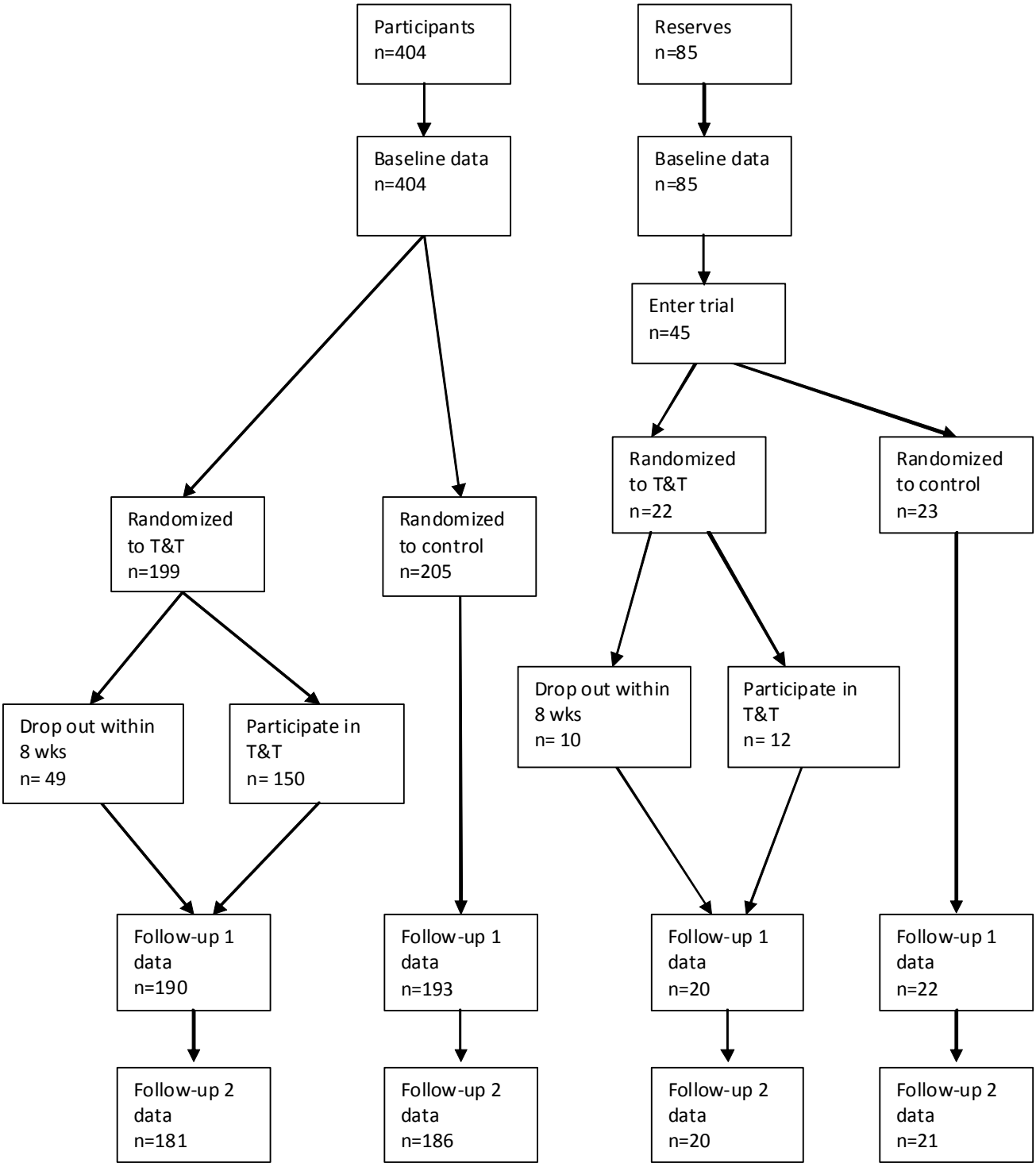
Base: All teenagers interviewed at baseline

	Follow-up one			Follow-up two		
	N	%		N	%	%
Number of girls who entered trial	449	-	100	449	-	100
Number of cases issued	449	100	100	440	100	98
Productive interview/questionnaire	425	95	95	408	93	91
No contact	9	2	2	9	2	2
- moved school and could not be contacted	*9*	*2*	*2*	*9*	*2*	*2*
Refusal	4	1	1	10	2	2
- by respondent	*2*	*0*	*0*	*5*	*1*	*1*
- broken appointment	*2*	*0*	*0*	*5*	*1*	*1*
Other non-response	11	2	2	13	3	3
- ill/absent during fieldwork	*8*	*2*	*2*	*5*	*1*	*1*
- other non-response	*3*	*1*	*1*	*8*	*2*	*2*

The differential response for girls in the intervention, control and reserve groups can be seen in Figure A.1.

[32] At follow-up two, only 440 cases were issued. This is because, at follow-up one, we found that nine girls had moved school and could not be contacted. There was therefore no way of contacting them to invite them to take part in follow-up two.

Figure A.1 Flow diagram of participation in the T&T trial



Teachers were also asked to provide attendance data for the girls who took part in the trial. In total these data were provided for 17 of the 32 groups who took part in the trial which represented a 53% response rate. These data covered 52% of the girls who took part in follow-up two of the trial (212 out of 408).

Quality control

All interviewers who work for NatCen Social Research complete a rigorous training programme to ensure that they can deliver work to the high standards NatCen Social Research requires in terms of research ethics and the quality of data collection. Furthermore, to ensure ongoing quality, interviewers are accompanied by a supervisor at different times throughout the year to monitor and develop the standard of their work. Interviewers are briefed on each project before starting work (see the section on interviewers and briefings earlier in this appendix for more information about the briefings on this project) and the quality of fieldwork on every project is monitored by NatCen Social Research.

In this study fieldwork was primarily conducted in schools and, as such, the quality of fieldwork was monitored via feedback from teachers. The contact teacher in each school was sent a letter to thank them for support throughout fieldwork. This thank you included £70 of High Street vouchers and a short evaluation form with a pre-paid return envelope. Teachers were asked whether the interviewer visited the school, whether they showed their ID card, how long the sessions took, how satisfied the teacher was with the interviewer's conduct, and whether they had any other comments about the fieldworker and manner in which fieldwork was conducted. No significant problems were revealed by the back-checking of this survey: teachers reported high satisfaction with the conduct of fieldwork, and the feedback was overwhelmingly positive.

Editing and coding

CAPI

The CAPI programme ensured that the correct routing was followed throughout the questionnaire and applied range checks, which prevented invalid values from being entered in the program, as well as consistency checks, which prompted interviewers to check answers that were inconsistent with information provided earlier in the interview. These checks allowed interviewers to clarify and query any data discrepancies directly with the respondent and were used throughout the questionnaire.

The data collected during interviews were coded and edited. The main coding/editing tasks included:

1. Back-coding of 'other' answers (this is carried out when a respondent provides an alternative answer to those that are pre-coded; this answer is recorded verbatim during the interview and is coded during the coding stage using the original list of pre-coded responses and sometimes additional codes available to coders only);

2. Checking notes that interviewers made during interviews.

These tasks were completed by a team of coders who were blind to which arm of the trial each girl had been allocated to. All the coders were briefed on the survey and given an opportunity to go through examples. If the coder could not resolve a query, this was referred to the research team.

Self-completion questionnaires

Before the data were keyed the questionnaires were visually inspected to ensure that there were no problems of completion that would affect the data entry process. After data entry, each batch of data was then submitted to a comprehensive edit programme that exhaustively checked valid ranges and routing, and made additional checks on consistency and plausibility. Error reports were referred back to the original questionnaire documents by experienced editing staff and individual corrections were specified until re-runs of the edit programme confirmed that the data were clean. At the same time the coders back-coded 'other' answers as appropriate.

Analysis

Analysis was undertaken after the data set had been edited and coded, by researchers who were aware of which arm of the trial each girl had been allocated to. The analysis file was set up in PASW 18 and all the questions and answer codes were labelled. The analysis itself, including the generation of some derived variables, tables and regression modelling, was undertaken using STATA 10.1.

The next section describes the rationale for analysing the data using an intention-to-treat approach. For more details the about analysis approach, including the adjustment for baseline differences, see Chapter 4.

Rationale for intention-to-treat analysis

The main trial analysis has been conducted on an intention-to-treat basis. This means that all intervention and control girls who were originally randomised as part of the trial were included in the analysis regardless of how many sessions of the T&T programme they attended. This approach is widely recommended as the preferred analysis strategy because it preserves the huge strengths of randomisation (see the Consolidated Standards of Reporting Trials (CONSORT Statement 2010) allowing strong inferences about cause and effect that are not justified with other study designs, and minimising bias compared to alternative approaches.

To give an example of one of the main alternatives, an 'on treatment' analysis would only include intervention girls who completed the T&T programme, and would compare them to control girls. The disadvantage of this approach is that it is subject to self-selection bias, and the findings from this trial illustrate that this would be true for any on treatment analysis of the T&T programme. This is because the girls who completed the programme were systematically different from those in the control group at baseline. In this instance, the girls who completed the T&T programme were typically less at risk at baseline than

the control group as a whole (because the girls who dropped out of T&T were some of the more risky ones, which left the less risky girls behind – whereas the equivalent girls did not drop out of the control group, see Chapter 2). This means that we would not be comparing like with like and the approach would be biased in the direction of over-estimating any benefits of the intervention. It is not possible to correct for this bias, because it is not possible to identify which girls in the control group would have completed an 18 to 20 week teenage pregnancy programme and which girls would have dropped out. For although the girls who dropped out of the T&T programme generally had a riskier profile than the girls who completed the T&T programme, no particular set of characteristics *defines* these girls and so their counterparts cannot be identified from within the control group. Also, the differences are so large that no amount of adjustment or method of statistical matching could enable a fair comparison between the two groups.[33]

The main disadvantage of conducting an intention-to-treat analysis is that the size of an intervention's impact is diluted by any girls assigned to the intervention group who dropped out of the programme. However, the impact of an intervention would still be apparent in an intention-to-treat analysis - it would just be smaller than it would have been otherwise. Moreover, in this evaluation, 73% of the girls allocated to the intervention completed the programme, which is a reasonably high retention rate. So, on balance, the disadvantage of this approach is outweighed by the substantial problems with bias that are associated with on treatment analysis. This constitutes the rationale for the analysis method employed in this study.

Registration and ethics

The trial was registered with http://clinicaltrials.gov/ on 4th March 2011 and given the ID number NCT01310543. It was finally updated on 11th April 2011.

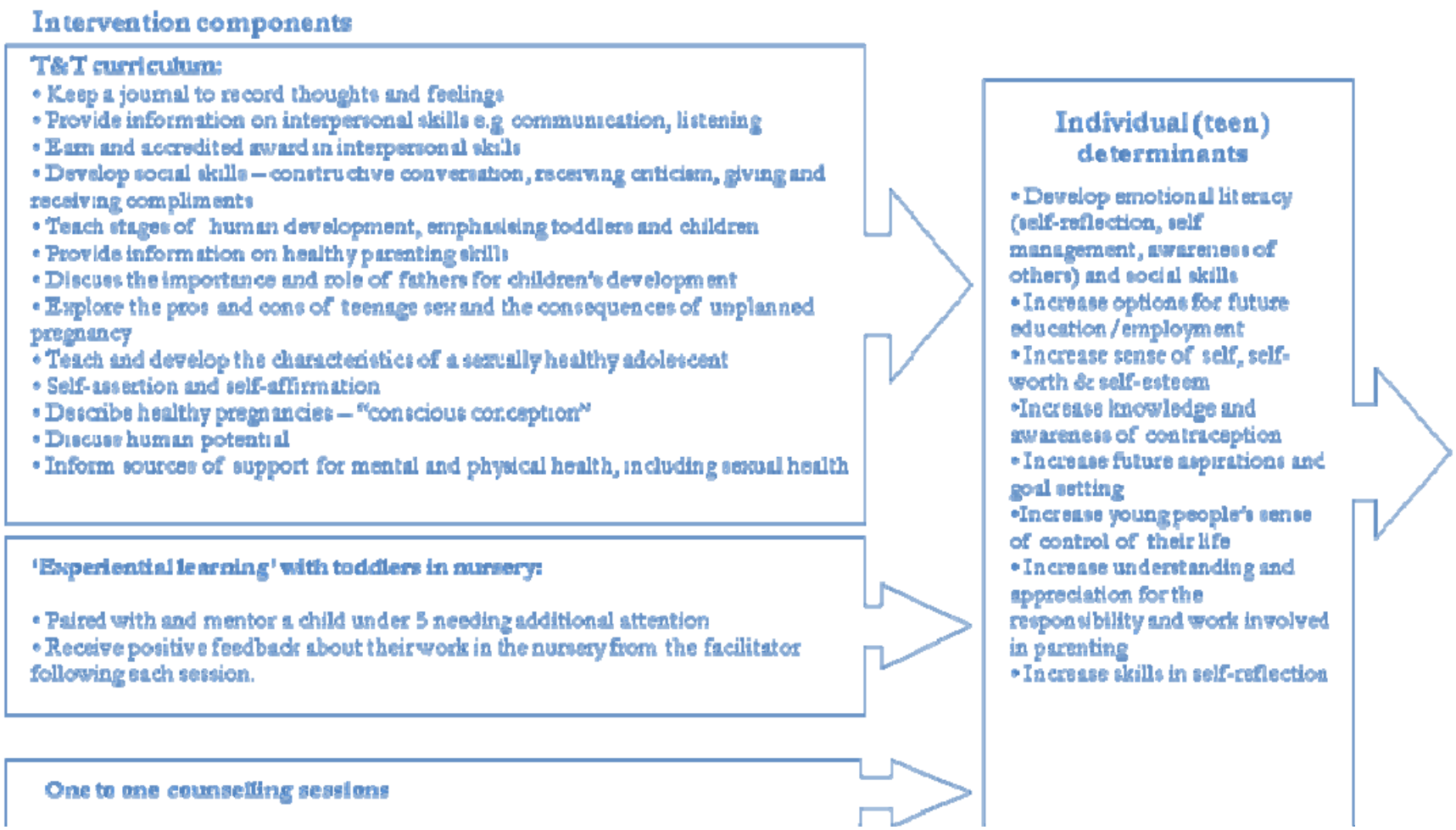
The complete list of outcomes was agreed with the DfE and the study steering group by November 2010, that is, prior to the registration of the trial in March-April 2011 and before collection of the follow-up two data.

The trial was approved by a research ethics committee at NatCen Social Research (ref P2922) and by a separate committee at the London School of Hygiene and Tropical Medicine (ref 5932).

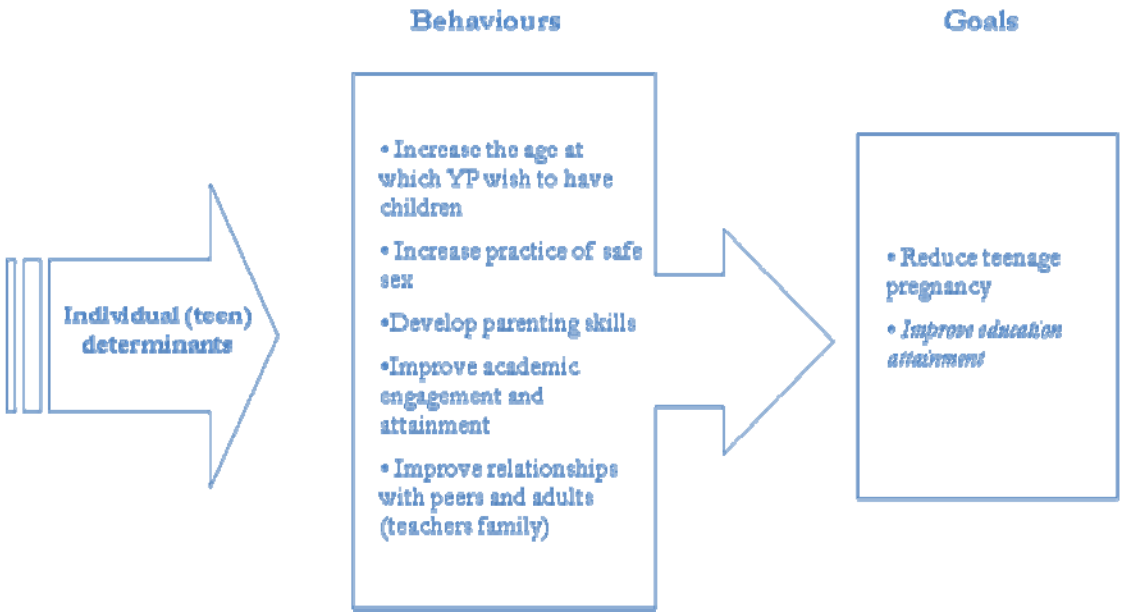
[33] Their match from the randomisation process does not constitute an equivalent counterpart because their partner is only alike in three ways (i.e. the school they attend, their age, their sexual experience). There are many other ways in which the two girls in each matched pair would differ from each other and so, whilst the intervention and control groups are balanced on a group level, the characteristics of the girls would not be balanced on an individual level.

Appendix B Logic model

Logic model


Intervention components
T&T curriculum:
• Keep a journal to record thoughts and feelings
• Provide information on interpersonal skills e.g. communication, listening
• Earn and accredited award in interpersonal skills
• Develop social skills – constructive conversation, receiving criticism, giving and receiving compliments
• Teach stages of human development, emphasising toddlers and children
• Provide information on healthy parenting skills
• Discuss the importance and role of fathers for children's development
• Explore the pros and cons of teenage sex and the consequences of unplanned pregnancy
• Teach and develop the characteristics of a sexually healthy adolescent
• Self-assertion and self-affirmation
• Describe healthy pregnancies – "conscious conception"
• Discuss human potential
• Inform sources of support for mental and physical health, including sexual health
'Experiential learning' with toddlers in nursery:
• Paired with and mentor a child under 5 needing additional attention
• Receive positive feedback about their work in the nursery from the facilitator following each session.
One to one counselling sessions
Individual (teen) determinants
• Develop emotional literacy (self-reflection, self management, awareness of others) and social skills
• Increase options for future education / employment
• Increase sense of self, self-worth & self-esteem
• Increase knowledge and awareness of contraception
• Increase future aspirations and goal setting
• Increase young people's sense of control of their life
• Increase understanding and appreciation for the responsibility and work involved in parenting
• Increase skills in self-reflection

Logic model continued…



Appendix C Additional tables

Table C2.1 Baseline characteristics (all at baseline), by trial arm

Base: All at baseline

	Control group			Intervention group			
	%	n	N	%	n	N	P value
Socio-demographic characteristics							
Age	mean= 13.5 SD=0.6 median= 14		226	mean= 13.5 SD=0.6 median= 13		220	0.436
Non-owner housing tenure	75	152	203	78	147	189	0.500
Single or reconstituted parent family	60	136	226	58	127	220	0.599
Household worklessness	31	70	226	34	75	221	0.503
Non-white ethnicity	50	114	226	51	113	220	0.846
Family's main language not English	21	48	227	21	47	220	0.955
Receives free school meals	42	96	228	46	100	218	0.423
Primary outcomes							
No contraception use at last sex in last three months	2	5	227	1	2	216	0.450
Expectation of teenage parenthood	18	41	227	26	56	216	**0.045**
Youth development score	mean= 49.5 SD=8.2 median= 50.5		222	mean= 48.8 SD=8.9 median= 49		217	0.379
Secondary outcomes							
No condom use at last sex in last three months	4	8	228	2	5	218	0.446
More than one episodes of no condom use in last three months	5	11	228	5	10	221	0.880
Best age to have sex under 16	19	39	210	22	44	203	0.431
Favourable to not using protection sometimes in sex	8	18	225	10	22	215	0.415
Low self-reflection	7	17	227	14	30	218	**0.031**
Low emotional vocabulary	12	27	226	15	33	219	0.335

Low self-esteem	15	33	226	14	30	219	0.785
Dislike of school	33	76	228	31	69	220	0.656
Lack of expectation of post-16 education, training or employment	2	5	222	3	6	219	0.743
Low sexual health knowledge	83	184	223	92	195	212	**0.003**
Difficulty of discussing sex with boyfriend	54	123	226	57	124	216	0.528
Difficulty of discussing the pill with a doctor	65	146	225	60	130	216	0.308
Has been pregnant	2	4	228	3	6	221	0.539
Lack of awareness of impact of parenthood on social life	15	34	224	15	33	215	0.960
Other characteristics							
Talking to mother or father about personal things difficult	21	46	217	23	49	212	0.633
No other adult to talk to about personal things	25	56	227	28	60	218	0.493
Not important to get good marks in school work	1	2	228	2	4	220	0.442
Misses school without permission	33	75	227	36	79	217	0.456
Has been suspended or temporarily excluded from school in last six months	11	26	227	11	23	220	0.735
Has been expelled from school	8	18	228	6	14	219	0.538
Current boy/girlfriend	35	79	227	33	72	221	0.619
Boy/girl friend pushes to do unwanted things	2	4	227	2	4	220	1.000
Kiss/cuddle with a boy/man	79	176	223	77	768	219	0.576
Genital/hand contact with a boy/man	18	41	222	23	49	216	0.275
Sexually active (had sex with a boy/man)	13	30	228	13	29	218	0.964
More than one sexual partner	6	13	228	8	17	221	0.398
Sex monthly or more	4	9	228	5	10	221	0.761
Age of sexual debut	mean= 13.1 SD=0.7 median= 13		29	mean= 13.1 SD=0.8 median= 13		29	0.863
First sex unplanned	9	20	222	6	13	213	0.253

First sex regretted or too young	8	17	219	9	19	215	0.685
First sex involved pressure from one partner	3	6	227	3	6	219	0.950
First sex without contraception	1	2	225	1	2	221	1.000
Most recent pregnancy unplanned	1	2	228	2	4	220	0.442
Most recent pregnancy terminated	1	2	228	1	2	221	1.000
Favourable to sex early in relationship	3	6	206	2	5	210	0.735
Favourable to not using condom with new partner	8	18	223	6	14	216	0.522
Favourable to smoking during pregnancy	2	5	222	2	4	217	1.000
Favourable to heavy drinking during pregnancy	0	1	228	0	1	218	1.000
Condom use difficult	50	112	226	51	111	216	0.700
Drunk ever	36	81	225	38	84	220	0.634
Drunk monthly or more	18	41	227	20	43	220	0.688
Worries a lot	68	153	225	78	170	218	**0.018**
Recently unhappy, downhearted or tearful	38	84	222	42	92	217	0.330
Parents only sometimes or never know where and/or with whom children are when out	27	62	228	33	72	218	0.179
Parents only sometimes or never set time for coming home	13	29	228	17	36	218	0.256
Does not discuss school with parents very often	54	122	228	53	115	218	0.873
Parents rarely or never say done something well	12	28	227	15	32	218	0.469
Not proud of friends	26	60	228	22	47	217	0.251
Friends do not support	21	49	228	21	45	217	0.846
Friends don't work hard at school	48	109	228	49	107	217	0.751
Parents don't know most or all friends	32	74	228	39	85	218	0.150
Best age to become a mum under 19	7	15	228	5	11	217	0.497
Family member had baby before age 20	60	134	224	68	146	214	**0.067**
Believes baby in teens would make them feel more	26	60	227	28	60	215	0.727

grown up							
Believes baby in teens would given them someone to love	43	96	225	42	90	215	0.864
Believes baby in teens would make life easier in some way	24	53	222	24	51	212	0.964
Believes being a good parent easy	38	86	226	34	73	216	0.351
Attended a personal development or education programme for young people other than T&T*	25	49	198	20	39	195	0.259

* Based on data from follow-ups one and two and on time period during and post T&T intervention.

Notes: The base for all questions is the whole baseline sample, including for sexually-related behaviours.
P values refer to: chi-squared tests where the baseline characteristic is a categorical variable with an expected cell count of five or more, t tests where it is continuous, and Fisher's exact test (two-sided) where the baseline variable is categorical but at least one of the expected cell counts is under five.

Table C2.2 Baseline characteristics (follow-up one participants), by trial arm							
Base: Those at baseline who also took part in follow-up one							
	Control group			Intervention group			P value
	%	n	N	%	n	N	
Socio-demographic characteristics							
Age	mean= 13.5 SD=0.6 median= 13		213	mean= 13.5 SD=0.6 median= 13		209	0.684
Non-owner housing tenure	75	142	190	77	139	180	0.576
Single or reconstituted parent family	60	128	213	58	122	209	0.719
Household worklessness	30	64	213	34	72	210	0.351
Non-white ethnicity	51	109	214	52	108	209	0.879
Family's main language not English	21	45	214	21	43	209	0.908
Receives free school meals	43	93	215	46	95	207	0.586
Primary outcomes							
No contraception use at last sex in last three months	2	4	214	1	2	205	0.686
Expectation of teenage parenthood	18	38	215	24	49	206	0.122
Youth development score	mean= 49.7 SD=7.9 median= 51		210	mean= 49.0 SD=8.9 median= 49.5		206	0.417
Secondary outcomes							
No condom use at last sex in last three months	3	6	215	2	5	207	0.809
More than one episode of no condom use in last three months	4	8	215	5	10	210	0.594
Best age to have sex under 16	19	38	198	21	40	194	0.724
Favourable to not using protection sometimes in sex	8	18	212	10	21	204	0.528
Low self-reflection	7	15	215	13	27	207	**0.037**
Low emotional vocabulary	12	26	214	15	31	208	0.408
Low self-esteem	14	30	214	14	30	208	0.905
Dislike of school	33	70	215	29	61	209	0.453
Lack of expectation of post-16 education, training or	2	5	210	2	4	208	1.000

employment							
Low sexual health knowledge	83	175	210	92	185	201	**0.007**
Difficulty of discussing sex with boyfriend	55	118	213	59	120	205	0.517
Difficulty of discussing the pill with a doctor	66	140	212	62	127	205	0.385
Has been pregnant	1	3	215	3	6	210	0.333
Lack of awareness of impact of parenthood on social life	15	31	211	16	32	205	0.794
Other characteristics							
Talking to mother or father about personal things difficult	19	40	206	23	47	201	0.329
No other adult to talk to about personal things	23	49	214	27	56	208	0.339
Not important to get good marks in school work	1	2	215	2	4	209	0.444
Misses school without permission	31	67	214	35	73	206	0.370
Has been suspended or temporarily excluded from school in last six months	12	25	214	10	21	209	0.589
Has been expelled from school	8	17	215	6	13	208	0.507
Current boy/girlfriend	35	75	214	32	67	210	0.493
Boy/girl friend pushes to do unwanted things	2	4	214	2	4	209	1.000
Kiss/cuddle with a boy/man	79	166	210	75	157	208	0.384
Genital/hand contact with a boy/man	18	37	210	22	46	206	0.229
Sexually active (had sex with a boy/man)	12	26	215	14	28	207	0.659
More than one sexual partner	5	10	215	8	17	210	0.146
Sex monthly or more	3	7	215	5	10	210	0.428
Age of sexual debut	mean= 13.1 SD=0.8 median= 13		25	mean= 13.1 SD=0.8 median= 13		28	0.968
First sex unplanned	9	18	210	6	12	202	0.304
First sex regretted or too young	7	15	207	9	19	205	0.456
First sex involved pressure from one partner	2	4	214	3	6	208	0.539

First sex without contraception	1	2	212	1	2	210	1.000
Most recent pregnancy unplanned	0	1	215	2	4	209	0.210
Most recent pregnancy terminated	0	1	215	1	2	210	0.620
Favourable to sex early in relationship	3	6	193	3	5	200	0.715
Favourable to not using condom with new partner	7	15	210	6	13	205	0.745
Favourable to smoking during pregnancy	2	5	209	2	4	206	1.000
Favourable to heavy drinking during pregnancy	0	1	215	0	1	207	1.000
Condom use difficult	50	106	213	54	111	205	0.370
Drunk ever	35	75	212	36	76	209	0.833
Drunk monthly or more	18	39	214	19	39	209	0.908
Worries a lot	67	142	213	77	159	207	**0.021**
Recently unhappy, downhearted or tearful	38	79	210	41	84	206	0.509
Parents only sometimes or never know where and/or with whom children are when out	27	57	215	32	67	207	0.187
Parents only sometimes or never set time for coming home	13	28	215	16	33	207	0.394
Does not discuss school with parent very often	53	115	215	51	107	208	0.674
Parents rarely or never say done something well	11	24	215	14	30	208	0.315
Not proud of friends	24	52	215	20	43	207	0.401
Friends do not support	20	43	215	19	40	207	0.861
Friends don't work hard at school	47	101	215	48	100	207	0.784
Parents don't know most or all friends	32	69	215	38	80	208	0.170
Best age to become a mum under 19	6	13	215	4	9	207	0.433
Family member had baby before age 20	61	128	211	68	138	204	0.138
Believes baby in teens would make them feel more grown up	28	60	214	28	57	205	0.958
Believes baby in teens would given them someone to love	43	92	212	41	85	205	0.690

Believes baby in teens would make life easier in some way	23	49	210	24	48	203	0.940
Believes being a good parent easy	38	82	214	34	69	206	0.303
Attended a personal development or education programme for young people other than T&T*	25	49	198	20	39	195	0.259

* Based on data from follow-ups one and two and on time period during and post T&T intervention.

Notes: The base for all questions is the whole baseline sample, including for sexually-related behaviours.

P values refer to: chi-squared tests where the baseline characteristic is a categorical variable with an expected cell count of five or more, t tests where it is continuous, and Fisher's exact test (two-sided) where the baseline variable is categorical but at least one of the expected cell counts is under five.

Table C2.3 Baseline characteristics (follow-up two participants), by trial arm							
Base: Those at baseline who also took part in follow-up two							
	Control group			Intervention group			P value
	%	n	N	%	n	N	
Socio-demographic characteristics							
Age	mean= 13.5 SD=0.6 median= 14		205	mean= 13.5 SD=0.6 median= 13		200	0.377
Non-owner housing tenure	75	137	183	78	134	172	0.500
Single or reconstituted parent family	60	124	206	58	115	200	0.581
Household worklessness	29	59	206	34	68	201	0.259
Non-white ethnicity	50	103	206	51	102	200	0.840
Family's main language not English	21	44	206	21	41	200	0.832
Receives free school meals	41	85	207	46	92	199	0.294
Primary outcomes							
No contraception use at last sex in last three months	2	5	206	1	2	196	0.450
Expectation of teenage parenthood	19	39	207	25	50	197	0.113
Youth development score	mean= 49.6 SD=8.1 median= 51		203	mean= 49.0 SD=8.8 median= 49		197	0.510
Secondary outcomes							
No condom use at last sex in last three months	3	7	207	3	5	198	0.611
More than one episodes of no condom use in last three months	5	10	207	5	10	201	0.946
Best age to have sex under 16	18	34	191	21	39	185	0.421
Favourable to not using protection sometimes in sex	8	17	204	10	20	195	0.508
Low self-reflection	8	17	207	14	27	198	**0.080**
Low emotional vocabulary	13	26	206	15	30	199	0.474
Low self-esteem	15	32	207	14	28	199	0.693
Dislike of school	33	69	207	31	61	200	0.540
Lack of expectation of post-16 education, training or	2	5	203	2	4	199	1.000

employment							
Low sexual health knowledge	83	170	204	92	179	194	**0.007**
Difficulty of discussing sex with boyfriend	55	114	206	58	115	197	0.539
Difficulty of discussing the pill with a doctor	66	135	205	61	121	197	0.356
Has been pregnant	1	3	207	3	6	201	0.332
Lack of awareness of impact of parenthood on social life	15	31	204	16	31	197	0.881
Other characteristics							
Talking to mother or father about personal things difficult	21	42	198	24	46	194	0.553
No other adult to talk to about personal things	25	52	206	28	56	198	0.490
Not important to get good marks in school work	0	1	207	2	3	200	0.365
Misses school without permission	32	66	206	36	71	197	0.397
Has been suspended or temporarily excluded from school in last six months	12	25	206	10	20	200	0.493
Has been expelled from school	8	16	207	6	12	199	0.499
Current boy/girlfriend	35	73	206	32	64	201	0.443
Boy/girl friend pushes to do unwanted things	2	4	206	2	4	200	1.000
Kiss/cuddle with a boy/man	79	160	203	75	150	199	0.412
Genital/hand contact with a boy/man	19	38	203	23	45	197	0.309
Sexually active (had sex with a boy/man)	13	27	207	14	28	198	0.747
More than one sexual partner	6	12	207	8	17	201	0.296
Sex monthly or more	3	7	207	5	10	201	0.421
Age of sexual debut	mean= 13.2 SD=0.7 median= 13		26	mean= 13.1 SD=0.8 median= 13		28	0.688
First sex unplanned	9	19	202	6	12	193	0.239
First sex regretted or too young	8	15	198	10	19	196	0.454
First sex involved pressure from one partner	2	5	206	3	6	199	0.716

First sex without contraception	1	2	204	1	2	201	1.000
Most recent pregnancy unplanned	0	1	207	2	4	200	0.209
Most recent pregnancy terminated	0	1	207	1	2	201	0.619
Favourable to sex early in relationship	3	5	188	3	5	192	1.000
Favourable to not using condom with new partner	8	16	202	6	12	196	0.483
Favourable to smoking during pregnancy	2	5	202	2	4	197	1.000
Favourable to heavy drinking during pregnancy	0	1	207	1	1	198	1.000
Condom use difficult	50	103	206	54	106	197	0.445
Drunk ever	36	75	206	38	75	200	0.820
Drunk monthly or more	18	37	206	19	38	200	0.787
Worries a lot	67	139	206	78	152	198	**0.038**
Recently unhappy, downhearted or tearful	40	82	204	42	83	197	0.694
Parents only sometimes or never know where and/or with whom children are when out	29	59	207	34	67	198	0.246
Parents only sometimes or never set time for coming home	14	29	207	17	33	198	0.458
Does not discuss school with parent very often	54	111	207	53	106	199	0.943
Parents rarely or never say done something well	12	24	207	15	30	199	0.302
Not proud of friends	25	52	207	21	42	199	0.338
Friends do not support	22	45	207	20	39	199	0.594
Friends don’t work hard at school	48	99	207	47	93	199	0.826
Parents don’t know most or all friends	30	63	207	40	79	199	**0.050**
Best age to become a mum under 19	7	14	207	5	9	198	0.335
Family member had baby before age 20	60	122	203	67	130	195	0.174
Believes baby in teens would make them feel more grown up	27	56	206	27	54	197	0.959
Believes baby in teens would given them someone to love	43	88	205	42	83	196	0.907

Believes baby in teens would make life easier in some way	24	49	204	25	48	195	0.890
Believes being a good parent easy	37	77	206	33	65	197	0.357
Attended a personal development or education programme for young people other than T&T*	25	49	198	20	39	195	0.259

* Based on data from follow-ups one and two and on time period during and post T&T intervention.

Notes: The base for all questions is the whole baseline sample, including for sexually-related behaviours.

P values refer to: chi-squared tests where the baseline characteristic is a categorical variable with an expected cell count of five or more, t tests where it is continuous, and Fisher's exact test (two-sided) where the baseline variable is categorical but at least one of the expected cell counts is under five.

Table C3.1 How many times teenagers talked to the counsellor

Base: Teenagers randomised to intervention

	%
None	8
Once	11
2-3 times	37
4-5 times	24
More than 5 times	19
Base	*193*

Table C3.2 Elements of the programme that teenagers enjoyed

Base: Teenagers randomised to intervention

	Elements of the programme				
	Toddler time	Group work	Writing journal	Counselling	T&T overall
	%	%	%	%	%
A lot	77	49	34	49	77
A little	18	27	33	27	17
Not much	3	17	23	11	4
Not at all	3	6	9	13	2
Bases	*190*	*190*	*180*	*179*	*191*

Table C3.3 Elements of the programme that teenagers found difficult or challenging					
Base: Teenagers randomised to intervention					
	Elements of the programme				
	Toddler time	Group work	Writing journal	Counselling	T&T overall
	%	%	%	%	%
Never	24	33	37	56	37
Not very often	53	48	42	24	46
Fairly often	19	15	16	13	14
Very often	3	4	4	7	3
Bases	*188*	*189*	*182*	*176*	*189*

Table C3.4 Aspects of the programme that teenagers thought were the best	
Base: Teenagers randomised to intervention	
	%
My toddler	81
Time spent in nursery	74
Things learnt through doing T&T	73
Made me feel good about myself	63
Spending time with other girls	56
Getting the qualification	51
Role playing	45
Check-in time	44
Counselling	44
Group work	44
Nursery staff	39
The T&T facilitator	39
Writing journal	30
Other	8
Base	*192*

Table C3.5 Aspects of the programme that teenagers found difficult or challenging	
Base: Teenagers randomised to intervention	
	%
Role playing	27
Writing journal	27
Too much work	23
Counselling	23
Group work	20
T&T facilitator	18
Check-in time	17
Nursery staff	16
Feeling anxious, angry, sad or bored	16
My toddler	13
Time spent in nursery	8
Spending time with other girls	8
Not enough help and support	5
Work was too hard	4
Other	9
Base	*166*

Table C3.6 Positive feelings about the programme	
Base: Teenagers randomised to intervention	
	%
Responsible	65
Happy	64
Positive about things	61
Interested	59
Excited	46
None of these	6
Base	*193*

Table C3.7 Negative feelings about the programme	
Base: Teenagers randomised to intervention	
	%
Bored	37
Irritated or annoyed	20
Frustrated	16
Sad or down	14
Angry	7
Anxious	7
None of these	44
Base	*182*

Table C3.8 Lessons missed through doing T&T	
Base: Teenagers randomised to intervention	
	%
Science	27
Physical Education	23
English	22
Art	20
Mathematics	20
Design Technology	17
History	14
Languages	13
Geography	12
Drama	9
Religious Education	9
Food Technology	7
Music	7
Information Communication Technology	6
Personal, Social and Health Education	6
Other	11
Base	*183*

Table C3.9 Whether teenagers fell behind with their school work	
Base: Teenagers randomised to intervention	
	%
Agree strongly	14
Agree	17
Neither agree nor disagree	24
Disagree	20
Disagree strongly	25
Base	*185*

Table C3.10 How often teenagers talked to their peers about T&T	
Base: Teenagers randomised to intervention	
	%
Always	20
Usually	25
Sometimes	37
Never	18
Base	*186*

Table C4.1 Primary and secondary outcomes at follow-up one (with details), by trial arm				
Base: All at follow-up one				
	Control group	**Intervention group**	**Unadjusted**	**Adjusted***
	n/N (%)	**n/N (%)**	**OR /coef. (95% CI) P-value**	**OR /coef. (95% CI) P-value**
Primary outcomes				
No contraception at last sex in last three months	1/213 (0)	5/207 (2)	OR 5.2 (0.6-45.3) 0.132	n/a
Expects teenage pregnancy	54/207 (26)	61/205 (30)	OR 1.2 (0.8-1.8) 0.407	OR 1.2 (0.7-1.8) 0.516
Youth development score	N=208 mean=47.6 SD=8.6 median=48	N=203 mean=47.6 SD=8.8 median=48	Coef. -0.03 (-1.7-1.7) 0.973	Coef. 0.4 (-1.2-2.1) 0.606
Secondary outcomes				
No condom use at last sex in last three months	8/214 (4)	9/208 (4)	OR 1.2 (0.4-3.1) 0.759	n/a
>1 episode of no condom use in last three months	11/215 (5)	15/210 (7)	OR 1.4 (0.6-3.2) 0.385	n/a
Best age to have first sex under 16	36/198 (18)	37/196 (19)	OR 1.0 (0.6-1.7) 0.859	OR 1.1 (0.6-1.9) 0.741
Favourable to not using protection sometimes in sex	18/212 (8)	21/204 (10)	OR 1.2 (0.6-2.4) 0.529	n/a
Low self-reflection	25/209 (12)	31/208 (15)	OR 1.3 (0.7-2.3) 0.379	OR 1.1 (0.6-1.9) 0.843
Low emotional vocabulary	25/211 (12)	39/207 (19)	OR 1.7 (1.0-3.0) **0.049**	OR 1.6 (0.9-2.8) 0.098
Low self-esteem	50/212 (24)	34/207 (16)	OR 0.6 (0.4-1.0) 0.068	OR 0.5 (0.3-0.9) **0.010**
Dislike of school	91/215 (42)	85/208 (41)	OR 0.9 (0.6-1.4) 0.761	OR 0.9 (0.6-1.3) 0.536
Lack of expectation of post-16 education, training or empl.	3/212 (1)	6/206 (3)	OR 2.1 (0.5-8.5) 0.302	n/a
Low sexual health knowledge	165/209 (79)	150/205 (73)	OR 0.7 (0.5-1.1) 0.169	OR 0.5 (0.3-0.9) **0.017**
Difficulty of discussing sex with a boyfriend	80/210 (38)	88/204 (43)	OR 1.2 (0.8-1.8) 0.296	OR 1.2 (0.8-1.7) 0.465
Difficulty of discussing the pill with a doctor	117/209 (56)	90/204 (44)	OR 0.6 (0.4-0.9) **0.016**	n/a
New pregnancy since baseline	4/215 (2)	4/210 (2)	OR 1.0 (0.3-4.1) 0.973	n/a
Lack of awareness of impact of parenthood on social life	33/213 (15)	30/204 (15)	OR 0.9 (0.5-1.6) 0.823	OR 0.9 (0.5-1.5) 0.614

* For details of adjustments, see Table C4.2 below.

Table C4.2 Covariates in models for outcomes at follow-up one			
	Covariates at baseline		
Outcomes at follow-up one	Low self-reflection	Low sexual health knowledge	Worries a lot
Primary outcomes			
No contraception at last sex in last three months			
Expects teenage pregnancy	√		
Youth development score	√		√
Secondary outcomes			
No condom use at last sex in last three months			
>1 episodes of no condom use in last three months			
Best age to have first sex under 16		√	√
Favourable to not using protection sometimes in sex			
Low self-reflection	√		√
Low emotional vocabulary			√
Low self-esteem	√		√
Dislike of school	√		
Lack of expectation of post-16 education, training or employment			
Low sexual health knowledge		√	
Difficulty of discussing sex with a boyfriend		√	
Difficulty of discussing the pill with a doctor			
New pregnancy since baseline			
Lack of awareness of impact of parenthood on social life	√		

Note: The three baseline variables in this table were checked as potential covariates for impacts at follow-up one because there were significant differences between the intervention and control groups at baseline on these parameters (see Table C2.2). The decision about whether to include these variables as covariates in the models for outcomes at follow-up one was based on whether there was a significant bivariate association (at $p<0.1$) between the prospective confounder and the outcome. If the bivariate association was not significant, it was deemed not necessary to include the baseline variable in question in the model for that outcome.

Table C4.3 Primary and secondary outcomes at follow-up two (with details), by trial arm				
Base: All at follow-up two				
	Control group	**Intervention group**	**Unadjusted**	**Adjusted***
	n/N (%)	**n/N (%)**	**OR /coef. (95% CI) P-value**	**OR /coef. (95% CI) P-value**
Primary outcomes				
No contraception at last sex in last three months	9/202 (4)	11/197 (6)	OR 1.3 (0.5-3.1) 0.606	n/a
>1 episodes of no contraception in last three months	25/207 (12)	26/200 (13)	OR 1.1 (0.6-2.0) 0.779	n/a
Expects teenage pregnancy	52/207 (25)	49/200 (25)	OR 1.0 (0.6-1.5) 0.885	n/a
Youth development score	N=205 mean=49.4 SD=8.1 median=51	N=198 mean=47.3 SD=9.2 median=48	Coef. -2.1 (-3.8--0.4) **0.015**	Coef. -1.4 (-3.1--0.3) 0.111
Secondary outcomes				
No condom use at last sex in last three months	22/203 (11)	24/199 (12)	OR 1.1 (0.6-2.1) 0.700	OR 1.1 (0.6-2.0) 0.804
>1 episodes of no condom use in last three months	29/206 (14)	32/201 (16)	OR 1.2 (0.7-2.0) 0.603	OR 1.1 (0.6-1.9) 0.786
Best age to have first sex under 16	34/193 (18)	30/185) (16)	OR 0.9 (0.5-1.6) 0.717	n/a
Favourable to not using protection sometimes in sex	21/202 (10)	17/199 (9)	OR 0.8 (0.4-1.6) 0.527	n/a
Low self-reflection	19/205 (9)	29/200 (15)	OR 1.7 (0.9-3.1) 0.106	OR 1.5 (0.8-2.9) 0.169
Low emotional vocabulary	31/206 (15)	36/201 (18)	OR 1.2 (0.7-2.1) 0.437	OR 1.1 (0.6-1.8) 0.806
Low self-esteem	47/205 (23)	31/201 (15)	OR 0.6 (0.4-1.0) 0.056	OR 0.6 (0.3-1.0) **0.032**
Dislike of school	95/207 (46)	93/201 (46)	OR 1.0 (0.7-1.5) 0.939	OR 1.0 (0.7-1.4) 0.881
Lack of expectation of post-16 education, training or employment.	2/206 (1)	2/200 (1)	OR 1.0 (0.1-7.4) 0.976	n/a
Low sexual health knowledge	134/202 (66)	137/197 (70)	OR 1.2 (0.8-1.8) 0.493	OR 1.0 (0.6-1.5) 0.843
Difficulty of discussing sex with a boyfriend	65/204 (32)	63/196 (32)	OR 1.0 (0.7-1.5) 0.952	OR 0.9 (0.6-1.5) 0.775
Difficulty of discussing the pill with a doctor	83/205 (40)	86/195 (44)	OR 1.2 (0.8-1.7) 0.465	OR 1.1 (0.7-1.6) 0.696
New pregnancy since baseline	12/207	9/201	OR 0.8 (0.3-1.8)	n/a

	(6)	(4)	0.548	
Lack of awareness of impact of parenthood on social life	27/205 (13)	20/200 (10)	OR 0.7 (0.4-1.4) 0.321	n/a
Number of school days missed	N=112 mean=2.3 SD=3.4 median=1	N=100 mean=2.2 SD=3.9 median=1	Coef. -0.1 (-1.1-0.8) 0.786	Coef. -0.2 (-1.2-0.8) 0.647

* For details of adjustments, see Table C4.4 below.

Table C4.4 Covariates in models for outcomes at follow-up two				
	Covariates at baseline			
Outcomes at follow-up two	**Low self-reflection**	**Low sexual health knowledge**	**Worries a lot**	**Parents don't know most or all friends**
Primary outcomes				
No contraception at last sex in last three months				
>1 episodes of no contraception in last three months				
Expects teenage pregnancy				
Youth development score	√		√	√
Secondary outcomes				
No condom use at last sex in last three months			√	
>1 episodes of no condom use in last three months			√	
Best age to have first sex under 16				
Favourable to not using protection sometimes in sex				
Low self-reflection	√			√
Low emotional vocabulary	√		√	√
Low self-esteem	√	√	√	
Dislike of school	√			
Lack of expectation of post-16 education, training or employment				
Low sexual health knowledge	√	√	√	
Difficulty of discussing sex with a boyfriend		√		
Difficulty of discussing the pill with a doctor	√			
New pregnancy since baseline				
Lack of awareness of impact of parenthood on social life				
Number of school days missed		√		

Note: The four baseline variables in this table were checked as potential covariates for impacts at follow-up two because there were significant differences between the intervention and control groups at baseline on these parameters (see Table C2.3). The decision about whether to include these variables as covariates in the models for outcomes at follow-up two was based on whether there was a significant bivariate association (at $p<0.1$) between the prospective confounder and the outcome. If the bivariate association was not significant, it was deemed not necessary to include the baseline variable in question in the model for that outcome.

Table C4.5 Proportion of teenagers who have had sex with a boy (man) at three waves, by trial arm				
Base: All at baseline, follow-up one or follow-up two				
Sexually active	**Control group** n/N (%)	**Intervention group** n/N (%)	**Unadjusted** OR /coef. (95% CI) P-value	**Adjusted*** OR /coef. (95% CI) P-value
At baseline	30/228 (13)	29/218 (13)	OR 1.0 (0.6-1.8) 0.964	n/a
At follow-up one	42/211 (20)	41/209 (20)	OR 1.0 (0.6-1.6) 0.941	OR 1.1 (0.7-1.8) 0.761
At follow-up two	76/205 (37)	65/199 (33)	OR 0.8 (0.5-1.2) 0.353	OR 0.9 (0.6-1.4) 0.611

*Both follow-up one and follow-up two results are adjusted for baseline sexual health knowledge and worry.

Table C4.6Self-esteem: Whether agreed with the sentence "I like myself" at three waves, by trial arm						
Base: All at baseline, follow-up one or follow-up two						
	Baseline		**Follow-up one**		**Follow-up two**	
Whether agreed	Control	Interv.	Control	Interv.	Control	Interv.
Agreed strongly	28	29	31	30	28	21
Agreed	30	32	25	30	23	30
Neither agreed nor disagreed	28	25	20	24	26	33
Disagreed	7	7	12	11	11	7
Disagreed strongly	8	6	12	6	12	8
Bases	*226*	*219*	*212*	*207*	*205*	*201*

Table C4.7	Youth development score components at follow-up two, by trial arm					
Base: All at follow-up two						
Youth development score components	**Control**		**Intervention**		**Coef. (CI)**	**P value**
	Mean	**SD**	**Mean**	**SD**		
Overall most things I do turn out well	5.2	1.4	5.0	1.6	-0.3 (-0.6-0.0)	0.091
I know I have the ability to do anything I want to do	6.0	1.7	5.7	2.0	-0.3 (-0.7-0.1)	0.106
My own efforts and actions are what will shape my future	6.8	1.5	6.6	1.6	-0.2 (-0.5-0.1)	0.252
I work hard at solving my problems	6.1	1.6	5.9	1.7	-0.2 (-0.6-0.1)	0.140
Personal goals are important to me	6.5	1.7	6.5	1.6	-0.03 (-0.4-0.3)	0.848
I respect other people	6.8	1.4	6.6	1.5	-0.2 (-0.5-0.1)	0.130
I sort out my conflicts with other people peacefully	5.5	2.1	5.0	2.0	-0.6 (-1.0--0.2)	**0.007**
People understand me when I am talking	6.4	1.6	6.1	1.8	-0.3 (-0.6-0.04)	0.089
Bases		*206-207*		*199-201*		

Note: The score for each item is based on the scale from 1 “false – not like you” to 8 “true – like you”. The coefficients, CIs and p values are from linear regression models (not adjusting for any relevant baseline differences).

Appendix D Study materials

Published as a separate annex accompanying this report – Randomised controlled trial of the ‘Teens and Toddlers’ programme –Study Materials (Appendix D)

Ref: DFE-RR211

ISBN: 978-1-78105-105-4

May 2012