BJPsych The British Journal of Psychiatry (2010)
196, 372–376. doi: 10.1192/bjp.bp.109.066050

Online supplement

Details of analysis to identify confounders

Imbalances in initial Lambeth Early Onset (LEO) trial

Factors that were unbalanced at baseline (due to failure of randomisation to produce balanced groups when using small sample sizes) between the two randomised groups were entered into the regression model. This adjustment was performed by the first LEO study and we replicate this analysis to increase comparability of the two studies. Both crude and adjusted ratios are given as adjusting for variables that are known to be unbalanced at baseline and could potentially introduce error into the analyses. Randomisation potentially removes confounding for both known and unknown confounders but may have been ineffectual in the original LEO study because of insufficient size. Adjusting for imbalances in known confounders could potentially introduce confounding by creating imbalances in unknown confounders.

Confounders in second trial

Potential confounders were chosen after review of the literature identified variables that could potentially be related to both outcome and chance of being traced for the LIFE study. As contactability lies on the causal pathway between exposure and outcome of interest, confounders of the relationship between contactability and outcome are very likely to also act in the relationship between randomisation group and outcomes. The following confounders were identified: ethnicity;[20] gender;[21] age;[22] duration of untreated psychosis;[23] whether or not the person was in a relationship at baseline;[24] whether or not the person was working at baseline;[25] whether or not the person was in education at baseline;[26] migrant status.[27]

Although the list is not exhaustive, these factors had the strongest relationship with outcomes of those considered. Specialist psychiatrists were consulted about other factors they considered important. The following factors are therefore also considered: discharge to general practitioner; and total time with psychiatric services.

Cannabis misuse at baseline was a potential confounder but could not be assessed as this information had not been collected in the original study. Duration of untreated psychosis is a potential confounder and was recorded at baseline. However, the recorded values were judged to have poor reliability and so were not used in further analyses.

All potential confounders were entered sequentially into a logistic regression model to ascertain strength of relationship with contactability at follow-up. Any variables thus identified were considered as potential confounders. There was insufficient sample size to examine the potential role for interaction between variables that were entered one at a time. Relationship with outcome (where known for participants) was similarly assessed by sequential entry of any variables identified in the first analysis into a logistic regression model with 'ever been admitted' as the primary outcome of interest.

Table DS1 shows the results of modelling for an association between contactability and the potential confounding variables.

Details of statistical modelling for missing data-sensitivity analysis

We were not able to contact 30% of participants. It is possible that this cohort may have had a systematically better or worse outcome than those who had been traced and thus the overall results obtained may have been biased. We modelled for missing data from uncontactable participants using sensitivity analyses in which we considered the possibilities of this missing cohort having different admission rates from those who had been contactable. The underlying assumption is that the missing individuals were 'not missing at random' (i.e. that missingness was not as a result of completely random loss of data (which would be termed 'missing completely at random'). Examples of mechanisms underlying missingness are demonstrated in Appendix DS1.

For missing 'not at random' data we assume that participants who were not followed up were less or more likely not to have ever been admitted potentially because they had the best or worse outcomes after the end of the LEO study. Data were modelled using a complete data-set that included both values of 'missing' as well as predicted outcome values.

Sensitivity analysis was employed to assess the robustness of the outcomes by randomly substituting possible outcome values for missing data. A fixed proportion of data from uncontactable participants was randomly chosen to be assigned the most severe outcome (ever admitted to hospital), whereas the missing data from the other participants was simultaneously assigned the best outcome (i.e. 'not admitted'). An alternative strategy would have been to set a proportion of the missing values to 'admitted' while simultaneously setting the other missing values to 'missing'. This analysis was not selected as the odds ratios thus produced for each permutation are less directly comparable.

Initially, 20% of patients for whom outcome data were missing in the care as usual arm were randomly selected and assigned the worse outcome value (ever admitted in the follow-up period). Participants were randomly selected by the computer programme STATA on Windows Vista. The subsequent odds ratio and 95% intervals of uncertainty of ever been admitted by randomised group were then calculated. This procedure was carried out 100 times and the average odds ratio and the associated 95% intervals of uncertainty were reported. Next, the percentage of missing data for the care as usual arm was increased by 20% and the odds ratio recalculated as before. This procedure was continued in intervals of 20% until all of the care as usual participants for whom data were missing had been assigned to the worst outcome. Then, 20% of participants for whom data were missing in the specialist care arm were randomly assigned the worst outcome and resultant mean odds ratios and associated 95% intervals of uncertainty were calculated after 100 iterations of each of the possible six groups of 'care as usual' participants with missing data (i.e. those with 0%, 20%, 40%, 60%, 80% and 100% of missing data randomly assigned to the worst outcome). The participants in the specialist care arm were randomly reselected ten times and the resultant odds ratios of ever being admitted calculated for each category of participant in the care as usual arm with missing data for each selection. The percentage of individuals in the specialist group with missing data were increased by another 20% and the odds ratios recalculated until 100% of the missing data in the specialist group had been assigned the worst outcome.

In the second step, the selection process was reversed and 20% of people in the specialist arm with missing data were randomly selected to have the worst outcome. The procedure for the initial sensitivity analysis was then repeated until all of the participants with missing data (in both arms) were again eventually assigned to the worst outcome.

In the third step, the results from the first and second iterations were compared to see if they differed significantly. In

no case did the results from the two sets of analyses differ by more than 1% and the results from the first set of analyses are presented. The results are shown in Table DS2.

Additional references

20 Shi L, Ascher-Svanum H, Zhu B, Faries D, Montgomery W, Marder SR. Characteristics and use patterns of patients taking first-generation depot antipsychotics or oral antipsychotics for schizophrenia. *Psychiatr Serv* 2007; **58**: 482–8.

21 Hopper K, Wanderling J. Revisiting the developed versus developing country distinction in course and outcome in schizophrenia: results from ISoS, the WHO collaborative followup project. International Study of Schizophrenia. *Schizophr Bull* 2000; **26**: 835–46.

22 Hartmann E, Milofsky E, Vaillant G, Oldfield M, Falke R, Ducey C. Vulnerability to schizophrenia. Prediction of adult schizophrenia using childhood information. *Arch Gen Psychiatry* 1984; **41**: 1050–6.

23 Bhugra D, Corridan B, Rudge S, Leff J, Mallett R. Social factors and first onset schizophrenia among Asians and whites. *Int J Soc Psychiatry* 1999; **45**: 162–70.

24 Harrow M, Westermeyer JF, Silverstein M, Strauss BS, Cohler BJ. Predictors of outcome in schizophrenia: the process-reactive dimension. *Schizophr Bull* 1986; **12**: 195–207.

25 Ikebuchi E. Support of working life of persons with schizophrenia [in Japanese]. *Seishin Shinkeigaku Zasshi* 2006; **108**: 436–48.

26 Isohanni I, Jones PB, Järvelin MR, Nieminen P, Rantakallio P, Jokelainen J, et al. Educational consequences of mental disorders treated in hospital. A 31-year follow-up of the Northern Finland 1966 Birth Cohort. *Psychol Med* 2001; **31**: 339–49.

27 Cooper B. Schizophrenia, social class and immigrant status: the epidemiological evidence. *Epidemiol Psichiatr Soc* 2005; **14**: 137–44.

Table DS1 Regression models to assess relationship of potential confounders with contactability

Variable	Contactable	Not contactable	OR (or regression coefficient) of being contactable	95% CI	*P*
Ethnicity, *n* (%)			1.196[a]	0.59–2.42	0.620
White	27 (27)	16 (32)			
Black	28 (29)	10 (22)			
Caribbean	26 (27)	16 (35)			
Black African	4 (4)	0 (0)			
Mixed/other	13 (13)	4 (9)			
Male, *n* (%)	60 (61)	33 (72)	0.559	0.258–1.21	0.141
In a relationship, *n* (%)	19 (19)	22 (48)	3.343	1.16–9.66	0.026[b]
Working, *n* (%)	15 (15)	11 (23)	1.822	0.758–4.38	0.180
Employment, *n* (%)			1.13[c]	0.549–2.32	0.743
Full time	8 (8)	9 (20)			
Part time	7 (7)	2 (4)			
Unemployed	81 (85)	35 (75)			
Not migrant, *n* (%)	57 (58)	18 (39)	0.331	0.158–0.689	0.003[d]
Discharge to GP[e]			0.048	0.005–0.404	0.005[f]
Log of total time with psychiatric services[e]			4.05	1.20–13.7	0.024[g]
Age at baseline, years: median (IRQ)	25.0 (20.0–31.3)	27.0 (22.0–30.0)	0.964	0.930–1.04	0.583[h]

a. Recategorised as White or other. White is the reference category.
b. There was no significant association between 'being in a relationship' and the outcome of 'ever been admitted' (OR = 0.542, 95% CI 0.113–2.59, *P* = 0.443). Being in a relationship was not further analysed as a potential confounder.
c. Odds of contactability for unemployed *v.* those who were employed.
d. There was no significant association between being a migrant and the outcome of 'ever been admitted' (OR = 1.12, 95% CI 0.449–2.82, *P* = 0.802). Being in a migrant was not further analysed as a potential confounder.
e. Discharge to general practitioner (GP) and time in psychiatric services not measured at baseline.
f. Being discharged to a GP was also strongly related to the outcome 'ever been admitted' (OR = 0.124, 95% CI = 0.027–0.569, *P* = 0.007). This variable was included in future analyses as a potential confounder.
g. The total length of time spent in psychiatric care was highly skewed. Time was therefore log adjusted. The log of time ranged from 4.40 to 7.70 (mean 7.33, s.d. = 0.485). The relationship between contactability and the log of total time spent in psychiatric care is presented. Graphical analysis showed six outliers below the median. Statistical analysis of the relationship with the outliers removed did not alter the odds ratio significantly. However, log of total time spent in psychiatry (entered as a continuous variable) was not significantly associated with the outcome and so was not considered further as a potential confounder (OR = 1.74, 95% CI 0.534–6.01, *P* = 0.345).
h. There was a relatively small age span between the oldest and youngest participants. We therefore treated the potential relationship between odds of being contactable and age as linear; it is unlikely that the odds of being contactable would change significantly over such a relatively small age range.

Appendix DS1

Mechanisms of missingness

Assumption of missingness	Mechanism of missingness
Missing completely at random (MCAR)	An example of a MCAR mechanism would be if a completely random process underlies the missingness of data (i.e. data lost in transit). If data are MCAR, then approximately equivalent results would be obtained by performing the analyses if there had been no missing data although there would be some loss of precision. Thus, the analysis of only those units with complete data gives valid inferences. This is not a valid analysis for this data-set as the missing data are not due to MCAR mechanisms.
Missing at random (MAR)	This assumption would be valid if there were predictor of missingness. Statistical analysis of this data-set identified no predictors of missingness and thus this assumption is not valid for this data-set.
No missing at random (NMAR)	This is the most suitable analysis if the assumptions for MCAR and MAR are not met and is used in the analysis of this data-set.

Table DS2 Sensitivity analysis for ever been admitted (odds ratio (OR) of ever being admitted with 95% interval of uncertainty)[a]

Percentage of missing participants in specialist care assigned to 'admitted'	Percentage of missing participants in care as usual assigned to 'admitted'					
	0% ($n=0$)	20% ($n=4$)	40% ($n=7$)	60% ($n=11$)	80% ($n=14$)	100% ($n=18$)
0% ($n=0$)	1.26 0.549–2.89	1.63 0.733–3.62	2.07 0.956–4.50	2.44 1.14–5.23	2.89 1.36–6.10	3.26 1.56–6.84
20% ($n=5$)	1.03 0.460–2.31	1.33 0.615–2.89	1.70 0.802–3.59	2.00 0.959–4.17	2.36 1.14–4.87	2.67 1.30–5.45
40% ($n=10$)	0.872 0.396–1.92	1.12 0.530–2.40	1.44 0.691–2.98	1.69 0.827– 3.46	2.00 0.990–4.04	2.26 1.13–4.52
60% ($n=16$)	0.782 0.370–1.92	1.01 0.530–2.40	1.29 0.691–2.98	1.52 0.827–3.46	1.79 0.989–4.04	2.02 1.13–4.52
80% ($n=22$)	0.687 0.318–1.48	0.889 0.426–1.85	1.13 0.557–2.29	1.33 0.666–2.67	1.58 0.797–3.11	1.78 0.907–3.48
100% ($n=26$)	0.613 0.287–1.31	0.793 0.383–1.64	1.00 0.501–2.03	1.19 0.599–2.36	1.41 0.718–2.75	1.59 0.817–3.08

a. Grey shaded area identifies analyses that would have resulted in a change in the conclusion, i.e. that the efficacy of the intervention had been diluted. The white area identifies sensitivity analyses that would have made no difference to the overall conclusions.

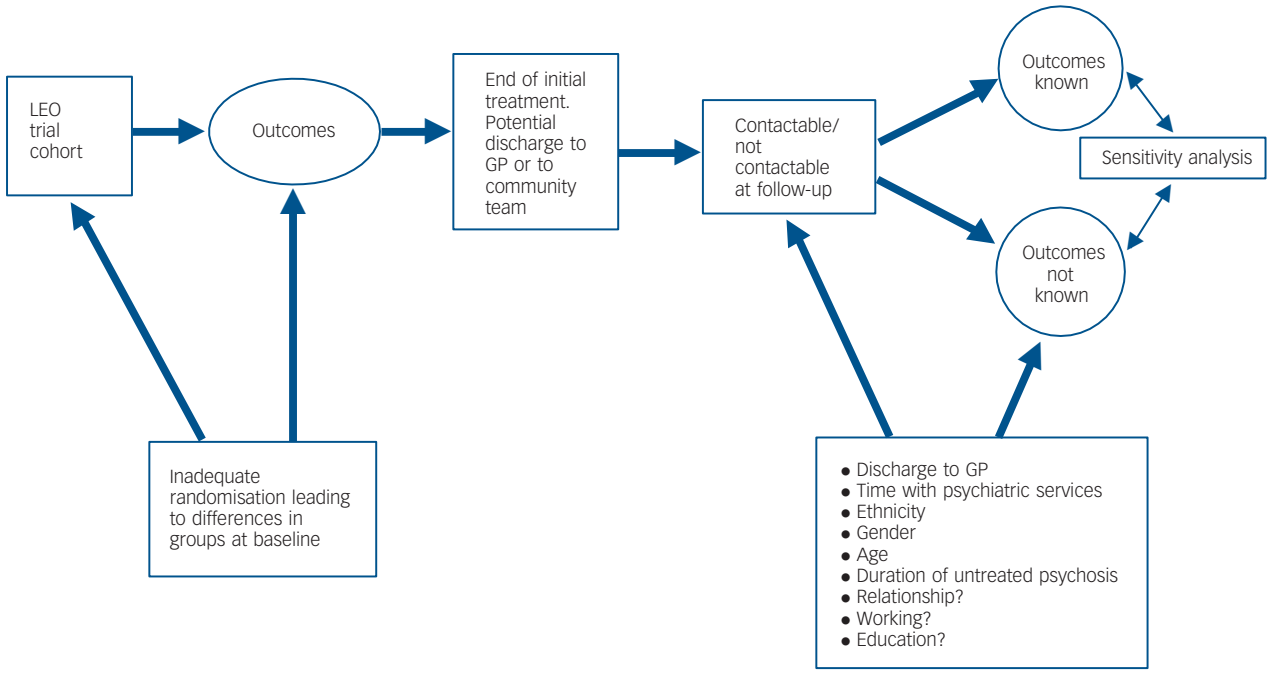

Fig. DS1 Schematic analysis for potential confounds.