

## Records

[Title Page\(s\)](#)   [Whole Article/Book](#)   [JLL Article\(s\)](#)   [Context](#)

[Download key passages/title pages as a PDF](#)

[Download article as a PDF](#)

**Gray R. Wheatley K (1991).** How to avoid bias when comparing bone marrow transplantation with chemotherapy. Proceedings of the International Course on Bone Marrow Transplantation in Children, 28 Nov-1 Dec 1990, Erice, Italy. Bone Marrow Transplantation 7, Suppl 3:9-12.

**Whole Article**

# How to avoid bias when comparing bone marrow transplantation with chemotherapy

Richard Gray, Keith Wheatley, MRC Clinical Trial Service Unit, Radcliffe Infirmary, Oxford OX2 6HE, England

## Summary

It is important to know whether the survival of patients receiving an allogeneic bone marrow transplant (BMT) is better than that of patients receiving "conventional" treatment and, if BMT is better, to know how much better. Unfortunately, this information is surprisingly difficult to obtain accurately. Most studies that have attempted to define the value of BMT have been subject to varying degrees of bias because of the problems of identifying a "conventional chemotherapy" control group with which the outcome of patients who have received BMTs can be compared.

A common bias arises when disease-free survival of BMT patients is compared with that of all other remitters. Early failures are then automatically assigned to the chemotherapy group even if they have donors and so would have gone on to a BMT had they not relapsed or died. Since some patients receive BMTs many months (or even years) after achieving remission - when their prognosis is already much improved - the definition of "early failure" is problematic. Nevertheless, although it is very rarely used, an adequate statistical method does exist to overcome this problem. Careful analysis cannot, however, overcome the problem of selection bias: patients selected for BMT are likely to have better (or worse) prognosis than patients who are treated conventionally.

The only really satisfactory way of assessing the value of BMT is to conduct randomised trials comparing BMT with no BMT - or with extra chemotherapy. Several such studies are currently being undertaken assessing the role of autologous BMT in AML. But, allogeneic BMT is not being assessed in the same way since only a minority of patients have HLA-matched donors and it is usually thought unethical to withhold BMT from those who do have donors. The value of allogeneic BMT can, however, be assessed unbiasedly using "Mendelian randomisation", i.e. comparing patients whose siblings are HLA-compatible with those who are not. This information is difficult to obtain reliably from treatment centres in multi-centre trials but can be obtained directly from tissue-typing laboratories. A pilot study of this method of assessing allo-BMT has demonstrated its feasibility and has provided perhaps the first unbiased evidence of the benefit of allo-BMT in AML.

## Observational data: the problems

Most papers on BMT report analyses of observational data: data on BMT patients are collected at individual centres, by multicentre trial organisations or by BMT registries and the survival of these patients is then compared with that of a reference population to assess the value of BMT in this particular group of patients. Unfortunately, such "observational studies" are likely to suffer from important biases, the magnitude and direction of which are difficult to estimate. Taking BMT in first remission of acute myeloid leukaemia (AML) as an example, this paper outlines some of these potential biases and suggests ways of avoiding them.

The most obvious example of a biased comparison of BMTs with chemotherapy would be to compare the survival of BMT patients with that of previous patients with the same disease. For instance, it would be a mistake to claim that since survival in AML is normally only 10 or 20% then survival of over 50% with

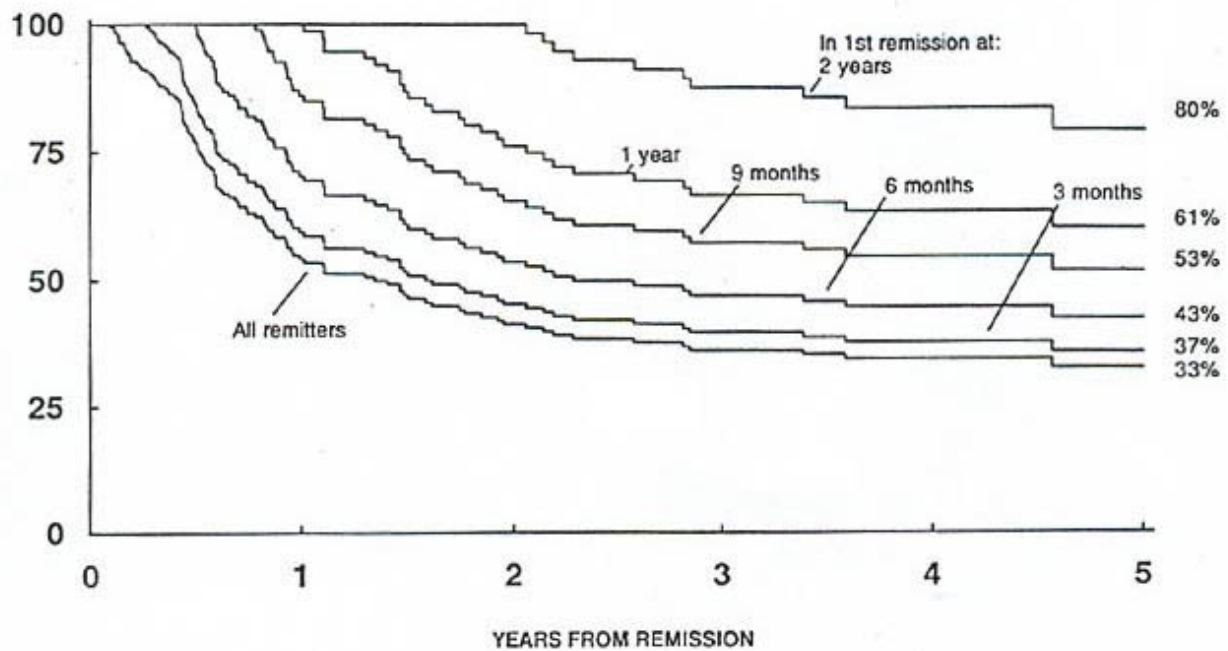
BMT must be an improvement. Such a comparison is bound to be misleading since the low survival "historical control" group will include many patients who did not achieve complete remission (CR). It may also include patients too old or too frail to be considered for BMT who, again, are likely to have poor survival. Quite apart from these already serious biases, there will be the well-recognised problem that historical comparisons are potentially unreliable.

Even if the "conventional chemotherapy" control group is restricted to a contemporaneous group of younger patients who have all achieved remission, similar potential for bias remains. Patients who relapse or die in the first few months of CR are often assigned to the chemotherapy group even if they had donors, and so would have gone onto a BMT had they stayed in remission. Since some patients are transplanted many months (or even years) after achieving remission - when their prognosis is already much improved - dealing with these early failures is problematic. For example, as in many other reported series, the early relapse rate in the Medical Research Council's AML 9 trial was high and nearly one-half (43%) of patients achieving remission relapsed within one year. The relapse rate fell steadily in subsequent years, with one third of those in remission at one year relapsing in the second year, one in six in the third year and few relapses thereafter. Figure 1 shows the probability of AML 9 chemotherapy-treated patients aged under 50 reaching 5 years without disease recurrence as a function of the time that they had already been in CR. For all 325 remitters, the disease-free survival at 5 years was 33%. However, the longer patients remained in CR, the greater was the probability of still being in first CR at 5 years. Patients who avoided relapse in the first 3 months of CR had a 37% probability of being disease-free at 5 years. If they were still in CR at 6 months, the chance of 5-year disease-free survival (DFS) improved further to 43%. The relapse-rate was about 5% per month at this stage and so by 9 months the expectation of 5-years DFS had risen to 53%. By the end of the first year of CR the chances of chemotherapy-treated patients remaining disease-free at 5 years were substantially better than they had been when CR was achieved (61% compared to 33%). Patients who were still disease-free two years after achieving remission had an even better (80%) chance of remaining disease-free at five years. Since prognosis improves so rapidly during the first year or so of CR, it is clearly important to allow for this when evaluating BMT.

The indications for and timing of BMTs in AML 9 were decided by the participating clinicians, and by October 1989 a total of 96 patients had received BMTs in first CR (45 allogeneic and 51 autologous). The timing of BMTs varied considerably (Table 1).

Table 1: Timing of BMTs and probability of remission failure in the Medical Research Council's AML 9 trial.

Duration of remission	Number of allografts in this period	Number of autografts in this period	Prob. of relapse or death during this period
1-3 months	14	4	7%
4-6 months	21	21	15%
7-9 months	6	10	15%
10-12 months	3	12	15%
13+ months	1	4	—



**Figure 1: AML 9 chemotherapy patients: Rapidly improving expectation of 5-year disease-free survival for patients who avoid early relapse**

Most BMTs were given in the first year of remission — although one patient received an autologous BMT as late as 4 years after achieving CR. The average duration of remission before BMT was about 5 months for allografts and about 7 months for autografts. As can be seen from Figure 1, patients who reached this stage of remission already had much better than average prognosis even without BMT and so, if they are removed from the chemotherapy arm, this will adversely affect the survival of this group. If a true comparison is to be made of patients who received BMTs in AML 9 with those who did not, then the chemotherapy group must be given due credit for the time at risk before BMT when patients who subsequently receive BMTs are effectively “chemotherapy” patients — none of whom relapse. Fortunately, although it is rarely used, a statistical method does exist to allow for this. Using this

method, all patients young enough to be candidates for BMT start off in the chemotherapy group when they enter remission. Patients who then go on to receive a BMT cease to be “at risk” in the chemotherapy group at that point and instead become “at risk” in the BMT group. An example of these calculations applied to the AML 9 data for the first few months of remission is shown in Table 2. In month 1, all patients were “at risk” in the chemotherapy group, but there were no relapses or deaths. By month 2, 3 patients had received allogeneic BMTs and were now at risk in this group. There were 17 failures in the second month, all in the chemotherapy group. However, since all but 3 patients were in the chemotherapy group during month 2 this was not surprising, and indeed the expected number of failures if chemotherapy and BMT were equivalent was 16.8 (17 x 322/325). By month 7 of CR there were 29 patients at risk in

**Table 2: AML 9: Numbers of remission failures and the expected numbers if BMT and chemotherapy were equivalent.**

Treatment group		Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7
Allogeneic BMT	a/b	0/0	0/3	0/9	1/14	1/20	4/22	1/29
	Exp.	0.0	0.2	0.4	0.5	0.9	1.9	1.6
Autologous BMT	a/b	0/0	0/0	0/0	0/4	0/13	0/17	1/25
	Exp.	0.0	0.0	0.0	0.2	0.6	1.4	1.4
Chemotherapy	a/b	0/325	17/322	13/299	10/277	11/251	19/233	12/195
	Exp.	0.0	16.8	12.6	10.3	10.6	19.7	11.0

**Key:** a/b: b = No. of patients in this treatment group alive and relapse-free at the beginning of the month.  
a = No. of these patients dying or relapsing during the month.

Exp.: The expected numbers in Table 2 are derived from the fact that if allogeneic BMT, autologous BMT and chemotherapy are equivalent then all the patients are equally likely to relapse or die during that month.

the allogeneic BMT group (6 of the early allograft patients had already died and so were no longer "at risk"), 25 patients in the autograft group and 195 chemotherapy patients. There were 14 failures in month 7 and, again, the expected number in each group is calculated based on each group's exposure to the risk of failure during this month. Since by this stage the BMT groups were contributing a substantial part of the exposure to risk, the expected proportion of failures among BMT patients was now much higher than in the early months. So, by the end of the seventh month, there had been a total of 7 failures among the 35 patients who had received allografts (total of 5.5 expected), 1 of the 25 autograft patients had failed (3.6 expected) and 82 chemotherapy patients had relapsed or died with 81.0 failures expected. The high number of events expected in the chemotherapy group reflected their much greater exposure to risk during the early months of CR. These calculations are then repeated for the subsequent months and, in this way, the expected number of events in each group can be calculated — preferably using exact days of events (not months) and stratifying by age, etc. if desired — with all patients contributing their exposure to risk to the appropriate group at the appropriate time after remission has been achieved. (The calculations are slightly different if BMT registry patients are compared with an external series of chemotherapy-treated patients, for example patients in a multi-centre trial. In this case the BMT patients have not contributed any prior exposure to risk as chemotherapy patients in the chemotherapy treated reference group and so their exposure to risk prior to BMT is ignored.) Finally, P-values are calculated using standard log-rank methods, and thus an unbiased comparison can be made of patients who received BMTs with those who did not.

Unfortunately, however, this analysis does not answer the crucial medical question of whether a policy of offering BMT to suitable patients is better than a policy of not doing so, because of the problem of selection bias. Patients who receive BMTs are selected for the procedure and are likely to differ from chemotherapy patients in several ways: they may have better (or worse) prognosis than patients who are treated conventionally. For example, frailer patients — those who have suffered cardiac, renal or hepatic toxicity during induction treatment — are likely to have a worse prognosis and are also less likely to be offered BMTs. Older patients with HLA-compatible siblings may receive BMTs only if they relapse and achieve second remissions, whereas the better prognosis younger patients are more likely to receive a first CR BMT if they have a donor. High risk patients — e.g. those with adverse chromosomal abnormalities or those with particularly high leukocyte counts — may be selectively offered BMTs whereas standard risk patients may not. And there may be additional unknown reasons why the two groups differ. Despite all of these problems, in some circumstances, the benefit of BMT can be so clear that selection bias is relatively unimportant. However, in many other circumstances this is not the case and then the selective nature of BMT patients can produce biases of unknown magnitude and direction and thus increase the likelihood of important errors.

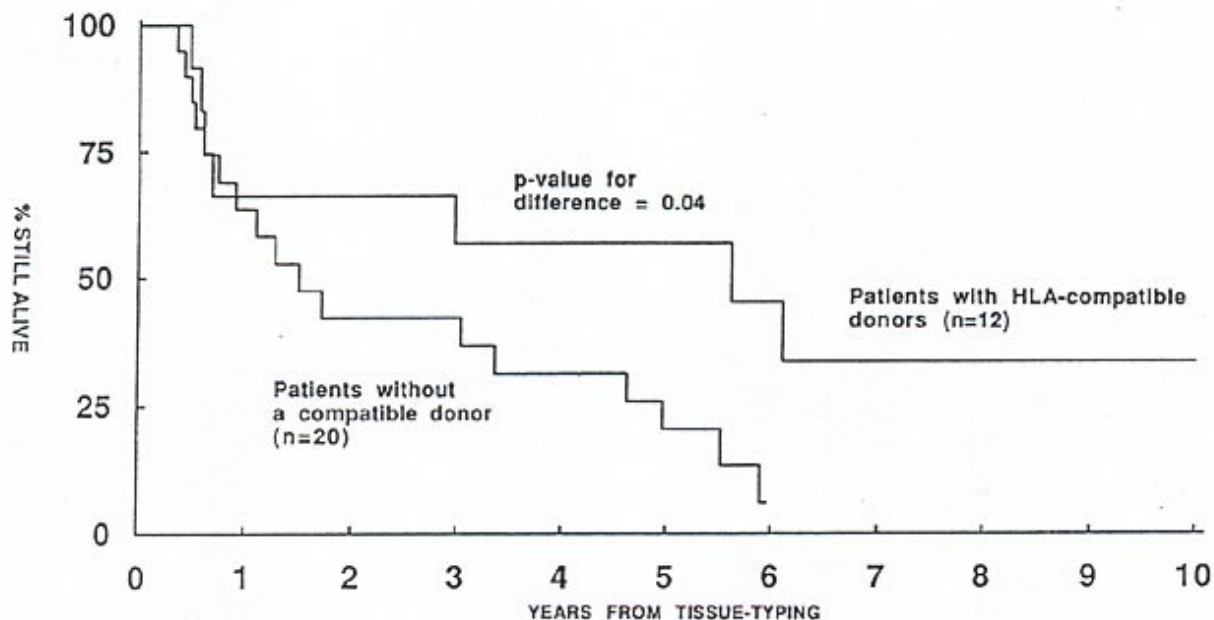
#### How to avoid bias

To avoid selection bias, the only really satisfactory way of assessing the value of BMT is to conduct controlled randomised trials that compare BMT directly with no BMT — or with extra chemotherapy. Indeed, several such randomised studies are currently being undertaken assessing the role of autologous BMT in first remission AML and their results are awaited with great interest. But, allogeneic BMT is not being assessed in the same way since only a minority of patients have

HLA-matched donors and it is usually thought unethical to withhold BMT from those younger patients who do have donors. Fortunately, however, the value of allogeneic BMT can be assessed unbiasedly without the need for formal trials by taking advantage of "Mendelian randomisation" and comparing the survival of patients whose siblings are HLA-compatible with that of patients whose siblings are not HLA-compatible. Since the "allocation" to one or other of these two groups is decided many years earlier, no selection bias can occur. All that is needed for this comparison to be valid is to obtain really complete data on HLA-typing, for most suitable patients with donors to go on to BMTs and then a totally unbiased assessment of whether allogeneic BMT is preferable to conventional treatment can be made. This approach has considerable practical advantages over alternative randomised trial designs, such as randomising those patients with matched donors to allogeneic BMT or to no BMT: the size of the study is increased substantially because all patients with siblings can be included in the study not just those whose siblings are HLA-compatible; there is no ethical necessity to obtain informed consent to a randomisation between BMT and no BMT; no patients with donors are denied BMTs and so clinicians who are convinced of the benefits of BMT can participate; and, where available, retrospective data can be used as well as prospective data.

The main disadvantage is that information on the results of tissue-typing can be difficult to obtain reliably, particularly from the busy clinicians who participate in multicentre trials; leukaemia patients' notes are copious and the information is not usually recorded systematically. For example, in the MRC's AML 8 trial, 40 patients received allogeneic BMTs. Data on tissue-typing had not been requested routinely and, in an attempt to obtain this information, all participants were written to and asked to supply data on patients whose siblings had been typed but were HLA mismatched. Only 9 such patients were identified by this means whereas 60 or more might have been expected. Such retrospective data collection is clearly hopelessly unreliable. Even when information on tissue-typing is sought systematically as part of trial documentation, it can be difficult to get really complete and reliable information. In the MRC's current AML 10 trial where these data are requested routinely, some of the essential information on the outcome of typing is missing, particularly for patients who relapse or die early in remission. There is sometimes a failure to distinguish between patients with no siblings and those with no HLA-matched siblings. (This distinction is important because bias is introduced if patients with no siblings who relapse early — or, equivalently, patients who relapse or die before they have a chance to be tissue-typed — are included in the no donor arm). Other important information, e.g. on the exact date(s) of typing and on whether the intention was to proceed with an early BMT if an HLA-match were found, is even more difficult to obtain reliably. If complete data cannot be obtained, there remains a definite possibility that moderate biases could exist. If moderate improvements in treatment are to be evaluated reliably then moderate biases must be avoided.

Fortunately, there is a better way of getting reliable and complete data on tissue-typing which is to obtain it directly from the tissue-typing laboratories where it is generated. These laboratories usually keep excellent records of names of patients and their clinicians, dates and outcome of HLA-typing, relationship of potential donors to patients, and they also often have data on the patient's disease and on the reasons for HLA-typing. These records have the considerable advantage that, even if some data are missing or if some HLA-typed patients cannot be traced, no material bias is introduced since the



**Figure 2: Oxford Tissue Typing Study: Survival of AML patients in first remission by HLA-match/no match**

outcome of treatment is usually not known to the laboratory staff and so cannot influence the availability of data. Careful analysis is still needed. In particular, all patients must be analysed according to their "allocated" group even if they have donors but do not receive BMTs or, alternatively, have no HLA-compatible donor but receive mismatched or autologous transplants. Similarly, when there is a substantial time lag between the first and last sibling being tissue-typed, patients who at first have no match but finally find an HLA-compatible sibling should be included in the no donor arm during this intervening period (the group that they would have been in if they had relapsed or died in the interim).

The feasibility of this method has been demonstrated in a pilot study using the Oxford regional tissue-typing laboratory records. Both tissue-typing laboratories in the region have routinely maintained detailed records on all requests for tissue-matching over the last ten or so years. Many of the patients who were typed with a view to BMT had been entered into MRC trials and so full information on subsequent events was readily available. For the remaining patients, further information could, in nearly all cases, be obtained from the referring consultant, sometimes with help from the local cancer registry. Patients with donors — most of whom went on to BMTs — have been

compared with patients without donors for an unbiased assessment of the benefits of having an HLA-matched sibling. This information is currently being prepared for publication and provides probably the first completely unbiased evidence of the benefit of BMT in first remission of AML. The 12 first CR AML patients with HLA-compatible donors (9 of whom received BMTs in first CR) had significantly better survival (Figure 2) than the 20 whose siblings proved to be HLA mismatched. The number of young patients with leukaemia in the Oxford region is quite small and so no meaningful analysis of the influence of other factors (e.g. age, stage of disease, etc.) or of HLA-typing in ALL and other diseases has been possible. What is needed now to clarify the role of BMT in haematological malignancy is to use the same approach at many other tissue-typing laboratories. First, existing records — with their inevitable limitations — could be used to evaluate BMT in previous series of patients. Second, whenever future requests for HLA-matching are made, tissue-typing laboratories should record basic data on the patient, his disease, the stage of the disease and whether the intention was to proceed with a BMT if a match were found. These patients could then be followed up and, with almost no extra work, valuable data on the efficacy of BMT could be generated. The authors are currently undertaking such a study and would be pleased to hear from any centres who wished to participate.