High-dose chemotherapy and autotransplantation in myeloma: Retraction of a study and lessons for the future

Jayesh Mehta MD

From: The Robert H Lurie Comprehensive Cancer Center
Northwestern University
Chicago

676 N St Clair Street, Suite 850
Chicago, IL 60611
j-mehta@northwestern.edu
312-695-4438

No conflict of interest information to disclose.

Abstract

A randomized study of 1 versus 2 autografts in myeloma showed that the event-free and overall survival after a single transplant was better; an unusual finding in light of published data. The survival of patients relapsing after transplantation was far shorter than expected. The outcome of the single transplant arm was better than any reported in the literature, and that of the double transplant arm worse than any. Investigation of the study based upon information available in the study suggested the possibility of serious errors, and a review of the primary data suggested research misconduct. Two publications reporting the study were retracted, and additional work from the group is under scrutiny. It is the author's opinion that primary data should be submitted when a manuscript is submitted for publication, and de-identified data made available to readers after publication to verify and replicate the analysis. This is a strong step towards reducing misconduct and encouraging greater debate – both highly desirable to ensure patient welfare and scientific integrity.

Background

Multiple myeloma is the commonest indication for high-dose chemotherapy and autologous hematopoietic stem cell transplantation (AHSCT) worldwide – benefiting young as well as old patients [1]. While there is evidence of superiority of tandem transplantation over single in myeloma, the subpopulation of patients that benefits the most and the exact extent of the benefit remain to be defined [2].

A prospective, randomized multi-center study from Tunisia [3] was designed to see if a single cycle of high-dose melphalan and AHSCT followed by 6 months of thalidomide maintenance therapy was equivalent to tandem cycles of high-dose melphalan and AHSCT without maintenance therapy. All patients received thalidomide-dexamethasone as induction therapy. By design, salvage therapy was to be high-dose melphalan and a second AHSCT in those relapsing after one AHSCT, and thalidomide in those relapsing after tandem AHSCT. The incomplete study, showing equivalence of the two approaches, was greeted with enthusiasm at the 2006 annual meeting of the American Society of Hematology (ASH) [4]. We felt caution was warranted because the study was unpublished and the follow-up was too short [2].

Online publication

The completed study appeared online in *Blood* as a First Edition paper on 17 September 2007 [4]. It showed that single transplantation followed by thalidomide

maintenance for 6 months (Arm B) resulted in significantly better progression-free (PFS) and overall (OS) survival than tandem transplantation (Arm A). While equivalence may have been unsurprising, superiority of single transplantation was unexpected based on extensive published literature with mature follow-up [2].

The article raised a number of other questions:

1. The PFS and OS curves were almost superimposable; implying very short post-relapse survival (PRS). This is very unusual because myeloma patients usually survive for a long time after relapse following autotransplantation. Figure 1A shows the outcome of 378 myeloma patients autografted at the Royal Marsden Hospital, UK, between February 1985 and March 2000; a subset of patients reported previously [2, 6]. 31 March 2000 was chosen as the reference date for Figure 1A – before the patients had access to thalidomide, bortezomib or lenalidomide in UK. Despite this, the median PRS is 2.1 years. The short PRS of patients in the Tunisian study (median 0.3 years) has been shown in Figure 1B for comparison.

2. The single transplantation arm outcome was superior to any other published single transplantation series, and better than most published tandem transplantation data [7].

3. The tandem transplantation arm outcome was inferior to most published tandem transplantation series, and indeed to a number of single transplantation studies [7].

4. Lack of a trial profile depicting patient flow and the absence of censor marks on survival curves hampered critical evaluation and interpretation of the data.

5. The study accrued patients from April 2003 to December 2006. As of April 2006 [4], 140 patients (70%) had been enrolled, suggesting that the remaining 30% of the patients were enrolled in the last 8 months and had limited follow-up.

6. There were discrepancies between the results reported in the ASH abstract [4] and the paper [5]. According to the ASH abstract [4], the 2-year PFS/OS in the single and tandem transplantation arms was 70%/75% and 41%/55% respectively. According to the paper [5], the 3-year PFS/OS in the single and tandem transplantation arms was 85%/88% and 57%/63%. While survival projections do change with time, the magnitude of the differences seemed unusual.

I thought of innocuous explanations such as erroneous data analysis or interpretation. The manuscript stated that the first author, Dr Abderrahman Abdelkefi, was solely responsible for analyzing and writing the paper. Without an adequate statistical background or professional help, mistakes could easily occur. There may also have been unrecognized imbalance in the two arms in terms of tumor burden and the biological nature of the disease. I felt that the paper required a closer scrutiny; starting with a review of the original data which would have revealed any errors.

I shared my concerns with *Blood* on 7 October 2007 suggesting a formal data review and consultation with an independent myeloma expert. The editorial leadership of *Blood* forwarded my concerns to the two original reviewers of the paper and an independent expert (whose identities, appropriately, remain unknown because of confidentiality). My concerns were apparently addressed by the authors to the satisfaction of these experts because *Blood* felt that a data audit was not indicated, and a second version [8] of the First Edition paper was posted online on 16 October 2007.

Online publication – Version 2

The second version included a chart showing patient flow. The authors explained that lack of access to newer salvage therapy agents (lenalidomide and bortezomib), and delays between relapse and presentation for evaluation were responsible for the short PRS. The lack of availability of new agents could not explain the poor PRS based on the data illustrated in Figure 1A and discussed above.

The survival curves were still missing censor marks, the patient flow chart did not provide precise information, and there were minor inconsistencies in the numbers. I corresponded directly with Dr Abdelkefi seeking clarification on patient numbers, subgroup outcomes, exact numbers of patients relapsing, time to relapse, duration of salvage therapy, and PRS.

He responded promptly, but the information he provided was at variance with the two online versions of the paper [5, 8] in terms of patient numbers, time to relapse, and time from relapse to death – suggesting that there were errors in data entry or analysis. The inconsistencies were relatively minor, but, in each case, favored the single transplantation arm or disfavored the tandem transplantation arm. The statistical differences between the outcomes of the two arms hinged upon a small number of events – and a few errors could have resulted in apparently significant differences in outcome when there really were none.

The paper described the beneficial effect of thalidomide in a statement that I could not understand: "In arm B, the effect of thalidomide on PFS differed according to the response achieved at three months after the first transplant. Patients who had at least a VGPR did not benefit from thalidomide (P = .5). Patients who did not have at least a VGPR had a significant benefit from thalidomide (P < .003)." Two P values implied that 4 subgroups were being compared. However, no identifiable subgroups could explain this statement. Additionally, with only 10 total events in Arm B, the highly significant P value of <0.003 was difficult to explain – even assuming it came from a valid comparison.

A closer look at the patient population revealed some odd features. There were no patients with renal dysfunction (maximum serum creatinine 126 μmol/L) although moderate renal dysfunction was not an exclusion criterion. Approximately 24% patients had light chain disease which is more likely to be associated with renal dysfunction. In

each arm, the maximum serum calcium was 4 mmol/L (16 mg/dL). It is unusual to have such severe hypercalcemia with normal renal function. The maximum ß2-microglobulin in Arm A was 20 mg/L. Such a high level usually occurs in the setting of renal dysfunction, and is very uncommon with normal renal function. The maximum ß2-microglobulin in Arm B was 4 mg/L. 40% of patients in Arm B had International Staging System stage 3 disease which is defined as ß2-microglobulin >5.5 mg/L. These discrepancies, once again, provided evidence of unreliability.

I brought all of these to the attention of *Blood* suggesting that there was strong evidence of poor data management as well as analytical errors. For the first time, I raised the possibility of fraud; once again requesting a review of the original data. I felt that only a review of the primary data could address the concerns, and that mere communication of the concerns and description of inconsistencies would permit data modification if there was misconduct involved. In response, a third version [9] of the First Edition paper was posted online on 8 November 2007.

Online publication – Version 3

The shapes of the survival curves had changed noticeably in the third version in response to the questions I had raised. All other discrepancies that had been noted by me – and had presumably been shared with the authors by *Blood* – had also been resolved. However, the poor survival of relapsing patients remained worrisome – as did

the persistence of the inexplicable statement quoted above ("In arm B, the effect of thalidomide…").

I independently requested Dr Abdelkefi to provide me with primary data to see if I could understand the meaning of the statement on the effect of thalidomide by replicating the analysis. He declined, saying the data had already been provided to *Blood* and had been found to be satisfactory. It was then that I found out that *Blood* had decided to publish Version 3 because the final data – presumably submitted with Version 3 and corrected twice based upon my concerns communicated to the authors by *Blood* – were reviewed by an independent statistician who generated survival curves that were identical to those submitted by the authors in Version 3. A clinical trial expert felt that there was no evidence of data manipulation or fabrication. I was informed by *Blood* that the matter was closed, and the paper was published [10].

Aftermath and impact on patient care

Blood published a letter from me [7] illustrating the disparity between the outcomes reported in the Tunisian study and the published literature sounding “a strong note of caution before its unusual findings are used as a basis to alter clinical practice. Because of its unusual findings, not only does the study need longer follow-up, but replication of the study (or a similar concept) is essential…” However, it became evident over the next several months that clinical practice *was* being altered based upon the Tunisian report. I therefore decided to investigate the study further on my own in March 2009.

The investigation

I requested Dr Abdelkefi to provide original primary data, and an update reflecting 2 more years of follow-up. His responded that he was in sole possession of all the data – and they were lost because of computer problems resulting in file corruption. I suggested re-creating the data from original source records (195 patients from 10 centers). He agreed to do so and sent me the original primary data 12 days later.

My analysis of the data showed that the PFS and OS curves were identical to those in the final paper [10]. However, contrary to what had been stated in every single version of the paper [5, 8-10], the multivariate analysis I performed showed that there was *no* significant difference in OS between the two arms (P=0.12) and the difference in PFS was borderline (P=0.057). There were no data that could be analyzed to evaluate – let alone support – the statement on the effect of thalidomide being related to the extent of response in Arm B.

More importantly, there were disparities in dates and timeline (Table 1) as mentioned in the study registration [3], the *Blood* paper [10], a previous publication from the group reporting on a subset of patients from the same study [11], and in the primary data. The dates of randomization shown in the primary data were highly improbable given the time that ought to have been taken for induction therapy (75 days) and stem cell collection after chemotherapy (2-3 weeks). These timeline inconsistencies were what gave me the most cause to worry about possible data manipulation.

The corrected β2-microglobulin levels (maximum 20 mg/L) and the age range (youngest 34 years and oldest 59 years) in the final version [7, 8] were inconsistent with the numbers from *Bone Marrow Transplantation* (maximum β2-microglobulin 52 mg/L, youngest 33 years, and oldest 60 years) [10].

At my request, Dr Abdelkefi had also provided the date on which each relapsing patient was known to be stable (i.e. not to have progressed) – which was the visit prior to the visit on which relapse was documented. I calculated PRS based on that date – and it was still very poor (higher by a median of 82 days, range 23-186 days, compared to the actual date of documented relapse).

I contacted Dr Abdelkefi requesting explanations for the discordant statistical findings and the statement on the effect of thalidomide in Arm B, and contact information for coauthors to seek direct clarification. He responded by sending me data updated to March 2009, but did not answer any of the other questions.

Analysis of the updated data showed that the survival of the single transplantation arm had worsened considerably whereas that of the tandem transplantation arm had not changed appreciably – resulting in the outcomes of the two arms becoming almost identical. Such a dramatic change in survival with a couple of additional years of follow-up was very unusual; again suggesting the possibility of problems with the data. The PRS had improved marginally – but was still much poorer than expected based upon

the outcome of the UK patients who did not have access to any novel agents at all including thalidomide. Figure 1B compares the PRS of the UK patients with the original and updated Tunisian data.

The paper had mentioned that karyotyping with Giemsa-banding was performed on 75% of the patients at one of two reference laboratories. Since the paper only provided data on chromosome 13 deletion, I asked Dr Abdelkefi for additional karyotype information in the hope that unrecognized differences in the arms that could still explain the findings. He said that no further cytogenetic data were available – and that he was the only person who possessed all the trial data, which had been collected and registered at Centre National de Greffe de Moelle Osseuse, Tunis.

At this point, I revealed the timeline discrepancies to Dr Abdelkefi. “With a sense of regret” I voiced “my concern about possible fraud”, and suggested 3 options: an off-site audit, an investigation by the Tunisian authorities, or retraction of the article. I also contacted Drs Ali Saad, Hechmi Louzir, and Koussay Dellagi, all coauthors and representatives of the two central laboratories that performed cytogenetic analysis for the study, requesting raw data. I did not hear back from them. Within a few hours of my email, Dr Abdelkefi chose to withdraw the paper and assumed sole responsibility for “all errors.”

Retraction

Both papers reporting the study [10, 11] were retracted [12, 13]. The retraction notice in *Bone Marrow Transplantation* [12], written by me as a Co-Editor, said: “Errors in data collection and entry have compromised the accuracy of the report.” The notice in *Blood* [13] said: “The initial data available at the time the paper was published were inappropriately collected. The first author, Abderrahman Abdelkefi, was solely responsible for the errors. The authors present their deepest apologies to the scientific community and to the readers, reviewers, and editors of Blood for publishing these data.”

On-site audits of the now-infamous breast cancer autotransplantation studies from South Africa [14, 15] revealed specific details of the nature of the fabrication and misconduct. No such information is available for the Tunisian study which, to the best of my knowledge, has never been audited.

In my opinion, in addition to all the points shown above, Figure 2 illustrates why the study is not credible. Figure 2 compares the updated outcome of all the Tunisian patients with 195 consecutive patients treated at the Royal Marsden Hospital starting April 1993 – exactly a decade prior to the Tunisian study. I chose the follow-up date to be 31 March 1999 to match the 6-year follow-up of the updated Tunisian data. Note that the steep fall in the initial part of the survival of the UK patients is from early treatment-

related mortality (bone marrow as source of stem cells, no growth factors post-transplant); something that would be unexpected today.

Assuming that all 195 patients described in the Tunisian study did exist and undergo the treatment they were supposed to, based upon the shape of the curves in Figure 2, I remain to be convinced that the original publication did not misrepresent the outcome of the single transplantation arm by making it seem better and that of the tandem transplantation arm by making it seem worse. Subsequently, the updated data may have been manipulated to show outcomes that were comparable for both arms to lay the controversy to rest. This remains conjectural.

Questions on other publications

In October 2007, while analyzing the *Blood* paper, I had studied additional publications from the same group because I was unfamiliar with their work. Three randomized studies on indwelling vascular catheters [16-18] stood out as being logistically impossible to perform. While I had done nothing about this in 2007, I notified the *Journal of Clinical Oncology* of my concerns in March 2009 when I was told by Dr Abdelkefi that all data from the myeloma study were lost. *Journal of Clinical Oncology* recently published an Expression of Concern [19] on the study because neither Dr Abdelkefi nor the Centre National de Greffe de Moelle Osseuse, Tunis could provide “evidence that the study was conducted, including primary source documents demonstrating that data

were appropriately collected and analyzed". The journal was told by Dr Abdelkefi that "the primary data were lost and not recoverable."

I notified *Thrombosis Haemostasis* and *Journal of Supportive Oncology* of all the events and my concerns in September 2009. *Thrombosis Haemostasis* contacted Dr Abdelkefi for information, and upon lack of satisfactory information, published an Expression of Concern [20]. The *Journal of Supportive Oncology* is in the process of obtaining clarification from Dr Abdelkefi in a manner similar to the *Journal of Clinical Oncology* and *Thrombosis Haemostasis*.

A number of papers the group has published in *Bone Marrow Transplantation* are being studied by the journal's editors. A paper by Abdelkefi et al published in the *Journal of Clinical Apheresis* [21] included patients from the now-discredited study. This was been brought to the attention of the journal in September 2009, but no response has been received on the plan of action as of December 2010.

Detecting misconduct

The outcome of this case shows that it is possible to investigate a study and ask probing questions based on data available in the public domain. However, conclusive proof of misconduct can come only with review of the primary data. Therefore, I believe that primary data from clinical publications must be submitted with all manuscripts, and, in the event of publication, made available in de-identified form to readers who want to

reproduce the analysis. The question of access to primary data has been addressed [22, 23]. According the US Office of Research Integrity (ORI) [22], “Respondents in several misconduct cases have been unable to produce the data reported in their manuscripts or articles. Some journals require authors to place the data supporting their manuscripts in depositories. Requiring that the data supporting all submitted manuscripts be deposited may not be feasible. However, authors could be explicitly informed that their data may be requested during the review process or if questions arise following publication.” I would expect a body such as the ORI to take a stronger stand on the matter.

In the context of the South African breast cancer studies [23], the *Lancet* editorialized that the approach of frequent, random requests for primary data by editors would undermine the honor system, and that the “research culture had not become that diseased.” I disagree: accuracy, honesty and transparency are not negotiable commodities in science – and access to primary data is an important way of ensuring that the likelihood of mistakes and misconduct is reduced, and wider debate is encouraged. Availability of primary data to all readers would also take the burden of ensuring accuracy away from journals to a significant extent.

Authors, statisticians, reviewers, other experts, and editors determine what is published. Mistakes or misconduct anywhere along the way eventually adversely affect the most important constituency we serve – our patients. My hope is that all individuals involved

in the Tunisian episode have learnt lessons that will make them discharge their responsibilities better in the future.

However, a critical and vigilant reader is the final line of defense between dubious research and clinical practice – often the only one. If the findings of a study do not make biological sense, they must be questioned vigorously. Until proven incontrovertibly, such approaches should certainly not be used to alter patient management.

Table 1: Timeline inconsistencies in the Tunisian study

Data	Statement	Source	Contradictory finding or statement	Source
Study start date	April 2003	*Blood* [4, 7-9]	May 2003	Study registration with NIH [2], ASH abstract [3], *Bone Marrow Transplantation* [10]
Time to stem cell mobilization	Mobilization chemotherapy given 90-182 days (median 102) after starting thalidomide	*Bone Marrow Transplantation* [10]	Time from enrollment to randomization was 2-4 months (median 3 months)	*Blood* [4, 7-9]
Enrollment	60 patients enrolled between May 2003 and September 2004	*Bone Marrow Transplantation* [10]	95 patients randomized as of 1 September 2004	Primary data
Enrollment	140 patients enrolled as of April 2006	ASH abstract [3]	173 patients randomized as of 1 April 2006	Primary data

References

1. Mehta J, Cavo, M, Singhal S. How I treat elderly patients with myeloma. Blood 2010. DOI 10.1182/blood-2009-10-163329.

2. Mehta J, Singhal S. High-dose chemotherapy and autologous hematopoietic stem cell transplantation in myeloma patients under the age of 65 years. Bone Marrow Transplant 2007; 40:1101-1114.

3. clinicaltrials.gov/ct2/show/NCT00207805 (Accessed 14 August 2010).

4. Abdelkefi A, Ladeb S, Ben Othman T, et al. Timing of second autologous transplantations in multiple myeloma: results of a multicenter sequential randomized clinical trial. Blood (ASH Annual Meeting Abstracts) 2006; 108: Abstract 59.

5. bloodjournal.hematologylibrary.org/cgi/reprint/blood-2007-07-101212v1 (Accessed 14 August 2010).

6. Sirohi B, Powles R, Mehta J, et al. An elective single autograft with high-dose melphalan: single-center study of 451 patients. Bone Marrow Transplant 2005; 36:19-24.

7. Mehta J. One or two autografts for myeloma? Blood 2008; 111:3899-3900.

8. bloodjournal.hematologylibrary.org/cgi/reprint/blood-2007-07-101212v2 (Accessed 14 August 2010).

9. bloodjournal.hematologylibrary.org/cgi/reprint/blood-2007-07-101212v3 (Accessed 14 August 2010).

10. Abdelkefi A, Ladeb S, Torjman L, et al. Single autologous stem-cell transplantation followed by maintenance therapy with thalidomide is superior to double autologous transplantation in multiple myeloma: results of a multicenter randomized clinical trial. Blood 2008; 111:1805-1810.

11. Abdelkefi A, Torjman L, Ben Romdhane N, et al. First-line thalidomide-dexamethasone therapy in preparation for auto-SCT in young patients (<61 years) with symptomatic multiple myeloma. Bone Marrow Transplant 2005; 36:193-198.

12. Retraction notice for: Blood 2008; 111:1805-1810. Blood 2009; 113:6265.

13. Retraction notice for: Bone Marrow Transplant 2005; 36:193-198. Bone Marrow Transplant 2009; 43:893.

14. Weiss RB, Rifkin RM, Stewart FM, et al. High-dose chemotherapy for high-risk primary breast cancer: an on-site review of the Bezwoda study. Lancet 2000; 355:999-1003.

15. Weiss RB, Gill GG, Hudis CB. An on-site audit of the south african trial of high-dose chemotherapy for metastatic breast cancer and associated publications. J Clin Oncol 2001; 19:2771-2777.

16. Abdelkefi A, Ben Othman T, Kammoun L, et al. Prevention of central venous line-related thrombosis by continuous infusion of low-dose unfractionated heparin in patients with haemato-oncological disease: A randomized controlled trial. Thromb Haemost 2004; 92:654-661.

17. Abdelkefi A, Torjman L, Ladeb S, et al. Randomized trial of prevention of catheter-related bloodstream infection by continuous infusion of low-dose unfractionated heparin in patients with hematologic and oncologic disease. J Clin Oncol 2005; 23:7864-7870.

18. Abdelkefi A, Achour W, Ben Othman T, et al. Use of heparin-coated central venous lines to prevent catheter-related bloodstream infection. J Support Oncol 2007; 5:273-278.

19. Expression of Concern. Abdelkefi et al. J Clin Oncol 2005; 23:7864-7870. J Clin Oncol 2009; 27:4631.

20. Expression of Concern. Abdelkefi et al. Thromb Haemost 2004; 92:654-661. Thromb Haemost 2010; 104:1082.

21. Abdelkefi A, Maamar M, Torjman L, et al. Prospective randomised comparison of the COBE spectra version 6 and haemonetics MCS(+) cell separators for hematopoietic progenitor cells leucapheresis in patients with multiple myeloma. J Clin Apher 2006; 21:111-115.

22. Managing allegations of scientific misconduct: a guidance document for editors. Washington: Office of Research Integrity, 2000. ori.hhs.gov/documents/masm_2000.pdf (Accessed 14 August 2010).

23. Horton R. After Bezwoda. Lancet 2000; 355:942-943.

Figure 1A: Outcome of 378 patients with myeloma undergoing autotransplantation after 200 mg/m^2 melphalan at the Royal Marsden Hospital, UK. The last follow-up date is March 2000 – before relapsing patients had access to novel salvage therapy agents (see text for explanation). The blue and red lines show overall and progression-free survival of all 378 patients from the time of transplantation. The green line shows the survival of the 205 patients relapsing after transplantation.

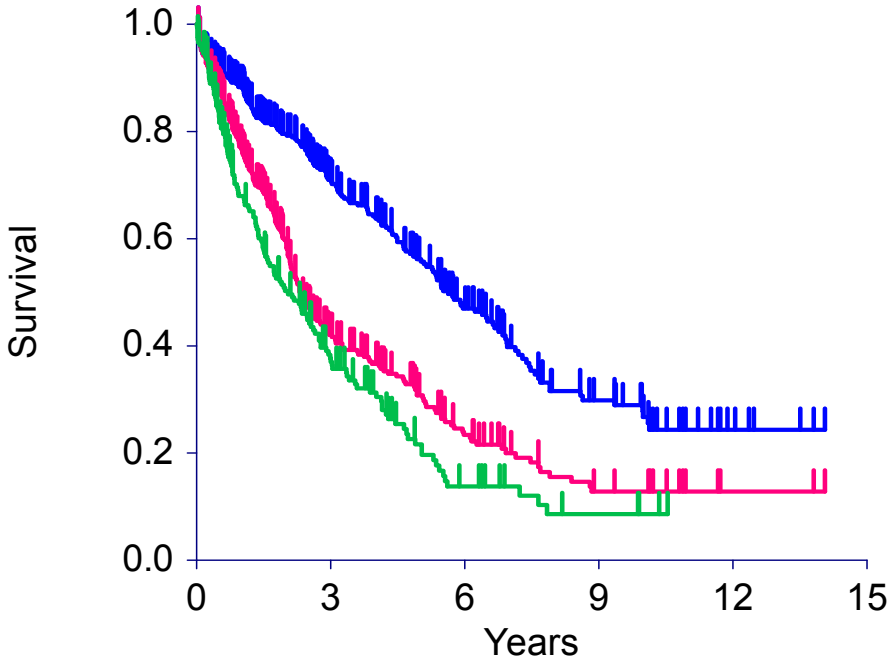

Figure 1B: Survival of patients relapsing after transplantation. The green line shows the Royal Marsden Hospital patients (n=205), the red line shows the Tunisian patients as published (n=27), and the blue line shows the updated outcome of the Tunisian patients (n=113).

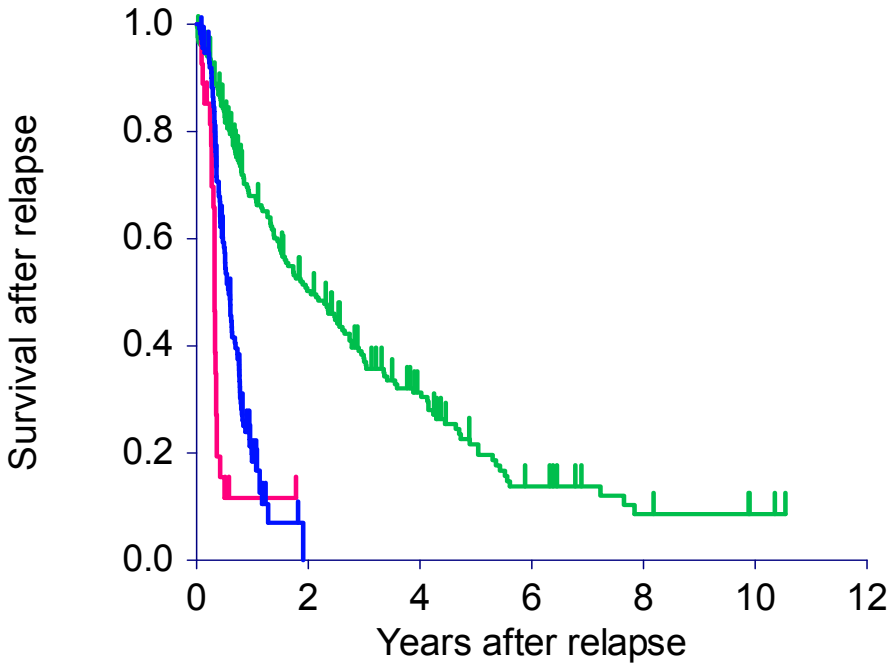

Figure 2: Illustration of the difference in the survival pattern of 195 patients from the Royal Marsden Hospital, UK (195 consecutive patients treated from April 1993 onwards; exactly 10 years prior to the Tunisian study – and followed up for 6 years) and the updated outcomes of the 195 patients in the Tunisian study. Censor marks have been removed to facilitate comparison. The blue and red lines show the overall and event-free survival of the UK patients, and the purple and green lines show the corresponding outcomes for the Tunisian patients.