

Transplants for acute myeloid leukaemia in 1st remission: statisticians, magicians and the rest of us

Robert P. Gale MD, PhD DSc(hc), FACP, FRSM

Haematology Research Centre, Division of Experimental Medicine, Department of Medicine, Imperial College London, London, UK SW7 2AZ

Robert Peter Gale, MD, PhD, DSc(hc), FACP, FRSM, Haematology Research Centre, Division of Experimental Medicine, Department of Medicine, Imperial College London, London, UK SW7 2AZ

Phone: +1-908-656-484

Fax: +1-310-388-1320

E-mail: robertpetergale@alumni.ucla.edu

Summary

The article concerns multiple factors influencing selection of patients with acute myeloid leukemia (AML) for hematopoietic stem cell transplantation. A number of prognostic and predictive variables may determine better probability of AML relapse, i.e., ROC analysis, thus allowing more accurate evaluation expressed in terms of concordance, or C-statistics. The final results are, however, subject to unexplainable variance.

Keywords

Hematopoietic stem cell transplantation, acute myeloid leukemia, relapse risk, treatment options, statistics, ROC analysis.



Figure 1. The Great Persky also considers transplants for AML in 1st remission

Professor [Joseph] Munro reminded him of an old saying which he rather reluctantly proposed, in that company, to repeat. It was to the effect that there were three gradations of inveracity – there were lies, there were d-d lies, and there were statistics.

Arthur James Balfour, 1st Earl of Balfour (Manchester Guardian, 29th June 1892)

If everyone in the world with acute myeloid leukaemia (AML) in 1st remission received a haematopoietic cell transplant we would know precisely how they fared. Forget sta-

tistics, confidence intervals, p-values, meta-analyses and the like, the outcome is the outcome. The problem is we do not have these data but data only from a subset of persons receiving a transplant and no data on many did not receiving a transplant in 1st remission. So, we need statistics applied to a small, selected sample of transplant recipients to try to estimate a larger truth: *what would be the outcome where everyone with AML in 1st remission received a transplant.* And with this approach come many assumptions, limitations and substantial uncertainty. As it turns out, people generally hate statistics but they hate uncertainty even more.

How can we rationally decide who should receive a transplant in 1st remission and who not. The answer hinges to a great extent on accurately estimating the probability of relapse in a person with AML in 1st remission. To make this estimate haematologists use *prognostic* and *predictive* variables, alone or combined into a score such as high-, intermediate- or low risk. Accuracy of this approach is best evaluated using a receiver-operator characteristic (ROC) curve with accuracy expressed as a concordance or C-statistic. The C-statistic is derived from the area under the curve (AUC) derived from the ROC curve. A C-statistic of 0.5 indicates no predictive accuracy and a value of 1, perfect predictive accuracy (*i.e.* no false-positives nor -negatives). However, the C-statistic has

limitations. For example, its value depends on the prevalence and/or distribution of covariates in the population being studied. Other estimators of accuracy include positive and negative predictive indices and net reclassification index.

Many variables and co-variables are associated with likelihood of relapse in someone with AML in 1st remission such as cytogenetics, WBC, numbers of cycles of induction therapy to achieve a complete remission, duration of complete remission at the time of assessment, results of *measurable* residual disease (MRD)-testing, expression of so-called leukaemia stem cell (LSC) associated genes *etc.* However, scores derived from these variables alone or combined explain only about one half of the variance in outcomes with C-statistics of about 0.65-0.75. The question is what accounts for the remaining unexplained variance. There are 3 sources: (1) unknown but potentially knowable (*latent*) co-variables; (2) measurement error; and (3) chance.

The issue of whether a person with AML in 1st remission should receive a haematopoietic cell transplant hinges on several assumptions: (1) we can predict which persons will relapse with *reasonable accuracy*; (2) a transplant can overcome the adverse biological features of high-risk AML; (3) there is an advantage to doing a transplant before relapse rather than waiting to see if a person relapses and then doing it *if needed*; and (4) we cause no harm if we predict leukaemia relapse incorrectly and transplant someone already cured by chemotherapy.

One demon confounding our estimates of outcomes and applicability of conclusions from a small sample to a wider population is *selection bias*. Selection bias sounds terrible, politically incorrect, like racial profiling. Perhaps something Donald Trump might suggest. However, selection biases operate in every aspect of our lives. For example, our old clothes dryer recently began making terrible noises. Death seemed imminent and a *do not resuscitate* order was written. I rushed online to read the *Consumer Reports* analysis of new dryers, let's say the universe of dryers (you would be amazed what's out there; forget targeted therapy). However, my wife Laura quickly ended my research. She wanted a Maytag (which was, sadly for me, expensive and low-rated by *Consumer Reports*). But she had a reason. Her mother wanted a Maytag but her father, a mechanical engineer, said he found a cheaper, better-rated brand in *Popular Mechanics*. According to Laura's mother (an involuntary but not impartial participant in the dryer experiment) the non-Maytag was a *loser*. She complained for the rest of her life, especially after the substitute dryer met an untimely end. *It never worked right* she pronounced. Who am I to argue; *happy wife, happy life*. Our Maytag is working great (6 months old; fingers crossed) and based on these data Laura pronounced the Maytag the greatest dryer in Earth. Reasonable? No, but happy Wife, happy life.

A more statistically-orientated definition of selection bias is a bias which occurs when the association between exposure (for example, an allotransplant) and a disease or condition (for example, AML) is different for those who complete a study compared with those in the target population, the overall population for which the measure of effect size is be-

ing calculated and from which study members are selected. What do you do with these limitations? My advice: *Be humble*. I am reminded of a line from a Woody Allen article in the *New Yorker* [1]. Kugelmass, an English professor at City University of New York (CCNY), is married to the now overweight Daphne and is seeing a psychiatrist, Dr. Mandel. He tells Mandel he is unhappy and dreams of romance, perhaps an affair with Emma Bovary. The psychiatrist thinks awhile and says: *Kugelmass, you need a magician, not a psychiatrist*. Viola! Enter the Great Persky, a Coney Island magician who accomplishes the task (but with a few amusing twists and turns. Strongly recommended).

Statisticians, like magicians, have lots of tricks up their sleeves. One is to analyze the data you have rather than the data you don't have. Terms like heterogeneity, random- and fixed-effects models, Cochran Q test, I² statistic, funnel plots, Egger test *etc.* magically appear. These manipulations, of course, greatly impress the non-statistician much like rabbits appearing in a hat or seeming to saw a beautiful woman in half. However, there is always a need for another non-statistical and imperfect but useful test: common sense (which, oddly, is distinctly uncommon). Can we rely on data from a very small sample of selected subjects to impute a higher truth? Does this make sense? Does it ring true? Psychologists and philosophers refer to this process as *thin slicing* [2]. Usually, your 1st impression is correct. Sometimes it's not, something referred to as the Herbert Hoover effect. This type of mistake can have tragic consequences: witness President Donald Trump.

Which brings us to the ability of physicians to predict how their patients will do. Prediction is imperfect, as Niels Bohr pointed out: *especially about the future*. Consider Field Marshal Ferdinand Foch in 1914: *Airplanes are interesting toys but of no military value*. However, physicians are somewhat better than Foch in predictions. Above I have discussed uses and limitations of analyses of prediction accuracy using a ROC curve and C-statistic. However, physicians claim to have the 6th sense, a bit like umami, which we cannot quantify, at least not yet, and which they think allows them to add something to these *predictive* scores. Unfortunately, this seems wrong. When formally-tested the C-statistic for physicians' estimates is only about 0.6, substantially worse than objective *prognostic* and *predictive* scores (Estey; unpublished). So much for MSG.

Let's return to the assumptions underlying the consideration of whether to do a transplant with someone with AML in 1st remission and see how many are proved:

(1) We can predict which persons will relapse with *reasonable accuracy*.

As discussed above our ability to accurately predict whether someone with mission will relapse is accurate only about one half other time. For example, results of MRD-testing are associated with about 30 percent false-positives and 30 percent – negatives. So, the answer hinges on how comfortable a haematologist is in being wrong in about 1 in every 3 people he/she treats. If an intervention is not dangerous, say giving aspirin, these error rates might be acceptable. Whether they are acceptable in the context of a transplant is more complex.

(2) A transplant can overcome the adverse biological features of high-risk AML.

This is unproved and requires a clinical trial in which persons predicted to be at substantial relapse risk are randomly assigned to conventional therapy or a transplant. No such data are reported but a trial with this design is underway in the UK. However, generally the poor outcomes associated with adverse biological features are only modestly overcome by more intensive interventions.

(3) There is an advantage to doing a transplant before relapse rather than waiting to see if someone relapses and then doing it *if needed*.

This is also unproved requiring data from a randomized clinical trial. However, data from controlled, non-randomized trials suggest waiting for relapse and transplanting only persons who relapse results in the same survival as transplanting larger numbers of persons in 1st remission. Whether persons who relapse and are in good clinical condition need to receive therapy to try to achieve a 2nd remission before proceeding to a transplant is also unproved and unlikely to be correct.

(4) We cause no harm if we predict leukaemia relapse inaccurately and transplant someone already cured by chemotherapy.

Obviously wrong. A transplant can kill someone already cured by chemotherapy.

The bottom line is most assumptions underlying doing a transplant for persons with AML in 1st remission are un-

proved and/or wrong. How do we then explain why so many of these persons receive a transplant in 1st remission? Perhaps we need to ask the Great Persky who got Kugelmass from Coney Island to Charles and Emma Bovary's bedroom in Yonville, France *where before him was a beautiful woman, standing alone with her back turned to him as she folded some linen. I can't believe this, thought Kugelmass, staring at the doctor's ravishing wife.* His caution was warranted.

Acknowledgement

RPG acknowledges support from the National Institute of Health Research (NIHR) Biomedical Research Centre funding scheme.

Conflict of interest

No conflict of interest is declared.

References

1. Allen W. The Kugelmass Episode, *New Yorker*, May 2, 1977; p. 34 and thereon.
2. Gladwell M. *Blink: The power of thinking without thinking*. New York, NY. Back Bay Books, Little Brown and Company, 2005.

Трансплантация в 1-й ремиссии острого миелобластного лейкоза: статистики, маги и прочие

Роберт П. Гэйл

Гематологический исследовательский центр, Отдел экспериментальной медицины, Департамент медицины, Имперский Колледж, Лондон, Великобритания

Резюме

В статье обсуждается многообразие факторов, влияющих на отбор пациентов с острым миелобластным лейкозом (ОМЛ) для трансплантации гемопоэтических клеток. Чтобы уточнить расчеты вероятности рецидива в 1-й ремиссии ОМЛ, оцениваются соответствующие прогностические и предиктивные переменные величины, например, применяется построение и анализ кривых ROC с точностью, выражаемой в виде конкордантности или C-статистики. Вопрос состоит в том, за счет чего возникает остающаяся необъяснимой изменчивость результатов. У нее есть 3 источника: (1) неизвестные, но потенциально выявляемые (скрытые) переменные; (2) ошибка измерений; и (3) случайность. Поэтому мы не можем сейчас с достаточной точностью предсказать, у кого из больных разовьется рецидив.

Например, результаты тестирования минимальной остаточной болезни ассоциированы примерно с 30% ложно-положительных и ложно-негативных ответов. Также невозможно заключить о том, есть ли выгода от проведения трансплантации до рецидива, или же ждать развития рецидива, а потом выполнять трансплантацию по необходимости. Некоторые из этих посылок нужно доказать в рандомизированных клинических исследованиях.

Ключевые слова

Трансплантация гемопоэтических стволовых клеток, острый миелобластный лейкоз, риск рецидива, выбор лечения, статистика, ROC-анализ.