

Monitoring Therapy-Response in acute myeloid leukemia

Robert Peter Gale

Haematology Research Centre, Division of Experimental Medicine, Department of Medicine, Imperial College London, London, UK.

Professor Robert Peter Gale, MD, PhD, DSc(hc), FACP, Haematology Research Centre, Division of Experimental Medicine, Department of Medicine, Imperial College London, Exhibition Road, South Kensington, London SW7 2AZ, United Kingdom

Phone: +44 20 7589111
Fax: +44 20 7594 8759
E-mail: rgale@celgene.com

Summary

Quantification of measurable residual disease (MRD) is increasingly used to determine whether a person with acute myeloid leukaemia (AML) in the 1st complete remission should receive a haematopoietic cell transplant. But how accurate are results of MRD-testing in this setting in predicting likelihood of leukaemia-relapse? There are 2 major determinants of accuracy of results of MRD-testing: (1) adequacy of sampling (sometime referred to as sampling-error); and (2) precision of the MRD-testing being used reflecting sensitivity and specificity of the assay.

Most analyses of the accuracy of MRD-testing are on the cohort level. E.g., most analyses of the accuracy of MRD-testing are on the cohort level. Typically, an outcome such as cumulative incidence of leukaemia-relapse, relapse-free survival for MRD⁺ vs MRD⁻ cases. The problem, for example, is that some subjects in a favourable risk cohort may have a worse prognosis than some subjects in an unfavourable risk cohort. One should mean a loss of predictive power with time post-treatment, i. e., the risk-determining variables are less powerful in a person remaining in remission at 3-4 months when a transplant decision is typically made. Other questions arise due to unidentified prognostic variables. Under these conditions, the ROC analysis shows rather low prediction accuracy. This is because of different stochastic events determining probable outcomes of the therapy applied.

Moreover, even a perfect MRD-test (100 percent sensitivity, 100 percent specificity) will have limited precision in predicting leukaemia-relapse in persons with AML in 1st complete remission. This is because of unavoidable sampling error which is further confounded by the high likelihood of leukaemia cells, and especially those causing leukaemia relapse in the blood and bone marrow. This impact would be associated with more common false-negative MRD-test results.

Moreover, AML is a genetically complex neoplasm at diagnosis and even more at the leukaemia relapse. Recently, some marker mutations previously thought to be typical of AML have been found in normal, older persons (e. g., DMNT3A, TET2 and IDH mutations). Their lack of specificity for AML would result in a substantial rate of false-positive MRD-test results.

In conclusion, although results of MRD-testing are correlated with probability of relapse in cohorts of persons with AML in 1st complete remission, there are substantial barriers to applying results of MRD-testing to recommendations for individual therapy.

Keywords

Acute myeloid leukemia, hematopoietic stem cell transplantation, relapse, minimal residual disease, prognostic values

Introduction

Recently quantification of measurable residual disease (MRD) has been increasingly-used to determine whether a person with acute myeloid leukaemia (AML) in 1st complete remission should receive a haematopoietic cell transplant. But how accurate are results of MRD-testing in this setting in predicting likelihood of leukaemia-relapse? I consider this issue in this commentary. I will provide data indicating results of MRD-testing in this setting are associated with substantial rates of false-negative and -positive predictions. The level of tolerance to be wrong and level of precision required or desired determine how useful results of MRD-testing are in deciding whether or not to recommend a person with AML in 1st complete remission should receive a haematopoietic cell transplant.

There are 2 major determinants of accuracy of results of MRD-testing: (1) adequacy of sampling (sometime referred to as sampling-error); and (2) precision of the MRD-testing being used reflecting sensitivity and specificity of the assay.

Predicting leukemia relapse in a subject is different than predicting leukaemia relapse in a cohort

Most analyses of the accuracy of MRD-testing are on the cohort level. Typically, an outcome such as cumulative incidence of leukaemia-relapse, relapse-free survival or survival is compared between cohorts of subjects with a positive or negative MRD-test result. Almost all of these studies report a greater risk of leukaemia-relapse in the cohort with a positive MRD-test result.

This association is often confirmed in multivariable regression analyses. Frequently results of MRD-testing are a stronger predictor of leukaemia-relapse than other leukaemia-associated variables such as age, gender, cytogenetic and/or molecular risk cohort. This is not surprising and indicates the predictive value of results of MRD-testing. However, more detailed analyses of hazards of leukaemia relapse indicate most of the difference in leukaemia relapse risk occurs within the 1st few months after the MRD-test is done. Afterwards, hazards of leukaemia relapse risk are similar. Reasons for this non-proportional hazard of leukaemia relapse relates primarily of residual numbers of leukaemia cells and are confounded with sampling error.

It is important to consider that predicting leukaemia relapse in a person is a different challenge than predicting leukaemia relapse in a cohort. Risk predictions of cohorts are typically given as point estimates with confidence intervals. Often these confidence intervals of cohorts are wide and may overlap even when the p-value for trend is significant. The con-

sequence, for example, is some subjects in a favourable risk cohort may have a worse prognosis than some subjects in an unfavourable risk cohort. This observation should begin to caution the reader to the hazards of apply cohort predictions to predictions in cohort members. For example, although a cohort may have a 30 percent risk of leukaemia relapse, no member of the cohort has a 30 percent risk. The individual risk is, in contrast, a binary, leukaemia-relapse or no leukaemia-relapse, 0 or 100 percent.

Loss of predictive power with time post-treatment

Most of the variables correlated with outcomes of people with AML such age, gender, WBC and/or cytogenetic or molecular risk cohort are most powerful at diagnosis. This is because they are associated with likelihood of achieving complete remission. Once a person achieves complete remission the prognostic power of these variables decreases. Furthermore, most variables are also associated with a risk of early leukaemia-relapse in persons achieving complete remission. Consequently, the variables are less powerful in a person remaining in remission at 3-4 months when a transplant decision is typically made.

Unidentified prognostic variables and chance

A key question is how much of the variance associated with outcomes in a population with AML are explained by known prognostic variables? This question is typically analyzed using a receiver-operator characteristic (ROC) curve and the C-statistic derived from this curve. A C-statistic of 0.5 implies the variable or combination of variables has no predictive value whereas a C-statistic of 1,0 implies perfect prediction. When the combination prognostic variables typically used in AML analyses is applied to persons with AML in complete remission for 3-4 months the C-statistic is about 0,7. This means prediction accuracy is better than random but far from perfect.

Why is this so? There are several possible answers but 2 most-likely are: (1) impact of potentially knowable but currently unknown variables; and (2) chance. Most scientists want to believe we can explain all uncertainty. The implication is we will eventually find other prognostic variables which allow us greater predictive precision. However, the expectation that all variance will be explainable is unlikely. This is because, much to most scientists' dismay, stochastic events are important determinants of outcomes of therapy-interventions. Failure to acknowledge this fact is a major obstacle to understanding and accepting limitations of prediction models. Put otherwise, reality is the leading cause of stress amongst those in touch with it. (Jane Wagner.)

Sampling error

Most MRD-tests used in AML have only modest levels of sensitivity and specificity (see below). However, even a perfect MRD-test (100 percent sensitivity, 100 percent specificity) will have limited precision in predicting leukaemia-relapse in persons with AML in 1st complete remission. This is because of unavoidable sampling error. For example, it is common to use a 5-10 ml sample of blood or bone marrow for MRD-testing. This is probably OK when there are many leukaemia cells in the body (say $10E+11$) i. e. every sample, even small sample volumes, are likely to contain ≥ 1 leukaemia cells. If so, a perfect MRD-test will detect them unambiguously with no false-negative or -positive results. (An unambiguous MRD-test result is different than a correct prediction.) However, when numbers of leukaemia cells in the blood and bone marrow decrease as a result of induction and post-remission therapies the MRD-test result will be a false-negative because there may not be a leukaemia cell in the sample whereas there may be many in the person being sampled. The probability of having ≥ 1 leukaemia cells in a small sample is reflected in a Poisson probability distribution (Figure 1). Examination of Figure 1 shows at low frequencies of leukaemia cells in a person, say $10E+3$, there is a substantial probability there will be no leukaemia cell in a small sample resulting in a false-negative MRD-test result. This unavoidable sampling error limitation is further confounded by the high likelihood leukaemia cells, and especially those with the biological ability to cause leukaemia relapse, are uniformly-distributed in the blood and bone marrow. The impact would be to increase the likelihood of false-negative MRD-test results. Naturally this error is confounded when a MRD-test has less than perfect sensitivity and/or specificity. An example of a substantial rate of false-negative MRD-test results is shown in Figure 2. This approximately 30 percent rate of false-negative MRD-test results in persons with AML in 1st complete remission after completing consolidation therapy likely results from a combination of sampling error and an imperfect M-test (see below).

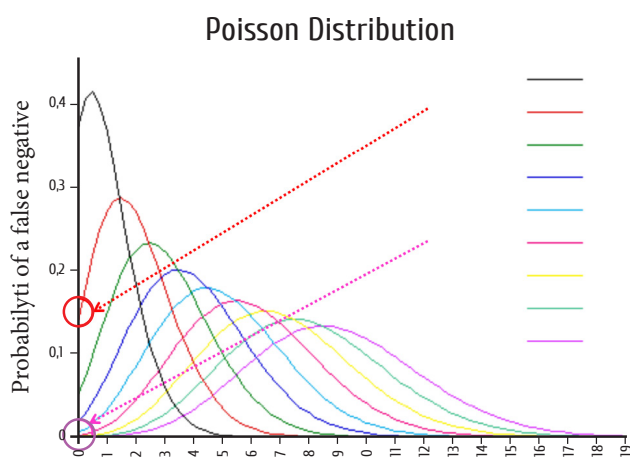


Figure 1. Increased probability of false-negative MRD-test from a small sample of blood or bone marrow when numbers leukaemia cells in a person is decreased by therapy.

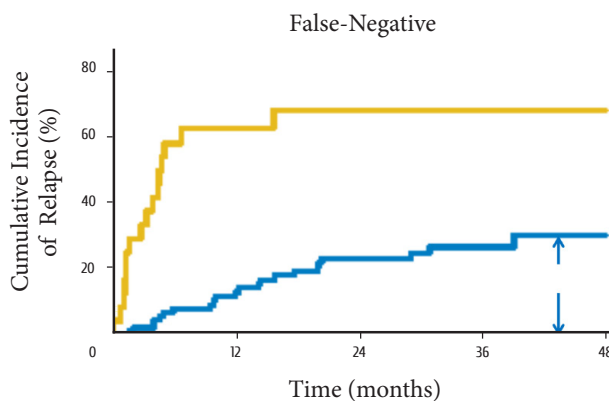


Figure 2. False-negative MRD-test results in persons with AML in 1st complete remission after completing consolidation therapy. About 30% of persons with a negative MRD-test result had leukaemia relapse. This is likely the result of sampling error and imperfect MRD-test sensitivity (from Terwijn et al., 2013).

Imperfect MRD-test sensitivity and specificity

It is unlikely that any MRD-test in AML can have 100 percent sensitivity and 100 percent specificity. For example, it is clear not every AML cell has the biological capability to cause leukaemia-relapse. Presently, we lack sensitive tests to distinguish leukaemia cells which can and cannot cause leukaemia relapse. The obvious consequence of this lack of discrimination is a high likely likelihood of false-positive MRD-test results. For example, a MRD-test which detects an AML cell which cannot cause leukaemia relapse will result in a false-positive MRD-test. Figure 3 shows data from the same study shown in Figure 2 indicating an about 30 percent frequency of false-positive MRD-test results.

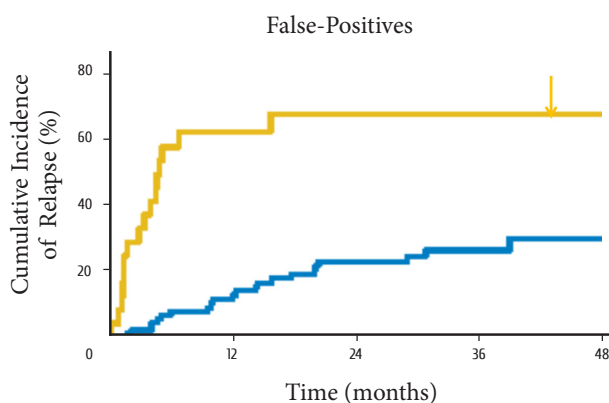


Figure 3. False-positive MRD-test results in persons with AML in 1st complete remission after completing consolidation therapy. About 30% of persons with a positive MRD-test result did not have leukaemia relapse during the observation interval. This is likely the result of imperfect MRD-test specificity (from Tei-jwin et al., 2013).

Complexity of AML

AML is a complex neoplasm with considerable genetic and clonal complexity at diagnosis and even more so when there is leukaemia relapse. There is substantial genotypic diversity between people with AML. Data from whole exome and whole genome sequencing indicate every case of AML analyzed at diagnosis is genotypically unique; no 2 share an identical mutational profile. This makes it difficult or unlikely there can be a uniform genotype-based MRD-test in newly-diagnosed persons with AML. This heterogeneity is further increased at relapse between and within persons with AML. The situation with phenotypic analyses is equally complex. In addition to phenotypic diversity between person with AML considerable data suggest phenotypic diversity with a person with AML at diagnosis and especially at relapse. These considerations make it unlikely tests for MRD in persons with AML can achieve the specificity of MRD tests in acute lymphoid leukaemia (ALL) where the usual testing target is the B- or T-cell clone and not the neoplasms. We can have some appreciation of the complexity of MRD-testing in AML if we consider precision of MRD-testing in chronic myeloid leukaemia (CML) a much simpler neoplasm caused by 1 mutation (BCR/ABL) and for which we have highly sensitive and specific mRNA-based tests for MRD. Here the rate of false-negative MRD tests in the context of stopping imatinib therapy is about 60 percent (Figure 4). This situation is further complicated by recent reports of the detection of mutations previously thought to be typical of AML in normal, older persons not developing AML in their lifetime. Examples include mutations in DMNT3A, TET2 and IDH1 and IDH2. If these mutations were used in MRD-testing their lack of specificity for AML would result in a substantial rate of false-positive MRD-test results.

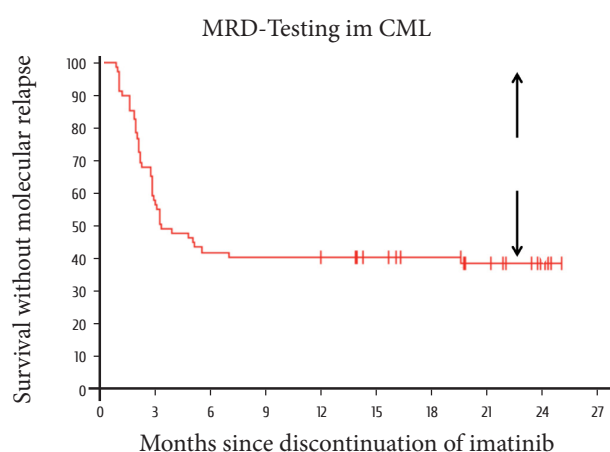


Figure 4. Rate of false-negative results in persons with CML discounting imatinib after having a negative MRD-test for BCR/ABL. The about 60% false-negative rate reflects imperfect sensitivity of the MRD-test.

Even accurate results may not be actionable

For results of an MRD-test to be clinically-useful the results must be accurate and actionable. For example, if one were to use a positive MRD-test as the basis for recommending a haematopoietic cell transplant one would need convincing data doing a transplant would result in a better outcome than an alternate such as no further therapy or more conventional chemotherapy. Proof of efficacy can only be obtained in a randomized trial in which persons with a positive MRD-test result are assigned to a transplant or an alternate. No such study is reported. However, a study underway in the UK may inform the question of whether a positive MRD-test result is actionable.

Conclusions

Although results of MRD-testing is correlated with probability of leukaemia relapse in cohorts of people with AML in 1st complete remission there are substantial barriers to applying results of MRD-testing to individual therapy recommendations. For reasons discussed a positive MRD-test result is likely to be wrong in 1 in 3 instances. Similarly, a negative MRD-test result is likely to be wrong in 1 of 3 instances. Thus, using MRD-test results as the basis for recommending a transplant to someone with AML in 1st remission requires a high tolerance level for being wrong. This is especially concerning in persons with a positive MRD-test about 25-35 percent are already cured and cannot benefit but can be substantially harmed by a transplant. Further, we presently lack convincing data the adverse prognostic impact of a positive MRD-test can be reversed by doing a transplant. I hope this commentary will cause physicians to more carefully weigh results of MRD-testing in their prognostic metric. Comments welcome.

Acknowledgement

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

References

1. Terwijn M, van Putten WL, Kelder A, van der Velden VH, Brooimans RA, Pabst T, et al. High prognostic impact of flow cytometric minimal residual disease detection in acute myeloid leukemia: data from the HOVON/SAKK AML 42A study. *J Clin Oncol.* 2013 Nov 1; 31 (31): 3889-3897.

Мониторинг ответа на терапию при остром миелобластном лейкозе

Роберт П. Гэйл

Центр гематологических исследований, отделение экспериментальной медицины, департамент медицины, Лондонский Имперский колледж, Лондон, Великобритания

Резюме

Количественная оценка определяемой остаточной болезни (МОБ) все чаще используется для решения о том, должна ли проводиться пересадка кроветворных клеток больному с острым миелоидным лейкозом (ОМЛ) в первой полной ремиссии. Однако насколько точны результаты тестирования МОБ в этой ситуации для прогнозирования вероятности рецидива лейкоза? Имеются две основные детерминанты точности результатов: (1) адекватность (или ошибка) сбора биоматериала, и (2) точность тестирования МОБ, в плане чувствительности и специфичности диагностического метода.

Так, например, большинство оценок точности тестирования МОБ проводится на уровне когорт пациентов. Обычно исходы оценивают в виде кумулятивной частоты рецидивов лейкоза и безрецидивной выживаемости для МОБ-положительных и негативных пациентов. Проблема, в частности, состоит в том, что некоторые больные из благоприятной группы риска могут иметь худший прогноз, нежели некоторые лица из неблагоприятной группы риска. Следует иметь в виду возможность утраты прогностического потенциала со временем после лечения, т.е. переменные, определяющие риск, имеют меньшую силу у человека в ремиссии в течение 3-4 мес., в тот срок, когда решается вопрос о трансплантации. Дополнительные проблемы возникают из-за невыясненных прогностических факторов. В этих случаях результаты ROC-анализа показывают значительное снижение точности прогнозирования рецидива, что связано с различными стохастическими событиями, определяющими возможные исходы

применяемой терапии. Кроме того, даже вполне совершенный тест на наличие МОБ со 100%-ной чувствительностью и 100%-ной специфичностью будет иметь ограничения по точности прогнозирования рецидивов лейкоза у лиц с ОМЛ в 1-й полной ремиссии. Это связано с неизбежной ошибкой сбора материала, которая затем еще более искажает результат ввиду высокой вероятности лейкозных клеток и, в особенности, клеток – источников рецидива в крови и костном мозге. Их влияние может быть связано с учащением ложно-негативных результатов теста на МОБ. Кроме того, ОМЛ является заболеванием со сложным генетическим составом и, в еще большей мере – в рецидиве заболевания. Недавно было показано, что некоторые маркерные мутации, ранее считавшиеся типичными для ОМЛ (например, мутации DNMT3A, TET2 и IDH), обнаруживают также у здоровых лиц старших возрастов. Их недостаточная специфичность при ОМЛ приводила бы к значительному снижению частоты ложно-положительных результатов тестирования МОБ.

В заключение, хотя результаты тестирования МОБ коррелируют с вероятностью рецидива в группах больных АМЛ в первой полной ремиссии, имеются существенные объективные препятствия к использованию результатов оценки МОБ для рекомендаций по индивидуальной терапии.

Ключевые слова

Острый миелобластный лейкоз, трансплантация гемопоэтических стволовых клеток, рецидивы, минимальная остаточная болезнь, прогностическая значимость.