EDITORIAL

Acute myeloid leukemia

New study-designs to address the clinical complexity of acute myeloid leukemia

Elihu Estey[1] · Megan Othus[2] · R. P. Gale [ID] [3]

For many years, studies of new drugs for acute myeloid leukemia (AML) and other cancers have used a three-step approach moving from phase-1 to −2 to −3. Phase-1 trials typically estimate the new drug's maximum tolerated dose (MTD) or a dose maximally impacting its presumed target (optimal biologic dose, OBD). Phase-2 trials evaluate efficacy, often defined as *response*. A conclusion the response rate justifies further study frequently leads to a phase-3 trial, randomly assigning subjects to receive the new drug or conventional therapy to determine which is better. The focus in phase-3 is generally on one primary outcome such as event-free survival (EFS) or survival. In each type of trial, the endpoint not considered primary is termed secondary, viz, response in phase-1 or toxicity in phase-2 and often not formally evaluated, especially if the trial fails to meet the primary endpoint.

Much of the current treatment of AML has evolved from this approach, resulting in regulatory approval of seven new drugs for AML in 2017–2018 [1–3], although the improvements afforded are modest, particularly in absolute rather than relative terms, and their applicability to all adults with the disease is uncertain [4]. In any event, the standard approach to trials ignores the complexity of AML and, very likely, what researchers and research subjects want to know. Here we discuss: (1) the focus on one primary endpoint; (2) disregard in phase-1 and −2 studies for the heterogeneity of AML; (3) use of generic false-positive and -negative rates; and (4) use of study-designs insufficiently adaptive in phase-3.

Physicians are often interested in multiple outcomes, for example, not only *safety* but also *response*, not only *response* but also *survival*, not only *survival* but also *quality-of-life* and so forth. We doubt many research subjects envision that the sole purpose of a phase-1 trial is identifying the MTD or OBD for future studies. Rather, most participate for a tangible personal benefit such as improved survival. But, because phase-1 trials often evaluate efficacy as a secondary objective, discordance arises with the investigator (seemingly) primarily interested in safety and the subject interested in safety but also efficacy. As typically only 6–20 subjects are treated at the MTD in phase-1, relatively little is known about toxicity after one phase-1 trial; nonetheless, phase-2 trials usually only informally monitor toxicity. Likewise, in phase-3 trials, arguments can be made for the primacy of survival or EFS as criteria for regulatory approval a new therapy [5]. However, only one of these endpoints is usually considered primary. Not only can the distinction between primary and secondary endpoints be arbitrary but much less attention is paid to the latter. In AML, for example, survival rather than complete remission is typically the primary endpoint of phase-3 trials. Although there are discordances between complete remission and survival [6, 7], many clinicians would argue that there is value in achieving a complete remission, for example, the possibility of fewer transfusions, less time in hospital, or increased psychological well-being, even if survival is not improved. Because complete remission rate is typically a secondary endpoint, current study-designs provide little encouragement to explore these possibilities. In all this, there is a loss of information inconsistent with subjects' expectations after giving informed consent.

It is obvious that different subjects are a priori at different risks of toxicity, often motivating exclusion of persons with ECOG performance scores of 3–4, for example, from phase-1 trials. However, Rogatko et al. [8] reported that subject-

✉ Elihu Estey
eestey@uw.edu

1 Divisions of Hematology and Clinical Research, University of Washington and Fred Hutchinson Cancer Research Center, Seattle, WA, USA

2 M. Othus - Public Health Division and SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

3 R. P. Gale - Haematology Research Centre, Division of Experimental Medicine, Department of Medicine, Imperial College London, London, UK

specific variables interact with dose in determining toxicity among subjects who are routinely eligible for phase-1 trials . Nonetheless, if two of the first three subjects at a dose-level in a phase-1 study have an adverse event in a trial conducted using the conventional $3 + 3$ design that dose-level is declared unsafe and never re-visited, regardless of whether the subjects were 30 or 70 years old, had an ECOG performance score of 0 or 2 or had a bilirubin of 0.6 or 1.4 mg/dL, values typically consistent with trial-entry. Single-arm phase-2 trials are inherently comparative: the worse the estimated efficacy compared with a perceived standard treatment, the less the motivation to start a phase-3 study. Many heterogeneous biologic covariates are associated with efficacy outcomes in AML [9]. Nonetheless, phase-2 trials typically assume the only effect being measured is the drug being tested rather than subject- and disease-related variables [10], measurement error and chance.

Phase-3 trials in AML routinely stipulate false-positive and -negative rates of, respectively, 5% and 10–20%, a metric common to trials in many diseases. However, these error rates seem more acceptable in diseases with effective therapies, for example, in hypertension or diabetes but less so in a disease without effective therapy such as poor-prognosis AML. Here, the consequences of a false-negative result are more, and of a false-positive result less, substantial. Phase-3 trials in AML might allow false-positive rates similar to the 20%, often stipulated in randomized phase-2 trials. Certainly, the time needed to (eventually) discover the false-positive is time that might otherwise be spent studying other new therapies. Nonetheless, we believe use of generic false- positive and -negative rates is difficult to defend in AML.

Given our limited ability to accurately predict outcomes of subjects receiving a new therapy had they received an older one [11, 12], there is little doubt of the need for randomized trials. Many new therapies turn out no better or even worse than older ones. However, intuition suggests many people properly informed (i.e., beyond the brief, standardized alternative therapy section of the usual informed consent document) of the likely unsatisfactory outcome with conventional therapies would reason "how much worse than the conventional therapy can the new therapy be" and decline randomization. Furthermore, very few AML trials use current outcomes data to influence randomization probabilities. This practice challenges subjects' expectation their physicians are constantly learning to improve their care.

Modern study-designs can address some of these problems. Examples are designs that (a) simultaneously monitor multiple outcomes and make adaptive decisions based on more than one outcome [13, 14], (b) account for covariates in phase-1 and −2 trials [15, 16], or (c) allow repeated outcome-adaptive randomization [17]. Although available for years and using either frequentist or Bayesian frameworks [18, 19], these study-designs are rarely used; only 2% of 1235 phase-1 trials conducted between 1991 and 2006 used an innovative statistical design [20]. We doubt the situation is substantively different today.

New study-designs often require more subjects, time, and resources than current phase-1, −2, and −3 designs. Outcome-adaptive randomization typically requires ~ 15% more subjects, comparable to the increase in sample size with 2:1 rather than 1:1, randomization [21]. The resultant longer trials might delay approval of an effective therapy. Hence, balancing between subjects' preference and public health benefit is necessary. In contrast, larger sample sizes in phase-1 and −2 trials might result in fewer expensive, time-consuming, negative phase-3 trials [22].

There are many reasons why phase-3 trials in AML often fail to confirm the promise of earlier trials [4, 23]. We discuss several reasons here. We argue the disruption and perceived inconvenience of using more-modern adaptive trial designs may be justified. Our hypothesis can only be tested if these newer study-designs find wider use.

Compliance with ethical standards

Conflict of interest RPG is a part-time employee of Celgene Corp.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. https://www.fda.gov/drugs/informationondrugs/approveddrugs/ucm279174.htm; accessed 23 Noveber 2018
2. Blumenthal G, Kim G, Pazdur R. Setting the record straight on new drug approvals in oncology. JAMA Intern Med. 2017;177:122.
3. Farrell A, Goldberg K, Pazdur R. Flexibility and innovation in the FDA's novel regulatory approval strategies for hematologic drugs. Blood. 2017;130:185–89.
4. Estey E, Gale R, Sekeres M. New drugs in AML: uses and abuses. Leukemia. 2018;32:1479–81.
5. Estey E, Othus M, Lee S, Appelbaum F, Gale R. New drug approvals in acute myeloid leukemia: what's the best endpoint? Leukemia. 2016;30:521–25.
6. Burnett AK, Russell NH, Hunter AE, Milligan D, Knapper S, Wheatley K, et al. Clofarabine doubles the response rate in older patients with acute myeloid leukemia but does not improve survival. Blood. 2013;122:1384–94.
7. Burnett AK, Hills RK, Hunter AE, Milligan D, Kell J, Wheatley K, et al. The addition of gemtuzumab ozogamicin to low-dose Ara-C improves remission rate but does not significantly prolong

survival in older patients with acute myeloid leukaemia: results from the LRF AML14 and NCRI AML16 pick-a-winner comparison. Leukemia. 2013;27:75–81.

8. Rogatko A, Babb J, Wang H, Slifker M, Hudes R. Patient characteristics compete with dose as predictors of dose in early phase clinical trials. Clin Cancer Res. 2004;10:4645–51.

9. Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood. 2017;129:424–47.

10. Estey E, Gale R. Acute myeloid leukemia and the chosen people. Leukemia. 2016;31:269–71.

11. Estey E, Gale R. How good are we at predicting the fate of someone with acute myeloid leukaemia? Leukemia. 2017; 31:1255–58.

12. Othus M, Wood B, Stirewalt D, Estey E, Petersdorf S, Appelbaum F, et al. Effect of measurable("minimal") residual disease(MRD) information on prediction of relapse and survival in adult acute myeloid leukemia. Leukemia. 2016;30:2080–83.

13. Thall P, Russell K. A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. Biometrics. 1998;54:251–64.

14. Thall P, Simon R, Estey E. New statistical strategy for monitoring safety and efficacy in single arm clinical trials. J Clin Oncol. 1996;14:296–303.

15. Rogatko A, Babb J, Tighiouart M, Khuri F, Hudes G. New paradigm in dose-finding trials: patient-specific dose finding and beyond phase 1. Clin Cancer Res. 2005;11:5342–46.

16. Wathen J, Thall P, Cook J, Estey E. Accounting for patient heterogeneity in phase 2 clinical trials. Stat Med. 2008; 27:2802–15.

17. Giles F, Kantarjian H, Cortes J, Garcia-Manero G, Verstovsek S, Faderl S, et al. Adaptive randomized study of idarubicin and cytarabine versus troxacitabine and cytarabine versus troxacitabine and idarubicin in untreated patients 50 years or older with adverse karyotype acute myeloid leukemia. J Clin Oncol. 2003;21:1722–27.

18. Berry D. Bayesian clinical trials Nature Reviews Drug Discovery.. 2006;5:27–36

19. Yuan Y, Hess K, Hilsenbeck S, Gilbert M. Bayesian optimal interval design: a simple and well-performing design for phase 1 oncology trials. Clin Cancer Res. 2016;22:4291–301.

20. Rogatako A, Schoeneck D, Jonas W, Tighiouart M, Khuri F, Porter A. Translation of innovative designs into phase 1 trials. J Clin Oncol. 2007;25:4982–86.

21. Schoenfeld D. "Sample-size formula for the proportional-hazards regression model.". Biometrics. 1983;39:499–503.

22. Zia M, Siu L, Pond G, Chen E. Comparison of outcomes of phase 2 studies and subsequent randomized control studies using identical chemotherapeutic regimens. J Clin Oncol. 2005; 23:6982–91.

23. Walter R, Appelbaum F, Tallman M, Weiss M, Larson R, Estey E. Shortcomings in the clinical evaluation of new drugs: acute myeloid leukemia as paradigm. Blood. 2010;116:2420–28.