Received: 27 January 2025 | Accepted: 14 May 2025

DOI: 10.1111/bjh.20185

BJHaem
BRITISH JOURNAL OF HAEMATOLOGY

ORIGINAL PAPER

Haematological Malignancy – Clinical

Comparability of external and internal control patients for the prospective randomized HOVON-103 trial in older AML patients

Sjoerd J. F. Hermans[1] | Jurjen Versluis[1] | Erik D. van Werkhoven[1,2] | Yvette van Norden[1,2] | Jeroen J. W. M. Janssen[3] | Gerwin A. Huls[4] | Thomas Pabst[5] | Dimitri A. Breems[6] | Elizabeth Berkx[7] | Avinash G. Dinmohamed[7] | Peter C. Huijgens[7] | Eric Sträng[8] | Jesús María Hernández Rivas[9] | Marta Sobas[10] | Rosa Ayala Diaz[11] | Joaquin Martinez Lopez[11] | Klaus H. Metzeler[12] | Torsten Haferlach[13] | Christian Thiede[14] | Carin A. Uyl-de Groot[15] | Lars Bullinger[16] | Bob Löwenberg[1] | Gert J. Ossenkoppele[17] | Francesco Pignatti[18] | Jan J. Cornelissen[1,18]

Correspondence
Jurjen Versluis, Erasmus University Medical Center Cancer Institute, Dr. Molewaterplein 40, Rotterdam 3015GD, The Netherlands.
Email: j.versluis.1@erasmusmc.nl

Summary

Real-world data (RWD) previously contributed to post-marketing regulatory decision-making, but are currently also considered as external controls to single-arm trials. The use of RWD control data may be compromised by methodological issues, urging validation of RWD control cohorts. Two external control cohorts of newly diagnosed acute myeloid leukaemia patients, one registered by the HARMONY Alliance (HA) and one by the Netherlands Cancer Registry (NCR), were compared to the control arm of the randomized HOVON-103 trial (H103 controls). All patients, aged >65 years with a WHO performance score of 0–2 (or missing), received standard induction chemotherapy. 1:1 propensity score calliper matching (PSM) was applied to improve comparability, and overall (OS) and relapse-free survival (RFS) were assessed. Fewer data elements were available in external cohorts compared to H103 controls, specifically in the NCR cohort. Baseline characteristics of the external cohorts differed from H103 controls; missing data were also more frequent and predominantly concerned WHO performance score. After PSM, HA patients demonstrated non-significantly different OS and RFS to H103 controls at 2 years (26 ± 4% vs. 31 ± 5%, $p = 0.59$; 24 ± 5% vs. 30 ± 6%, $p = 0.52$), while NCR patients had 12% lower OS (28 ± 4% vs. 40 ± 4%, $p = 0.21$). Validation of external control cohorts is needed before incorporating RWD control data into comparative analyses, as missing data, specifically comorbidities, and residual confounding may limit comparability.

KEYWORDS
acute myeloid leukaemia, clinical trials, external control arm, external controls, non-randomized studies, real-world data

Francesco Pignatti and Jan J. Cornelissen: The views expressed in this article are the personal views of the author and may not be understood or quoted as being made on behalf of or reflecting the position of the agencies and organizations with which the author is employed or affiliated.

Presented in abstract form at the 65th Annual Meeting of the American Society of Hematology, 9–12 December 2023.

For affiliations refer to page 186.

INTRODUCTION

Real-world data (RWD) have earlier been used in post-marketing safety studies and in health technology assessments (HTA).[1,2] RWD might also be useful by providing supportive evidence to phase II single-arm trials in clinical settings where randomized controlled trials (RCTs) are not feasible.[3,4] One such setting is acute myeloid leukaemia (AML), in which novel drugs are currently being developed for tightly defined subgroups on the basis of specific molecular abnormalities or physical fitness. Regulatory agencies, such as the European Medicines Agency (EMA) and US Food and Drug Administration, recognize the potential of RWD for regulatory decision-making at the marketing authorization (MA) phase.[5–8] EMA's interest in RWD has also resulted in the recent development of the Data Analysis and Real World Interrogation Network (DARWIN EU) (https://www.darwin-eu.org/) to improve RWD sources. The DARWIN EU project aims to deliver over 140 studies yearly by 2025.[9]

While RWD cohorts have been useful for HTA and safety monitoring, their use as external control cohorts for efficacy assessments is still debated.[10,11] A recent systematic review by our group identified that non-randomized RWD comparisons in haemato-oncology often encountered methodological and statistical limitations, such as differences in patient selection, matching for a limited number of prognostic parameters, and missing data.[12] These limitations urge caution in the use of RWD controls in comparative studies and necessitate validation of RWD beforehand. One approach for validation may be to assess the comparability of RWD to a control arm in a concurrent prospective randomized study. In order to test that approach, we set out to evaluate the comparability of two distinct (RWD) control cohorts, one registered by the European HARMONY Alliance (HA) and another by the Netherlands Cancer Registry (NCR), with the control arm of the recently completed prospective randomized HOVON-103 (H103) trial in older AML patients.[13–15] It was hypothesized that either cohort might serve as a potential external control cohort for a clinical trial in older AML patients if outcome would be similar to the HOVON-103 control arm.

PATIENTS AND METHODS

Patient selection criteria

All H103-control patients received intensive induction chemotherapy and were compared with patients from the external cohorts (Supplementary Methods). In brief, external control patients (1) received intensive induction/consolidation chemotherapy, (2) had a WHO performance score (WHO-PS) ≤2 (or missing) at diagnosis and (3) had available survival data (Figure S1).

Harmonization of baseline and outcome parameters

To allow for comparative analyses between the H103 controls and the external patient cohorts, we harmonized data definitions for baseline and outcome parameters (Table S1). For example, registration of cytogenetics and molecular data was mandatory in the H103 protocol and these were available for all patients, provided that cytogenetic and mutational analyses could successfully be performed. While these data were available for HA patients, the NCR cohort included only the first three reported cytogenetic and molecular aberrations per patient. To derive the European LeukemiaNet (ELN) 2022 risk classification,[16] patients were classified as 'favourable' if a favourable aberration was present, and as 'intermediate' or 'adverse' when a patient had aberrations corresponding to these risk categories. Complete remission (CR) was based on local assessment but considered for this analysis if it occurred within 120 days after start of treatment. Because event-free survival (EFS) was not uniformly assessed in the external cohorts and failure to achieve CR as such was not explicitly registered in both databases, we applied a relapse-free survival (RFS) definition for patients in CR1 in the H103 and HA cohorts including relapse and death without relapse as events. RFS could not be estimated for the NCR cohort as relapse dates were not routinely monitored.

Statistical analysis

H103 controls were matched 1:1 using nearest neighbour propensity score calliper matching (PSM) with patients from the HA ($n = 315$) and NCR ($n = 320$). The applied calliper width was 0.2 standard deviation of the estimated propensity score.[17] Variables for PSM included age, ELN 2022 risk classification, white blood cell count at diagnosis and WHO-PS. Missing values of WHO-PS were considered as a distinct category. Patients with missing values for age, ELN 2022 risk classification and white blood cell count were excluded from the analysis (HA: $n = 19$; H103 controls: $n = 1$; NCR: $n = 1$). End-points included overall survival (OS) and RFS, which were both estimated by the Kaplan–Meier method and Cox proportional hazards regression. Median follow-up was estimated using the inverse Kaplan–Meier estimator. Sensitivity analyses were performed with different matching methods using coarsened exact matching,[18] exact matching on ELN 2022 risk classification and WHO-PS followed by calliper PSM on age and leucocyte count, and inverse probability of treatment weighting (IPTW). The distribution of propensity scores was inspected using histograms, whereas matching was more formally assessed using standardized mean differences and variance ratios. The strength of residual confounding was determined using E-values.[19,20] These

E-values quantify the minimum hazard ratio of an unobserved confounder to nullify the observed outcome effect estimate, which was calculated using the approximation for common outcomes.[19]

Data management and patient selection were performed using STATA (STATACorp 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC). All statistical analyses were done in R version 4.3.0 or higher (R Core Team, 2021). Propensity score matching was implemented using package *MatchIt* version 4.5.4 or higher.[21]

RESULTS

Patient characteristics

Before matching, H103 controls had a similar age at diagnosis as patients in the HA cohort (median age: 69 vs. 70 years, $p = 0.07$) and as NCR patients (69 vs. 69 years, $p = 0.46$, Table 1). WHO-PS was missing in the majority of HA and NCR cohorts compared with H103 controls (66% vs. 1%; 54% vs. 1%). As a result, a WHO-PS of 0 or 1 was observed in 49% and 41% of H103 controls, respectively, which was less frequently observed in HA patients (5% and 18%, respectively, $p < 0.01$). WHO-PS of 0 and 1 was also relatively infrequent for NCR patients compared with H103 controls (26% and 18% vs. 49% and 41%, respectively, $p < 0.01$). 41% of HA patients were previously included in a clinical trial. A favourable AML risk as determined by the ELN 2022 risk classification was more frequently found in H103 controls than in HA and NCR patients (24% vs. 17%, $p < 0.01$; 24% vs. 17%, $p < 0.01$). Lastly, median white blood cell count at diagnosis was significantly lower in H103 controls than in patients from the HA cohort (3.7 vs. 16.7, $p < 0.01$) or NCR patients (3.7 vs. 13.1, $p < 0.01$) respectively.

Treatment response

Response data were monitored after induction and/or consolidation treatment in the HOVON trial, but response data for the HA and NCR cohorts were not specifically registered in relation to treatment phase (Table S1). 69% of H103 controls obtained a CR within 120 days, which was observed in 58% of HA patients ($p < 0.01$) and in 50% of NCR patients ($p < 0.01$) respectively (Table 2). Allogeneic stem cell transplantation (alloSCT) was similarly applied across H103 and NCR cohorts, with 26% of H103 controls receiving alloSCT and 28% of the NCR patients ($p = 0.60$), while 11% of HA patients received an alloSCT ($p < 0.01$).

TABLE 1 Patient characteristics of H103 controls, HARMONY Alliance and NCR patients in the unmatched cohorts.

Parameter	H103 controls ($n = 279$)	HARMONY Alliance ($n = 315$)	NCR ($n = 320$)	*p*-value	
				Versus HARMONY Alliance	Versus NCR
Age at diagnosis, median (range)	69 (65–84)	70 (65–86)	69 (66–89)	0.07	0.46
Male sex, *n* (%)	172 (62)	166 (53)	197 (62)	0.03	1.00
Diagnosis year, median (range)	2014 (2010–2018)	2009 (2000–2018)	2016 (2014–2018)		
Trial participation, *n* (%)	279 (100)	128 (41)	0 (0)		
WHO performance score, *n* (%)					
0	136 (49)	15 (5)	84 (26)	<0.01	<0.01
1	114 (41)	57 (18)	58 (18)		
2	25 (9)	34 (11)	7 (2)		
Unknown	4 (1)	209 (66)	171 (54)		
ELN 2022 risk classification, *n* (%)					
Favourable	68 (24)	53 (17)	53 (17)	<0.01	<0.01
Intermediate	98 (35)	149 (47)	168 (52)		
Adverse	113 (41)	113 (36)	99 (31)		
White blood cell count (10^9/L), median (range)	3.7 (0–301)	16.7 (0.6–340)	13.1 (0.3–437)	<0.01	<0.01
Missing	1 (0.4)	19 (6.0)	1 (0.3)		
Blasts in peripheral blood (%), median (range)	10 (0–100)	38 (0–100)	30 (0–99)	<0.01	<0.01
Missing	4 (1)	159 (50)	27 (8)		
Blasts in bone marrow (%), median (range)	43 (1–100)	66 (2–100)	55 (1–100)	<0.01	<0.01
Missing	17 (6)	11 (3)	24 (8)		

Abbreviations: ELN, European LeukemiaNet; NCR, Netherlands Cancer Registry; WHO, World Health Organization.

TABLE 2 Outcome of H103 controls, HARMONY Alliance patients and NCR patients in the unmatched cohorts.

Parameter	H103 controls	HARMONY Alliance	NCR	*p*-value	
				Versus HARMONY Alliance	Versus NCR
Unmatched (279 vs. 315 vs. 320 patients)					
CR rate[a], *n* (%)	192 (69)	182 (58)	159 (50)	<0.01	<0.01
AlloSCT, *n* (%)	72 (26)	36 (11)	90 (28)	<0.01	0.60
2-year cumulative incidence of relapse (estimate ± SE)	49 ± 4	39 ± 3	—	0.03	—
2-year cumulative incidence of NRM (estimate ± SE)	16 ± 3	23 ± 3	—	0.02	—
2-year RFS (estimate ± SE)	35 ± 3	38 ± 4	—	0.64	—
2-year OS (estimate ± SE)	39 ± 3	33 ± 3	28 ± 3	0.20	0.01

Abbreviations: AlloSCT, allogeneic stem cell transplant; CR, complete remission; NCR, Netherlands Cancer Registry; NRM, non-relapse mortality; OS, overall survival; PSM, propensity score matching; RFS, relapse-free survival.

[a]CR rate was calculated as the proportion of patients that obtained a CR within 120 days after diagnosis, without accounting for censoring.

Survival outcomes

Median follow-up was 63 [95% CI: 61–67], 69 [95% CI: 61–72] and 48 [95% CI: 42–59] months for the H103 control, HA and NCR cohorts respectively. OS at 2 years after diagnosis (estimate ± standard error [SE]) estimated 39 ± 3% for H103 controls, which was non-significantly different from the HA patients, in whom 2-year OS estimated 33 ± 3% ($p = 0.20$) (Table 2). A significantly lower 2-year OS of 28 ± 3% was observed for patients in the NCR cohort, compared to H103 controls ($p = 0.01$). In addition, RFS was calculated for patients who obtained first complete remission. Two-year RFS estimated 35 ± 3% and 38 ± 4% ($p = 0.64$) for H103 controls and HA patients respectively. The 2-year cumulative incidence of relapse estimated 49 ± 4% for H103 controls and 39 ± 3% for HA patients ($p = 0.03$) respectively. The cumulative incidence of non-relapse mortality (NRM) at 2 years after diagnosis was estimated at 16 ± 3% and 23 ± 3% for H103 controls and HA patients respectively ($p = 0.02$). NRM and relapse data were not available for the NCR patients, which precluded the investigation of the survival differences between the NCR cohort and H103 controls.

Treatment response and survival outcomes in the matched cohorts

To account for the differences in cohort characteristics, 1:1 nearest neighbour calliper PSM was applied, which resulted in 103 HA patients being matched to 103 H103 controls, and 148 NCR patients matched to 148 H103 controls respectively (Figures S2 and S3, Tables S2–S4). CR rates were 53% and 62% in the matched HA and H103 control cohorts ($p = 0.26$) and 54% and 63% in the matched NCR and H103 control cohorts respectively (Table 3). OS at 2 years of H103 controls was 31 ± 5%, whereas the 2-year OS of HA patients was 26 ± 4% (HR = 0.93 [95% CI: 0.72–1.21], $p = 0.59$; Figure 1A, Table 3). NCR patients had a 2-year OS of 28 ± 4%, which was lower than H103 controls (40 ± 4%, HR = 0.84 [95% CI: 0.64–1.10], $p = 0.21$; Figure 1B, Table 3). 60-day and 100-day mortality were significantly different between HA patients and H103 controls (11% vs. 25%, respectively, and 14% vs. 31%, respectively), while these estimates were non-significantly different between NCR patients and H103 controls (14% vs. 18%, respectively, and 22% vs. 22%, respectively, Table 3). The cumulative incidences of relapse and NRM were non-significantly different between cohorts, which were estimated at 63 ± 6% vs. 53 ± 6% ($p = 0.21$) and 13 ± 4% vs. 17 ± 5% ($p = 0.40$) for matched HA patients and H103 controls respectively (Table 3). RFS was not observed to be significantly different between the HA cohort and H103 controls (24 ± 5% and 30 ± 6%, respectively; HR = 0.88 [95% CI: 0.60–1.29], $p = 0.52$; Figure 2, Table 3). Coarsened exact matching, exact matching followed by calliper PSM and IPTW were also applied to determine whether the matching algorithm impacted the observed results. These sensitivity analyses yielded consistent results, which indicated the robustness of the choice for the matching algorithm (Table S5).

Residual confounding

As the comparison of the H103-control population with the external cohorts was made in the absence of randomization, residual confounding, quantified by E-values, might influence the observed results in the matched cohorts. Regarding OS, we estimated E-values at 1.28 and 1.51 for comparing H103 controls versus HA patients and H103 controls versus NCR patients respectively. For RFS, the E-value estimated 1.41 comparing H103 controls and HA patients. E-values for the upper limits of the 95% CIs of OS and RFS were estimated to be 1 for all comparisons in accordance with Table 1 in VanderWeele et al.[19]

DISCUSSION

An increasing interest of regulatory agencies to apply RWD as (supportive) evidence during the MA process has become

TABLE 3 Outcome of H103 controls, HARMONY Alliance patients and NCR patients in the calliper matched cohorts.

Parameter	H103 controls	HARMONY Alliance	*p*-value
Matched 1:1 using calliper PSM (103 vs. 103 patients)			
CR rate[a], *n* (%)	64 (62)	55 (53)	0.26
60-day mortality[b], *n* (%)	26 (25)	10 (11)	0.005
100-day mortality[b], *n* (%)	32 (31)	14 (14)	0.004
2-year cumulative incidence of relapse (estimate ± SE)	53 ± 6	63 ± 6	0.21
2-year cumulative incidence of NRM (estimate ± SE)	17 ± 5	13 ± 4	0.40
2-year RFS (estimate ± SE)	30 ± 6	24 ± 5	0.52
2-year OS (estimate ± SE)	31 ± 5	26 ± 4	0.59
Parameter	**H103 controls**	**NCR**	***p*-value**
Matched 1:1 using calliper PSM (148 vs. 148 patients)			
CR rate[a], *n* (%)	93 (63)	80 (54)	0.16
60-day mortality[b], *n* (%)	27 (18)	20 (14)	0.34
100-day mortality[b], *n* (%)	33 (22)	33 (22)	1.0
2-year OS (estimate ± SE)	40 ± 4	28 ± 4	0.21

Abbreviations: AlloSCT, allogeneic stem cell transplant; CR, complete remission; NCR, Netherlands Cancer Registry; NRM, non-relapse mortality; OS, overall survival; PSM, propensity score matching; RFS, relapse-free survival.

[a]CR rate was calculated as the proportion of patients that obtained a CR within 120 days after diagnosis, without accounting for censoring.

[b]60-day and 100-day mortality were calculated as the proportion of patients that died within 60 or 100 days since the start of treatment, without accounting for censoring.

apparent, for example as supporting evidence to single-arm trials.[5,6] However, two recent studies by EMA identified that only a minority of MA applications and extensions of indications, submitted to EMA in 2018–2019, contributed efficacy or effectiveness data into regulatory decision-making during the pre-authorization phase.[7,8] Furthermore, previous studies incorporating RWD for haemato-oncology indications may have suffered from methodological and statistical limitations,[12] warranting precise validation of external cohorts. Here, the comparability of two external (RWD) cohorts, that is, the HA database and RWD from the Dutch national registry, with the control arm of the randomized phase II H103 trial for older AML patients was investigated. The HA and NCR control cohorts appeared statistically significantly different from H103 controls at baseline, and missing data were frequently observed in these cohorts. After matching, non-significantly different OS and RFS were observed for HA patients compared with H103 controls. In contrast, NCR patients demonstrated a 12% lower OS, but the lack of relapse monitoring and precise registration of cause of death in the NCR precluded a further investigation of these differences.

RWD have previously been applied as external control cohorts in haemato-oncology trials. Recently, we performed a systematic review of all phase I–III haemato-oncology trials published between 2000 and 2023 in which an RWD cohort was used to supplement or replace the control arm, and investigated to what extent these studies methodologically compared to RCTs.[12] For example, out of 32 identified studies, two studies incorporated prospectively monitored data, and only seven studies applied all trial inclusion criteria to the RWD cohort. Furthermore, the same end-point definitions were applied to the trial intervention arm and RWD cohort in only eight studies. None of the identified studies adhered fully to the methodological and statistical rigour that is currently applied in RCTs, which resulted in biased comparisons. In the current study in older AML patients, similar concerns were observed, such as missing data and differences in end-point definitions. It highlights that RWD quality needs to be carefully assessed before incorporating RWD into indirect efficacy assessments.

Randomization is the gold standard for creating similar patient groups by balancing both known and unknown confounders. Indirect efficacy estimates might suffer from confounding as randomization cannot be performed. It is important that non-randomized comparisons are balanced by other methods to obtain valid estimates. A careful quality assessment of the use of external control cohorts in RCTs may be such an approach.[22] Key items include the similarity of patient characteristics, outcome definitions, contemporaneity of the cohorts, and statistical adjustment.[22] We aimed to adhere to this quality assessment by applying the available H103-trial eligibility criteria to the external control cohorts, selecting similar diagnosis years and harmonizing outcome definitions. After matching for the most important prognostic parameters,[23–26] baseline characteristics of HA patients more closely resembled those of the H103 controls. Matched NCR patients were less similar, but matching could not be improved as WHO-PS were missing, and comorbidity data to further align with trial eligibility criteria were not available. It suggests that calliper PSM was not sufficient to guarantee a comparable external control cohort from the NCR data. Additionally, the strict nature of calliper PSM resulted in excluding >50% of H103 controls for both HA and NCR analyses, highlighting the problem of missing data in external control cohorts.[12]

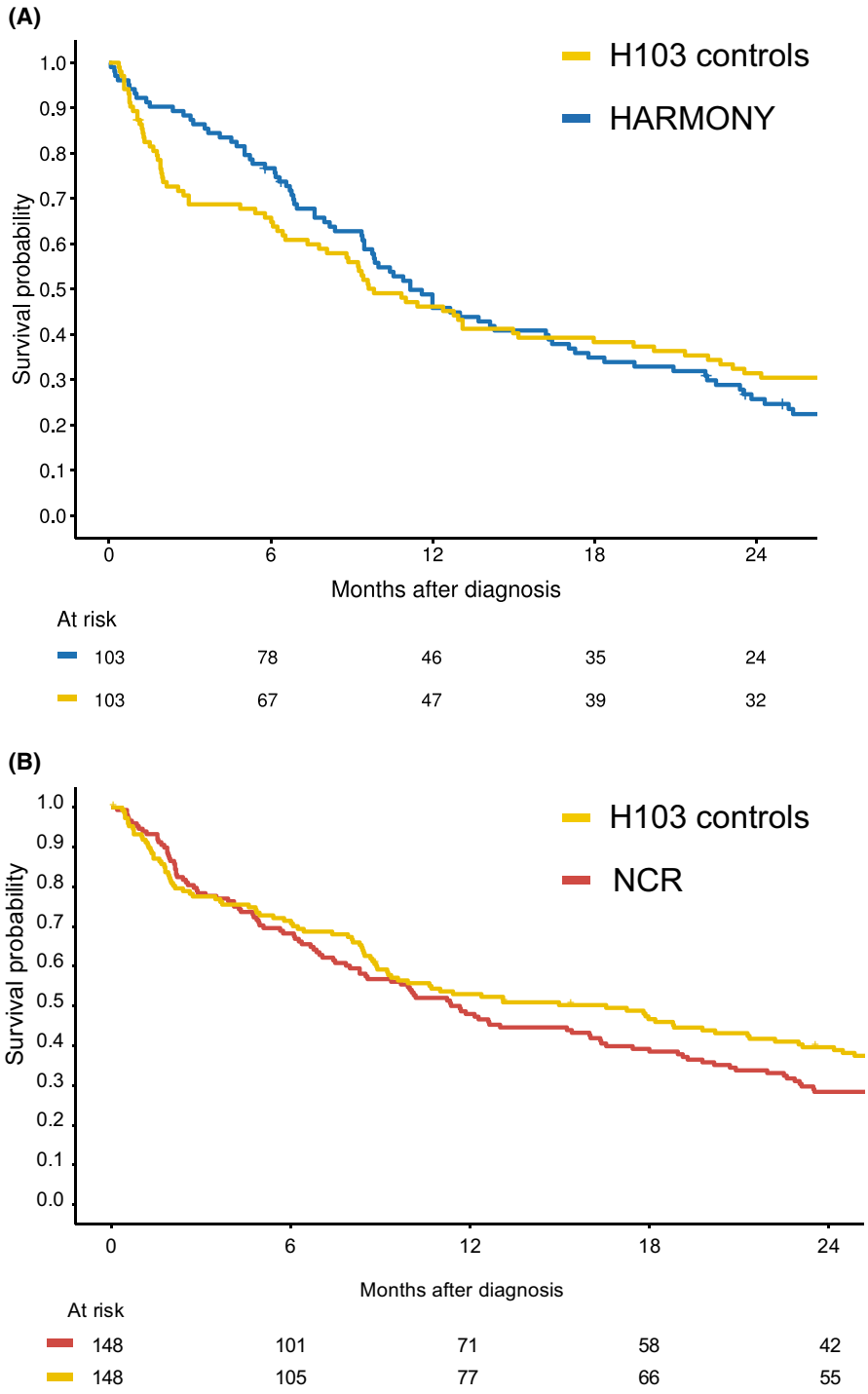

FIGURE 1 Overall survival for (A) H103 controls versus HARMONY Alliance patients, and (B) H103 controls versus NCR patients in the calliper matched cohorts. Cohorts were matched 1:1 using NN propensity score calliper matching for age, ELN 2022 risk classification, white blood cell count at diagnosis and WHO performance status. Patients with missing values for WHO performance score were included as a distinct category, while patients with missing values for age, ELN 2022 risk classification and white blood cell count were excluded from the analysis. The 2-year OS (estimate±SE) was 26 ± 4% versus 31 ± 5% for HARMONY Alliance patients matched with H103 controls ($p = 0.59$), and 28 ± 4% vs. 40 ± 4% for NCR patients matched with H103 controls ($p = 0.21$) respectively. ELN, European LeukemiaNet; NCR, Netherlands Cancer Registry; NN, nearest neighbour; OS, overall survival; SE, standard error.

External control cohorts have been previously investigated, both in clinical trials for solid tumours and in AML,[27–29] but encountered similar methodological issues as identified here. For example, Döhner et al. evaluated the addition of midostaurin to intensive induction chemotherapy in AML patients with *FLT3* mutations in a single-arm phase II trial. Midostaurin demonstrated an improved EFS and OS compared with five historical trials using IPTW, but the historical cohort lacked contemporaneity.[27] Similarly, control patients of the IMblaze370

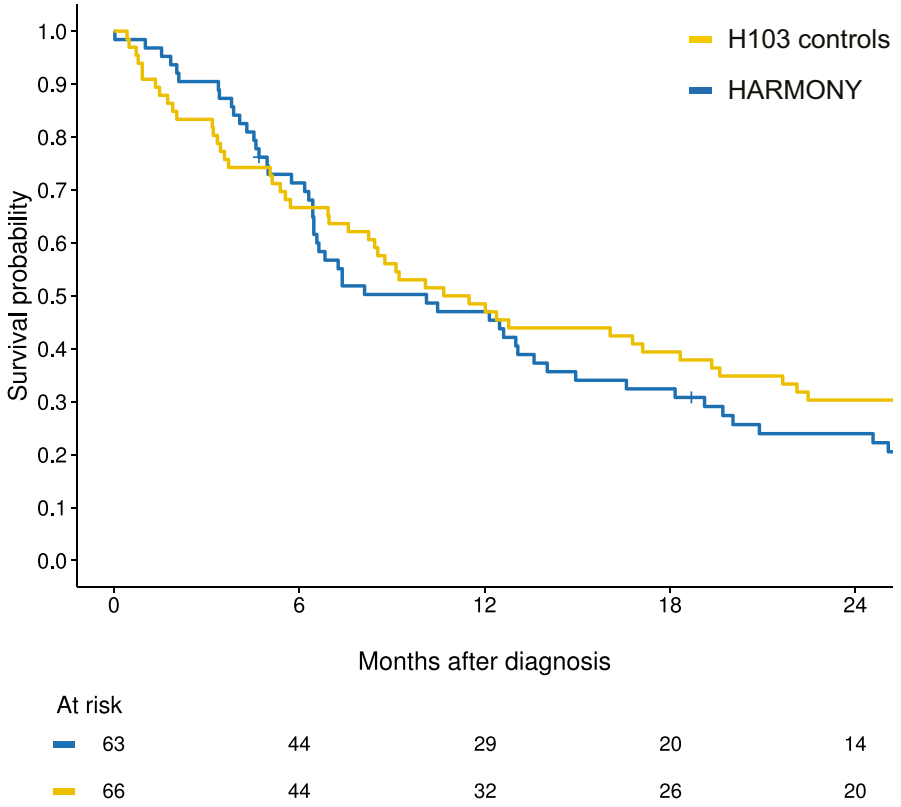

FIGURE 2 Relapse-free survival for H103 controls versus HARMONY Alliance patients in the calliper matched cohorts. H103 control and HARMONY Alliance cohorts were matched 1:1 using NN propensity score calliper matching for age, ELN2022 risk classification, white blood cell count at diagnosis and WHO performance status. Patients with missing values for the WHO performance score were included as a distinct category, while patients with missing values for age, ELN 2022 risk classification and white blood cell count were excluded from the analysis. The 2-year RFS (estimate±SE) was 24 ± 5% versus 30 ± 6% for HARMONY Alliance patients and matched H103 controls ($p = 0.52$) respectively. ELN, European LeukemiaNet; NCR, Netherlands Cancer Registry; NN, nearest neighbour; OS, overall survival; SE, standard error.

trial for metastatic colorectal cancer were compared with matched patients from the Flatiron Health database and Foundation Medicine Clinico-Genomic Database (CGDB).[28] The external cohorts demonstrated similar baseline characteristics and OS as the trial control arm after applying trial eligibility criteria, including performance status and organ function, and propensity score weighting. The observed HR for OS (trial intervention arm vs. external control cohorts) was compared to the HR originally reported. Notably, there were no missing data for ECOG performance score. The latter analysis indicated that external control cohorts might be feasible if Real-World databases also contain detailed information on comorbidities and performance status.

Monitoring of relapse, second-line treatment, and follow-up were performed for the H103 controls according to protocol. While response data were recorded for every treatment cycle in the HOVON trial, response data were not specifically registered for each treatment phase in the HA and NCR cohorts. Although EFS was included in the HA cohort, its definition was not standardized between data providers. Monitoring of relapse was altogether not performed in the NCR. This precluded incorporating response and relapse into the analysis of real-world NCR patients. Evaluation of short-term treatment response, EFS and relapse are important secondary or co-primary end-points in clinical trials, which are important to assess in external comparator cohorts.[30]

Regulatory agencies acknowledge the role that RWD may serve when RCTs are not feasible, such as diseases with a high unmet medical need.[4,5] However, informing regulatory decision-making by RWD remains challenging due to observed methodological and statistical issues. To improve RWD sources, EMA has recently initiated DARWIN EU. Within DARWIN EU, patient-level data of approximately 130 million patients from 20 (national) RWD sources are linked for comparative analyses. So far, DARWIN EU studies have focused on drug utilization patterns and disease history, but pilot projects have started to leverage clinical trial data as well. Similarly, others, such as GetReal Institute or ISPOR, also aim to improve the adoption of RWD. Given the data limitations observed, it is proposed that RWD sources to be used in haemato-oncological comparative studies should follow as much as possible the rigour applied in RCTs when the aim is to expand or substitute trial control arms. This may improve the validity of RWD comparisons, contributing to the potential of large-scale RWD initiatives such as DARWIN EU.

External control cohorts are becoming increasingly important in the era of novel treatments developed in AML. Although the '7 + 3' regimen has been the cornerstone of

intensive treatment in older AML patients for decades,[13,14,31,32] it is also associated with substantial toxicity which established the need for less intensive therapeutic options specifically in older patients or patients with comorbidities. Recently, a randomized phase III trial in older, newly diagnosed AML patients fit for intensive treatment showed similar OS for '7+3' compared with the hypomethylating agent (HMA) decitabine, whereas fewer toxicities were observed in patients receiving decitabine.[33] Outcomes might be even further improved by adding drugs to HMA, such as venetoclax or other targeted treatments.[34–37] In parallel, efforts are being made to improve survival with intensive '7+3' regimens, by adding drugs that target specific mutations, such as *FLT3*,[38–40] and *IDH1/2*,[41–43] or the regulatory protein menin.[23,44,45] These novel targeted drugs require sufficiently sized phase III trials but are applied in smaller subsets of AML patients, which may significantly extend study enrolment times. These trials might be accelerated using external control data, which need to reflect the current treatment algorithm.

In conclusion, we assessed the comparability of the HA and NCR cohorts as external control cohorts to the control arm of the randomized phase II H103-trial. It was observed that important baseline and outcome parameters had different definitions, or were frequently missing, such as comorbidities in the NCR cohort. It limited the extent to which the H103-trial eligibility criteria and end-points could be imposed on or be estimated in the external cohorts. After matching, OS, while non-significantly different for both comparisons, was more comparable between H103 controls and HA patients than for H103 controls and NCR patients. While HA might potentially serve as an external control cohort for intensively treated older AML patients, the absence of baseline and certain outcome parameters also questions the applicability of that external (RWD) control cohort for comparison with an experimental single-arm study group receiving a new intervention.

AUTHOR CONTRIBUTIONS

Contribution: S.J.F.H. and Y.v.N. designed the study and analysed the data. S.J.F.H. wrote the manuscript; E.D.v.W. and Y.v.N. provided statistical oversight and guidance. J.V. and J.J.C. supervised the study. All other authors provided clinical data and/or carefully reviewed and approved the final version of this manuscript.

AFFILIATIONS

[1]Erasmus University Medical Center Cancer Institute, Rotterdam, The Netherlands
[2]HOVON Foundation, Rotterdam, The Netherlands
[3]Department of Haematology, University Medical Centre Nijmegen, Nijmegen, The Netherlands
[4]Department of Hematology, University Medical Center Groningen (UMCG), Groningen, The Netherlands
[5]Department of Medical Oncology, Inselspital, Bern University Hospital, Bern, Switzerland
[6]Ziekenhuis Netwerk Antwerpen, Antwerp, Belgium
[7]Department of Research and Development, Netherlands Comprehensive Cancer Organization (IKNL), Utrecht, The Netherlands
[8]Department of Hematology, Oncology, Tumor Immunology, Charité University Medicine, Berlin, Germany
[9]Servicio de Hematología, Hospital Universitario de Salamanca, Universidad de Salamanca, Salamanca, Spain
[10]Department of Hematology, Blood Neoplasms, and Bone Marrow Transplantation, Wroclaw Medical University, Wroclaw, Poland
[11]Hematology Department, Hospital 12 de Octubre (i+12), Centro Nacional de Investigaciones Oncológicas (CNIO), Complutense University, Madrid, Spain
[12]Department of Hematology, Cellular Therapy, Hemostaseology and Infectious Diseases, University Medical Center Leipzig, Leipzig, Germany
[13]MLL Munich Leukemia Laboratory, Munich, Germany
[14]Department of Medicine I, University Hospital Carl Gustav Carus, Dresden, Germany
[15]Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, The Netherlands
[16]Department of Hematology, Oncology, and Cancer Immunology, Charité-Universitätsmedizin Berlin, Berlin, Germany
[17]Department of Hematology, Cancer Center Amsterdam, Amsterdam University Medical Centers, Vrije Universiteit University Medical Center, Amsterdam, The Netherlands
[18]Oncology and Hematology Office, European Medicines Agency, Amsterdam, The Netherlands

ACKNOWLEDGEMENTS

The authors thank the HARMONY Healthcare Alliance Consortium, and the registration team of the Netherlands Comprehensive Cancer Organisation (IKNL) for the collection of data for the Netherlands Cancer Registry.

CONFLICT OF INTEREST STATEMENT

All authors have no competing financial interests to declare related to this work.

DATA AVAILABILITY STATEMENT

The HOVON-103 data that support the findings of this study are available on request from the corresponding author. The HARMONY Alliance and the Netherlands Comprehensive Cancer Organization maintain the HARMONY Alliance and Netherlands Cancer Registry data, respectively, and require separate data requests. All datasets are not publicly available due to privacy or ethical restrictions. The HOVON-103 trial is registered in the Netherlands Trial Register (https://onderzoekmetmensen.nl) as #NTR2294, #NTR5902 and #NTR2477.

ORCID

Sjoerd J. F. Hermans https://orcid.org/0000-0002-3906-4584
Thomas Pabst https://orcid.org/0000-0002-6055-5257
Avinash G. Dinmohamed https://orcid.org/0000-0002-4767-6716
Marta Sobas https://orcid.org/0000-0003-0781-8668
Rosa Ayala Diaz https://orcid.org/0000-0002-2699-8353
Joaquin Martinez Lopez https://orcid.org/0000-0001-5904-0902
Klaus H. Metzeler https://orcid.org/0000-0003-3920-7490
Christian Thiede https://orcid.org/0000-0003-1241-2048

REFERENCES

1. Sola-Morales O, Curtis LH, Heidt J, Walsh L, Casso D, Oliveria S, et al. Effectively leveraging RWD for external controls: a systematic literature review of regulatory and HTA decisions. Clin Pharmacol Ther. 2023;114(2):325–55.
2. Brown JP, Wing K, Evans SJ, Bhaskaran K, Smeeth L, Douglas IJ. Use of real-world evidence in postmarketing medicines regulation in the

European Union: a systematic assessment of European medicines agency referrals 2013-2017. BMJ Open. 2019;9(10):e028133.
3. Cave A, Kurz X, Arlett P. Real-world data for regulatory decision making: challenges and possible solutions for Europe. Clin Pharmacol Ther. 2019;106(1):36–9.
4. Jansen MS, Dekkers OM, le Cessie S, Hooft L, Gardarsdottir H, de Boer A, et al. Real-world evidence to inform regulatory decision making: a scoping review. Clin Pharmacol Ther. 2024;115(6):1269–76.
5. Eichler H-G, Pignatti F, Schwarzer-Daum B, Hidalgo-Simon A, Eichler I, Arlett P, et al. Randomized controlled trials versus real world evidence: neither magic nor myth. Clin Pharmacol Ther. 2021;109(5):1212–8.
6. Prilla S, Groeneveld S, Pacurariu A, Restrepo-Méndez MC, Verpillat P, Torre C, et al. Real-world evidence to support EU regulatory decision making-results from a pilot of regulatory use cases. Clin Pharmacol Ther. 2024;116:1188–97.
7. Flynn R, Plueschke K, Quinten C, Strassmann V, Duijnhoven RG, Gordillo-Marañon M, et al. Marketing authorization applications made to the European medicines agency in 2018–2019: what was the contribution of real-world evidence? Clin Pharmacol Ther. 2022;111(1):90–7.
8. Bakker E, Plueschke K, Jonker CJ, Kurz X, Starokozhko V, Mol PGM. Contribution of real-world evidence in European medicines Agency's regulatory decision making. Clin Pharmacol Ther. 2023;113(1):135–51.
9. Arlett P, Kjær J, Broich K, Cooke E. Real-world evidence in EU medicines regulation: enabling use and establishing value. Clin Pharmacol Ther. 2022;111(1):21–3.
10. Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the use of nonrandomized real-world data analyses for regulatory decision making. Clin Pharmacol Ther. 2019;105(4):867–77.
11. Skovlund E, Leufkens HGM, Smyth JF. The use of real-world data in cancer drug development. Eur J Cancer. 2018;101:69–76.
12. Hermans SJF, van der Maas NG, van Norden Y, Dinmohamed AG, Berkx E, Huijgens PC, et al. Externally controlled studies using real-world data in patients with hematological cancers: a systematic review. JAMA Oncol. 2024;10:1426–36.
13. Ossenkoppele GJ, Breems DA, Stuessi G, van Norden Y, Bargetzi M, Biemond BJ, et al. Lenalidomide added to standard intensive treatment for older patients with AML and high-risk MDS. Leukemia. 2020;34(7):1751–9.
14. Janssen JJWM, Löwenberg B, Manz M, Biemond BJ, Westerweel PE, Klein SK, et al. Addition of the nuclear export inhibitor selinexor to standard intensive treatment for elderly patients with acute myeloid leukemia and high risk myelodysplastic syndrome. Leukemia. 2022;36(9):2189–95.
15. Janssen J, Löwenberg B, Manz M, Bargetzi M, Biemond B, von dem Borne P, et al. Inferior outcome of addition of the aminopeptidase inhibitor tosedostat to standard intensive treatment for elderly patients with AML and high risk MDS. Cancer. 2021;13(4):672.
16. Döhner H, Wei AH, Appelbaum FR, Craddock C, DiNardo CD, Dombret H, et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. Blood. 2022;140(12):1345–77.
17. Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. Pharm Stat. 2011;10(2):150–61.
18. Iacus SM, King G, Porro G. Causal inference without balance checking: coarsened exact matching. Polit Anal. 2012;20(1):1–24.
19. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med. 2017;167(4):268–74.
20. Thorlund K, Duffield S, Popat S, Ramagopalan S, Gupta A, Hsu G, et al. Quantitative bias analysis for external control arms using real-world data in clinical trials: a primer for clinical researchers. J Comp Eff Res. 2024;13(3):e230147.
21. Ho DE, Imai K, King G, Stuart EA. MatchIt: nonparametric preprocessing for parametric causal inference. J Stat Softw. 2011;42(8):1–28.
22. Thorlund K, Dron L, Park JJH, Mills EJ. Synthetic and external controls in clinical trials—A primer for researchers. CLEP. 2020;12:457–67.
23. Kantarjian HM, Short NJ, Fathi AT, Marcucci G, Ravandi F, Tallman M, et al. Acute myeloid leukemia: historical perspective and progress in research and therapy over 5 decades. Clin Lymphoma Myeloma Leuk. 2021;21(9):580–97.
24. de Leeuw DC, Ossenkoppele GJ, Janssen JJWM. Older patients with acute myeloid leukemia deserve individualized treatment. Curr Oncol Rep. 2022;24(11):1387–400.
25. Laribi K, Sobh M, Ghez D, Baugier de Materre A. Impact of age, functional status, and comorbidities on quality of life and outcomes in elderly patients with AML: review. Ann Hematol. 2021;100(6):1359–76.
26. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41–55.
27. Döhner H, Weber D, Krzykalla J, Fiedler W, Wulf G, Salih H, et al. Midostaurin plus intensive chemotherapy for younger and older patients with AML and FLT3 internal tandem duplications. Blood Adv. 2022;6(18):5345–55.
28. Schröder C, Lawrance M, Li C, Lenain C, Mhatre SK, Fakih M, et al. Building external control arms from patient-level electronic health record data to replicate the randomized IMblaze370 control arm in metastatic colorectal cancer. JCO Clin Cancer Inform. 2021;5:450–8.
29. Carrigan G, Whipple S, Capra WB, Taylor MD, Brown JS, Lu M, et al. Using electronic health records to derive control arms for early phase single-arm lung cancer trials: proof-of-concept in randomized controlled trials. Clin Pharmacol Ther. 2020;107(2):369–77.
30. Griffith SD, Tucker M, Bowser B, Calkins G, Chang CH(J), Guardino E, et al. Generating real-world tumor burden endpoints from electronic health record data: comparison of RECIST, radiology-anchored, and clinician-anchored approaches for abstracting real-world progression in non-small cell lung cancer. Adv Ther. 2019;36(8):2122–36.
31. Löwenberg B, Beck J, Graux C, van Putten W, Schouten HC, Verdonck LF, et al. Gemtuzumab ozogamicin as postremission treatment in AML at 60 years of age or more: results of a multicenter phase 3 study. Blood. 2010;115(13):2586–91.
32. Löwenberg B, Zittoun R, Kerkhofs H, Jehn U, Abels J, Debusscher L, et al. On the value of intensive remission-induction chemotherapy in elderly patients of 65+ years with acute myeloid leukemia: a randomized phase III study of the European Organization for Research and Treatment of cancer leukemia group. J Clin Oncol. 1989;7(9):1268–74.
33. Lübbert M, Wijermans PW, Kicinski M, Chantepie S, Van der Velden WJFM, Noppeney R, et al. 10-day decitabine versus 3 + 7 chemotherapy followed by allografting in older patients with acute myeloid leukaemia: an open-label, randomised, controlled, phase 3 trial. Lancet Haematol. 2023;10(11):e879–e889.
34. DiNardo CD, Jonas BA, Pullarkat V, Thirman MJ, Garcia JS, Wei AH, et al. Azacitidine and Venetoclax in previously untreated acute myeloid leukemia. N Engl J Med. 2020;383(7):617–29.
35. Yilmaz M, Kantarjian H, Short NJ, Reville P, Konopleva M, Kadia T, et al. Hypomethylating agent and venetoclax with FLT3 inhibitor "triplet" therapy in older/unfit patients with FLT3 mutated AML. Blood Cancer J. 2022;12(5):1–9.
36. Maiti A, DiNardo CD, Daver NG, Rausch CR, Ravandi F, Kadia TM, et al. Triplet therapy with venetoclax, FLT3 inhibitor and decitabine for FLT3-mutated acute myeloid leukemia. Blood Cancer J. 2021;11(2):1–6.
37. Pratz KW, Jonas BA, Pullarkat V, Thirman MJ, Garcia JS, Döhner H, et al. Long-term follow-up of VIALE-A: venetoclax and azacitidine in chemotherapy-ineligible untreated acute myeloid leukemia. Am J Hematol. 2024;99:615–24.
38. Stone RM, Mandrekar SJ, Sanford BL, Laumann K, Geyer S, Bloomfield CD, et al. Midostaurin plus chemotherapy for acute myeloid leukemia with a FLT3 mutation. N Engl J Med. 2017;377(5):454–64.
39. Quizartinib plus chemotherapy in newly diagnosed patients with FLT3-internal-tandem-duplication-positive acute myeloid leukaemia (QuANTUM-first): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet. 2023;401(10388):1571–83.
40. Perl AE, Martinelli G, Cortes JE, Neubauer A, Berman E, Paolini S, et al. Gilteritinib or chemotherapy for relapsed or refractory -mutated AML. N Engl J Med. 2019;381(18):1728–40.

41. Montesinos P, Recher C, Vives S, Zarzycka E, Wang J, Bertani G, et al. Ivosidenib and Azacitidine in -mutated acute myeloid leukemia. N Engl J Med. 2022;386(16):1519–31.
42. Pollyea DA, Tallman MS, de Botton S, Kantarjian HM, Collins R, Stein AS, et al. Enasidenib, an inhibitor of mutant IDH2 proteins, induces durable remissions in older patients with newly diagnosed acute myeloid leukemia. Leukemia. 2019;33(11):2575–84.
43. Stein EM, DiNardo CD, Fathi AT, Mims AS, Pratz KW, Savona MR, et al. Ivosidenib or enasidenib combined with intensive chemotherapy in patients with newly diagnosed AML: a phase 1 study. Blood. 2021;137(13):1792–803.
44. DiNardo CD, Erba HP, Freeman SD, Wei AH. Acute myeloid leukaemia. Lancet. 2023;401(10393):2073–86.
45. Issa GC, Aldoss I, DiPersio J, Cuglievan B, Stone R, Arellano M, et al. The menin inhibitor revumenib in KMT2A-rearranged or NPM1-mutant leukaemia. Nature. 2023;615(7954):920–4.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hermans SJF, Versluis J, van Werkhoven ED, van Norden Y, Janssen JJWM, Huls GA, et al. Comparability of external and internal control patients for the prospective randomized HOVON-103 trial in older AML patients. Br J Haematol. 2026;208(1):179–188. https://doi.org/10.1111/bjh.20185