

Autonomous artificial intelligence prescribing a drug to prevent severe acute graft-versus-host disease in HLA-haploidentical transplants

Received: 18 March 2025

Accepted: 1 August 2025

Published online: 25 September 2025

 Check for updates

A list of authors and their affiliations appears at the end of the paper

Autonomous artificial intelligence (AI) models for deciding treatment strategies are available but rarely applied prospectively in clinical settings. Here we present a prospective study of deploying daGOAT, an algorithm we have developed, as a conditional autonomous AI agent to prescribe a drug to prevent severe (grade 3–4) acute graft-versus-host disease (acute GvHD) following human leukocyte antigen (HLA)-mismatched haematopoietic cell transplantation (ClinicalTrials.gov, NCT05600855). During the enrollment period physicians invite 85% of eligible patients to participate and 88% of the invited patients agree. Among the 110 enrolled participants who receive HLA-haploidentical transplants, daGOAT predicts intermediate to high risk of severe acute GvHD in 57 participants between days +17 and +23 posttransplant and prescribes ruxolitinib in addition to the existing regimen to intensify immune suppression. The initial compliance with AI prescription is 98% (56/57), with dose and/or schedule deviating from the AI prescription within one month in a total of eight participants. In conclusion, we show that many physicians and patients are receptive to using conditional autonomous AI to prescribe a drug and that the decision for pharmaceutical intervention could be facilitated by autonomous AI.

Artificial intelligence (AI) can be classified into ‘assistive’ or ‘autonomous’¹. With assistive AI, the physician is prompted with AI-generated insights after which he/she decides whether and how to use the prompt(s). With autonomous AI, the AI agent suggests a course of action which is the default action plan unless vetoed by the physician (‘conditional automation’). When patient care is at stake, acceptance of an AI algorithm for treatment decision-making in prospective, real-world clinical settings may deviate from the estimated level of acceptance based on retrospective or simulated evaluation². To date, prospective trials of autonomous AI are often limited to image-based diagnosis, radiation therapy planning, behavioral nudges, or a chatbot for post-procedural follow-up^{2–12}. Previously, AI models for deciding treatment strategies have been constructed, but they are rarely

applied prospectively in clinical settings^{13–15}. We hypothesize autonomous AI may be useful in pharmaceutical intervention when the physician believes that there is otherwise no reliable means to determine whether to initiate medication, that the consequence of inaction can be dangerous, and that the intervention may be harmful or costly.

There are many unanswered questions regarding use of autonomous AI to prescribe drugs. For instance, how do we embed an autonomous AI algorithm into the patient care pathway? Does the autonomous AI agent make decisions on which drug, which dose, and/or which schedule? Will physicians and patients agree to participate in a trial wherein an autonomous AI agent makes drug prescription decisions? In addition, we anticipate automatic ingestion of multimodal co-variate data directly into the model would be necessary for

✉ e-mail: chenjunren@ihcams.ac.cn; caoyigeng@ihcams.ac.cn; jiangerlie@ihcams.ac.cn

acceptance and successful implementation of autonomous AI. Whether this is broadly feasible is challenged by the complex organization of hospital information systems¹⁶.

In this work, to test the feasibility of using autonomous AI to prescribe a drug we choose a setting after a haematopoietic cell transplant (HCT) where the aim is to prevent severe (grade 3–4) acute graft-versus-host disease (acute GvHD). Accurately predicting risk of developing severe acute GvHD is challenging with no widely-accepted model. Survival of patients with severe acute GvHD could be <50% within one year¹⁷. The question is how to identify intermediate- and high-risk patients and intervene.

We conduct a proof-of-concept, phase-2 study of using ‘daGOAT’, an algorithm we have developed¹⁸, to monitor dynamic changes in 141 common clinical co-variates and prescribe a drug when the model identifies a participant to be at risk of developing severe acute GvHD. Our data show that physicians agree to use daGOAT in 85% of eligible patients and, in turn, 88% of these potential participants agree to use

the model. Compliance with AI prescription is 98% initially, with few deviations from the AI-prescribed dose and/or schedule within one month. To sum up, many physicians and patients are receptive to conditional autonomous AI prescription.

Results

Model update

Retrospectively we used data of 723 human leukocyte antigen (HLA)-mismatched transplants from the NICHE-GOAT cohort at the Institute of Hematology, Chinese Academy of Medical Sciences (IHCAMS) to optimize the daGOAT model¹⁹.

After optimization, daGOAT (Fig. 1a) includes 14 peri-transplant features (recipient age, sex, body mass index, primary disease, donor type [HLA-haploidentical versus HLA-mismatched unrelated donor [MMUD]], sex match, ABO match, HLA match, graft source, pre-transplant conditioning regimen, peri-transplant use of anti-thymocyte globulin [ATG], baseline regimen for acute GvHD

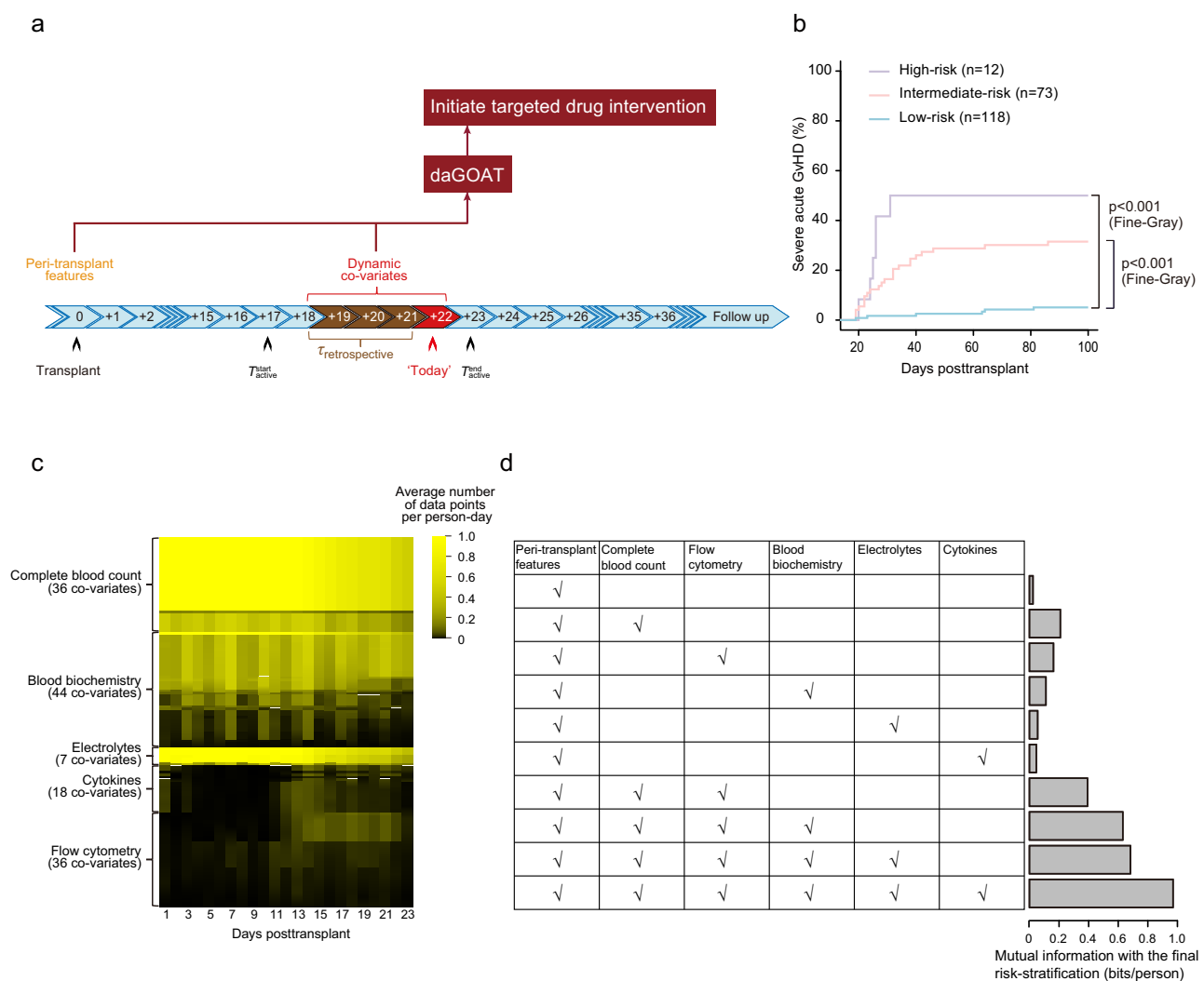


Fig. 1 | Model update using retrospective data. **a** The daGOAT model evaluates 14 peri-transplant features and 141 dynamic co-variates with three hyperparameters including $T_{\text{retrospective}} = 3$ days (how far back the model looks at past values of dynamic co-variates when calculating risk score for severe acute graft-versus-host disease [acute GvHD]), $T_{\text{start active}} = +17$ days posttransplant (the day the model starts reporting severe acute GvHD risk), and $T_{\text{end active}} = +23$ days posttransplant (the last day the model reports severe acute GvHD risk). For instance, if ‘today’ were day +22 posttransplant, daGOAT would integrate peri-transplant features and dynamic co-variates from day +19 to day +22 to compute the risk of developing severe acute

GvHD. **b** Performance of daGOAT-computed risk-stratification on the validation set (1 December 2020–31 December 2021; $n = 203$) whilst the model was fitted using only data from the training set (1 April 2012–30 November 2020; $n = 519$). Purple, high-risk ($n = 12$); pink, intermediate-risk ($n = 73$); blue, low-risk ($n = 118$). Pairwise comparisons are performed using the two-sided Fine-Gray test. p -values are not adjusted for multiple-hypothesis testing. **c** Data densities of dynamic co-variates in the retrospective cohort. **d** Information contributed by peri-transplant features and separate categories and various combinations of categories of dynamic clinical co-variates. Source data for (b–d) are available online.

prophylaxis, CD34-positive cell dose, and total nucleated cell dose) and 141 dynamic co-variables (complete blood count [CBC; 36 co-variables], blood biochemistry [44 co-variables], electrolytes [7 co-variables], blood cell flow cytometry [36 co-variables], and cytokines [18 co-variables]; Supplementary Data 1). In retrospective cross-validation experiments, AI-designated ‘high-risk’ had a subdistribution hazard ratio (sHR) = 14.3 (95% confidence interval [CI], 4.5–45.1) for severe acute GvHD compared with ‘low-risk’ in the validation set whilst for ‘intermediate-risk’ sHR = 7.3 (95% CI, 3.0–17.6) (Fig. 1b).

Mutual information between co-variables and risk-stratification

Despite the high number of dynamic co-variables included in daGOAT, low data density was normal for most of the co-variables in the retrospective dataset (Fig. 1c; Supplementary Data 1). When only one solitary category of dynamic co-variables was chosen to be included in the model (that is, rendering the model into a highly truncated version), choosing CBC, blood cell flow cytometry, or blood biochemistry data led to higher mutual information between the risk-stratification computed by the truncated model and the risk-stratification computed by the complete model (Fig. 1d).

Conditional autonomous AI-driven prescription

We deployed daGOAT as a conditional autonomous AI agent in the hospital intranet and connected to the hospital information system (Fig. 2a). From day +17 to day +23, each day at 1720 h the model would autonomously extract participant data from the hospital information system and classify each participant as low-, intermediate-, or high-risk to develop severe acute GvHD (Fig. 2b). If a participant were classified as high-risk, his/her risk-stratification would stay at high-risk regardless of subsequent events. If a participant were classified as intermediate-risk, his/her risk level would stay at intermediate-risk unless it were subsequently revised to high-risk. Any adjustment of risk-strata would be communicated to the clinical study coordinator and attending physician at 1721 h using a dashboard on the desktop computer daGOAT information system and alert messages via mobile-phone short message service (SMS; Fig. 3).

A model-designated high-risk participant was to be given oral ruxolitinib, 5 mg twice daily, until \geq day +60 posttransplant. A model-designated intermediate-risk participant was to be prescribed oral 2.5 mg ruxolitinib twice daily until \geq day +60 posttransplant; if his/her risk-stratification were later up-staged to high-risk the ruxolitinib dose would be increased to 5 mg twice daily. Physicians were directed to attempt discontinuing ruxolitinib by day +100 posttransplant. The protocol stipulated the physician would regain full control of all intervention decisions when severe acute GvHD occurred (Fig. 2b).

Participant enrollment

During the enrollment period 152 patients receiving HLA-mismatched transplants of granulocyte-colony stimulating factor (G-CSF)-mobilized blood-derived haematopoietic cells were eligible for participation. The physicians excluded 23 (15%) patients. The most common reason given was ‘slow haematopoietic recovery’ ($n = 10$), followed by ‘not in histologic complete remission (CR) or having detectable residual disease’ ($n = 6$ [acute leukaemia, $n = 4$; myelodysplastic syndromes, $n = 2$]), ‘unable to swallow medicine’ ($n = 5$), ‘human immunodeficiency virus (HIV) infection’ ($n = 1$), and ‘no given reason’ ($n = 1$). Of the remaining 129 eligible patients, 15 (12%) declined to participate in the trial. In summary, 114 (75%) of the 152 eligible patients participated in the trial whereas 38 did not participate. All the participants received myelo-ablative conditioning and baseline acute GvHD prophylaxis using ATG 2.5 mg/(kg·d) during days –4 to –1.

Because the predominant majority (96% [110/114]) of the transplants included in the trial were HLA-haploidentical whilst a small minority (4% [4/114]) were HLA-mismatched unrelated donor (MMUD) transplants, we focused on the 110 HLA-haploidentical transplants –

hereafter referred to as the ‘AI focus group’ – in our analyses (Fig. 4a; Table 1).

Dynamic co-variate data in the participants

Densities of autonomously-extracted dynamic co-variate data between days +14 and +23 posttransplant in the AI focus group (Fig. 4b) were largely comparable to data densities during the same posttransplant time interval in the retrospective data used to train the model (Fig. 1c). In the prospective trial, zeros in laboratory test results were occasionally recorded as ‘-’ rather than ‘0’ in the hospital information system and as a consequence daGOAT failed to extract all the zeros. daGOAT did not halt or give error messages because of these missing values; rather, it continued updating risk-stratification based on the data that were successfully extracted.

Risk-stratification of the participants

Median time of granulocyte and platelet recovery was +12 and +14 days posttransplant, respectively (Supplementary Fig. 2). Cumulative incidences of granulocyte and platelet recovery were 100% (110/110) and 98% (108/110) by day +100. No participant developed severe acute GvHD before (including) day +17. According to daGOAT, 53 (48%), 39 (35%), and 18 (16%) participants in the AI focus group were at low, intermediate, and high risk, respectively, of developing severe acute GvHD (Fig. 4c). In 56 (98%) of the 57 intermediate-to-high-risk participants, daily daGOAT-computed risk scores were steady or escalated progressively from day +17 to day +23. In one participant (#109), risk score fluctuated between ‘high’ and ‘low’ between days +17 and +23; by protocol his/her risk-stratification was designated by daGOAT as ‘high-risk’.

Compliance with autonomous AI drug prescriptions

None of the low-risk participants in the AI focus group took ruxolitinib except for two persons (Participants #52 and #53) who started ruxolitinib after they developed severe acute GvHD (Fig. 4d). Fifty-six (98% of 57) intermediate- to high-risk participants immediately started ruxolitinib when prescribed by daGOAT (Fig. 4d). In one intermediate-risk participant (#55), the physician started giving ruxolitinib one day after daGOAT started prescribing. In seven additional participants the physicians deviated from the AI-prescribed dose or schedule within one month after AI prescription started: In two intermediate-risk participants (#56 and #90), physicians increased their dose to ≥ 10 mg/d after grade-2 acute GvHD was diagnosed. In three high-risk participants, physicians prescribed ≤ 5 mg/d because of concerns about positive measurable residual disease (MRD)-status pretransplant (Participants #109 and #110) or pancytopenia that started pretransplant (Participant #103). In one intermediate-risk participant (#87), dose was decreased to 2.5 mg/d on day +31 because of *Stenotrophomonas* bacteraemia. In another intermediate-risk participant (#92), ruxolitinib was discontinued early because his/her MRD-status became positive on day +33.

In the AI focus group there were 52 participants with acute leukaemia and negative MRD-status pretransplant. Two of them became MRD-positive before day +100, including one low-risk and one intermediate-risk participants (#47 and #92, respectively). In contrast, in the co-variate-matched controls, three of the 121 acute leukaemia patients who were MRD-negative pretransplant became MRD-positive before day +100. The probability of ‘worsening’ or conversion from MRD-negative to MRD-positive before day +100 was comparable between the AI focus group and co-variate-matched controls (4% [2/52] versus 2% [3/121]; $p = 0.64$; two-sided Fisher’s exact test).

Outcomes in the participants

The pre-specified primary clinical endpoint was cumulative incidence of severe acute GvHD. Six participants (5.5% of 110) in the AI focus group developed severe acute GvHD, compared with 16% (41/252) in

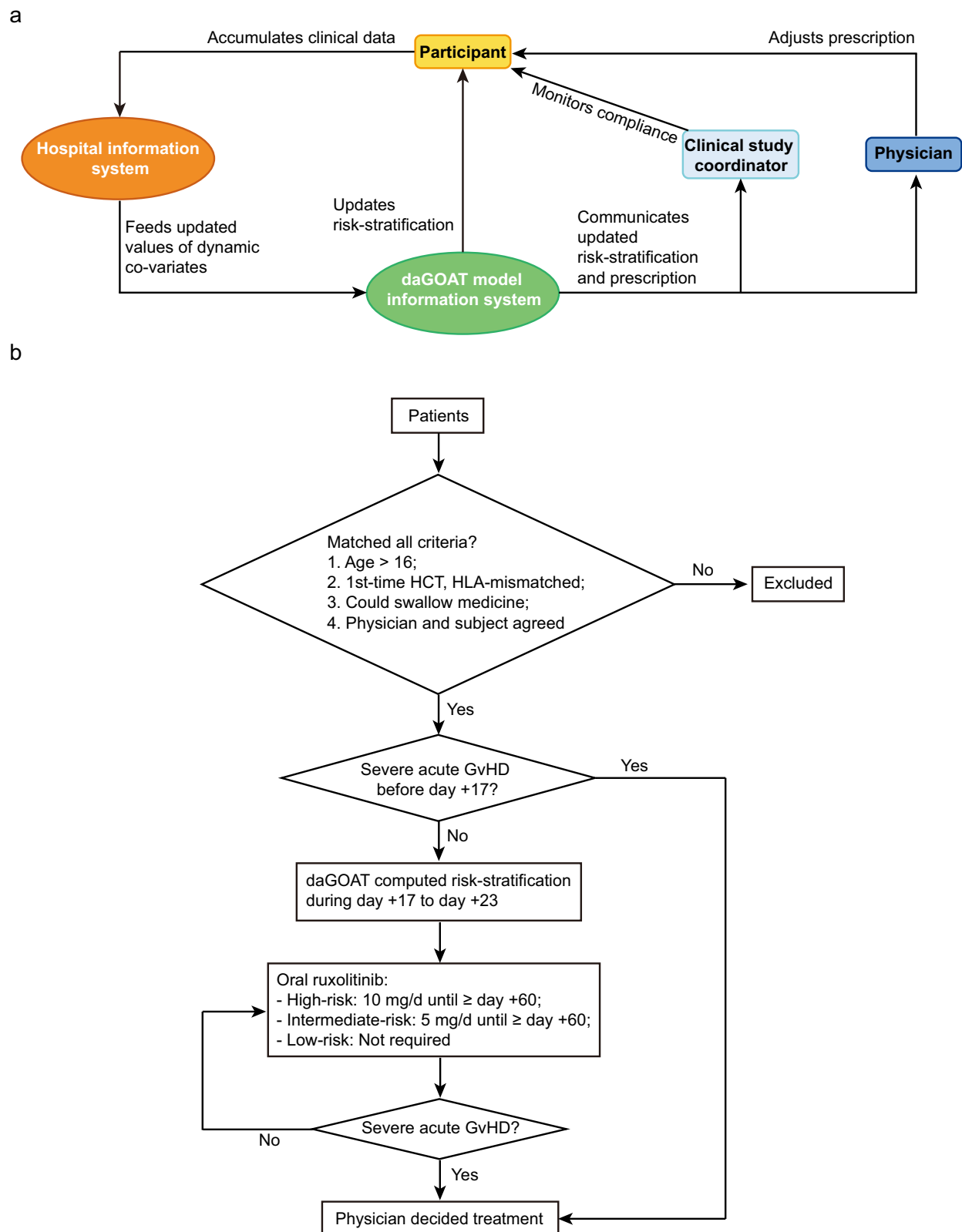


Fig. 2 | Deploying daGOAT as an autonomous artificial intelligence agent in the posttransplant care pathway. a Interactions among the participant, clinician, researcher, hospital information system, and daGOAT. **b** Work flow of the

prospective trial. Acute GvHD, acute graft-versus-host disease; HCT, haematopoietic cell transplant; HLA, human leukocyte antigen.

the co-variate-matched controls (Fig. 5a). Cumulative incidences of \geq stage 2 lower-gastrointestinal (GI), \geq stage 2 hepatic, and \geq stage 3 cutaneous acute GvHD were 5.5% (6/110), 3% (3/110), and 4% (4/110), respectively, in the AI focus group (Fig. 5b–d). Results of sensitivity

analyses and a sub-group analysis of HLA-haploidentical transplants with HLA-mismatch $\geq 3/10$ are displayed in Supplementary Fig. 3.

Between days +24 and +100 posttransplant, 66% (73/110) and 65% (72/110) of the participants in the AI focus group had haematologic and



Fig. 3 | Mock examples of daGOAT-generated alerts. **a** Alert via mobile-phone short message service. Translation: “Short message service (July 17, Wednesday, 1721 h): [daGOAT] Today the daGOAT model is monitoring three participants, of which one (patient name: Fa Hai) is high-risk, one (Xu Xian) is intermediate-risk, and one (Bai Suzhen) is low-risk.” **b** Alert via a dashboard on the desktop computer. “返

回患者中心”, back to the patient list; “姓名”, name; “性别”, sex; “就诊年龄”, age at transplant; “诊断”, primary disease; “修改信息”, edit patient information; “病程概览”, overview of the posttransplant clinical course; “中风险”, intermediate-risk; “高风险”, high-risk; “术后第_天”, day +_ posttransplant. Displayed patient names are made-up and do not correspond to real names of trial participants.

non-haematologic abnormalities, respectively, and 37% (41/110) had infection (Table 2; Supplementary Fig. 4). 18% (10/57) and 18% (10/57) of the daGOAT-designated intermediate-to-high-risk participants in the AI focus group had severe neutropenia and severe thrombocytopenia, respectively, between days +24 and +100, compared with 34% (18/53) and 45% (24/53) in the low-risk participants. The observed lower frequency of cytopenia in the intermediate-to-high-risk participants compared with the low-risk participants could be attributed to higher mean concentrations of neutrophils and platelets in blood during days +17 to +23 in the intermediate-to-high-risk participants compared with the low-risk participants (Supplementary Fig. 1), which – we speculate – might have offset the anticipated cytopenia side-effects of ruxolitinib. No immune flare (for example, cytokine storm) was observed after ruxolitinib-discontinuation in any person. Frequencies of most abnormalities and infection types were comparable between the AI focus group and co-variate-matched controls (Table 2). However, cumulative incidences of severe thrombocytopenia (31% versus 43%; $p = 0.04$), high aspartate transaminase level (7% versus 21%; $p = 0.001$), and haemorrhagic cystitis (18% versus 31%; $p = 0.01$) were lower in the AI focus group compared with the co-variate-

matched controls. These observed differences could be attributed to faster platelet recovery in the AI focus group compared with the co-variate-matched controls (Supplementary Fig. 2), chance, incomplete co-variate-matching, and/or other explanations.

In the AI focus group, one-year cumulative incidence of chronic GvHD was 36% (95% CI, 19–53%), 47% (95% CI, 29–63%), and 23% (95% CI, 7–45%) in the daGOAT-designated low-, intermediate-, and high-risk participants, respectively. One-year cumulative incidence of relapse was 8% (95% CI, 3–16%) in the AI focus group overall. One-year survival was 90% (95% CI, 84–96%; Supplementary Fig. 5).

Post-hoc analyses suggest the six participants with severe acute GvHD in the AI focus group could not be readily distinguished from the other 104 participants in the AI focus group based on their daGOAT-computed risk scores between days +17 and +23 (Fig. 5e). All the six incidences of severe acute GvHD were steroid-refractory, and their second-line treatment included ruxolitinib ($n = 6$), basiliximab ($n = 4$), mesenchymal stromal cells ($n = 3$), infliximab ($n = 1$), and vedolizumab ($n = 1$) (Fig. 6). One daGOAT-designated high-risk participant (#110) died on day +49 posttransplant, and one intermediate-risk participant (#63) died on day +190 posttransplant. The other four cases of severe

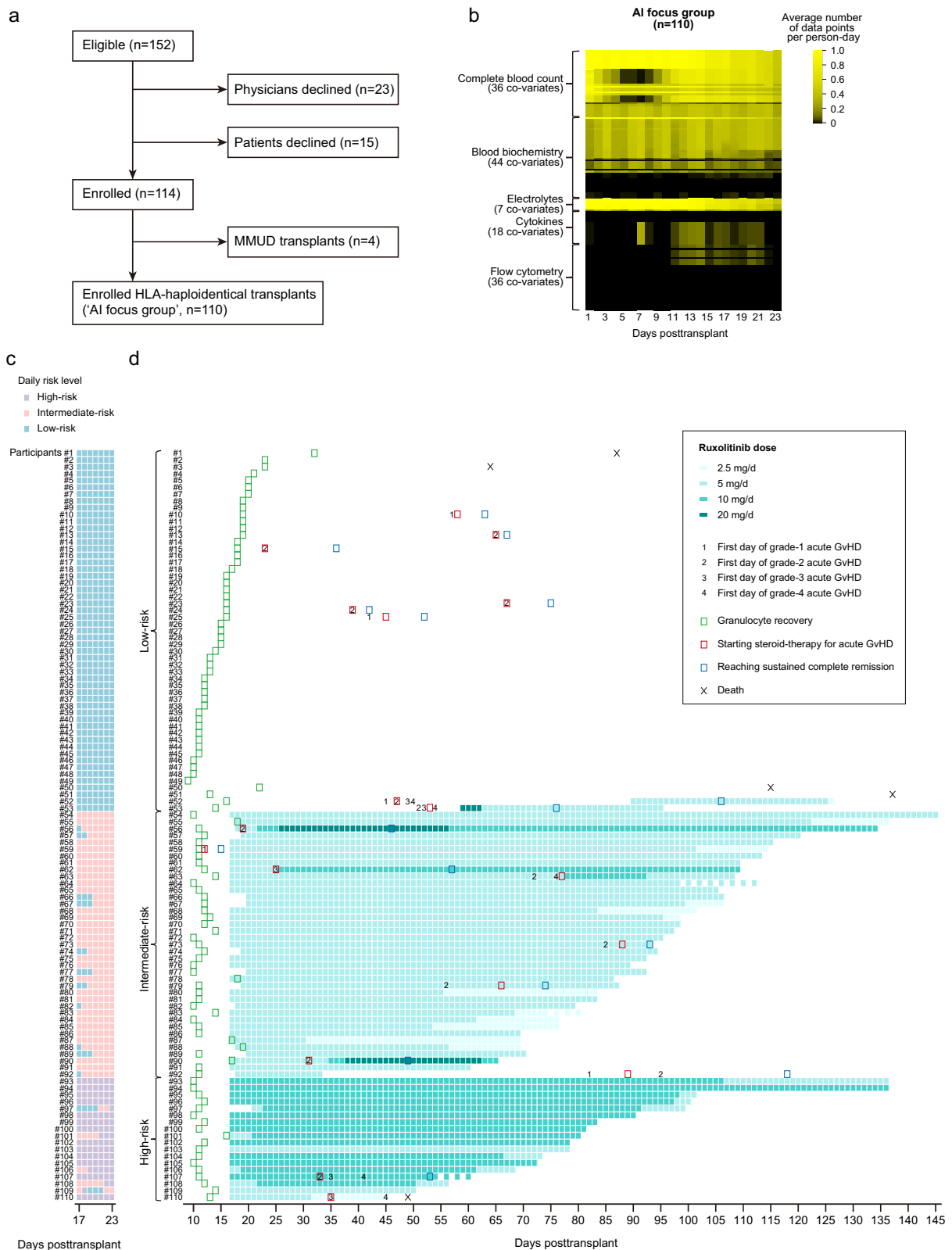


Fig. 4 | Execution of the prospective trial. a Patient screening and participant enrollment. AI, artificial intelligence; HLA, human leukocyte antigen; MMUD, HLA-mismatched unrelated donor. **b** Densities of autonomously-extracted dynamic covariate data in the AI focus group ($n = 110$). **c** daGOAT-computed risk-stratification

in the AI focus group. Participant numbering does not correspond to the participants' chronological order of receiving transplants. **d** Compliance with AI prescriptions in the AI prefocus group. Acute GvHD, acute graft-versus-host disease. Source data for (b) are available online.

Table 1 | Baseline characteristics of patients

	Prospective HLA-haploidentical transplants			Retrospective HLA-haploidentical transplants	
	AI focus group (n = 110)	Non-participating (n = 38)	p-value	Co-variate-matched controls (n = 252)	p-value (AI focus group versus co-variate-matched controls)
Recipient's age at transplant, median years (range)	42 (17–68)	40 (17–60)	0.14 ^a	40 (17–63)	0.05 ^a
Recipient's sex, n (%)			>0.99 ^b		0.81 ^b
Male	70 (64)	24 (63)		165 (65)	
Female	40 (36)	14 (37)		87 (35)	
Body mass index, median (range)	24 (15–47)	23 (17–32)	0.23 ^a	23 (13–36)	0.002 ^a
Primary disease, n (%)			0.13 ^b		0.26 ^b
Acute leukaemia	67 (61)	22 (58)		175 (69)	
MDS	33 (30)	8 (21)		57 (23)	
Others	10 (9)	8 (21)		20 (8)	
Acute leukaemia status pretransplant, n (%)			0.37 ^b		0.13 ^b
CR					
MRD-positive	6 (9)	2 (9)		31 (18)	
MRD-negative	52 (78)	14 (64)		121 (69)	
Unknown MRD-status	6 (9)	3 (14)		8 (5)	
Not CR	3 (4)	3 (14)		15 (9)	
Donor type, n (%)			0.82 ^b		0.57 ^b
First-degree relative	104 (95)	37 (97)		241 (96)	
Second-degree relative	3 (3)	0 (0)		3 (1)	
Third-degree relative	3 (3)	1 (3)		8 (3)	
HLA mismatch, n (%)			0.28 ^a		0.26 ^a
1/10	1 (1)	0 (0)		0 (0)	
2/10	2 (2)	0 (0)		4 (2)	
3/10	14 (13)	3 (8)		9 (4)	
4/10	21 (19)	7 (18)		66 (26)	
5/10	72 (65)	28 (74)		173 (69)	
ABO match, n (%)			0.47 ^b		0.26 ^b
Matched	69 (63)	25 (66)		130 (52)	
Major mismatch	16 (15)	6 (16)		52 (21)	
Minor mismatch	21 (19)	4 (11)		55 (22)	
Bidirectional mismatch	4 (4)	3 (8)		15 (6)	
Sex match (donor to recipient), n (%)			0.91 ^b		0.8 ^b
Female to Female	12 (11)	3 (8)		24 (10)	
Male to Male	47 (43)	18 (47)		121 (48)	
Male to Female	29 (26)	11 (29)		63 (25)	
Female to Male	22 (20)	6 (16)		44 (17)	
Stem cell source, n (%)			>0.99 ^b		>0.99 ^b
G-CSF mobilized blood	110 (100)	38 (100)		252 (100)	
Total nucleated cell dose in the graft, median 10⁹/kg (range)	12 (9–21)	12 (9–21)	0.65 ^a	12 (5–28)	0.83 ^a
Conditioning regimen, n (%)			>0.99 ^b		>0.99 ^b
MAC	110 (100)	38 (100)		252 (100)	
Baseline acute GvHD prophylaxis, n (%)			0.57 ^b		0.25 ^b
ATG + CsA	54 (49)	21 (55)		107 (42)	
ATG + TAC	56 (51)	17 (45)		145 (58)	
Time of granulocyte recovery (days post-transplant), median (range)	12 (9–32)	14 (9–24) [#]	0.08 ^a	13 (9–28)	0.04 ^a
daGOAT-computed risk-stratification, n (%)					0.19 ^b
Low-risk	53 (48)			123 (49)	
Intermediate-risk	39 (35)			104 (41)	
High-risk	18 (16)			25 (10)	

^a Two-sided Wilcoxon test; p-values are not adjusted for multiple-hypothesis testing.

^b Two-sided Fisher's exact test; p-values are not adjusted for multiple-hypothesis testing.

[#] Two non-participating patients died (at days +11 and +40 posttransplant, respectively) before granulocyte recovery. Reported median value and range of values are based on the other 36 non-participating patients.

AI artificial intelligence, ATG anti-thymocyte globulin, CR complete remission, CsA cyclosporine A, G-CSF granulocyte-colony stimulating factor, HLA human leukocyte antigen, MAC myelo-ablative conditioning, MDS myelodysplastic syndromes, MRD measurable residual disease, TAC tacrolimus.

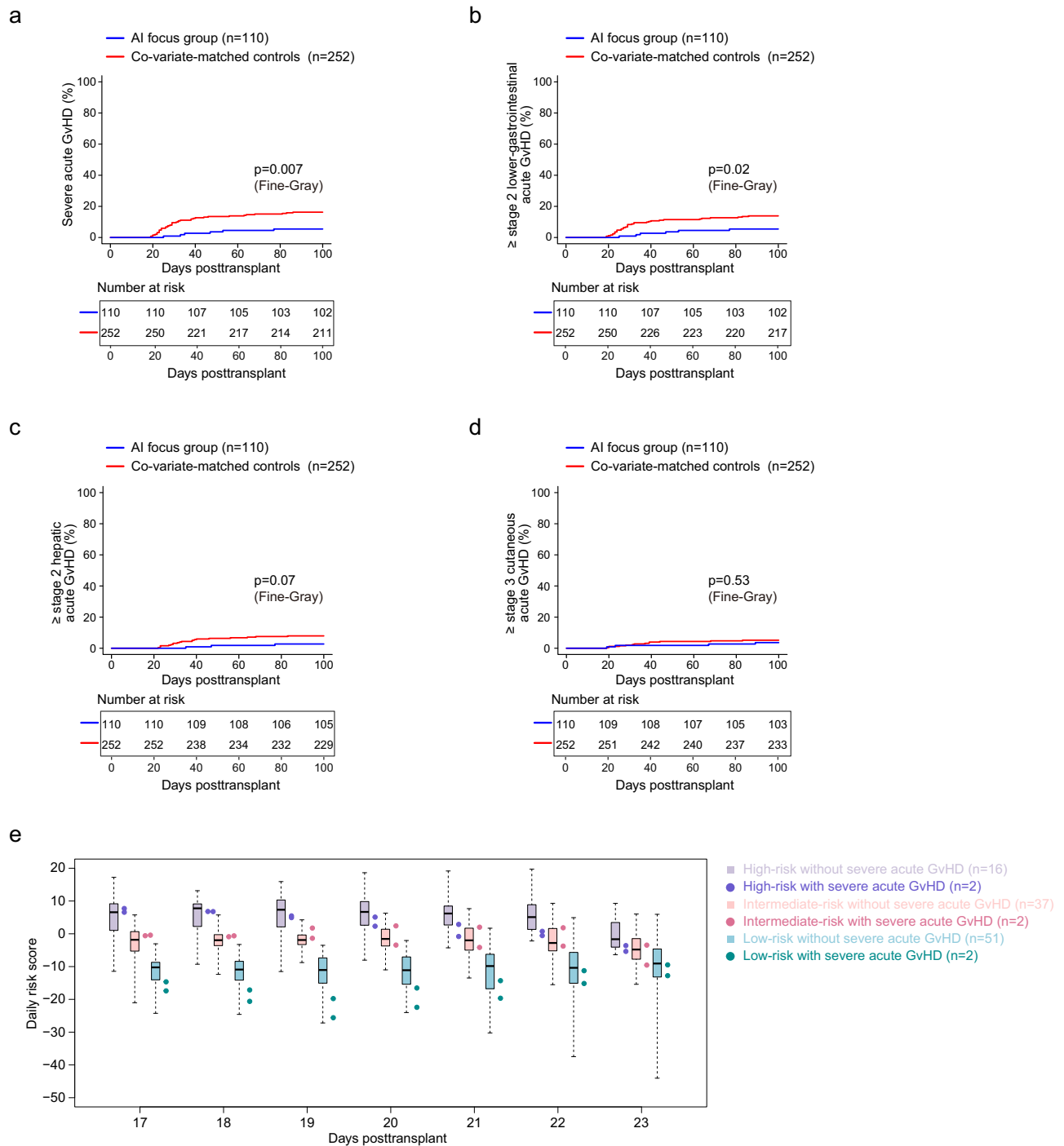


Fig. 5 | Clinical outcomes. **a** Severe (grade 3–4) acute GvHD. Blue, AI focus group ($n = 110$); red, co-variate-matched controls ($n = 252$). Pairwise comparison is performed using the two-sided Fine-Gray test. p-value is not adjusted for multiple-hypothesis testing. Acute GvHD, acute graft-versus-host disease; AI, artificial intelligence. **b** \geq stage 2 lower-gastrointestinal acute GvHD. **c** \geq stage 2 hepatic acute GvHD. **d** \geq stage 3 cutaneous acute GvHD. **e** Comparison of daGOAT-computed daily severe acute GvHD risk scores of the participants who developed severe acute GvHD in the AI focus group and the daily risk scores of the other participants who did not develop severe acute GvHD. Displayed daily risk scores are outputs of the ‘component model’ trained using severe acute GvHD as the regression target (Methods). The box-and-whisker plots indicate the distributions of daily risk scores

of the 104 participants who did not develop severe acute GvHD (disaggregated into three risk-strata: high-risk, $n = 16$ [purple]; intermediate risk, $n = 37$ [pink]; low-risk, $n = 51$ [blue]), whilst the solid circles indicate the individual daily scores of the six participants who later developed severe acute GvHD (disaggregated into three risk-strata: high-risk, $n = 2$ [purple]; intermediate risk, $n = 2$ [pink]; low-risk, $n = 2$ [blue]). In the box-and-whisker plots, the box indicates the 25th-, 50th-, and 75th-percentile values whilst whiskers, the ranges of values. Because only a small number of participants ($n = 6$) in the AI focus group developed severe acute GvHD, no statistical test is performed. Source data for (a–e) (including the individual data points underlying the box-and-whisker plots in (e)) are available online.

Table 2 | Haematologic and non-haematologic abnormalities and infections between +24 and +100 days posttransplant

	AI focus group (n = 110)	Co-variate-matched controls (n = 252)	p-value ^a
Haematologic abnormalities, n (%)			
Severe anaemia (haemoglobin <80 g/L)	67 (61)	140 (56)	0.36
Severe neutropenia (neutrophils <500/ μ L)	28 (25)	62 (25)	0.90
Severe thrombocytopenia (platelets <20000/ μ L)	34 (31)	108 (43)	0.04
Non-haematologic, n (%)			
Headache	13 (12)	17 (7)	0.15
Dizziness	9 (8)	17 (7)	0.66
Diarrhoea	28 (25)	77 (31)	0.38
Oedema	8 (7)	29 (12)	0.26
Severe hypercholesterolaemia (total cholesterol >17.1mmol/L)	0 (0)	1 (0.4)	>0.99
High alanine transaminase level (>120 U/L)	16 (15)	58 (23)	0.07
High aspartate transaminase level (>105 U/L)	8 (7)	53 (21)	0.001
High total bilirubin (>63 μ mol/L)	5 (5)	21 (8)	0.27
Petechiae, bruises, or purpura	5 (5)	28 (11)	0.05
Haemorrhagic cystitis	20 (18)	78 (31)	0.01
Bleeding at other loci (oropharynx, nose, lung, or digestive tract)	6 (5)	21 (8)	0.39
Cardiac event(s)	7 (6)	11 (4)	0.44
Thrombosis	0 (0)	3 (1)	0.56
Infection, n (%)			
CMV viraemia (DNA \geq 5000 copies/ml) [*]	7 (6)	112 (44)	<0.001
Pneumonia	4 (4)	19 (8)	0.24
Urinary tract infection	24 (22)	70 (28)	0.24
Infection at other loci (oropharynx, skin, soft tissue, or perianal region)	13 (12)	31 (12)	>0.99

^a Two-sided Fisher's exact test; p-values are not adjusted for multiple-hypothesis testing.

^{*} The lower rate of CMV viraemia in the AI focus group compared with the co-variate-matched controls might be attributable to recent adoption of antiviral letermovir, which was approved in the Chinese mainland in January, 2022.

AI artificial intelligence, CMV cytomegalovirus.

acute GvHD achieved sustained CR of acute GvHD on days +53 (Participant #107), +57 (#62), +76 (#53), and +106 (#52) posttransplant, respectively (Fig. 6).

During the enrollment period four participants receiving MMUD transplants also utilized daGOAT-driven drug prescription (Fig. 4a). daGOAT prescribed ruxolitinib for two of them; both immediately complied with AI prescriptions and did not develop severe acute GvHD. However, one low-risk participant developed severe acute GvHD (he/she reached sustained CR at day +116 posttransplant). Were we to include the four MMUD transplants in our analysis, cumulative incidence of severe acute GvHD was 6.1% (7/114; 95% CI, 2.7–11.6%) in all the participants who utilized AI prescription, compared with 16% (42/264) in co-variate-matched controls.

Post-hoc analyses of the non-participating patients

We also reviewed the 38 eligible patients who did not participate in the trial (Fig. 4a). All 38 received HLA-haploidentical transplants. Baseline characteristics of these non-participating patients were overall

comparable to the AI focus group (Table 1). However, exploratory analysis suggests that eligible patients whose blood granulocyte concentration did not recover by day +12 posttransplant were less likely to participate compared with those with earlier granulocyte recovery (66% participating [52/79] versus 85% participating [62/73]; $p = 0.008$; odds ratio [OR] for participation = 0.34). Two of the 38 patients died before granulocyte recovery. Among the 38 non-participating patients there were five (13%) who developed severe acute GvHD; three of them were excluded by the physicians from trial enrollment ('unable to swallow medicine', $n = 2$; 'acute leukaemia not in CR', $n = 1$) whereas the other two patients refused to participate.

Clinicians' attitudes towards AI-driven prescription

Sixteen physicians and 46 nurses participated in trial conduct. They were surveyed regarding their attitudes towards conditional autonomous AI-driven prescription (Table 3). Fifteen (94%) physicians stated that the main benefit of daGOAT-driven, targeted pre-emptive drug intervention was the decrease of severe acute GvHD incidence. Five (31%) physicians mentioned that because of autonomous-AI monitoring of severe acute GvHD risk they could focus on other clinical problems. Five (31%) physicians expressed that the main drawback they encountered in the daGOAT trial was the need to closely monitor the participants who took ruxolitinib. Thirty percent (3/10) of the senior physicians (with ≥ 10 years of transplantation experience) said they still did not believe an AI model can decide when to prescribe medication to pre-empt severe acute GvHD. Forty-four (96%) nurses said that the main drawback was the workload required of them to explain to the participants risks and benefits of using daGOAT ($n = 28$), train the participants on how to take low-dose ruxolitinib ($n = 32$), or monitor the participants' compliance with AI prescriptions ($n = 38$).

Discussion

Our proof-of-concept study interrogated two critical issues in medical AI.

First, how do we derive the maximal amount of value from the considerable data collected in a patient? New research indicates that even CBC, one of the most commonly-done laboratory tests in clinical medicine, has valuable information hitherto untapped²⁰. We reason that common laboratory test data in transplant recipients are plausibly under-utilized also.

Second, what is the maximally plausible level of AI autonomy? Even 'conditional automation', the lowest level of AI autonomy, is rarely tested in drug prescription. Using the keyword phrase 'autonomous artificial intelligence' to query the PubMed database we identified 40 English-language medicine-related articles published before 24 February, 2025, of which ten were prospective clinical trials^{6,8–11,21–25}. One study used an autonomous AI conversational agent for follow-up assessment of post-cataract surgery, whilst the other nine studied the role of autonomous AI in image-based diagnosis of diabetic retinopathy or colorectal polyps. However, not all studies of autonomous AI use the phrase 'autonomous artificial intelligence'. A systematic review¹² found that studies investigating the use of AI in cancer care following diagnosis are often about radiation therapy planning or behavioral nudges^{2–5,7}. Currently, little is known about the feasibility of autonomous AI in pharmaceutical intervention. This contrasts with the vision that future clinicians 'will increasingly interact with task-specific and domain-specific AI systems across a continuum of automation'¹.

In this study we evaluate a conditional autonomous AI agent monitoring high-dimensional laboratory data and prescribing a drug to prevent severe acute GvHD in the setting of HLA-haploidentical transplants. Our success in trial conduct suggests the possibility of establishing trusting and working relationships among patients, clinicians, researchers, and an autonomous AI agent, provided that the algorithm is transparent, that its deployment does not cause

Table 3 | Clinicians' attitudes towards conditional autonomous AI-driven prescription

	Self-identified as		
	Senior physician (n = 10) ^a	Junior physician (n = 6) ^{a*}	Nurse (n = 46)
1. I perceive the daGOAT model has the following benefit(s):			
Severe acute GvHD incidence rate decreased, n (%)	9 (90)	6 (100)	39 (85)
The healthcare team could focus on other clinical problems, n (%)	3 (30)	2 (33)	13 (28)
Other benefit(s), n (%)	1 (10)	0 (0)	1 (2)
2. I perceive the daGOAT model has the following shortcoming(s):			
I needed to explain to the patients risks and benefits of AI prescriptions, n (%)	1 (10)	2 (33)	28 (61)
I needed to verify medication was prescribed in accordance with AI directives, n (%)	2 (20)	0 (0)	2 (4)
I needed to confirm that the patients complied with AI prescriptions, n (%)	1 (10)	1 (17)	38 (83)
AI violated my autonomy, n (%)	0 (0)	0 (0)	1 (2)
I still doubt AI can prescribe a drug to pre-empt severe acute GvHD, n (%)	3 (30)	0 (0)	1 (2)
I needed to provide trainings to the patients on how to take low-dose ruxolitinib, n (%)	0 (0)	1 (17)	32 (70)
I needed to closely monitor the patients who were on ruxolitinib, n (%)	3 (30)	2 (33)	11 (24)
Other shortcoming(s), n (%)	0 (0)	0 (0)	0 (0)
3. (For physicians) When I enrolled a patient in the daGOAT trial, it was mainly because: [Select any statement most fitting to describe ≥1 of your patients]			
My patient matched the enrollment criteria, n (%)	4 (40)	1 (17)	–
I anticipated my patient to have good compliance, n (%)	1 (10)	1 (17)	–
I thought my patient was easier to treat, n (%)	2 (20)	0 (0)	–
I believed ruxolitinib is safe, n (%)	4 (40)	3 (50)	–
I believed AI is an important research topic, n (%)	5 (50)	4 (67)	–
I did what my superiors told me to do, n (%)	1 (10)	1 (17)	–
Enrolling the patient would not increase my workload, n (%)	2 (20)	0 (0)	–
My colleague(s) encouraged me to enroll my patient, n (%)	0 (0)	0 (0)	–
My colleague(s) were enrolling their patients proactively, n (%)	0 (0)	0 (0)	–
Patients who had already participated appeared to be doing well, n (%)	2 (20)	2 (33)	–
Other reason(s), n (%)	0 (0)	0 (0)	–
4. (For physicians) When I excluded a patient from the trial, it was mainly because: [Select any statement most fitting to describe ≥1 of your patients]			
I did not know how to enroll my patient, n (%)	0 (0)	1 (20)	–
My patient refused to participate, n (%)	2 (20)	0 (0)	–
My patient did not meet the enrollment criteria, n (%)	4 (40)	2 (40)	–
I anticipated my patient to have poor compliance, n (%)	3 (30)	3 (60)	–
I thought my patient was more difficult to treat, n (%)	5 (50)	1 (20)	–
I was unsure that ruxolitinib is safe, n (%)	0 (0)	0 (0)	–
I believed other directions of research are more important than AI, n (%)	0 (0)	0 (0)	–
I had my own research priorities, n (%)	1 (10)	0 (0)	–
Enrolling the patient would increase my workload, n (%)	0 (0)	0 (0)	–
My colleague(s) dissuaded me from enrolling my patient, n (%)	1 (10)	0 (0)	–
My colleague(s) did not enroll their patients proactively, n (%)	1 (10)	0 (0)	–
Patients who had already participated appeared to be doing poorly, n (%)	0 (0)	0 (0)	–
Other reason(s), n (%)	1 (10)	0 (0)	–

^a Senior physician, ≥10 years of transplantation experience; junior physician, <10 years.

* One person who identified himself/herself as a junior physician did not answer questions #3 and #4.

it is also plausible the model could be further optimized by narrowing the monitoring time window or reducing the number of co-variables included in the model.

The current study has limitations. First, our study was not designed to critically evaluate efficacy of ruxolitinib at the dose and schedule prescribed by daGOAT to prevent severe acute GvHD. Also, it was not designed to see if daGOAT outperforms physicians in deciding when and how to intervene to prevent severe acute GvHD. Second, we

have not addressed the crucial issues of liability (that is, determining who should be held accountable when errors occur) and the institution of a closed-loop mechanism whereby the AI agent adjusts itself in response to physician-reported anomalies⁵⁸. Third, median age of the participants in the AI focus group was 42 years (close to the median age of transplant recipients in China⁵⁹). Whether the investigated strategy is applicable to patients >60 years old is undetermined. Fourth, biology of acute GvHD in haploidentical transplants could have differences

across different transplant protocols and/or ethnic groups, and it might be necessary to refit or fine-tune the model for each separate situation.

In conclusion, our proof-of-concept study suggests physicians and patients can be receptive to conditional autonomous AI drug prescription. Generalizability of our conclusion requires testing of other drugs and/or in other clinical settings.

Methods

Clinical definitions

An HLA-mismatched donor is defined as a donor who is $\leq 9/10$ HLA-matched at HLA-A, -B, -C, -DR and -DQ. An HLA-haploidentical donor is defined as a donor who is a first- to third-degree relative $\geq 5/10$ HLA-matched other than an HLA-identical sibling. Time of attaining granulocyte recovery posttransplant is defined as the first day of ≥ 3 consecutive days of a granulocyte concentration of $\geq 500/\mu\text{L}$ in blood. Time of attaining platelet recovery posttransplant is defined as the first day of ≥ 3 consecutive days of a platelet concentration of $\geq 20,000/\mu\text{L}$ in blood. Severe acute GvHD is defined as acute GvHD with a peak severity of grade 3–4 according to the MAGIC criteria, that is, \geq stage 2 lower-gastrointestinal (GI) acute GvHD, \geq stage 2 hepatic acute GvHD, or stage 4 cutaneous acute GvHD⁶⁰.

Model update

The daGOAT model¹⁸ updates the risk score $\varphi(t)$ for severe acute GvHD according to

$$\varphi(t) = \rho(\mathbf{z}) + \sum_{\tau=t-\tau_{\text{retrospective}}}^{\tau=t} \sum_k I_{k\tau} \theta_k(x_k(\tau), \tau), \quad (1)$$

where \mathbf{z} is a vector of peri-transplant features; $x_k(\tau)$ is the k -th dynamic co-variate's value at time τ ; $I_{k\tau} = 0$ if the value of $x_k(\tau)$ is unavailable, and $I_{k\tau} = 1$ otherwise; $\theta_k(x_k(\tau), \tau)$ is a 'smooth' function that describes the relationship between the k -th dynamic co-variate and severe acute GvHD risk (technical details of the smoothing function have been described elsewhere¹⁸); $\rho(\mathbf{z})$ and $\theta_k(x_k(\tau), \tau)$ are contributions of \mathbf{z} and $x_k(\tau)$, respectively, to the risk score $\varphi(t)$; and $\tau_{\text{retrospective}}$ (unit: days) denotes how far back the model looks at past values of dynamic co-variables when computing the risk score $\varphi(t)$. Furthermore, two additional hyperparameters are introduced: $T_{\text{active}}^{\text{start}}$ (the day the model starts computing the risk score for severe acute GvHD [unit: days posttransplant]) and $T_{\text{active}}^{\text{end}}$ (the last day the model computes the risk score for severe acute GvHD [unit: days posttransplant]).

The original daGOAT model included 14 peri-transplant features and 194 dynamic co-variables. Values of the hyperparameters were originally set to be: $T_{\text{active}}^{\text{start}} = +1$, $T_{\text{active}}^{\text{end}} = +100$, and $\tau_{\text{retrospective}} = 14$.

While planning for the prospective trial, we realized the original daGOAT model had several weaknesses that would inconvenience its prospective application:

First, data collection for some of the dynamic co-variables (for example, frequency of defecation) was not easy to automate.

Second, the original model's ability to detect severe acute GvHD peaked at day +23 posttransplant whereas accuracy was considerably worse before and after¹⁸. As a consequence, we reasoned that after transplantation it might be beneficial to silence daGOAT initially and only allow the model to start giving alerts for severe acute GvHD risk after a certain amount of elapsed time.

Third, for at least some patients it would be too late to wait until day +23 to attempt to pre-empt severe acute GvHD.

Fourth, daGOAT originally classified patients into two risk-strata. However, having three risk-strata might be better than having two, because this would permit a more granular approach to pharmaceutical intervention that gives lower-intensity intervention to intermediate-risk participants and higher-intensity intervention to high-risk participants.

To remedy these weaknesses, we took a multi-pronged approach we describe below. We used retrospective data from the NICHE-GOAT cohort at the IHCAMS¹⁸ to assist model optimization. Five hundred and nineteen patients receiving HLA-mismatched HCT at the IHCAMS during 1 April 2012–30 November 2020 were used as the training set whilst 204 patients receiving HLA-mismatched HCT during 1 December 2020–31 December 2021, the validation set.

First, we eliminated some dynamic co-variables that were deemed impractical or unnecessary in prospective applications. For instance, whilst there were some bone marrow cell cytometry data points in the retrospective dataset, bone marrow cell cytometry is rarely done during the first month posttransplant, and even when it is occasionally done results are rarely available on the same day of sample collection. In addition, we removed co-variables whose evaluation could not be easily automated; examples included symptomatic complaints (for example, nausea), frequency of defecation, volume of defecation, urine output, etc. We also deleted co-variables that are no longer routinely reported by the clinical laboratory at the IHCAMS; examples included high fluorescence lymphocyte cells count, neutrophil forward scatter mean intensity, and monocyte fluorescent light mean intensity.

Although evidence was scanty for association between acute GvHD and many of the clinical co-variables (for example, plasma electrolytes), we did not perform additional 'variable selection' beyond the trimmings described above. We decided to let each co-variate speak for itself based on data. We expect co-variables not informative for predicting severe acute GvHD would have small weight values at the conclusion of model-fitting.

After being streamlined, daGOAT includes 14 peri-transplant features (same as the list in the original model) and 141 dynamic co-variables. The 141 dynamic co-variables are: CBC (36 co-variables), blood biochemistry (44 co-variables), electrolytes (7 co-variables), blood cell flow cytometry (36 co-variables), and cytokines (18 co-variables) (Supplementary Data 1).

Value of $\tau_{\text{retrospective}}$ was chosen so that the following two criteria were met in cross-validation experiments: (1) the AUROC score between $\varphi(t)$ (a numeric variable; computed using severe acute GvHD as the regression goal in Eq. 1) and severe acute GvHD onset peaked earlier between days +17 and +23 and (2) mean AUROC score between days +17 and +20 was near-maximized. These two criteria were designed to encourage the AUROC score to rise faster without sacrificing its absolute peak value. The two criteria were simultaneously met when $\tau_{\text{retrospective}} = 3$, with the AUROC score peaking at day +17 whilst mean AUROC score between days +17 and +20 was 0.67, close to the maximum mean value 0.68 (Supplementary Fig. 6).

We then developed a two-component model architecture for daGOAT to further improve its performance (Supplementary Fig. 7). We fitted the first 'component model' using severe acute GvHD as the regression target in Eq. 1 (as usual). For any given day between $T_{\text{active}}^{\text{start}}$ and $T_{\text{active}}^{\text{end}}$, we classify this component model's output on that day as 'high' if it is larger than the 97th-percentile value in the training set for that day, 'low' if it is smaller than the 70th-percentile value, and 'intermediate' otherwise. Then, we fitted a second 'component model' using grade 2–4 acute GvHD as the regression target (that is, interpreting $\varphi(t)$ in Eq. 1 as the risk score for grade 2–4 acute GvHD rather than severe acute GvHD) and likewise classify its daily outputs into 'high', 'intermediate', and 'low'. We reasoned that 'true' higher risk should be higher risk according to both the component models. With the two-component model architecture, a patient's risk level on a given day would be classified as \geq intermediate-risk if outputs from both the component models were \geq intermediate on that day; if, in addition, output from ≥ 1 component model were high, then the patient's risk level on that day would be classified as high. Thus, from $T_{\text{active}}^{\text{start}}$ to $T_{\text{active}}^{\text{end}}$, each day the combined model classifies each person's daily risk level as low, intermediate, or high. Finally, we designate each patient's final

risk-stratification to be his/her highest daily risk level between $T_{\text{active}}^{\text{start}}$ and $T_{\text{active}}^{\text{end}}$.

With $\tau_{\text{retrospective}} = 3$ (its optimal value), values of $T_{\text{active}}^{\text{start}}$ and $T_{\text{active}}^{\text{end}}$ were chosen so that both concordance (between the maximum risk level between $T_{\text{active}}^{\text{start}}$ and $T_{\text{active}}^{\text{end}}$ [a categorical variable: low-, intermediate-, or high-risk; computed using the two-component model architecture for daGOAT] and severe acute GvHD onset) and recall scores were near-maximized in the validation set. This happened when $T_{\text{active}}^{\text{start}} = +17$ and $T_{\text{active}}^{\text{end}} = +23$, with concordance score = 0.77 and recall score = 0.83 (Supplementary Fig. 6).

In summary, based on retrospective cross-validation experiments, concordance, recall, and AUROC scores and the AUROC score's rising speed were simultaneously near-optimized with $\tau_{\text{retrospective}} = 3$, $T_{\text{active}}^{\text{start}} = +17$, and $T_{\text{active}}^{\text{end}} = +23$. In the validation set, with daGOAT with the two-component model architecture, the concordance score for predicting severe acute GvHD was 0.77. In contrast, using the original daGOAT model, one had to wait until day +23 for the concordance score to peak at 0.78¹⁸.

Cumulative incidence of severe acute GvHD in daGOAT-designated low-, intermediate-, and high-risk patients was 5%, 32%, and 50%, respectively, in the validation set (Fig. 1b). Eighty-three percent of the patients who later developed severe acute GvHD were classified as intermediate- or high-risk whilst 67% of the patients who did not develop severe acute GvHD were classified as low-risk. Cumulative incidence of \geq stage 2 lower-GI acute GvHD was 4%, 25%, and 50% in the low-, intermediate- and high-risk patients, respectively; \geq stage 2 hepatic acute GvHD, 2%, 16%, and 0%; and \geq stage 3 cutaneous acute GvHD, 2%, 7%, and 8%.

Finally, fixing $\tau_{\text{retrospective}} = 3$, $T_{\text{active}}^{\text{start}} = +17$, and $T_{\text{active}}^{\text{end}} = +23$ (that is, at their respective optimal values) and leveraging data from the entire retrospective cohort (that is, 723 transplants), we refitted the values of $\rho(\mathbf{z})$ and $\theta_k(x, \tau)$. The model was last updated on 29 September 2022.

Correlation of co-variables with risk-stratification

For many dynamic co-variables, their correlation with daGOAT-computed risk-stratification became apparent when we aggregated data across patients. Of the 141 dynamic co-variables included in the updated daGOAT model, 121 (86%) had ≥ 30 data points between days +1 and +23 in each of the three risk-strata in the retrospective cohort of 723 transplants. The mean temporal profiles of these 121 dynamic co-variables are displayed in Supplementary Fig. 1. Some dynamic co-variables had 'crescendo' patterns, with the lowest mean values in the low-risk stratum and highest mean values in the high-risk stratum; examples included lymphocyte count (LYMPH#), neutrophil count (NEUT#), platelet count (PLT), T cell count, T cell percentage in lymphocytes, interleukin (IL)-2R, IL-5, IL-6, IL-8, IL-10, bilirubin (BIL), aspartate transferase (AST), and alanine transaminase (ALT). On the other hand, some dynamic co-variables had 'decrecendo' patterns; examples included albumin (ALB), AST-to-ALT ratio, and natural killer (CD3⁺CD56⁺/CD16⁺) cell percentage in lymphocytes.

Mutual information between co-variables and risk-stratification

Entropy $H(z)$ of a random variable z and mutual information $I(z_1; z_2)$ between two random variables z_1 and z_2 were computed using logarithmic base 2 (unit: bits) as in common practice⁶¹.

Let Y denote risk-stratification based on the full-version daGOAT model (that is, including all the 14 peri-transplant features and 141 dynamic co-variables), Y_0 denote risk-stratification based on only the 14 peri-transplant features, and Y_Ω denote the risk-stratification were we to add a subset of dynamic co-variables, Ω , to the 14 peri-transplant features.

A priori (that is, without any prior information on the peri-transplant features or dynamic co-variables), the entropy or uncertainty of risk-stratification was $H(Y)$ bits/person. Out of this total uncertainty $I(Y_0; Y)$ bits/person could be resolved by the 14 peri-transplant

features. Adding a subset of dynamic co-variables, Ω , to the 14 peri-transplant features would add $I(Y_\Omega; Y) - I(Y_0; Y)$ bits/person.

Among the 719 (99% of 723) patients who received HLA-mismatched transplants in the retrospective cohort and did not develop severe acute GvHD before day +17, using only the 14 peri-transplant features resolved 0.02 bits/person uncertainty in risk-stratification (that is, $I(Y_0; Y) = 0.02$). Adding CBC-, flow cytometry-, biochemistry-, electrolyte-, or cytokine-related dynamic co-variables (that is, adding only one of the five categories of dynamic co-variables) would add 0.18, 0.14, 0.09, 0.03, or 0.02 bits/person incremental information whereas including data of all the dynamic co-variables added 0.95 bits/person total incremental information (Fig. 1d).

Note that contributions to 'realized' risk-stratification should not be interpreted as contributions to 'true' risk-stratification, which is unknown to us. For example, when cytokines were added at the last (that is, the last row in Fig. 1d), the cumulative mutual information with the final risk-stratification suddenly 'jumped'; we interpret this 'jump' merely reflected that once cytokines – the last group of co-variables to be included in the model – were also included, at last the computed risk-stratification became fully identical to the final daGOAT-computed risk-stratification.

Required sample size for the prospective trial

Cohort size of the prospective trial was designed so that there would be sufficient power for comparing cumulative incidences of severe acute GvHD between the prospective-trial cohort and co-variate-matched controls.

We assumed there would be two intensity levels of add-on pharmaceutical intervention for enhancement of acute GvHD prophylaxis: higher- and lower-level doses. We further assumed that, with 'add-on prophylactic medication at the higher-level dose', cumulative incidence of severe acute GvHD in high-risk participants would become comparable to that in intermediate-risk patients receiving only baseline prophylaxis, whilst, with 'add-on prophylactic medication at the lower-level dose', cumulative incidence of severe acute GvHD in intermediate-risk participants would become comparable to that in low-risk patients receiving only baseline prophylaxis. Thus, we estimated that, overall, cumulative incidence of severe acute GvHD would fall from 18% (the cumulative incidence in the recipients of HLA-mismatched transplants at the IHCAMS during 1 December 2020–31 December 2021) to 8% in the prospective-trial cohort. Assuming the prospective-trial cohort would be compared with co-variate-matched control cases selected at the target ratio of 3:1 from electronic medical records, to attain a 0.05 significance level and a 0.8 power at a presumed 5% dropout rate, we estimated that at least 102 participants would need to be enrolled in the prospective trial.

Dosing of intensified immune suppression

For participants identified by daGOAT to be at risk of severe acute GvHD, the IHCAMS Ethics Committee approved use of ruxolitinib, a selective JAK1/JAK2 inhibitor effective in treating steroid-refractory acute GvHD^{62–64}, for severe acute GvHD prophylaxis in the context of the daGOAT trial. It was approved that 'add-on prophylactic medication at the higher-level dose' would be 'oral 5 mg ruxolitinib twice daily until at least day +60' and 'add-on prophylactic medication at the lower-level dose' would be 'oral 2.5 mg ruxolitinib twice daily until at least day +60'.

Because 20 mg/d ruxolitinib is the standard dose for treating steroid-refractory acute GvHD, it was reasoned a lower dose might be sufficient for prophylaxis⁶². Because 5 mg/d has been tested in HLA-matched allogeneic transplants for acute GvHD prophylaxis, it was reasoned a higher dose might be needed for at least some patients receiving HLA-haploidentical transplants⁶⁵.

Because daGOAT was trained on clinical co-variate data obtained before the start of therapy for acute GvHD, the study protocol did not

stipulate to de-escalate ruxolitinib dose even if a participant's risk score improved after he/she started taking ruxolitinib. However, the trial protocol stipulated that in participants with neutrophil concentration $<100/\mu\text{L}$ in blood the ruxolitinib dose could be reduced or discontinued until the physician judged it safe to restart.

Participant enrollment

Eligibility criteria included the following: (1) age >16 ; (2) first transplant and HLA-mismatched; (3) able to swallow medicine; (4) consent from the physician and patient. Sex was not a considered factor in the study design. From 17 January 2023 to 30 June 2024 ('enrollment period') two transplant wards at the IHCAMS recruited participants. Physicians and potential participants were explicitly told that an AI model would decide whether, when, and at what dose to prescribe ruxolitinib.

Dynamic co-variate data collection in the participants

The daGOAT model requested from the hospital information system only data of the co-variables required by the model and only from the participants who enrolled in the trial. Because daGOAT permits missing data¹⁸, physicians were told to order laboratory tests in accordance with their own standard practice. If a physician decided to order blood cell flow cytometry analysis, we required sample to be collected before 0700 h so that results would be available before 1620 h the same day. There was no other stipulation on data collection.

Outcome measures

Acceptance rate of trial participation and compliance with AI prescriptions were quantified. Primary clinical endpoint pre-specified in the trial protocol was the cumulative incidence of severe acute GvHD at day +100 posttransplant.

Participant follow-up

The last follow-up date was 11 March 2025. The median duration of follow-up was +327 days posttransplant.

Clinician survey

On 27 November 2024 we conducted an anonymous survey of physicians and nurses participating in the study to interrogate how they perceived daGOAT.

Statistics and software

We used greedy nearest neighbor matching with caliper constraint and the target matching ratio of 3:1 to select co-variate-matched controls from the 428 HLA-mismatched transplants that were done at the IHCAMS during 1 January 2019–31 December 2021 and did not use ruxolitinib before acute GvHD onset. Co-variables used for case-control matching were: recipient age, sex, primary disease, disease status (CR versus not in CR) and MRD-status pretransplant, donor type (HLA-haploidentical versus MMUD), HLA match, graft source (blood versus bone marrow), infused cell dose, conditioning regimen (myelo-ablative versus reduced-intensity), and baseline regimen for acute GvHD prophylaxis ('ATG + cyclosporine A' versus 'ATG + tacrolimus').

R code for executing the daGOAT model is available at GitHub¹⁹. The following R functions were used in statistical analyses: two-sided Wilcoxon test for comparing median values, 'wilcox.test'; two-sided Fisher's exact test for comparing proportions between two groups, 'fisher.test'; two-sided log-rank test for comparing survival curves, 'ggsurvplot'; computing hazard ratio when comparing survival curves or calculating concordance between risk score and acute GvHD onset, 'coxph'; two-sided Fine-Gray test (treating mortality as a competing risk; single-variable model, that is, without controlling for additional co-variables) for comparing two cumulative incidence functions (of acute GvHD, chronic GvHD, or relapse), 'crr'; and nearest neighbor matching for identifying the control cohort, 'matchit'. Statistical significance is defined as $p < 0.05$. All the reported p-values are two-sided

and not adjusted for multiple-hypothesis testing. In the plots of cumulative incidence functions of acute GvHD, event time is the first day of steroid-therapy.

Ethics approval and trial registration

This study was approved by the Ethics Committee (IIT2022034-EC-1) of the IHCAMS and registered in ClinicalTrials.gov (NCT05600855). All participants gave written informed consent (**Supplementary Notes**) consistent with precepts of the Declaration of Helsinki.

Role of the funding sources

The funders of this study had no role in data collection, analysis, interpretation, writing of the report, or the decision to submit for publication.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw data of 584 patients in the retrospective cohort are publicly accessible at the People's Republic of China (PRC) National Genomics Data Center (NGDC) with the accession number [OMIX001095](https://ngdc.cncr.ac.cn/omix/001095). Source data for Figs. 1, 4, and 5 and Supplementary Figs. 1–6 are provided with this paper. Request of raw data on the prospective AI focus group should be addressed to Junren Chen and will be subject to review by the PRC Administration of Human Genetic Resources before approval of data transfer.

Code availability

Computational code for the updated daGOAT model is publicly available¹⁹. Alternatively, a user can upload patient data to a cloud-based service (https://skirt-calculator.shinyapps.io/shiny_GOAT/) to generate the daGOAT-computed risk-stratification.

References

1. Bitterman, D. S., Aerts, H. & Mak, R. H. Approaching autonomy in medical artificial intelligence. *Lancet Digit Health* **2**, e447–e449 (2020).
2. McIntosh, C. et al. Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer. *Nat. Med* **27**, 999–1005 (2021).
3. Nicolae, A. et al. Conventional vs machine learning-based treatment planning in prostate brachytherapy: Results of a Phase I randomized controlled trial. *Brachytherapy* **19**, 470–476 (2020).
4. Wong, J. et al. Implementation of deep learning-based auto-segmentation for radiotherapy planning structures: a workflow study at two cancer centers. *Radiat. Oncol.* **16**, 101 (2021).
5. Hassoon, A. et al. Randomized trial of two artificial intelligence coaching interventions to increase physical activity in cancer survivors. *NPJ Digit Med.* **4**, 168 (2021).
6. Abramoff, M. D. et al. Autonomous artificial intelligence increases real-world specialist clinic productivity in a cluster-randomized trial. *NPJ Digit Med.* **6**, 184 (2023).
7. Manz, C. R. et al. Long-term effect of machine learning-triggered behavioral nudges on serious illness conversations and end-of-life outcomes among patients with cancer: a randomized clinical trial. *JAMA Oncol.* **9**, 414–418 (2023).
8. Wolf, R. M. et al. Autonomous artificial intelligence increases screening and follow-up for diabetic retinopathy in youth: the ACCESS randomized control trial. *Nat. Commun.* **15**, 421 (2024).
9. Abramoff, M. D. et al. Mitigation of AI adoption bias through an improved autonomous AI system for diabetic retinal disease. *NPJ Digit Med.* **7**, 369 (2024).

10. Djinbajian, R. et al. Autonomous artificial intelligence vs artificial intelligence-assisted human optical diagnosis of colorectal polyps: a randomized controlled trial. *Gastroenterology* **167**, 392–399 e392 (2024).
11. Meinert, E. et al. Accuracy and safety of an autonomous artificial intelligence clinical assistant conducting telemedicine follow-up assessment for cataract surgery. *EClinicalMedicine* **73**, 102692 (2024).
12. Macheka, S. et al. Prospective evaluation of artificial intelligence (AI) applications for use in cancer pathways following diagnosis: a systematic review. *BMJ Oncol.* **3**, e000255 (2024).
13. Gerstung, M. et al. Precision oncology for acute myeloid leukemia using a knowledge bank approach. *Nat. Genet* **49**, 332–340 (2017).
14. Komorowski, M., Celi, L. A., Badawi, O., Gordon, A. C. & Faisal, A. A. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nat. Med.* **24**, 1716–1720 (2018).
15. Bill, M. et al. Precision oncology in AML: validation of the prognostic value of the knowledge bank approach and suggestions for improvement. *J. Hematol. Oncol.* **14**, 107 (2021).
16. Yang, H., Guo, X., Peng, Z. & Lai, K.-H. The antecedents of effective use of hospital information systems in the chinese context: A mixed-method approach. *Inf. Process Manag.* **58**, 102461 (2021).
17. El-Jawahri, A. et al. Improved treatment-related mortality and overall survival of patients with grade IV acute GVHD in the modern years. *Biol. Blood Marrow Transpl.* **22**, 910–918 (2016).
18. Liu, X. et al. Dynamic forecasting of severe acute graft-versus-host disease after transplantation. *Nat. Comput Sci.* **2**, 153–159 (2022).
19. Chen, J., Qi, S., Feng, Y., Wang, Y. & Cui, M. daGOAT (v2.0). *Zenodo*, <https://doi.org/10.5281/zenodo.14898554> (2025).
20. Foy, B. H. et al. Haematological setpoints are a stable and patient-specific deep phenotype. *Nature* **637**, 430–438 (2025).
21. Grzybowski, A. & Brona, P. A pilot study of autonomous artificial intelligence-based diabetic retinopathy screening in Poland. *Acta Ophthalmol.* **97**, e1149–e1150 (2019).
22. Wolf, R. M. et al. The SEE study: safety, efficacy, and equity of implementing autonomous artificial intelligence for diagnosing diabetic retinopathy in youth. *Diab. Care* **44**, 781–787 (2021).
23. Sedova, A. et al. Comparison of early diabetic retinopathy staging in asymptomatic patients between autonomous AI-based screening and human-graded ultra-widefield colour fundus images. *Eye (Lond.)* **36**, 510–516 (2022).
24. Nolan, B., Daybranch, E. R., Barton, K. & Korsen, N. Patient and provider experience with artificial intelligence screening technology for diabetic retinopathy in a rural primary care setting. *J. Maine Med Cent.* **5**, 2 (2023).
25. Bhambhwani, V. et al. Feasibility and patient experience of a pilot artificial intelligence-based diabetic retinopathy screening program in Northern Ontario. *Ophthalmic Epidemiol.* <https://doi.org/10.1080/09286586.2024.2434738> (2024).
26. Lee, C. et al. Prediction of absolute risk of acute graft-versus-host disease following hematopoietic cell transplantation. *PLoS One* **13**, e0190610 (2018).
27. Arai, Y. et al. Using a machine learning algorithm to predict acute graft-versus-host disease following allogeneic transplantation. *Blood Adv.* **3**, 3626–3634 (2019).
28. Vander Lugt, M. T. et al. ST2 as a marker for risk of therapy-resistant graft-versus-host disease and death. *N. Engl. J. Med.* **369**, 529–539 (2013).
29. Socie, G. et al. Prognostic value of blood biomarkers in steroid-refractory or steroid-dependent acute graft-versus-host disease: a REACH2 analysis. *Blood* **141**, 2771–2779 (2023).
30. Akahoshi, Y. et al. Flares of acute graft-versus-host disease: a mount sinai acute GVHD international consortium analysis. *Blood Adv.* **8**, 2047–2057 (2024).
31. DeFilipp, Z. et al. The MAGIC algorithm probability predicts treatment response and long-term outcomes to second-line therapy for acute GVHD. *Blood Adv.* **8**, 3488–3496 (2024).
32. Tang, S. et al. Predicting acute graft-versus-host disease using machine learning and longitudinal vital sign data from electronic health records. *JCO Clin. Cancer Inf.* **4**, 128–135 (2020).
33. Bayraktar, E. et al. Data-driven grading of acute graft-versus-host disease. *Nat. Commun.* **14**, 7799 (2023).
34. Henry, K. E., Hager, D. N., Pronovost, P. J. & Saria, S. A targeted real-time early warning score (TREWScore) for septic shock. *Sci. Transl. Med* **7**, 299ra122 (2015).
35. Johnson, A. E. et al. MIMIC-III, a freely accessible critical care database. *Sci. Data* **3**, 160035 (2016).
36. Hyland, S. L. et al. Early prediction of circulatory failure in the intensive care unit using machine learning. *Nat. Med.* **26**, 364–373 (2020).
37. Im, A. et al. Risk Factors for Graft-versus-Host Disease in Haploidentical Hematopoietic Cell Transplantation Using Post-Transplant Cyclophosphamide. *Biol. Blood Marrow Transpl.* **26**, 1459–1468 (2020).
38. Nagler, A. et al. Post-transplant cyclophosphamide versus antithymocyte globulin for graft-versus-host disease prevention in haploidentical transplantation for adult acute lymphoblastic leukemia. *Haematologica* **106**, 1591–1598 (2021).
39. Battipaglia, G. et al. Impact of the addition of antithymocyte globulin to post-transplantation cyclophosphamide in haploidentical transplantation with peripheral blood compared to post-transplantation cyclophosphamide alone in acute myelogenous leukemia: a retrospective study on behalf of the acute leukemia working party of the European Society for Blood and Marrow Transplantation. *Transpl. Cell Ther.* **28**, 587 e581–587.e587 (2022).
40. Baron, F. et al. GVHD occurrence does not reduce AML relapse following PTCy-based haploidentical transplantation: a study from the ALWP of the EBMT. *J. Hematol. Oncol.* **16**, 10 (2023).
41. Ling, Y. et al. Busulfan plus fludarabine compared with busulfan plus cyclophosphamide for AML Undergoing HLA-haploidentical hematopoietic cell transplantation: a multicenter randomized phase III Trial. *J. Clin. Oncol.* **41**, 4632–4642 (2023).
42. Nagler, A. et al. Matched related versus unrelated versus haploidentical donors for allogeneic transplantation in AML patients achieving first complete remission after two induction courses: a study from the ALWP/EBMT. *Bone Marrow Transpl.* **58**, 791–800 (2023).
43. Mehta, R. S. et al. Impact of Donor Age in Haploidentical-Post-Transplantation Cyclophosphamide versus Matched Unrelated Donor Post-Transplantation Cyclophosphamide Hematopoietic Stem Cell Transplantation in Patients with Acute Myeloid Leukemia. *Transpl. Cell Ther.* **29**, 377 e371–377.e377 (2023).
44. Kharfan-Dabaja, M. A. et al. Significance of degree of HLA disparity using T-cell replete peripheral blood stem cells from haploidentical donors with posttransplantation cyclophosphamide in AML in first complete hematologic remission: a study of the acute leukemia working party of the EBMT. *Hemasphere* **7**, e920 (2023).
45. Ye, Y. et al. Similar outcomes following non-first-degree and first-degree related donor haploidentical hematopoietic cell transplantation for acute leukemia patients in complete remission: a study from the global committee and the acute leukemia working party of the European Society For Blood And Marrow Transplantation. *J. Hematol. Oncol.* **16**, 25 (2023).
46. Mussetti, A. et al. Haploidentical versus matched unrelated donor transplants using post-transplantation cyclophosphamide for lymphomas. *Transpl. Cell Ther.* **29**, 184 e181–184.e189 (2023).
47. Yu, W. J. et al. Comparison of outcomes for patients with acute myeloid leukemia undergoing haploidentical stem cell

- transplantation in first and second complete remission. *Ann. Hematol.* **102**, 2241–2250 (2023).
48. Fukunaga, K. et al. HLA haploidentical stem cell transplantation from HLA homozygous donors to HLA heterozygous donors may have lower survival rates than haploidentical transplantation from HLA heterozygous donors to HLA heterozygous donors: a retrospective nationwide analysis. *Int. J. Hematol.* **119**, 173–182 (2024).
 49. Bazarbachi, A. H. et al. Posttransplant cyclophosphamide versus anti-thymocyte globulin versus combination for graft-versus-host disease prevention in haploidentical transplantation for adult acute myeloid leukemia: a report from the European Society For Blood And Marrow Transplantation Acute Leukemia Working Party. *Cancer* **130**, 3123–3136 (2024).
 50. Elmariah, H. et al. Sirolimus is an acceptable alternative to tacrolimus for graft-versus-host disease prophylaxis after haploidentical peripheral blood stem cell transplantation with post-transplantation cyclophosphamide. *Transpl. Cell Ther.* **30**, 229 e221–229.e211 (2024).
 51. Shimomura, Y. et al. Effect of graft-versus-host disease on outcomes of HLA-haploidentical peripheral blood transplantation using post-transplant cyclophosphamide. *Bone Marrow Transpl.* **59**, 66–75 (2024).
 52. Solomon, S. R. et al. Impact of graft-versus-host disease on relapse and nonrelapse mortality following posttransplant cyclophosphamide-based transplantation. *Transpl. Cell Ther.* **30**, 903 e901–903.e909 (2024).
 53. Fuji, S. et al. Low- versus standard-dose post-transplant cyclophosphamide as GVHD prophylaxis for haploidentical transplantation. *Br. J. Haematol.* **204**, 959–966 (2024).
 54. Nagler, A. et al. Young (<35 years) haploidentical versus old (>=35 years) mismatched unrelated donors and vice versa for allogeneic stem cell transplantation with post-transplant cyclophosphamide in patients with acute myeloid leukemia in first remission: a study on behalf of the Acute Leukemia Working Party of the European Society for Blood and Marrow Transplantation. *Bone Marrow Transpl.* **59**, 1552–1562 (2024).
 55. Moriguchi, M. et al. Comparison of HLA-haploidentical donors with post-transplant cyclophosphamide versus HLA-matched unrelated donors in peripheral blood stem cell transplantation for acute myeloid leukaemia. *Br. J. Haematol.* **205**, 2376–2386 (2024).
 56. Subbaswamy, A. et al. A data-driven framework for identifying patient subgroups on which an AI/machine learning model may underperform. *NPJ Digit Med.* **7**, 334 (2024).
 57. Norona, J. et al. Glucagon-like peptide 2 for intestinal stem cell and Paneth cell repair during graft-versus-host disease in mice and humans. *Blood* **136**, 1442–1455 (2020).
 58. Mathias, R. et al. Safe AI-enabled digital health technologies need built-in open feedback. *Nat. Med.* **31**, 370–375 (2025).
 59. Xu, L. P. et al. Hematopoietic stem cell transplantation activity in China 2019: a report from the Chinese Blood and Marrow Transplantation Registry Group. *Bone Marrow Transpl.* **56**, 2940–2947 (2021).
 60. Schoemans, H. M. et al. EBMT-NIH-CIBMTR Task Force position statement on standardized terminology & guidance for graft-versus-host disease assessment. *Bone Marrow Transpl.* **53**, 1401–1415 (2018).
 61. Cover T. M., Thomas J. A. *Elements of Information Theory*, 2nd Edition. John Wiley (2006).
 62. Zeiser, R. et al. Ruxolitinib for glucocorticoid-refractory acute graft-versus-host disease. *N. Engl. J. Med.* **382**, 1800–1810 (2020).
 63. Przepiorka, D. et al. FDA approval summary: ruxolitinib for treatment of steroid-refractory acute graft-versus-host disease. *Oncologist* **25**, e328–e334 (2020).
 64. Escamilla-Gomez, V. et al. Ruxolitinib in acute and chronic graft-versus-host disease: real life long-term experience in a multi-center study for adult and pediatric patients, on behalf of the GETH-TC. *Bone Marrow Transpl.* **60**, 353–362 (2025).
 65. Zhang, B. et al. Ruxolitinib early administration reduces acute GVHD after alternative donor hematopoietic stem cell transplantation in acute leukemia. *Sci. Rep.* **11**, 8501 (2021).

Acknowledgements

This study was supported, in part, by grants from the National Key R&D Program of China (2024YFC2510500 and 2024YFC2510505 to J.C.; 2023YFC2508900 to E.J.), National Natural Science Foundation of China (82370212 and 62306340 to J.C.; 82070192 to E.J.; 82300249 to Y.C.; 82341080 and 82270236 to H.W.; 82400271 to M.N.), Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (2021-I2M-1-001 and 2022-I2M-2-003 to J.C.; 2023-I2M-3-014, 2021-I2M-1-073, and 2023-I2M-2-007 to H.W.), Tianjin Municipal Science and Technology Commission Grant (24ZXRKSY00040 to J.C.), Tianjin Natural Science Foundation (23JCZXCJ00220 to E.J.), Fundamental Research Funds for the Central Universities (2023-RW320-12 to Y.C.), and Distinguished Young Scholars of Tianjin (22JCJQC00090 to H.W.). R.P.G. acknowledges support from the UK National Institute for Health and Care Research (NIHR). We thank Yao Wang and Mengxuan Cui (Yidu Cloud Technology Inc. [Beijing, China]) for assistance in interlinking the daGOAT model and the hospital information system.

Author contributions

J.C. is the lead corresponding author of this study. J.C. and E.J. co-led and designed the study. Y.C., Y.F., Q.S., Y.H., D.Y., A.P., M.H., Y.G. and X.Zhu contributed to study design. J.C., E.J., X.Zhu, Y.F., Y.C. and X.Liu prepared the submission for institutional ethics review. Y.F. and Y.H. were project managers on the data science side, whilst Y.C. was the project manager on the clinical side. J.C. designed the updated daGOAT model. Y.F. and S.Q. calibrated the model. Y.F., Y.H., Q.S., X.G., and J.C. designed the user interface for the daGOAT app. Y.F., S.Q., Y.H., and J.C. deployed daGOAT as an autonomous AI agent in the hospital information system with assistance from X.G., X.Liu, N.Z., Y.X., and Z.S. Y.F., Y.C., and J.C. provided personnel training in using the model in clinical settings. E.J., D.Y., A.P., Y.C., S.F., and J.C. led the execution of the prospective trial with assistance from R.Z., J.W., C.L., Weihua Zhai, Y.F., and Y.H. M.X., X.Zhang, and M.C. assisted in sample processing. H.W., M.N., D.Z., X.S., and R.L. assisted in data collection. Y.F. and Y.H. managed data quality during trial conduct. Y.F., S.Q., Y.H., X.G., X.Zhai, and J.C. performed clinical data analyses with assistance from J.L., Q.S., X.Zhang, W.Y., M.C., N.Z., X.Li., P.P., M.W., M.X., and W.Zhang. J.C. and Y.C., assisted by Y.F. and Y.H., designed the survey to interrogate potential factors influencing adoption and perception of conditional autonomous AI-driven prescription; Y.C. led the execution of the survey; X.Zhang, Y.H., Y.F., and J.C. analyzed the survey. Q.S., X.G., X.Zhai, Y.H., W.Zhang, S.Q., and J.C. coordinated the efforts for summarizing the study for publication. J.C., Y.F., and Y.H. prepared the typescript with assistance from Q.S., X.G., X.Zhai, R.P.G., S.Q., S.L., and Wangsong Zhai. Y.F., S.Q., Y.H., X.Zhai, X.G., and J.C. generated the tables and figures with assistance from X.Zhang. All the authors reviewed the typescript, take responsibility for the content, and agreed to submit for publication.

Competing interests

RPG is a consultant to Antengene Biotech LLC and Shenzhen TargetRx; Medical Director, FFF Enterprises Inc.; a speaker for Janssen Pharma, BeiGene, and Hengrui Pharma; Board of Directors, Russian Foundation for Cancer Research Support; and Scientific Advisory Board, StemRad Ltd. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-62926-0>.

Correspondence and requests for materials should be addressed to Junren Chen, Yigeng Cao or Erlic Jiang.

Peer review information *Nature Communications* thanks Arnon Nagler and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

Junren Chen ^{1,2,4} , Yigeng Cao ^{1,2,4} , Yahui Feng^{1,2,4}, Saibing Qi^{1,2,4}, Donglin Yang ^{1,2,4}, Yu Hu^{1,2,4}, Aiming Pang^{1,2,4}, Qiujin Shen^{1,2}, Jieya Luo^{1,2}, Xiaowen Gong^{1,2}, Rongli Zhang^{1,2}, Xiaolin Zhai^{1,2}, Xueqian Li^{1,2}, Wen Yan^{1,2}, Xianjing Zhang^{1,2}, Mengyun Chen^{1,2}, Mingming Niu^{1,2}, Jialin Wei^{1,2}, Chen Liang^{1,2}, Weihua Zhai^{1,2}, Ningning Zhao^{1,2}, Xueou Liu^{1,2}, Sichang Liu^{1,2}, Wangsong Zhai^{1,2}, Ruixin Li^{1,2}, Xianfeng Shao^{1,2}, Dong Zhang ¹, Mingyang Wang^{1,2}, Pan Pan^{1,2}, Mingyue Xu^{1,2}, Wei Zhang^{1,2}, Yunqiang Xu^{1,2}, Xiaofan Zhu ^{1,2}, Ye Guo^{1,2}, Hong Wang ^{1,2}, Zhen Song^{1,2}, Robert Peter Gale³, Mingzhe Han ^{1,2}, Sizhou Feng ^{1,2} & Erlic Jiang ^{1,2} 

¹State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin, China. ²Tianjin Institutes of Health Science, Tianjin, China. ³Centre for Haematology, Department of Immunology and Inflammation, Imperial College of Science, Technology and Medicine, London, UK. ⁴These authors contributed equally: Junren Chen, Yigeng Cao, Yahui Feng, Saibing Qi, Donglin Yang, Yu Hu, Aiming Pang.

 e-mail: chenjunren@ihcams.ac.cn; caoyigeng@ihcams.ac.cn; jiangerlie@ihcams.ac.cn