A novel algorithmic approach to generate consensus treatment guidelines in adult acute myeloid leukaemia

Thomas Coats1,2 | Daniel Bean2,3 | Aymeric Basset2 | Tamir Sirkis4 | Jonathan Brammeld5 | Sean Johnson6 | Ian Thomas6 | Amanda Gilkes7 | Kavita Raj5 | Mike Dennis8 | Steve Knapper7 | Priyanka Mehta9 | Asim Khwaja10 | Hannah Hunter11 | Sudhir Tauro12 | David Bowen13 | Gail Jones14 | Richard Dobson2,3 | Nigel Russell5 | Richard Dillon5

1Haematology Department, Royal Devon & Exeter NHS Foundation Trust, Exeter, UK
2Biostatistics and Health Informatics, King’s College London, UK
3Health Data Research UK London, University College London, UK
4Exeter Medical School, Exeter, UK
5Guys’ and St Thomas’ NHS Foundation Trust, London, UK
6Centre for Trials Research, Cardiff University, Cardiff, UK
7Haematology, Cardiff University School of Medicine, Cardiff, UK
8Haematology, The Christie NHS Foundation Trust, Manchester, UK
9Haematology, University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK
10University College London Hospitals NHS Foundation Trust, London, UK
11University Hospitals Plymouth NHS Trust, Plymouth, UK
12Haematology, Ninewells Hospital & School of Medicine, University of Dundee, Dundee, UK
13Haematology, Leeds Teaching Hospitals NHS Trust, Leeds, UK
14The Newcastle Upon Tyne Hospitals NHS Foundation Trust, Newcastle Upon Tyne, UK

Summary
Induction therapy for acute myeloid leukaemia (AML) has changed with the approval of a number of new agents. Clinical guidelines can struggle to keep pace with an evolving treatment and evidence landscape and therefore identifying the most appropriate front-line treatment is challenging for clinicians. Here, we combined drug eligibility criteria and genetic risk stratification into a digital format, allowing the full range of possible treatment eligibility scenarios to be defined. Using exemplar cases representing each of the 22 identified scenarios, we sought to generate consensus on treatment choice from a panel of nine UK AML experts. We then analysed >2500 real-world cases using the same algorithm, confirming the existence of 21/22 of these scenarios and demonstrating that our novel approach could generate a consensus AML induction treatment in 98% of cases. Our approach, driven by the use of decision trees, is an efficient way to develop consensus guidance rapidly and could be applied to other disease areas. It has the potential to be updated frequently to capture...
INTRODUCTION

Treatment options for adult acute myeloid leukaemia (AML) have become more complex with the approval of three new drugs for first-line treatment. Gemtuzumab ozogamicin (GO), midostaurin and CPX-351 are funded in the UK for, respectively, CD33+ non-adverse risk disease, FLT3-mutated disease and myelodysplastic syndrome (MDS)-related/treatment-related AML.1–3 For many years, front-line intensive treatment options consisted of only daunorubicin and cytarabine (DA) or FLAG-IDA (fludarabine, high-dose cytarabine, idarubicin and granulocyte colony-stimulating factor).1 In the UK, use of these newer treatments is governed by strict eligibility criteria; however there are a number of situations where patients are eligible for two or even three new therapies and these are mutually exclusive, i.e. they cannot currently be used in combination outside of clinical trials due to lack of safety data. Therefore, clinicians often have to make a decision about which novel agent to prioritise, without clear guidance. A second issue is that currently all approvals are age-agnostic; however the evidence underpinning the licensing of each of these agents was largely derived from single studies which only enrolled patients in specific age groups, leading to significant uncertainty outside the trial populations.1–6 The RATIFY study (upon which the approval of midostaurin is based) compared the addition of midostaurin or placebo DA for newly diagnosed FLT3-mutated AML in patients up to the age of 60 years.5 No randomised data exist in the older population, nevertheless, midostaurin is approved for use in adults of any age.2 For CPX-351, Study 301 enrolled patients with AML with myelodysplasia-related changes (AML MRC) and therapy-related AML (t-AML), but excluded patients aged <60 years.4 Nevertheless, use of CPX is approved in patients aged <60 years despite absence of data in this population.3 GO was licensed based on the findings of the ALFA-0701 study, which excluded patients with previous MDS/MPN (myeloproliferative neoplasm) and recruited only patients aged between 50 and 70.8 Overall, the specific approvals for each new agent do not fully reflect the current evidence base, and in the absence of comprehensive guidelines or other forms of decision support, this could lead to non-expert clinicians making suboptimal treatment selections as well as unacceptable geographic variations in care.

Similar problems exist worldwide: for example the National Comprehensive Cancer Network AML guideline generally reflects the treatment approvals in the UK. Treatment options are grouped into five very broad disease groups, of which three map to the novel agents described above, e.g. favourable risk for GO, AML-MRC/t-AML/antecedent MDS-CMML (chronic myelomonocytic leukaemia) for CPX and FLT3-mutated disease for midostaurin.7 However, no specific guidance is provided for patients who are eligible for more than one of these novel agents, for example the co-existence of an MDS-defining cytogenetic abnormality and FMS-like tyrosine kinase-3 internal tandem duplication (FLT3-ITD) mutation.

Clinical guidelines exist to improve health care outcomes.8 They can improve consistency of care, support the use of proven clinical interventions and help clinicians to make informed decisions based on the most up-to-date evidence.9 However, the creation of any clinical guidance is a time-consuming exercise and therefore between updates, the clinical evidence is liable to change rendering some of the recommendations out of date.10 It is also challenging for guidelines to be exhaustive, and situations expected to arise in less than or equal to 5% of cases may be omitted from guidelines.11

Here, we have attempted to address these problems from the perspective of the clinician faced with selecting the most appropriate treatment for an individual patient. Depending on the specific disease characteristics, the patient may be eligible for 0–3 of the recently licensed therapies, as well as the pre-existing options of DA or FLAG-IDA. By digitalising eligibility criteria and baseline clinical, molecular and cytogenetic features we identify 32 possible treatment eligibility scenarios. In many of these, there is a paucity of data and/or a choice between two or more treatment options which may lead to challenges and inconsistencies in clinical decision-making. We generate exemplar cases representing each of the possible treatment scenarios and use this as the basis to generate a digitised consensus guideline, providing clinicians with a pragmatic solution to increase the quality and consistency of treatment selection.

METHODS

Eligibility criteria

Whether a treatment will be routinely funded by the National Health Service in the UK, is contingent on criteria published by National Institute for Clinical Excellence (NICE) in England, Wales and Northern Ireland, and the Scottish Medical Council (SMC) in Scotland.1–3,12–14 Overviews of the treatments and eligibility criteria are as follows:
1. Daunorubicin and cytarabine (DA) + gemtuzumab (GO)—for untreated de novo CD33+ AML, if cytogenetic test confirms that the disease has favourable, intermediate or unknown cytogenetics.
2. DA + midostaurin—for newly diagnosed FLT3-mutation-positive AML.
3. CPX-351—for untreated t-AML or AML MRC defined as: (i) prior MDS/CMML; or (ii) a disease-defining cytogenetic abnormality.
4. DA and FLAG-IDA are routinely funded for all AML patients.

Clinical scenarios

We defined each clinical scenario according to the different combination of options of novel drugs that could be available to the clinician for any given patient—using the eligibility criteria outlined above—across different European Leukaemia Net (ELN) genetic risk groupings [Favourable with/without a core-binding factor (CBF) translocation, i.e. inv(16) or t(8;21), Intermediate and Adverse].

This gives rise to a matrix of 32 theoretically possible clinical scenarios (Figure 1).

Decision trees

We converted the eligibility criteria for novel induction AML treatments (DA + GO, DA + midostaurin and CPX-351) into a digital format, using an open-platform software, esyN. This enabled the criteria for each treatment to be visually represented as a decision tree model; a case that meets all of the necessary criteria for the treatment would be able to pass each branch point in the tree and therefore be deemed eligible. A decision tree was also designed to replicate the ELN risk groupings (Favourable with/without CBF, Intermediate and Adverse) based on the relevant inputted molecular and cytogenetic features.

Generating a representative clinical case for each scenario

One thousand in silico AML cases were randomly created to cover a broad variety of AML clinical and genetic features. Cases had between 0 and 4 cytogenetic changes (30% probability for each of 0–2 abnormalities and 5% for both 3 or 4 abnormalities) with a selection of 18 disease-defining changes e.g. del(5q), and 9 non-disease-defining changes e.g. monosomy 8. Cases were also randomly assigned to mutated or wild-type status for NPM1 (30% mutated, 70% WT) and FLT3-ITD mutations (25% ITDhigh, 25% ITDlow, 50% WT), prior history of MDS (10%) or prior chemo or radiotherapy (5%) (data available on request).

All cases were CD33+. The clinical information from each case was fed into the four decision trees and based on the drug eligibility and ELN risk group that was outputted, cases were classified into the 32 possible scenarios (Figure 1).

Delphi consensus survey (first round)

One representative case from each treatment scenario was identified for review by nine members of the UK National Cancer Research Institute (NCRI) AML working group, who were asked to select their preferred induction chemotherapy for a 40- and a 65-year old patient, both with good performance status and no major co-morbidities (Figure S1). NCRI trials have generally incorporated the use of FLAG-IDA for high-risk disease, therefore a follow-up question asked if FLAG-IDA was preferred over the initial choice. A threshold for establishing a strong consensus was set as ≥85% agreement on first-line choice in line with international practice. An additional threshold for moderate consensus was set at ≥75% but <85% agreement.

Delphi consensus survey (second round)

For a second round of the consensus survey, results and comments from the first were shared anonymously with all...
respondents for clinical scenarios where there was no, or moderate, consensus.

Additional options were given for the second round of the survey: (i) the option to recognise equivalence between different treatment options; and (ii) for cases where at least one respondent preferred FLAG-IDA, respondents were asked if FLAG-IDA was an acceptable treatment over the initial choice, even if not preferred by the respondent. Respondents were then given the opportunity to change their initial choice and the same thresholds for consensus applied.

**Comparison to an existing clinical guideline**

To compare the responses of the survey to an existing clinical guideline, the European Society for Medical Oncology (ESMO) 2020 guidelines were converted into a decision tree model (Figure 2). Each of the cases reviewed in our survey was classified by this decision tree to determine the ESMO-recommended treatment.

**Incidence of each clinical scenario in real-world data sets**

Real-world data sets were analysed to establish the incidences of the different clinical scenarios identified by the in silico cases by combining two cohorts. These were patients enrolled in the UK NCRI AML 17 trial (EudraCT 2007-003798-16, AML 17) and data routinely collected from patients treated with venetoclax, as an alternative to induction chemotherapy during the coronavirus pandemic, in the National Health Service (NHS) England scheme (NHS venetoclax scheme). A total of 2757 patients were analysed, 2550 patients from the NCRI AML 17 trial and 207 from the venetoclax cohort. All cases had complete information regarding de novo/secondary disease, cytogenetics and/or fluorescence in situ hybridisation (FISH) and FLT3/NPM1 status and were inputted into the same four decision trees as the in silico cases. Patients with AML secondary to myeloproliferative neoplasms were excluded.

**Validating the decision trees**

To assess the accuracy of decision trees, 238 real-world clinical cases were independently reviewed by a clinician to determine the ELN risk group and drug eligibility for each case. The same cases were then classified using the decision tree models. Any discrepancies were reviewed by a second independent clinician to ascertain if the decision trees were correct. Two hundred seven cases were from the NHS venetoclax scheme together with an additional 31 routinely collected cases so that all major branches of the decision trees were covered by these cases (Figure S2).

**RESULTS**

Clinical data from the 1000 in silico cases were inputted into the decision trees for drug eligibility and ELN risk classification. Cases were distributed across 22 of the 32 possible scenarios based on the combination of available therapies and genetic risk (Figure S3). Ten scenarios had no case assigned to them and on subsequent review were deemed not to be biologically plausible. One representative case for each scenario was

---

**FIGURE 2** Visual representation of the decision tree designed in esyN derived from the European Society for Medical Oncology (ESMO) clinical guideline. White circles, or nodes, represent decision points within the decision tree. The decision tree commences from the node ‘start’. Orange circles represent end nodes, or outputs, from the decision tree e.g. the recommended treatment. White boxes show the criteria necessary to be fulfilled to advance to the next decision point/node.
circulated among members of the UK AML NCRI working group to establish a preferred induction treatment consensus. A strong consensus was established in 13/22 scenarios for standard first-line treatment for a 40-year-old patient, following two rounds of surveying. A moderate consensus was found in 3/22 scenarios and no consensus could be agreed upon in 6/22 scenarios. DA + midostaurin was the preferred choice in eight scenarios, DA + GO in five, CPX-351 in two and DA in one. There were no instances where FLAG-IDA was preferred over daunorubicin-based regimens, but it was considered a reasonable first-line option in five scenarios. A consensus (combining strong and moderate) was reached for 94.7% of cases in the in silico cohort of 1000 cases.

For a 65-year-old patient with a good performance status and no major co-morbidities, a strong consensus was established in 15/22 scenarios and no consensus could be agreed upon in 7/22 scenarios. DA + midostaurin was the preferred choice in eight scenarios, DA + GO in five, CPX-351 in two and DA in one. There were no instances where FLAG-IDA was preferred over daunorubicin-based regimens, but it was considered a reasonable first-line option in five scenarios. A consensus (combining strong and moderate) was reached for 94.7% of cases in the in silico cohort of 1000 cases.

To compare the recommendations of this survey to an existing international guideline, the 22 cases used in our survey were run through a model representing the ESMO 2020 guidance (Figure 2). Across the surveys of the 40-year-old and 65-year-old patients there were a total of 28 scenarios with a strong consensus. In 19 of these 28, the ESMO recommendation is the same. In three scenarios, the consensus from the survey recommends a different treatment to the ESMO guidance. In a further four scenarios, where DA + GO is the preferred option in the survey, the ESMO guidance is less specific, with DA plus or minus GO recommended. The ESMO guideline does not include a recommendation for two of the scenarios covered by this algorithmic approach.

FLAG-IDA was considered reasonable as an induction therapy in five scenarios for a 40-year-old and in one scenario for a 65-year-old patient in our survey. In the ESMO guidance FLAG-IDA was suggested as an option in two and one scenarios respectively.

To understand the real-world incidence of the 22 different clinical scenarios, and by extension the proportion of cases for which our method could provide a strong or moderate consensus treatment recommendation, two validation cohorts were identified—2550 patients from the AML 17 trial (of whom 1970 were under 60 years of age and 580 were 60 or over) and 207 patients from the NHS venetoclax scheme (15 under the age of 60 and 192 were 60 or over). We first assessed the accuracy of the decision tree models to process real-life (as opposed to artificially generated) cases by inputting clinical data from 238 retrospective cases, consisting of the NHS venetoclax scheme cohort and a further 31
DISCUSSION

Here, we demonstrate a novel approach to generate consensus guidelines for induction treatments in AML by using decision tree models. Strong or moderate consensus, applicable to 98% of real-world cases of AML, was generated within two rounds of surveying. This work identifies scenarios with a lack of strong consensus, notably where there is a choice of multiple different drugs, highlighting areas where further clinical studies are required. This approach also highlights that in some scenarios there appear to be different treatment preferences in the UK compared to other published guidelines. This may reflect greater methodological rigour of our process, rather than differences in interpretation of published data between European and UK experts.

The clinical scenarios used here were generated using a digital framework which is now accessible through a companion web application that provides a user interface for clinicians to directly interrogate the guideline by inputting the relevant clinical details and outputting the specific consensus guidance. This is available at https://amlconsensus.rosalind.kcl.ac.uk/. To improve the transparency of the guideline recommendations, comments made by experts when completing the survey are also accessible (Figure S1). Use of this application could improve the accuracy and consistency of clinical decision-making.

A limitation of this approach is that cytogenetic results are not always available at the time of starting definitive induction treatment. In the UK, patients with unknown/failed cytogenetics are considered eligible for DA + GO, not eligible for CPX (unless there was prior MDS/CMML or prior therapy) whilst eligibility for midostaurin does not rely on cytogenetics. Therefore, based on eligibility criteria alone, ‘unknown’ cytogenetics would be handled the same as a normal karyotype. FLT3-tyrosine-kinase domain (TKD) results and Next-Generation-Sequencing mutations were not included in this guidance, the latter as they are currently rarely available at the time of treatment initiation. With improvements in sequencing technology and laboratory turnaround times, future versions of this algorithm could incorporate these variables.

The decision tree models used in this study have been framed by the specific licensing conditions found in the UK. They could however be adjusted to reflect the practice or funding environment in different countries or health networks. We believe that our approach could be adapted to tackle other areas of haematology where clinical decision-making is complex, or dependent on the integration of multiple variables. Examples include selecting the most effective sequential lines of myeloma therapy, or assisting with appropriate genetic testing in myeloproliferative (MPN) and MPN/MDS neoplasms.19,20

Finally, the algorithmic method we present can easily be updated at regular intervals, allowing changes in drug approvals, novel therapies and emerging clinical data to be captured, providing a living guideline.

ACKNOWLEDGEMENTS

We are grateful to Cancer Research UK (CRUK) for the research funding for the AML17 clinical trial and to Alan Burnett for his work as chief investigator of the trial. D.Be. holds a UK Research and Innovation (UKRI) Fellowship as part of Health Data Research UK (HDRUK) MR/S00310X/1. R.Do. is supported by the following: (1) NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London, London, UK; (2) Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome Trust; (3) The BigData@Heart Consortium, funded by the Innovative Medicines Initiative-2 Joint Undertaking under grant agreement No. 116074. This Joint Undertaking receives support from the European Union’s Horizon 2020 research and innovation programme and EFPIA; it is chaired by DE Grobbee and SD Anker, partnering with 20 academic and industry partners and ESC; (4) the National Institute for Health Research University College London Hospitals Biomedical Research Centre; (5) the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London; (6) the UK Research and Innovation London Medical Imaging & Artificial Intelligence Centre for Value Based Healthcare; (7)

routinely collected cases. The drug eligibility and genetic risk of these cases were independently assessed by two clinicians and compared to the decision tree outputs. 952/952 of the decision tree's eligibility outputs were confirmed to be accurate. Next, clinical data from 2757 cases from the AML 17 trial and the NHS venetoclax scheme were inputted into the validated decision trees, and cases were assigned to 21 of the 22 previously identified scenarios (no cases were found that had co-occurrence of prior history of therapy, a CBF fusion gene and an FLT3-ITD mutation, despite this being considered an established clinical scenario). Of these cases, we could identify a strong consensus in 1685/1985 (84.9%) and a moderate consensus in 274/1985 (13.8%) for a 40-year-old patient, i.e. an overall rate of consensus of 98.7% (Figure 3A). In a 65-year-old patient our method established a strong consensus in in 755/772 (97.8%) and there were no cases with moderate consensus (Figure 3B). The consensus guidance for induction treatment was different from the ESMO guideline in 298/1985 (15%) of cases for a younger patient, and 9772 (1.1%) of cases in an older patient (Figure S4). FLAG-IDA was recommended as a reasonable option in 415/1985 (20.9%) of cases for a younger patient in our survey and 351/1985 (17.7%) cases as per ESMO. There was no difference in the recommendation of the use of FLAG-IDA in the cohort of older patients between our consensus survey and ESMO with 11772 (1.4%) of cases in both.
CONFLICT OF INTERESTS

AUTHOR CONTRIBUTIONS
Thomas Coats, Daniel Bean, Richard Dobson, Richard Dillon, Nigel Russell designed the research. Thomas Coats, Aymeric Basset designed the in silico cases. Thomas Coats, Daniel Bean, Jonathan Brammell, Richard Dillon designed and validated the models. Sean Johnson, Ian Thomas, Amanda Gilkes collected the AML17 data. Thomas Coats, Daniel Bean, Tamir Sirkis designed the web-application. Asim Khwaja, Gail Jones, Kavita Raj, Hannah Hunter, David Bowen, Sudhir Tauro, Mike Dennis, Priyanka Mehta, Steve Knapper completed the surveys. Thomas Coats, Daniel Bean, Richard Dobson, Richard Dillon, Nigel Russell analysed and interpreted the data. Thomas Coats, Nigel Russell, Richard Dillon wrote the manuscript. All authors critically reviewed the final manuscript.

ORCID
Thomas Coats https://orcid.org/0000-0003-4341-4361
Mike Dennis https://orcid.org/0000-0002-2540-8673
Nigel Russell https://orcid.org/0000-0003-1893-8155
Richard Dillon https://orcid.org/0000-0001-9333-5296

REFERENCES

SUPPORTING INFORMATION
Additional supporting information may be found in the online version of the article at the publisher’s website.