SOP # 002

SOP Title:	HEMONC QUALITY CONTROL/QUALITY ASSURANCE
SOP #:	002
Approval Date:	October 27, 2021
Last Revised By:	Jeremy L. Warner, MD, MS
Last Revised:	February 1, 2024
Effective Date:	October 27, 2021

Table of Contents

1.0 Purpose

The objective of this Standard Operating Procedure (SOP) is to describe the methods for Quality Control (QC)/Quality Assurance (QA) for identification and annotation of clinical studies compiled on HemOnc.org, for maintenance of the HemOnc.org website, and for the creation of the HemOnc ontology.

2.0 Scope

This SOP pertains to all clinical studies described on HemOnc.org, to the general maintenance and upkeep of the website, and to all content within the derivative HemOnc ontology.

3.0 Responsibilities

The Chief Executive Officer (CEO) and Chief Technology Officer (CTO) are responsible for ensuring that requirements of this SOP are met.

4.0 Procedures

4.1 Screening for trials

To identify new regimens and study references for inclusion on HemOnc.org, we undertake several parallel screening methods. These are described in further detail here and the frequency of searches is summarized in Table 1 below:

4.1.1 Review of PubMed

4.1.1.1 *[Annually]* PubMed search for ("Phase 3" OR "Phase III") AND "neoplasms"[MeSH Terms]

4.1.1.2 *[Ongoing; Reviewed at least Monthly]* PubMed email alert for "Clinical Trial, Phase III "[Publication Type] AND "Neoplasms"[Mesh]

4.1.2 Review of select publications:

4.1.2.1 *[Quarterly]* Review of the entire Table of Contents (eTOC) of the following "top-tier" general medical and hematology/oncology journals:

- *JAMA*
- *The Lancet*
- *The New England Journal of Medicine*
- *Annals of Oncology* Official journal of the European Society for Medical Oncology (ESMO)
- *Blood* Official journal of the American Society of Hematology (ASH)
- *JAMA Oncology*
- *Journal of Clinical Oncology* Official journal of the American Society of Clinical Oncology (ASCO)
- *The Lancet Haematology*
- *The Lancet Oncology*

4.1.2.2 *[Biennally]* Review of all freely available Cochrane Library Reviews under the topics "Cancer" and "Blood disorders"

4.1.3 Review of additional key sources:

4.1.3.1 *[Monthly]* Review of submitted feedback containing any missing studies obtained via the REDCap feedback form linked on the HemOnc.org homepage

4.1.3.2 *[Quarterly]* Review of all studies cited on the FDA drug label ("package insert") section 14 (CLINICAL STUDIES) for all antineoplastic agents with new approvals and/or new indications

4.1.3.3 *[Annually]* Review of all non-active interventional phase 3 clinical trials with keyword “cancer” registered on ClinicalTrials.gov

4.1.3.4 *[Biennally]* Review of all ASCO, ESMO, and NCCN clinical practice guidelines

4.1.4 Input from Editorial Board members:

4.1.4.1 *[Once, at the time of Onboarding]* For the assigned page, perform the following tasks:

- Review the relevant content for any important missing trials and/or references which should be added
- Suggest new subsections/treatment scenarios (e.g., based on mutations, clinical scenario) to group regimens/references under
- Suggest useful links to websites, tools, or additional references for a disease type
- Review the disease page(s) for outdated regimens which should be moved to “historical” pages

4.1.4.2 *[Ongoing]* Editorial Board members and anyone else can make new addition requests as well as suggestions for guidelines or regimens that can be moved to a historical status via a REDCap form accessible through the banner at the top of each HemOnc.org page.

Table 1: Frequency at which trial screening procedures are executed.

	Frequency			
Screening Procedure	**Biennially**	**Annually**	**Quarterly**	**Monthly**
Review of PubMed				
PubMed search		X		
PubMed e-mail alerts				X
Review of select publications				
Top-tier Journals			X	

Cochrane Library Reviews	X			
Additional Review				
Review of REDCap feedback survey				X
FDA drug labels			X	
ClinicalTrials.gov		X		
Guidelines	X			
Editorial Members				
Suggested additions			X	

4.2 Prioritization

Because HemOnc.org is a voluntary effort driven by the contributors, it is not possible to add every study describing every treatment regimen, in anything close to real time. Priority is on randomized studies, but we do add high-quality non-randomized studies (or even retrospective case series) if they are used in practice. What follows is the prioritization list for adding new regimens to HemOnc.org. This prioritization schedule will be revisited from time to time and updated as required.

For each category, priority would be given to those studies published in top-tier journals and then those published in other journals (see Sources).

For all RCTs, annotation should be performed of both the control arm(s) of RCTs reported by the paper describing these regimens, if they are described in sufficient detail, as well as the experimental arm(s) of RCTs reported by the paper describing these regimens, if they have statistically superior findings.

Note: On a case-by-case basis, we add regimens/references which only have been presented in abstract form if they are practice-changing or anticipated to lead to FDA approval.

4.2.1 Regimens described in registration studies (FDA has priority; other approval agencies are considered) whether they are randomized or not. However, if a product is approved and labeled based on both randomized and non-randomized studies, the priority will be to add the randomized studies.

4.2.2 Clinically relevant regimens, evaluated in fully enrolling phase 3 RCTs (including trials not specified as phase 3 but having a statistical power of 90% or greater):

4.2.3 Clinically relevant regimens, evaluated in randomized phase 2 RCTs or incompletely enrolled phase 3 RCTs (e.g., those closed early due to poor accrual

4.2.4 Clinically relevant regimens, evaluated in non-randomized trials

4.2.5 Clinically relevant regimens, evaluated in case series

4.2.6 Clinically relevant regimens, evaluated in retrospective studies

4.2.7 Clinically relevant regimens, published in conference proceedings (see Sources):

4.2.7.1 Evaluated in fully enrolling phase 3 RCTs (including trials not specified as phase 3 but having a statistical power of 90% or greater

4.2.7.2 Evaluated in randomized phase 2 RCTs or incompletely enrolled phase 3 RCTs (e.g., those closed early due to poor accrual

4.2.7.3 Evaluated in non-randomized trials

4.2.7.4 Evaluated in case series

4.2.7.5 Evaluated in retrospective studies

4.3 Quality Assurance Checks

4.3.1 Wiki Maintenance Reports

[Quarterly] To ensure the functionality of the Wiki itself, several maintenance reports are generated and reviewed, as described below.

4.3.1.1 Broken redirects

4.3.1.2 Dead-end pages

4.3.1.3 Double redirects

4.3.1.4 Long pages

4.3.1.5 Oldest pages

4.3.1.6 Pages with the fewest revisions

4.3.1.7 Pages without language links

4.3.1.8 Protected pages

4.3.1.9 Protected titles

4.3.1.10 Short pages

4.3.1.11 Uncategorized categories

4.3.1.12 Uncategorized files

4.3.1.13 Uncategorized pages

4.3.1.14 Uncategorized templated

4.3.1.15 Unused categories

4.3.1.16 Unused files

4.3.1.17 Unused properties

4.3.1.18 Unused templates

4.3.1.19 Wanted categories

4.3.1.20 Wanted files

4.3.1.21 Wanted pages

4.3.1.22 Wanted properties

4.3.1.23 Wanted templates

4.3.2 Additional Checks

[Ongoing] Several additional QA checks are performed periodically to ensure the functionality and quality of the Wiki, as described below. Many of these are continual works-in-progress due to the large and extensive nature of the Wiki.

4.3.2.1 Confirm that external links to original manuscripts are working as expected

4.3.2.2 Confirm that internal wiki links are functional

4.3.2.3 Add content to pages categorized as stubs

4.3.2.4 Define regimens that are categorized as regimen stubs

4.3.2.5 Add missing pages to the wiki in cases where the link exists, but the page does not exist

4.3.2.6 Verify unverified dosing information

4.3.2.7 Monitor the REDCap feedback survey dashboard for new entries from users

4.3.2.8 Monitor HemOnc.org email and feedback form linked from the homepage

4.4 Ontology Generation

[Quarterly] The process of transforming the HTML website content and additional offline helper tables to the structured OMOP-conformant ontology and ancillary tables requires a combination of manual and automated processes. The process is accretive, i.e., a full new version of the ontology is created with each parse and differences are calculated between the last stable version of the ontology and the newly extracted information. The following sections summarize the general steps that are undertaken, in order:

4.4.1 Part 1: Main Wiki extraction routine

4.4.1.1 Download and generate a new local copy of the HemOnc.org HTML

4.4.1.2 Review new additions to page.table and check they are correctly formatted
4.4.1.3 Parse all regimen, drug, and category pages for the key components that will be used to make changes to the ontology
4.4.1.4 Check for potential missed chunks of pages and repeat parsing as necessary

4.4.2 Part 2: First phase of post-processing. Several automated routines are undertaken to clean and prepare the output from part 1 for subsequent manual review and processing.

4.4.3 Part 3: Review of ancillary tables (unless noted, these steps are carried out in the order in which they appear)

4.4.3.1 Review indications and drugs
4.4.3.1.1 Conduct manual review of new entries in indications
4.4.3.1.2 Review indications for now-obsolete (no changes in) "FDA indication" entries
4.4.3.1.3 Review drugs for no-longer-investigational "Investigational drug"
4.4.3.1.4 Fill in empty "first-in-class" entries in indications
4.4.3.1.5 Dissect indications
4.4.3.2 Review ref.table and author.table
4.4.3.2.1 Determine whether "new" references are actually new or renamed or false positives
4.4.3.2.2 Remove No-PMID errors from ref.table/author.table
4.4.3.2.3 For the confirmed new references, manually check for co-first/co-last authors on journal websites
4.4.3.2.4 Review all new modifications to ref.table and revert if needed, fix on HemOnc.org if possible
4.4.3.3 Review authors and affiliations in author.table
4.4.3.3.1 Go through all new entries and manually update blank or mismatched author forenames
4.4.3.3.2 Go through all new study and new update entries and merge authors as indicated
4.4.3.3.3 Check affiliations with semicolons as these may be dual affiliations
4.4.3.3.4 Check for blobs and mark them as such (mostly, NEJM blobs)
4.4.3.3.5 Crack the NEJM blobs and review the results
4.4.3.3.6 Review affiliations with numbers and "@" in them (addresses/emails)
4.4.3.3.7 Review invalid countries
4.4.3.3.8 Review invalid states/territories
4.4.3.3.9 Review potential new cities
4.4.3.3.10 Review invalid cities
4.4.3.3.11 Review invalid city-region-country triples

4.4.3.3.12 Review cities with a count of 0
4.4.3.3.13 Review sites file for newly invalidated sites due to change in city, region, or country
4.4.3.3.14 Review invalid sites for addition to the mapping table

4.4.3.4 Review study.table and study_results

4.4.3.4.1 Using merges.csv as a guide, manually update study names
4.4.3.4.2 Sort study.table by key and process duplicates
4.4.3.4.3 Review multi-condition studies to make sure that the duplication is appropriate
4.4.3.4.4 Review any new entries left in the study.table
4.4.3.4.5 Fill in any missing sponsor_type information
4.4.3.4.6 Highlight any newly added study_result rows, then sort by study and manually page through the spreadsheet (this step can be done out of order)
4.4.3.4.7 Review any rows in study_results with duplicated key
4.4.3.4.8 Review any rows in study_results with metric.num.this.arm = ERROR
4.4.3.4.9 Review all rows with an existing study notation in temp field
4.4.3.4.10 Manually add or update the metric_version for new additions
4.4.3.4.11 Manually break difficult multi-arm statements (e.g., those that mix 1. and 2a/b. formats)
4.4.3.4.12 Review "could not be determined" endpoints to see if they can in fact be determined
4.4.3.4.13 Check for missed splits of PFS/OS etc.
4.4.3.4.14 Review discrepancies between study_results and study.table

4.4.3.5 Review person.table

4.4.3.5.1 Check for names within names
4.4.3.5.2 Check for atomic digest equivalents
4.4.3.5.3 Check for atomic digest stringdist = 1
4.4.3.5.4 Manually determine binary gender for the newly added names
4.4.3.5.5 Bring in gender mapping from the manually curated work file
4.4.3.5.6 Manually determine gender for the "not yet determined" with 5+ publications
4.4.3.5.7 Update vital status for those authors known to be deceased

4.4.4 Part 4: Concept name post-processing

4.4.4.1 Several additional automated routines are undertaken to clean and prepare the output from part 3 for subsequent manual review and processing.

4.4.4.2 Review new contexts and curate context.table

4.4.4.3 Review new entries in the drug.table and add classes

4.4.5 Part 5: OMOP table post-processing

4.4.5.1 Review all potential concept additions

4.4.5.2 Review all potential concept deprecations

4.4.5.3 Review all potential relationship additions

4.4.5.4 Review all potential relationship deprecations

4.4.5.5 Review all potential new synonyms

4.4.5.6 Create updated OMOP concept, relationships, and synonym tables

4.4.6 Part 6: Add external mappings and synthetic regimens

4.4.6.1 Review for new HCPCS mappings

4.4.6.2 Review for new NCIT condition mappings

4.4.6.3 Review for new NCIT regimen mappings

4.4.6.4 Review for new OncoTree mappings

4.4.6.5 Review for new RxNorm mappings (make sure all drugs have a code in rxnorm_map, first)

4.4.6.6 Review for new SEER site recode mappings

4.4.7 Part 7: Quality control checks. Several QC routines are executed to identify any inconsistencies or errors in the near-final output. This includes the following routines:

4.4.7.1 Check whether there is any cui-description discordance

4.4.7.2 Concept table should have all unique concept_codes

4.4.7.3 Check for case-sensitive duplicates

4.4.7.4 Review concept name duplicates including case-insensitive duplicates

4.4.7.5 Verify that no "Other therapy" is allowed

4.4.7.6 Verify no self-referrals except for preceding/subsequent

4.4.7.7 Check that only canonical triples are present

4.4.7.8 Review concepts that have no instantiated relationships

4.4.7.9 Add new regimen stubs to the manual curation file

4.4.7.10 Check drugs with 2+ instantiated modalities for errors

4.4.7.11 All sigs$study must be in study.table

4.4.7.12 Check SIGs value sets

4.4.7.13 Check for missing SIG dose information

4.4.7.14 Check for missing SIG days/frequency information

4.4.8 Part 8: Final de-duplication, generation of a “what’s new” file, and file saving

5.0 Sources

Knowledge aggregation sites such as HemOnc.org could not exist without primary sources. We access primary sources through a variety of means, including subscription to eTOCs, conference proceedings, and customized e-mail alerts through PubMed and other sources. This is a list of the sources that we use most frequently; decisions on whether to include treatment regimens are informed by our eligibility criteria.

5.1 "Top-tier" general medical journals

- *JAMA*
- *The Lancet*
- *The New England Journal of Medicine*

5.2 "Top-tier" hematology/oncology journals

- *Annals of Oncology* Official journal of the European Society for Medical Oncology (ESMO)
- *Blood* Official journal of the American Society of Hematology (ASH)
- *JAMA Oncology*
- *Journal of Clinical Oncology* Official journal of the American Society of Clinical Oncology (ASCO)
- *The Lancet Haematology*
- *The Lancet Oncology*

5.3 Other hematology/oncology journals

- *Blood Advances*
- *British Journal of Haematology* Official journal of the British Society for Haematology
- *Cancer* Published on behalf of the American Cancer Society (ACS)
- *Clinical Cancer Research*
- *Clinical Lymphoma, Myeloma & Leukemia*
- *Haematologica* Journal of the European Hematology Association (EHA)
- *Journal of Hematology & Oncology* Official journal of the Chinese American Hematologist and Oncologist Network
- *Journal of the National Cancer Institute*
- *Leukemia*
- *Leukemia & Lymphoma*
- *Seminars in Oncology*

5.4 Conference proceedings

- ASCO Annual Meeting Proceedings
- ASH Annual Meeting Abstracts
- EHA Abstract Book
- ESMO Scientific Meeting Reports

5.5 E-mail Alerts

- ASH Practice Updates
- FDA Oncology Drug Approvals
- Research to Practice

6.0 History

Version	Effective date	Author(s)
001	2021-10-27	Catherine Del Vecchio Fitz, PhD; Jeremy L. Warner, MD, MS

7.0 Signatures

Jeremy Warner
Jeremy Warner (Feb 1, 2024 22:28 EST)

Author: Jeremy L. Warner, CTO Date: 01/02/2024

Peter Yang
Peter Yang (Feb 7, 2024 11:42 EST)

Approved: Peter C. Yang, President and CEO Date: 07/02/2024

20240201_HemOnc_SOP

Final Audit Report 2024-02-07

Created:	2024-02-02
By:	Jeremy Warner (jeremy_warner@brown.edu)
Status:	Signed
Transaction ID:	CBJCHBCAABAALZqOhzkuHYBlGZ-Fdj7sObAu4NtOlmRQ

"20240201_HemOnc_SOP" History

Document created by Jeremy Warner (jeremy_warner@brown.edu)
2024-02-02 - 3:22:17 AM GMT- IP address: 70.172.135.43

Document emailed to Jeremy Warner (jeremywarner.md@gmaill.com) for signature
2024-02-02 - 3:22:21 AM GMT

Email sent to Jeremy Warner (jeremywarner.md@gmaill.com) bounced and could not be delivered
2024-02-02 - 3:22:29 AM GMT

Jeremy Warner (jeremy_warner@brown.edu) replaced signer Jeremy Warner (jeremywarner.md@gmaill.com) with Jeremy Warner (jeremywarner.md@gmail.com)
2024-02-02 - 3:27:05 AM GMT- IP address: 70.172.135.43

Document emailed to Jeremy Warner (jeremywarner.md@gmail.com) for signature
2024-02-02 - 3:27:06 AM GMT

Email sent to Jeremy Warner (jeremywarner.md@gmaill.com) bounced and could not be delivered
2024-02-02 - 3:27:14 AM GMT

Email viewed by Jeremy Warner (jeremywarner.md@gmail.com)
2024-02-02 - 3:27:42 AM GMT- IP address: 74.125.210.68

Document e-signed by Jeremy Warner (jeremywarner.md@gmail.com)
Signature Date: 2024-02-02 - 3:28:28 AM GMT - Time Source: server- IP address: 70.172.135.43

Document emailed to Peter Yang (peteryang@hemonc.org) for signature
2024-02-02 - 3:28:29 AM GMT

Email viewed by Peter Yang (peteryang@hemonc.org)
2024-02-02 - 3:43:38 AM GMT- IP address: 74.125.210.68

Document e-signed by Peter Yang (peteryang@hemonc.org)
Signature Date: 2024-02-07 - 4:42:17 PM GMT - Time Source: server- IP address: 146.115.74.162

Agreement completed.

2024-02-07 - 4:42:17 PM GMT