



PARIS -
SACLAY
CANCER
CLUSTER
FRANCE INNOVATION CANCER

OSIRIS RWD

Guide d'implémentation (V2)

PRÉAMBULE

OSIRIS est un projet qui a pour but de favoriser le partage des données et poser des bases communes pour permettre l'interopérabilité en oncologie. OSIRIS RWD (pour Real World Data) est une adaptation aux données « en vie réelle » du modèle OSIRIS publié en 2021 ⁽¹⁾.

Le modèle OSIRIS RWD inclut un Minimum Data Set (MDS) de données cliniques, complété par un module génomique et un MDS de données concernant l'imagerie (OSIRIS – imagerie) et la radiothérapie (OSIRIS – RT).

Le MDS clinique a été défini sur la base d'un consensus d'experts. Il est complété du guide d'implémentation qui a été élaboré en parallèle à la liste des items.

Les MDS Imagerie et radiothérapie ont été définis sur la base d'un consensus d'experts (ingénieurs en imagerie médicale, médecins médicaux, radiothérapeutes). Ils sont complétés d'un modèle de donnée conceptuel et logique dérivé du projet CodeX d'HL7 : <https://hl7.org/fhir/us/codex-radiation-therapy/>. L'ensemble de la documentation est disponible sur le lien suivant : <https://github.com/InstitutNationalduCancer>

Avec la mise en place d'un modèle de données clinique (modèle conceptuel, modèle logique et modèle physique), l'objectif est une mise en œuvre opérationnelle dans une base de données « open source ». Le choix s'est porté sur PostgreSQL, système de gestion de base de données relationnelle (SGBDR).

En aval de la base de données OSIRIS RWD, il est prévu d'extraire et transformer ces données vers d'autres formats : OMOP (pour le partage) et FHIR (pour la communication et le retour au dossier médical).

A terme, un des objectifs du projet OSIRIS RWD est de permettre l'interopérabilité entre les systèmes d'information hospitaliers (SIH) et les entrepôts de données de santé (EDS) hospitaliers. L'interopérabilité s'entend sous tous ses aspects : technique, syntaxique, sémantique et juridique.

- Technique : les données sont indépendantes de l'infrastructure des systèmes et des protocoles de communication
- Syntaxique : le modèle de données, le format des tables et le nom de la variable sont documentés
- Sémantique : les données sont codées de manière standardisée et normalisée avec un même sens donné aux codes et à la description de la variable
- Juridique avec un même respect des règles et des politiques de confidentialité pour être conforme au RGPD.

¹ OSIRIS: A Minimum Data Set for Data Sharing and Interoperability in Oncology
Julien Guerin, Yec'han Laizet, Vincent Le Texier and all. JCO Clin Cancer Inform 2021, 5:256-265.

Dans ce document sont définies les règles d'alimentation de la base de données qui constituent le guide d'implémentation. L'alimentation de la base de données est sous la responsabilité des établissements de santé et des gestionnaires des EDS hospitaliers.

Ce document a pour objet de détailler les items du MDS et de proposer les modalités d'alimentation qui vont de la source (le SIH) vers la base de données pour une alimentation automatique. Différents points sont abordés en fonction de la complexité des items :

- Définition de la variable : à quelle entité elle appartient (patient, tumeur, traitement),
- Choix des terminologies (Ex : CIMO-3, LOINC, SNOMED CT).
- Définir la source : dossier médical textuel ou donnée structurée, issu de quel type de compte-rendu du dossier médical en tenant compte des possibilités de structuration par des techniques de traitement automatique du langage naturel (TALN).

Le guide sera complété de règles pour une alimentation « manuelle » pour permettre une mesure de la qualité des données alimentées automatiquement.

Descriptif du document

Le guide d'implémentation regroupe les recommandations qui ont émanées du groupe d'experts. Les recommandations ont été validées par consensus.

Les principes « FAIR »

Le modèle de données OSIRIS RWD répond aux principes dits FAIR (Findable, Accessible, Interoperable, Reusable) ; ces principes décrivent comment les données sont organisées pour être accessibles, comprises, échangeables et réutilisables.

- Le principe « Findable » a pour objectif de permettre le traitement de la donnée qui doit être comprise autant par des humains que par des systèmes informatiques. Ceci requiert une description des données et de tout ce qui sert à caractériser une donnée, comme sa définition et son format.
- Le principe « Accessible » nécessite de stocker durablement les données et les métadonnées et faciliter leur accès en tenant compte des conditions d'accès (ouvert ou restreint) et d'utilisation (licence).
- Le principe « Interoperable » signifie que la donnée est standardisée sur le plan national voire international.
- Le principe « Reusable » met en avant les caractéristiques qui rendent les données réutilisables à d'autres finalités que les soins c'est-à-dire pour la recherche, l'enseignement, l'innovation ou l'évaluation.

Les 10 principes spécifiques au modèle OSIRIS RWD

Dans le choix des items et des règles de structuration, plusieurs principes ont été adoptés que l'on peut résumer en dix points :

1. **Utile**, l'information collectée doit répondre à des objectifs de soins, de recherche, d'évaluation de pratique. Le MDS exclut des données pour effectuer des études de marché, comparer des hôpitaux entre eux dans le cadre d'une éventuelle concurrence.
2. La donnée doit pouvoir être collectée en **automatique** à partir des dossiers médicaux et des multiples sources des systèmes d'information hospitalier.
3. La mesure de **qualité** des données d'une base alimentée automatiquement à partir de données transformées est indispensable. Il est proposé la création de bases de données de référence avec une saisie manuelle (« gold standard ») pour mesurer la qualité des données extraites automatiquement ou structurées à partir de textes via des règles ou des systèmes de traitement automatique du langage.
4. **Implémentable** ce qui veut dire que l'item doit être bien défini. Toutes les données sont décrites afin que l'alimentation automatique se fasse sans ambiguïté, ce qui nécessite une description précise de chaque variable.
5. Dans le respect de la **réglementation** en vigueur (RGPD/CNIL). Le respect de la réglementation s'impose et doit permettre de gérer les données identifiantes et non identifiantes, à charge pour l'établissement de gérer une table de correspondance. Les règles concernant la non-opposition et leur traçabilité sont conservées dans la base de données.
6. Les recommandations doivent être **reproductibles** dans le temps (valable pour des dossiers médicaux récents et anciens) et dans l'espace (pour tout établissement hospitalier quel que soit son système d'information). Les établissements hospitaliers sont susceptibles de conserver des données pendant un temps long (plusieurs années). Malgré les modifications de prises en charge dues aux progrès dans les méthodes diagnostiques ou thérapeutiques, les recommandations du guide doivent s'appliquer pour fournir des résultats cohérents sur une longue période, et même si le SIH évolue dans le temps.
7. Privilégier les **données structurées** à la source quand elles sont disponibles dans les SIH, par rapport à une source de données non structurée. Dans la mesure du possible, les sources de données sont indiquées pour chaque item (donnée structurée par l'établissement, donnée récupérée par TALN à partir des documents internes, donnée récupérée par TALN à partir de documents provenant d'autres structures).
8. Lorsqu'on utilise un référentiel, on l'utilise avec les propres **règles du référentiel**. Par exemple, la CIMO-3, que nous utilisons pour gérer les codes morphologiques de la tumeur primaire, possède un certain nombre de règles, comme de faire un choix entre plusieurs codes morphologiques pour une même topographie.
9. Pour le choix entre deux **classifications**, il a été privilégié l'aspect interopérable avec l'existence ou non de tables de correspondance avec les autres classifications, le caractère international de la classification, les recommandations de l'Agence de

numérique en santé, les habitudes d'utilisation dans les établissements hospitaliers en France, la granularité sur le sujet traité, le maintien à jour de la classification.

10. Des **conventions** sont parfois nécessaires ; elles sont explicitées et justifiées. Par exemple, une date de début d'évènement tumoral ne correspond pas à une date réelle ; une maladie métastatique n'apparaît pas comme une maladie aiguë à un instant t mais correspond à un processus biologique qui peut s'étaler sur plusieurs années. La date de survenue ne peut être qu'une convention. On propose de prendre les conventions internationales quand elles existent.

Les sources de données « nécessaires » pour une alimentation automatique

Afin de structurer les informations d'OSIRIS RWD, les sources nécessaires sont les données suivantes : L'attendu minimum pour un établissement est de fournir :

- Les données d'état civil
- L'ensemble des comptes rendus médicaux ;
- Les données du PMSI
- Les données d'administration des molécules anticancéreuses
- Les données du PACS.

Le principe général est de considérer les sources de données structurées comme fiables et prioritaires par rapport aux concepts extraits par TALN.

OSIRIS RWD : des utilisations multiples

Il existe des utilisations et des finalités diverses avec des conséquences techniques et juridiques selon que les bases de données sont utilisées pour la recherche, pour l'évaluation des soins ou le pilotage de l'activité.

OSIRIS-RWD doit servir les besoins au niveau de chaque établissement, pour les soins, mais aussi permettre des études de recherche multicentriques.

OSIRIS RWD peut permettre la mise en œuvre d'une interopérabilité entre les systèmes d'information hospitaliers (SIH).

OSIRIS RWD peut devenir le socle des entrepôts de données de santé (EDS) hospitaliers en cancérologie.

OSIRIS RWD peut servir au niveau local avec un retraitement des données en consolidant des informations de différentes sources pour permettre :

- Le développement d'outils d'aide à la décision, d'aide à la prescription ou au diagnostic
- L'utilisation des données pour une analyse multimodale,
- La mise en application de jumeaux numériques
- Et permettre tout projet impliquant des algorithmes d'intelligence artificielle avec retour au dossier de soins dans une « boucle fermée ».

C'est une des raisons qui nécessitent une conversion des données d'OSIRIS RWD au format FHIR.

Gestion des dates dans le modèle

La gestion des dates est primordiale dans un modèle évènementiel qui reconstitue la chronologie de l'histoire de la maladie.

Les informations relatives à une date (année/mois/jour) sont enregistrées dans 3 champs séparés.

Si une valeur est non disponible sur le jour (DD), le mois (MM) ou l'année (YYYY), le champ correspondant sera remplacé par la valeur UN (Unknown). A titre d'exemple, si l'information du jour n'est pas disponible, elle ne sera pas remplacée par le 15 du mois ou si le mois n'est pas disponible, elle ne sera pas remplacée par la valeur 06.

Source de l'information

La source de données permet de distinguer si l'acte (ou la prise en charge) est réalisé dans l'établissement ou en dehors de l'établissement.

L'objectif est de distinguer une information structurée venant de l'établissement, d'un document interne à l'établissement ou d'un document externe.

On propose de créer une métadonnée attachée à certains items selon les informations de la source ;

- Source structurée dans le dossier du patient ou dans les progiciels de soins
- Autre source structurée de l'établissement (exemple : base de données)
- Source structurée externe à l'établissement (exemple : INSEE)
- TALN d'un document texte issu de l'établissement
- TALN d'un document issu d'un autre établissement

Les conséquences sont multiples :

- En termes de qualité de données, l'information est toujours meilleure lorsqu'elle provient de l'établissement. Elle est parfois structurée à la source (par exemple pour les actes chirurgicaux) et est plus complète.
- A l'utilisation, certaines études peuvent restreindre l'inclusion aux patients pris en charge dans l'établissement.
- La prise en charge dans l'établissement indique indirectement que des éléments du dossier ou des éléments dérivés des prélèvements sont disponibles (par exemple, si un prélèvement est réalisé dans l'établissement, celui-ci a l'obligation de les conserver comme les blocs de tumeurs inclus en paraffine).
- Une prise en charge diagnostique ou thérapeutique peut être réalisée en dehors de l'établissement ; elle est alors répertoriée sous la forme d'un compte-rendu conservé dans le dossier électronique (scanné puis interprété par reconnaissance optique des caractères), ou bien l'information est retranscrite dans un compte-rendu médical. Une des problématiques est qu'il n'est pas possible d'assurer l'exhaustivité de l'information.

Si la prise en charge d'un cancer est totalement réalisée en dehors de l'établissement, et avant la première venue du patient dans l'établissement, on peut considérer qu'il s'agit d'un antécédent de cancer. En général, les informations recueillies sont assez succinctes mais cette information mérite d'être recueillie, ne serait-ce que pour repérer les cancers multiples chez un même patient.

MDS et données complémentaires

Il est créé une table « observation » et une table « mesure » pour des données complémentaires non prévues dans le MDS. Ces tables pourront contenir des informations cliniques attachées au patient si cela concerne directement la personne : données obtenues dans le cadre d'un examen, d'un interrogatoire ou d'une procédure, sur le mode de vie, de la personne, etc.

Les données sources seront issues d'une paire clé-valeur auquel on ajoute l'identifiant du patient et une date. Pour certaines données, il sera utile de disposer du meilleur rattachement possible de la valeur, au patient, à la tumeur, à l'évolution tumorale, au traitement.

Evolution du document

Cette version du document dite 2025 V2 est amenée à évoluer.

Issu d'un groupe d'expert dont la liste figure en annexe, le guide d'implémentation doit être soumis à discussion auprès de différentes institutions ainsi qu'auprès de sociétés savantes. La première version du document a été élaborée avec une liste d'experts qui figure en annexe. Une relecture a été réalisée par un Comité de relecteur.

Il est prévu d'implémenter une première version du modèle dans plusieurs projets de recherche multicentriques afin d'en tester la pertinence, la robustesse et de préciser certains points.

Une traduction en langue anglaise est prévue.

PRINCIPES DES CONVENTIONS D'ALIMENTATION

Information et non-opposition du patient pour la réutilisation de ses données de santé

La « non-opposition du patient à l'utilisation de ses données » est un principe encadré par le Règlement Général sur la Protection des Données (RGPD). En matière de santé, cela signifie que les données personnelles d'un patient peuvent être utilisées, sous certaines conditions, sans qu'un consentement explicite soit requis, tant que le patient n'a pas exprimé d'opposition et sous réserve de respecter des conditions strictes comme l'anonymisation, la pseudonymisation et l'information du patient.

Il est de la responsabilité de l'établissement hospitalier d'informer ses patients si leurs données sont réutilisées. L'information des patients et la traçabilité de ces informations sont à recueillir en amont de la base OSIRIS RWD.

En cas d'opposition dès le début du processus de soins, il est recommandé de ne pas alimenter la base de données. Néanmoins, cette opposition peut survenir à une date ultérieure à une étude déjà conduite sans qu'il soit possible de revenir sur les résultats et sans qu'il soit possible de supprimer certaines données.

Afin d'assurer l'interopérabilité sous l'angle juridique, dans le cadre de l'échange ou du partage des données, il est nécessaire de tracer que les procédures ont été respectées et conservées avant l'utilisation des données.

Facteurs de risque

Le cancer est une maladie multifactorielle et il existe de nombreux facteurs de risque qui peuvent contribuer au développement de différents types de cancer. Certains sont identifiés et démontrés, de nombreux facteurs de risque sont suspectés.

Les facteurs de risque reconnus sont le tabagisme, la consommation d'alcool, un surpoids, l'alimentation trop riche, une exposition à certaines substances toxiques (amiante), aux rayonnements ionisants ou certaines infections virales et bactériennes (papillomavirus, virus de l'hépatite B et C).

Certains facteurs de risque seront aisés à structurer de façon automatique comme le « tabagisme » ou une « consommation d'alcool » car l'information existe très souvent sous forme textuelle ou structurée dans les dossiers (première consultation ou consultation d'anesthésie par exemple).

D'autres facteurs de risque seront certainement plus difficiles à récupérer comme une « alimentation déséquilibrée » ou une « activité physique insuffisante ».

Certains facteurs de risque sont liés à un seul cancer comme la durée insuffisante de l'allaitement infantile pour le cancer du sein ou une infection chronique par *Helicobacter pylori* pour les lymphomes de MALT et le cancer de l'estomac.

Une liste de 13 facteurs reconnus est proposée par le CIRC et repris par l'INCa ont été retenus

dans le modèle :

- Tabagisme (dont tabagisme passif)
- Consommation d'alcool
- Alimentation déséquilibrée
- Surpoids et obésité
- Activité physique insuffisante
- Utilisation d'hormones exogènes
- Durée insuffisante de l'allaitement infantile
- Infections (ex : hépatite B et C, papillomavirus, H. Pylori)
- Radiations ionisantes (radon, exposition diagnostic)
- Pollution atmosphérique, particules fines
- Rayonnement solaire (UV)
- Expositions professionnelles (ex : amiante)
- Exposition à certaines substances chimiques (ex : arsenic, benzène).

Les informations recueillies dans le modèle sont, pour chacun des 13 facteurs de risque reconnus par le CIRC, limités à Présent, absent, inconnu.

Si dans le cadre d'une étude particulière, il est nécessaire d'obtenir plus de précisions, il est proposé d'utiliser la table « observation » et la table « mesure ».

Statut vital du patient

Les informations sur le statut vital du patient peuvent être issues de plusieurs sources :

- Le dossier administratif du patient dans l'établissement de soins avec la date de dernière venue et la date de décès si celui-ci est enregistré.
- Le dossier médical par le biais d'un compte-rendu médical réalisé dans l'établissement.
- Des courriers externes donnant des nouvelles du patient dans des comptes rendus de consultations, des comptes rendus d'hospitalisation, des résultats d'examen, des lettres du patient ou de son entourage.
- L'INSEE.

Outre l'information sur le statut vital (vivant, décédé), la recommandation est de conserver quatre dates :

- La date de décès.
- La date de dernière venue physique dans l'établissement.
- La date de mise à jour de l'information issue de l'établissement ou de l'INSEE (dite « date de point »).
- La date de dernier contact issue des documents externes.

Chaque date est décomposée en année, mois, jour (YYYY, MM, DD).

Les items concernant le statut vital sont rattachés au patient.

Il est recommandé de mettre à jour les données en respectant une priorisation pour l'alimentation automatique :

- Si le patient s'est présenté dans l'établissement à une date donnée, cette date est appelée « date de dernière venue ».

- Si l'information de décès est structurée dans le dossier de l'établissement, il faut privilégier cette source (date de décès).
- En complément de ces 2 sources, il est possible de rechercher les informations de décès dans les documents du dossier médical pour alimenter la « date de dernier contact ».
- Il est recommandé d'interroger l'INSEE et de mettre à jour la date d'interrogation de l'INSEE pour toute étude de survie et faire un mapping avec les données administratives.

Antécédent médical, comorbidité et événement indésirable

Définitions

Un **antécédent médical** est la survenue d'une maladie ou d'un symptôme avant une date fixée, par exemple la date de diagnostic d'un cancer. Cette information est importante car cela peut influencer la décision thérapeutique. L'antécédent médical est un fait antérieur concernant la santé du patient qu'il est nécessaire de connaître pour prendre en charge la personne malade.

Une **comorbidité** est une maladie ou un symptôme, aiguë ou chronique, active au cours de la prise en charge d'une pathologie, en l'occurrence la maladie cancéreuse. C'est l'état, pour un patient, d'avoir une ou plusieurs maladies en même temps que la maladie cancéreuse.

Un **événement indésirable** est la conséquence, souvent prévisible, d'un traitement survenant en plus de son effet principal. Il existe des effets secondaires immédiats et des effets secondaires tardifs. Un effet secondaire peut persister longtemps après l'arrêt des traitements et, parfois, devenir définitif. L'effet indésirable a comme synonyme « toxicité » ou « effet secondaire ».

La difficulté est de les différencier, a priori. Ainsi, si certaines pathologies ou maladies peuvent être aisément classées en antécédent (un infarctus du myocarde, une pneumonie), pour d'autres, il est parfois difficile de classer en comorbidité ou en événement indésirable la survenue de certaines pathologies. Il faudrait pour cela considérer qu'elles sont attribuables à un traitement et être certain de l'imputabilité d'un effet secondaire à une molécule par exemple. De plus, des événements pathologiques, des comorbidités, sont, soit favorisés par la maladie cancéreuse, soit seraient apparus de toute façon. A l'évidence, il est difficile d'affirmer ce qu'il serait advenu indépendamment du cancer.

Exemples

Trois exemples peuvent illustrer cette problématique ; la maladie thrombo-embolique veineuse (MTEV), l'insuffisance rénale, des troubles digestifs plus ou moins sévères.

La maladie thrombo-embolique veineuse (MTEV)

La MTEV, embolie pulmonaire ou thrombose veineuse profonde, est une pathologie fréquente en population générale. Si elle survient avant la découverte du cancer, on la définira comme un antécédent médical qui pourra influencer la thérapeutique du cancer, par exemple le choix entre plusieurs hormonothérapies dont certaines augmentent le risque de MTEV.

Si la MTEV survient au cours de la maladie cancéreuse, on parlera de comorbidité, d'autant que l'on sait que pour certains cancers, le risque est multiplié par un facteur 2 à 3. Ceci est d'autant plus important que la MTEV est la deuxième cause de mortalité des patients atteints de cancer.

L'insuffisance rénale chronique (IRC)

L'insuffisance rénale chronique (IRC) est une maladie fréquente qui touche près de 15 % des adultes de plus de 65 ans. Or, certains médicaments du cancer favorisent la survenue d'une IRC et peuvent être considérés comme un effet indésirable dont la gravité est variable, allant d'une augmentation du taux de créatinine dans le sang à une insuffisance rénale aigüe.

En fonction de la date de survenue, on définira l'insuffisance rénale comme une comorbidité ou comme un effet indésirable. De plus, il ne peut être exclu qu'il s'agit d'une comorbidité qui serait apparu de toute façon, soit avec une cause connue (liée à une prise médicamenteuse comme des AINS), soit d'étiologie inconnue.

Les troubles digestifs

Les troubles digestifs, vomissements et diarrhées, sont des effets indésirables fréquents des traitements de chimiothérapie ou de radiothérapie. Ils peuvent aussi survenir à l'occasion d'une intoxication alimentaire. Si ces symptômes surviennent dans les suites immédiates d'un traitement, on parlera volontiers d'effet indésirable. S'ils surviennent à distance, c'est plus probablement une cause fortuite, indépendante de tout traitement.

Application au modèle OSIRIS RWD

Ces trois exemples montrent la difficulté, voire l'impossibilité, d'imputer des comorbidités ou des effets indésirables aux traitements du cancer, surtout lors d'associations thérapeutiques (bithérapie ou association radiothérapie-chimiothérapie).

C'est la raison pour laquelle les notions de comorbidités, d'évènements indésirables et d'antécédents sont regroupées, laissant à l'analyse des données la possibilité de les classer sur un faisceau d'argument (probabilité, date de survenue, etc.). Afin de les différencier au moment de l'analyse, la date de survenue doit pouvoir être repérée à chaque évènement repéré dans le dossier médical.

La convention adoptée est de recueillir l'information brute dans un seul item appelé **évènement morbide** sans préjuger du fait qu'il s'agit d'un antécédent médical, une comorbidité ou un effet indésirable.

La date de survenue, de découverte ou de recueil dans le dossier médical sera notée, à charge au moment de l'analyse d'interpréter les résultats comme un antécédent, une comorbidité ou un possible effet indésirable.

Tous les évènements morbides sont rattachés au patient.

Evènements indésirables et choix du référentiel

On propose d'utiliser une seule classification, la CIM-10, pour structurer maladies et symptômes, antécédent, comorbidité et effets indésirables.

Les effets indésirables sont le plus souvent des symptômes et on utilisera le chapitre 18 de la CIM-10 pour les « symptômes, signes et résultats anormaux d'examens cliniques et de laboratoire, non classés ailleurs ».

On ne propose pas de grader les événements indésirables car l'information est souvent manquante dans les dossiers médicaux « en vie réelle ».

Gestion des dates

Les antécédents médicaux ou les comorbidités sont notés dans le dossier médical mais la date de début est souvent indiquée de façon imprécise. L'information peut apparaître à n'importe quel moment, dans tout type de compte rendu médical. Si la date est précisée dans le compte-rendu, on utilisera des techniques de traitement du langage qui permettent d'attribuer la bonne date à l'évènement. S'il n'est pas possible d'affecter une date dans le corps du compte-rendu, on propose d'affecter à l'évènement, la date du compte-rendu.

De la même façon, pour les effets indésirables, la précision quant à la date d'apparition est inconstante. Si cela est précisée on note la date de survenue de l'évènement ; dans le cas contraire, c'est la date du compte-rendu qui fera office de date de survenue.

Les dates sont décomposées en année, mois, jour : YYYY, MM, DD

Données unitaires ou données répétitives

Une maladie chronique ne peut être répétitive comme l'insuffisance rénale chronique.

Une comorbidité aigüe peut récidiver et il est possible de présenter plusieurs épisodes d'embolie pulmonaire.

Les effets indésirables sont très souvent répétés dans le temps, surtout s'ils sont liés à une thérapeutique médicale qui se reproduit comme les cycles de chimiothérapie.

Par convention, on propose que les premiers événements morbides codés en CIM-10 soient notés avec une date et que tous les événements morbides à type de symptômes soient répétitifs avec toutes les dates.

Le cancer

La tumeur primitive

Un cancer est défini par :

- Une localisation tumorale (ou code topographique)
- Une anatomo-pathologie (ou morphologie)
- Une latéralité

D'autres données ont été retenues dans le cadre du MDS :

- L'occurrence du cancer, en raison des cancers multiples
- Un lieu de prise en charge
- Une date de diagnostic (anatomo-pathologique ou cytologique)

- Un bilan d'extension initial : TNM et/ou stade
- La taille de la tumeur primitive, la date de la mesure et la méthode de mesure
- Un grade histologique
- Des biomarqueurs prédictifs, pronostiques ou théranostiques (voir le chapitre biomarqueurs)
- Une date de début de traitement (voir le chapitre traitements)

Exemple : Patiente sans antécédent de cancer (occurrence 1), présentant un adénocarcinome galactophorique infiltrant du sein droit mesurant 18 mm avec une biopsie réalisée le 1/02/2022 ; tumeur classé T2N0M0 de stade IIA, de grade SBR II, RO + RP – HER2 - dont le traitement a débuté le 1/03/2022.

Pour les dates de début de la maladie cancéreuse, deux dates sont conservées :

- La date de diagnostic du cancer, le plus souvent histologique, cytologique, plus rarement sur des tests cytogénétiques ou un examen d'imagerie.
- La date de début de premier traitement du cancer (figure dans le chapitre « Traitement »).

La date dite de découverte du cancer (par exemple basée sur les premiers symptômes, un examen clinique, une mammographie de dépistage, un scanner thoracique pour un cancer du poumon) n'est pas conservée dans le MDS car sujet à trop de disparité.

Il est recommandé de ne pas intégrer dans le modèle les tumeurs bénignes (codes D10-D36) et codes morphologiques /0 (sauf pour certaines tumeurs du SNC).

Cependant, on recommande d'intégrer dans le modèle :

- Les tumeurs malignes /3 et les topographies de C00 à C97 (à l'exception de C77 et C78 et C79)
- Les tumeurs à évolution imprévisible /1 er de D37 à D48
- Les cancers « in situ » /2 et de (D00 à D09)

Pour le cas particulier des tumeurs du système nerveux central bénigne, sont conservés les codes D32, D33, D35.2 D35.3 et D35.4

Règles pour la création d'une tumeur

Un cancer est caractérisé par une topographie (sein), une morphologie (adénocarcinome canalaire infiltrant) et une latéralité (pour les organes doubles).

Une des difficultés est de créer un nouveau cancer alors qu'il s'agit d'un événement tumoral d'une tumeur primitive (récidive, métastase) ou l'inverse.

Pour éviter de créer une tumeur à tort, on tient compte des trois critères qui caractérisent un cancer et de l'appartenance ou non au même groupe :

Une tumeur est caractérisée par un groupe topographique (voir en annexe le tableau 2), un groupe morphologique (voir en annexe le tableau 1) ainsi qu'une latéralité pour les organes doubles.

- Si un des trois critères (latéralité, groupe topographique, groupe morphologique) diffère, on crée une autre tumeur
- Sinon, il s'agit de la même tumeur et on crée un évènement tumoral (voir chapitre 9)

Une exception concerne les cancers de l'ovaire, les cancers cutanés, les tumeurs de Wilms et les rétinoblastomes

Pour les organes considérés comme non doubles

La latéralité sera indiquée comme NA, ce qui revient à créer la règle suivante :

- Si au moins un des deux critères (groupe topographique, groupe morphologique) diffère, on crée une autre tumeur.
- Si même groupe topographique et même groupe morphologique, on crée un évènement tumoral.

En annexe

- Les groupes des tumeurs malignes considérées comme histologiquement différents dans le cadre de la définition des tumeurs multiples
(réf : IARC, Lyon, 2004, *Internal Report N02004/02*). Table 1. *Groups of topography codes considered a single site in the definition of multiple cancers*
- Les groupes de codes topographiques considérés comme un seul site pour la définition des tumeurs primitives multiples
(Réf : IARC, Lyon, 2004, *Internal Report N02004/02*). Table 2. *Groups of malignant neoplasms considered to be histologically 'different' for the purpose of defining multiple tumours (adapted from Berg JW. Morphologic classification of human cancer. In: Schottenfeld D & Fraumeni JF Jr. Cancer Epidemiology and Prevention, 2nd edition, Chapter 3 of Section 1: Basic Concepts. Oxford, New York, Oxford University Press, pp. 28-44, 1996).*

Le codage des localisations tumorales et des morphologies

Les cancers sont codés en CIMO-3, en utilisant l'axe topographique et l'axe morphologique.

Pour la topographie de la tumeur (localisation tumorale)

1. Le codage de la localisation anatomique du cancer se fait en CIMO-3 axe topographique
2. Les données structurées (PMSI ou autre) priment sur les données textuelles. Le code PMSI CIM-10 (en dehors des codes utilisés de tumeurs malignes secondaires) est transcodé en CIMO-3
3. En l'absence de données PMSI, on utilise un système de TALN pour coder la topographie de la tumeur en CIMO-3
4. Si le code topographique en CIMO-3 et le code PMSI en CIM10 (en dehors des codes C77, C78, C79) sont manquants ou lorsqu'on ne dispose pas des possibilités de traitement du langage pour repérer le code topographique dans les comptes rendus médicaux, on utilise le code NA.
5. On utilise le code C80.9 (primitif inconnu) uniquement pour les tumeurs dites CUP ou ACUP.

Pour le code morphologique de la tumeur

1. On priorise le code morphologie codé en CIMO-3 dans les « fiches tumeur structurées », puis dans les comptes rendus d'anatomo-cytopathologie, puis dans la fiche RCP (ou tout autre document qui utilise la codification CIMO-3)
2. Si la tumeur est codée en ADICAP, on réalise un transcodage ADICAP en CIMO-3
3. Pour le cas particulier de la CIM-10 qui code la morphologie (exemple : l'hématologie, les mélanomes, les mésothéliomes, le sarcome de Kaposi), on réalise un transcodage de la CIM-10 vers l'axe morphologique de la CIMO-3
4. En l'absence de données structurées, on utilise un système de TALN pour coder la tumeur en CIMO-3 (code morphologique) dans les comptes rendus d'anatomo-pathologie puis dans les fiches RCP (ou autre document, par exemple pour les antécédents de cancer).
5. Si le code morphologique en CIMO-3 et le code ADICAP sont manquants, on ne crée pas de code morphologie et on accepte un code morphologique « NA » par exemple pour les antécédents de cancer ou par les tumeurs cérébrales.

Pour les tumeurs de morphologies multiples, on propose de suivre la règle K de la CIMO-3 (page 41) afin, pour une tumeur donnée, d'avoir une information monovaluée :

- Etape 1 : prendre le code morphologique CIMO-3 s'il existe
- Etape 2 : si plusieurs codes CIMO-3, on privilégie toujours la tumeur maligne /3 plutôt qu'une tumeur bénigne /0 ou borderline /1
- Etape 3 : si les 5ème digit « / » sont identiques (par exemple /3), prendre le nombre le plus élevé pour les 4 premiers digits

Taille de la tumeur primitive au diagnostic

La taille de la tumeur primitive est un facteur pronostic et un élément de la décision thérapeutique.

La taille de la tumeur primitive est consignée dans plusieurs comptes rendus du dossier mais l'information est rarement structurée.

On dispose de plusieurs mesures :

- La taille clinique lorsque la tumeur est accessible à l'inspection ou à la palpation.
- La taille mesurée par les examens d'imagerie.
- La taille mesurée lors de l'examen anatomopathologique lorsqu'on dispose de la pièce opératoire.

La mesure de la taille clinique dépend du type de tumeur ; c'est possible dans le cancer du sein ou les cancers de la peau.

Les mesures objectives réalisés en imagerie dépendent du type d'examen (radiologique, TDM, IRM, échographie, médecine nucléaire).

Le mesure des pathologistes est disponible si le patient est opéré et l'information précise est retrouvée dans le compte rendu anatomo-pathologique.

Une mesure du T (classification TNM) est présente dans le dossier et elle est parfois structurée. La classification TNM classe la taille de la tumeur (le T du TNM) mais ne tient pas compte que de la taille. Par exemple, un cancer du poumon est classé T2 si la tumeur mesure

plus de 3 cm et moins de 5 cm mais elle sera classée T2 quelle que soit la taille si d'autres critères sont présents :

- Un envahissement d'une bronche souche sans envahissement de la carène
- Ou un envahissement de la plèvre viscérale,
- Ou l'existence d'une atélectasie ou pneumonie obstructive.

Les informations recueillies dans le modèle sont :

- La taille de la tumeur et on ne conserve qu'une seule taille
- La date de la mesure
- La méthode de mesure : imagerie, pathologique.

L'information issue de la pathologie (donc la chirurgie) prime sur l'imagerie. En l'absence de chirurgie, seule l'imagerie est prise en compte avec une règle :

Tep > IRM > TDM > Echo > Radio standard (ex : mammographie)

En cas de tumeur multifocale dans le même compte-rendu, on prend la taille de la plus grande avec toujours la même unité (le millimètre)

Les seules mesures qui sont conservées sont les mesures réalisées avant tout traitement du cancer, lors du bilan diagnostic.

Stade et grade des cancers

Le stade du cancer est une évaluation de l'étendue de la maladie selon un système de classification qui peut varier d'un cancer à l'autre. Le stade est directement déduit de la classification TNM avec des modifications selon les versions au cours du temps. Certains cancers ont une stadification particulière car la classification TNM ne permet pas de séparer de façon suffisamment précise le pronostic du cancer.

On peut citer le stade FIGO gynécologique ou le stade de Masaoka-Koga pour les thymomes. Le type de stadification doit être spécifié ; si aucune méthode spécifique n'a été identifiée, on peut coder en NA ou UN en sachant que lorsque l'information n'est pas précisée c'est souvent le stade TNM qui prévaut.

Le grade du cancer (ou grade histologique) est une mesure de la différenciation des cellules et de l'agressivité de la tumeur selon une classification qui varie d'un cancer à l'autre. On peut citer l'indice de Breslow pour le mélanome, l'indice de Gleason pour le cancer de la prostate ou la classification SBR pour le cancer du sein.

De plus en plus souvent, les indices tiennent compte à la fois de l'étendue de la maladie (le stade) et de facteurs pronostiques notamment dans les cancers hématologiques comme la présence d'anomalies cytogénétiques (R-ISS dans le myélome).

Il n'existe aucune liste internationale qui répertorie tous les stades et les grades possibles pour toutes les tumeurs rencontrées. Il est proposé une liste fermée en récupérant toutes les valeurs de tous les référentiels, au fur et à mesure, de l'utilisation de la base. Des exemples figurent dans le tableau ci-joint.

	Stades (extension)	Grades (histologie)
Tumeurs gynécologiques	FIGO	
Mélanomes	TNM	Breslow
Prostate	TNM	Gleason ou ISUP
Neuroblastome	INSS ou INRGSS	
Thymome	Masaoka-Koga	Classification OMS
Sein	TNM	SBR
Néphroblastome	NWTS	

Evènements tumoraux

On appelle « évènement tumoral » les récurrences, métastases, transformations, en rapport avec la tumeur d'origine et qui surviennent après le traitement.

Les évènements tumoraux sont rarement structurés dans les dossiers médicaux ; il convient de bien les définir afin d'assurer une structuration automatique à partir de différents éléments du système d'information.

Même s'il existe des analogies entre tumeurs solides et cancers hématologiques, les termes recouvrent des concepts différents. Si le terme de récurrence peut être utilisé dans les deux cas, d'autres sont plus spécifiques aux cancers hématologiques (transformation, acutisation) et d'autres aux tumeurs solides (métastases).

Pour chaque évènement tumoral, on peut distinguer :

- Le type de l'évènement tumoral (récurrence ou rechute, métastase, transformation)
- La date de 1^{ère} évocation de l'évènement tumoral (clinique, biologique, imagerie)
- Les investigations réalisées pour établir le diagnostic des évènements tumoraux
- La date de diagnostic lorsqu'un prélèvement cytologique ou histologique est réalisé
- Le code anatomo-pathologique (en cas de prélèvement)
- Des biomarqueurs éventuels
- La topographie des métastases et la date de découverte de chaque métastase (qui surviennent à des dates distinctes).

Dans le chapitre « traitement », on disposera également de la date de 1^{er} traitement pour chaque évènement tumoral, lorsqu'un traitement est mis en œuvre dans la grande majorité des cas.

Par convention, on propose de structurer l'information que pour le 1^{er} évènement tumoral du même type, soit la première récurrence ou la première métastase pour chaque cancer en cas de cancers multiples.

La problématique des dates des évènements tumoraux

Un consensus sur une seule date est difficile car pour un cancer ou tout évènement tumoral, (comme toute maladie chronique), cela ne débute jamais à une date donnée contrairement à une maladie aiguë.

Pour chaque évènement tumoral, sauf pour la tumeur primaire, on propose de distinguer trois dates :

- La date de première évocation de l'évènement dans le dossier (clinique, biologie ou imagerie)
- La date de diagnostic histologique en cas de prélèvement
- La première date de traitement de l'évènement (donnée calculée)

Pour le cas particulier des métastases, on cherchera à identifier tous les sites métastatiques survenus au cours de l'évolution de la maladie ainsi que leur date d'apparition.

Les tumeurs solides

Une récurrence de cancer solide signifie que la tumeur réapparaît :

- Au même site tumoral que la tumeur initiale pour les organes non doubles
- Ou au même groupe topographique (voir en annexe le tableau 2) ET au même groupe morphologique (voir en annexe le tableau 1) et du même côté (pour les organes où une latéralité est applicable),
- Ou au niveau ganglionnaire dans le territoire de drainage par rapport à la tumeur primitive.

En d'autres termes, on considère une récurrence locale lorsque la tumeur réapparaît au même site ou par contiguïté à la tumeur primitive (c'est-à-dire faisant partie du même groupe tumoral).

Exemple : récurrence d'un cancer du col utérin (C53.9) étendu au corps utérin (C54). Le risque est de noter un deuxième cancer (corps utérin) alors qu'il s'agit d'une récurrence par contiguïté d'un cancer du col.

En ce qui concerne l'histologie de la récurrence, le code anatomopathologique de la tumeur peut être différent de la tumeur primitive sans pour autant qu'il s'agisse d'un deuxième cancer. Pour cela, on vérifiera que les résultats de l'analyse anatomocytopathologique font partie du même groupe morphologique (au sens de la classification CIMO-3).

Exemple :

Janvier 2010, adénocarcinome canalaire infiltrant du sein droit (8500/3).

Janvier 2015 ; mammographie suspecte. Récurrence locale confirmée avec biopsie positive le 15 janvier 2015 (code 8140/3 pour adénocarcinome).

Une difficulté concerne la récurrence ganglionnaire : une récurrence ganglionnaire dans une chaîne ganglionnaire de drainage de la tumeur primitive est considérée comme une récurrence locale ; lorsque le ganglion se trouve dans un territoire de drainage à distance du cancer (par exemple inguinale pour un cancer du sein), on considère qu'il s'agit d'une métastase. Pour cela, on se base sur les règles utilisées dans la classification TNM.

Exemple : janvier 2020, récurrence ganglionnaire axillaire droite d'un adénocarcinome du sein droit, RO -, RP -, HER2 -

Les règles pour déterminer une récurrence locale ou loco-régionale sont les suivantes :

- Même code topographique (topo) que la tumeur primitive
- ou même groupe topo que la tumeur primitive + même groupe morphologique (morpho) que la tumeur primitive + même latéralité que la tumeur primitive
- ou code topo ganglionnaire + même groupe topo que la tumeur primitive + même latéralité que la tumeur primitive
- ou identifié dans un document par NLP

Les récurrences métastatiques sont définies par :

- Une ou plusieurs localisations tumorales survenant à distance de la tumeur initiale

Dans le cas particulier des métastases, on cherchera à structurer les informations suivantes :

- Le site anatomique de chaque métastase
- La date de découverte de chaque métastase

Exemple :

Janvier 2023, apparition de métastases osseuses et pulmonaires d'un adénocarcinome du sein droit traitées par chimiothérapie.

Janvier 2024, apparition de métastases hépatiques.

La date de survenue de l'évènement métastatique est janvier 2023 avec métastases osseuses et pulmonaires. Survenue d'un autre site métastatique hépatique en janvier 2024.

Les règles pour déterminer une évolution métastatique sont les suivantes :

- Même groupe morpho
- Ou identifié dans un document par NLP

Les cancers hématologiques

La rechute (ou récurrence) est définie par :

- Une maladie hématologique qui est réapparue après traitement
- Après une phase de réponse complète (ou de rémission si la réponse est prolongée)
- La réapparition de signes cliniques, biologiques, à l'imagerie ou aux examens isotopiques.

Exemple :

Janvier 2022, découverte d'un lymphome non hodgkinien traité par chimiothérapie. Septembre 2022, réponse complète et mise en place d'une surveillance.

Janvier 2024, apparition d'adénopathies cervicales ; confirmation de la récurrence par TEP.

La transformation est définie par :

- La progression d'une maladie hématologique indolente vers une forme plus agressive et maligne
- Associée à des modifications cliniques, génétiques ou moléculaires.

On considère que l'acutisation (transformation d'une leucémie chronique en leucémie aigüe) correspond à une transformation.

Exemple : En 2020, découverte insidieuse sur une NFS d'une leucémie myéloïde chronique (LMC) asymptomatique. Confirmation avec découverte du chromosome Philadelphie. En mai 2024, transformation avec apparition de signes cliniques (splénomégalie, pâleur, des ecchymoses, adénopathies) et mise en évidence de blastes en NFS et à la biopsie médullaire le 20/05.2024. Début de traitement le 01/06.2024 par imatinib.

Autres exemples :

- *Transformation d'une leucémie lymphoïde chronique (LLC) en lymphome à grandes cellules (syndrome de Richter)*
- *Transformation de myélofibrose en leucémie aiguë*
- *Transformation d'une thrombocythémie essentielle en leucémie aiguë*

Les rechutes et transformations pour les cancers hématologiques seront identifiées par NLP avec une liste de termes.

Etat général et examen physique (taille, poids, PS)

Ces items sont des mesures survenant à différents moments de l'histoire de la maladie cancéreuse.

Pour chaque item, ce sont des données numériques ou alphanumériques qui se caractérisent par :

- Un type de mesure : poids, taille, performance status
- Une valeur ; poids en kilogrammes, taille en centimètre, performance status (code selon la classification OMS)
- Une date de la mesure décomposée en année, mois, jour (YYYY, MM, DD)

L'unité ou l'échelle de mesure sont définies par défaut selon le type de mesure.

Ces trois items sont tous rattachés au patient, avec un nom de variable (poids, taille, PS), une valeur et une date.

Si la source est non structurée, c'est la date du compte rendu qui est retenue.

Le poids

L'unité de mesure par défaut est le kilogramme (kg)

Chez l'adulte, un nombre entier pourrait suffire mais chez l'enfant, il est nécessaire d'avoir au moins un chiffre après la virgule (XXX.X)

La recommandation est de conserver toutes les valeurs avec les dates

La taille

L'unité de mesure par défaut est le centimètre (XXX)

Chez l'adulte, une seule valeur est recueillie avec une date

Chez l'enfant, l'adolescent et le jeune adulte (âge < 25 ans), toutes les valeurs sont conservées.

Le performance status (PS)

La classification utilisée chez l'adulte est l'échelle de Zubrod adoptée par l'ECOG et l'OMS avec une échelle de 0 à 4.

Si l'information est codée en Index de Karnofsky (IK) de 0 à 100, on propose un transcodage en utilisant la table de correspondance entre PS et IK.

Si un médecin code un PS avec une valeur ambiguë (exemple 1-2), on conserve la valeur la plus « péjorative » c'est-à-dire le chiffre le plus grand.

Les traitements du cancer

Les traitements sont classés en trois types :

- La chirurgie et la radiologie interventionnelle
- Les traitements médicaux ou systémiques à visée anti-cancéreuse (chimiothérapie, immunothérapie, thérapies ciblées, hormonothérapies, thérapie cellulaire)
- La radiothérapie (y compris la curiethérapie, la protonthérapie et les autres traitements utilisant les radiations ionisantes).

Les traitements chirurgicaux et les traitements médicaux sont décrits avec un nombre limité d'items. Par contre, la radiothérapie est décrite de façon très précise dans le modèle OSIRIS radiothérapie (OSIRIS – RT). Un lien logique est réalisé entre la radiothérapie décrite dans le modèle OSIRIS RWD clinique et le modèle OSIRIS RT.

Un traitement est défini par :

- Un type de traitement (chirurgie et radiologie interventionnelle, traitement médical, radiothérapie)
- Une date de traitement
- Une précision concernant le type de traitement (type d'acte chirurgical, molécules utilisées, technique de radiothérapie, etc.)
- Un lieu de réalisation de l'acte en différenciant un acte réalisé dans l'établissement d'un acte effectué dans une autre structure hospitalière.
- La source de la donnée (à chaque date) selon que l'information est structurée à la source, récupérée par TALN à partir d'un document de l'établissement ou récupérée d'un document extérieur.

Lorsque le traitement est réalisé dans l'établissement, l'information est structurée et est présente dans la source PMSI (hospitalisation, actes et séances) avec un code CCAM ; lorsque le traitement est réalisé en dehors, l'information peut être structurée par TALN à partir d'un compte rendu médical.

Un traitement est toujours lié au patient ; l'information sur les traitements est alimentée dès lors que les informations sur la tumeur sont présentes dans la base.

Précisions concernant les dates de traitements

Toutes les dates de tous les traitements sont recueillies.

Dans le MDS, la date du traitement ne fait pas de différence entre date de prescription et date d'administration ; en d'autres termes, on ne propose pas de récupérer les deux dates.

On privilégie les dates d'administration lorsque celles-ci sont connues (par exemple pour les traitements injectables) aux dates de prescription.

Lorsque la date d'administration n'est pas connue, par exemple lorsque le traitement est réalisé en dehors de l'établissement ou pour les traitements oraux, on récupère la date de prescription.

Les dates de début de traitement et de fin de traitement ne sont pas retenues dans le minimum data set mais sont calculées lors de l'analyse.

Chirurgie et imagerie interventionnelle

Utilisation de la CCAM dans OSIRIS RWD

La Classification Commune des Actes Médicaux (CCAM) est une classification d'actes nationale et généraliste des actes médicaux. Elle est maintenue à des fins de facturation et de suivi d'activité. Elle est très précise et regroupe plusieurs milliers d'actes répartis dans 19 catégories topographiques.

Cette classification englobe tous les actes de chirurgie et d'imagerie interventionnelle. C'est la raison pour laquelle les deux types d'acte sont regroupés dans une même table.

OSIRIS RWD a pour objectif de produire un modèle de données alimenté par des données collectées automatiquement à partir du dossier patient. Lorsque la chirurgie ou la radiologie interventionnelle sont réalisées dans l'établissement, les données sont structurées à la source en CCAM, avec des règles communes à toutes les structures de soins. De plus, des contrôles qualité sont réalisés dans les établissements, augmentant le niveau de confiance de ces données structurées à la source.

Lorsque la chirurgie ou la radiologie interventionnelle sont réalisées en dehors de l'établissement, la problématique est de ne pas disposer d'un codage à la source. Les informations présentes dans les comptes rendus ne permettent pas de structurer les actes avec le niveau de précision de la CCAM. A titre d'exemple, une « Pneumonectomie » peut être codée selon 8 codes différents :

- Pneumonectomie, par thoracotomie avec préparation par thoracoscopie
- Pneumonectomie, par thoracotomie
- Pneumonectomie avec résection-anastomose de la bifurcation trachéale, par thoracotomie
- Pneumonectomie avec exérèse totale de la plèvre [Pleuro-pneumonectomie], par thoracotomie
- Pneumonectomie avec résection de la paroi thoracique, par thoracotomie

- Pneumonectomie avec résection-anastomose de la veine cave supérieure ou résection de l'oreillette gauche, par thoracotomie
- Pneumonectomie avec résection et remplacement prothétique de la veine cave supérieure, par thoracotomie
- Pneumonectomie avec résection d'organe et/ou de structure de voisinage, par thoracotomie.

Cependant, il est possible de repérer à partir des textes l'acte de « pneumonectomie ». On propose d'utiliser le code de la hiérarchie du chapitre (soit dans notre exemple : 06.03.04.00 – Pneumonectomie).

Proposition pour OSIRIS RWD

Pour les actes réalisés en interne, on conserve le code CCAM le plus précis (par exemple *Pneumonectomie par thoracotomie avec préparation par thoracoscopie* codé en GFFA012) et on ajoute un code avec un transcodage ascendant, c'est-à-dire vers le code de la hiérarchie (06.03.04.00 – Pneumonectomie).

Lorsque l'acte est réalisé en dehors de l'établissement, il est difficile de trouver automatiquement le code CCAM à partir d'un texte. Il est plus facile de structurer l'information à partir du texte en utilisant un code moins précis (pneumonectomie dans notre exemple). Pour les actes réalisés à l'extérieur, on propose de structurer l'information vers le code de la hiérarchie (06.03.04.00 – Pneumonectomie).

Avantages

- On utilise la même classification et le même code pour les actes réalisés en interne et en externe.
- On utilise la même classification et les mêmes règles pour la chirurgie et la radiologie interventionnelle.
- On conserve le maximum de précision pour les actes réalisés en interne sans demander le même niveau de précision pour les actes réalisés à l'extérieur.
- On réalise un regroupement des actes réalisés en interne vers le code de la hiérarchie qui est parfois suffisant dans le cadre des études cliniques.
- Si un transcodage doit être proposé pour permettre une interopérabilité au niveau international (en SNOMED), cela semble plus atteignable en partant des regroupements que des codes CCAM précis.

Les traitements médicaux du cancer

Les traitements médicaux du cancer comportent l'ensemble des traitements à visée anti-cancéreuse administrés par voie systémique, par voie orale ou injectable. En pratique, ce sont des molécules qui sont regroupées dans de grandes classes pharmacologiques : chimiothérapie anti-cancéreuse, immunothérapie, thérapies ciblées, hormonothérapies, vaccins. Ne sont pas pris en compte dans le MDS les thérapies de support.

Trois sources peuvent être utilisées :

- Les logiciels de prescription et d'administration des traitements médicaux avec une information structurée
- Les comptes rendus médicaux de l'établissement avec une information en texte libre notamment pour les thérapeutiques orales ou les chimiothérapies systémiques administrées en dehors de l'établissement.
- Les documents récupérés de l'extérieur de l'établissement et scannés dans le dossier.

La source de données à privilégier est celle issue des outils de prescription de chimiothérapie (comme le logiciel CHIMIO®). On peut ainsi disposer de toutes les molécules administrées avec les dates d'administration : molécules en DCI en ne tenant compte que des médicaments administrés (et non des médicaments prescrits ou prévus). La source CHIMIO® dispose uniquement des molécules injectées ou qui nécessitent une reconstitution. Ceci ne concerne que les patients traités dans l'établissement par injection uniquement c'est-à-dire la chimiothérapie systémique, l'immunothérapie, les anticorps monoclonaux.

Par ailleurs, l'information sur les médicaments prescrits dans les outils de prescription est également présente dans les comptes rendus médicaux, ce qui pourrait occasionner des doublons. La recommandation est de privilégier les données structurées issues des logiciels de prescription, en tenant compte de la date. Afin d'éviter les doublons, il est proposé de dédoublonner lors de l'alimentation quand l'information sur un médicament est identique dans un délai déterminé. En cas d'impossibilité de dédoublonner à la source, l'information sera dédoublonner à l'analyse.

Certains traitements ne se trouvent pas dans la source structurée et il est nécessaire de reconstituer le traitement pour structurer l'information à partir des dossiers médicaux. Cela concerne les traitements oraux (hormonothérapie, chimiothérapie orale, thérapie ciblée) et les patients non traités dans l'établissement. Afin de les distinguer, et pour déterminer une mesure de la qualité des données, les informations sur la source de la donnée et le lieu d'administration sont recueillies.

Un changement de traitement désigne, qu'à partir d'une certaine date, un ou plusieurs médicaments ont été modifiés ou substitués. La variable changement de traitement et le calcul du nombre de lignes sont des variables calculées au moment de l'analyse, avec des règles communes établies par consensus. Les règles et recommandations figurent en annexe.

Les variables « début de traitement » et « fin de traitement » ne sont pas des variables stockées dans le MDS. Elles sont calculées lors de l'analyse.

Afin de calculer les changements de traitement et les numéros des lignes de traitement, il est nécessaire de reconstituer le traitement avec toutes les molécules administrées et toutes les dates pour structurer l'information. Autant, il est possible de disposer de toute l'information lorsque le patient est sous traitement injectable réalisé dans l'établissement, autant cette information est plus difficile à structurer pour les traitements par voie orale ou les traitements systémiques réalisés en dehors de l'établissement. Pour ceux-là, il est conservé la première date de prescription.

Réponse, résistance ou progression

La réponse à un traitement est rarement structurée dans les dossiers médicaux et il est difficile de la structurer par TALN pour plusieurs raisons :

- Le type de réponse n'est pas toujours spécifié : réponse complète, clinique, radiologique, histologique, métabolique, réponse partielle, stabilisation, meilleure réponse, etc.
- Les critères RECIST, à la base des études cliniques, sont rarement retrouvés dans les dossiers médicaux en situation de « vie réelle ».
- A la différence des essais cliniques, les évaluations sont réalisées à des moments variables, difficiles à définir avec des règles communes.
- Les mesures de la taille de la tumeur et de la taille des métastases ne permettent pas de structurer une progression. Par exemple, la survenue d'une nouvelle lésion signifie une progression même si la taille de la tumeur a diminué.

Il apparaît donc plus simple de structurer la « progression » ou la « résistance » aux traitements avec une date que la réponse. On notera aussi que c'est la première progression qui est fondamentale pour les projets de recherche, progression évaluée après chaque changement de traitement. Dans le modèle, nous nous intéresserons qu'à la situation métastatique.

Le terme "progression" désigne l'évolution d'un cancer lorsque la maladie évolue défavorablement malgré le traitement. Cela signifie que la tumeur augmente de volume, que de nouvelles lésions apparaissent ou que le cancer se développe avec l'apparition de métastases. Dans cette situation, il est décidé un changement de stratégie thérapeutique ou un changement de traitement (chimiothérapie, immunothérapie, radiothérapie, etc.). Le terme « progression » est préféré au terme « résistance ».

Dans la mesure où l'information est structurée par TALN, on cherchera de nombreux termes qui indique la notion de "progression" ou de « résistance au traitement » : dissémination (si le cancer se propage), extension tumorale, croissance tumorale, reprise évolutive (notamment après une période de rémission), progression métastatique, échappement tumoral, etc.

A noter qu'il faut différencier les cancers solides des cancers hématologiques où on parle plus souvent de « maladie réfractaire ». Un cancer réfractaire est défini par une maladie qui n'a pas répondu au traitement sans possibilité de définir précisément la date de début. Cette situation aboutit fréquemment à un changement de traitement. A l'inverse, une tumeur solide « stable » sous traitement ne conduit pas nécessairement à un changement de traitement.

Les principes pour structurer la progression dans OSIRIS RWD

Lors d'une progression, dans la quasi-totalité des situations, cela conduit à un changement de traitement ou à un arrêt définitif des traitements. L'information sur le changement de traitement peut être structurée de façon automatique avec des règles qui peuvent être partagées. Ceci concerne autant les chimiothérapies que les thérapies ciblées, l'hormonothérapie ou l'immunothérapie.

Le terme « changement de traitement » a été préféré à « changement de ligne ». Certains protocoles alternent des associations de différentes molécules, ce qui peut compliquer la structuration des lignes en automatique. Le calcul de changement de lignes pouvant être fait

au moment de l'analyse selon le projet, les localisations tumorales et les protocoles utilisés (voir en annexe).

Le principe de la structuration de la progression (ou de la résistance à un traitement), est donc de repérer tous les changements de traitement et la cause par une analyse des textes précédents ce changement, une information signifiant la progression (item progression) et l'item sur les événements indésirables.

Un événement indésirable peut nécessiter un changement de traitement et cela ne doit pas être confondue avec une progression. Il est donc nécessaire de structurer les événements indésirables les plus sévères pour ne pas confondre, changement de ligne par progression avec un changement de ligne pour événement indésirable ou toxicité. On ne parle ici que des traitements du cancer et pas des thérapies de support de type corticothérapie.

En d'autres termes, cette stratégie de détection de la progression grâce à la structuration des traitements doit être couplée avec la détection par TALN de la progression et des toxicités si ces deux données ne sont pas déjà structurées à la source.

Nous proposons donc de structurer deux items (item progression et la date de la progression ou de la maladie réfractaire), de tenir compte de l'item « événement indésirable ou toxicité » et de calculer le changement de traitement lors de l'analyse. Ainsi, on pourra définir la progression selon les règles suivantes :

- A : changement de traitement, structuré de façon automatique par des règles qui figurent en annexe
- B : progression, structurée par TALN,
- C : Les effets indésirables (ou toxicité), structurés par TALN.



A + B – C = Progression

A + B = Progression

A + B + C = Progression

A + C = Pas de progression

Ces différents items sont à structurer avec une date qui leur sont attachés :

- Le changement de traitement (A) avec la date du changement qui est la date de début du traitement suivant.

- La progression (B) repérée par TALN dans les CR médicaux ou d'imagerie avec une date.
- La toxicité (C) détectée par NLP avec une date (et un délai court juste avant le changement de traitement + une borne).

La date de fin de chaque traitement est difficile à repérer ; les dates de changement de traitement correspondent au début du traitement suivant.

On propose des règles de définition de la progression comme indiqué ci-dessus : $A + B$ ou $A + B - C$ ou $A + B + C$ (au même moment).

Le but est de ne pas passer à côté d'une progression avec le risque d'accepter des faux positifs (changement de traitement pour une autre raison que la progression considérée à tort comme une progression). Ces faux positifs seront à discuter au moment de l'analyse en fonction des projets.

Le grade des événements indésirables n'est pas structuré dans le cadre du projet OSIRIS ; la notion de gravité ne pourra pas être retenue pour identifier de potentiels changements de traitement liés à une toxicité.

Les changements de ligne à partir des changements de traitement et des autres informations disponibles (progression en NLP et toxicité) seront structurées au moment de l'analyse.

Cas particulier des cancers réfractaires en hématologie

En hématologie, les traitements conventionnels permettent d'obtenir un taux de réponse supérieur aux tumeurs solides. Alors que dans les tumeurs solides, une stabilité est parfois considérée comme un contrôle de la maladie, la stabilité sous traitement, et a fortiori une progression, fait référence à une « maladie réfractaire » qui ne répond pas ou répond mal aux traitements standards.

Une maladie réfractaire en hématologie se définit ainsi : maladie qui persiste ou progresse malgré l'administration de traitements appropriés. Elle peut être initialement réfractaire (ne répondant pas dès le début du traitement) ou devenir réfractaire après une période de réponse initiale.

Les critères d'une maladie réfractaire sont :

- Une maladie qui n'a pas répondu au traitement
- Après un traitement de 1^{er} ligne (ou de n^{ième} ligne)
- Avec la persistance de signes cliniques
- Ou la persistance de signes biologiques
- Ou d'anomalies à l'imagerie ou aux examens isotopiques (TEPscan)
- Sans possibilité de définir précisément la date de début de la phase réfractaire dans la mesure où la maladie évolue sous traitement.

Exemple : janvier 2023, découverte d'une leucémie aiguë myéloïde (LAM) traitée par plusieurs cycles de chimiothérapie intensive ; persistance des signes biologiques après 6 cycles ; devant le caractère réfractaire, proposition d'un essai clinique de phase 2.

Imagerie et médecine nucléaire

Plusieurs sources de données sont possibles :

- Les données issues des PACS
- Les comptes rendus d'imagerie issus de l'établissement
- Les comptes rendus issus de l'extérieur de l'établissement

Les données issues des PACS et le standard DICOM

Le PACS (Picture archiving and communication system) est un serveur d'images qui permet de stocker les images DICOM (Digital imaging and communication in medicine) dans une base de données. A partir du PACS, il est possible d'extraire, outre les images, des métadonnées au format DICOM correspondant aux images.

Le standard DICOM désigne à la fois un format de fichier et un protocole de transmission des données. Un fichier DICOM est composé de :

- L'image numérisée dans un champ particulier
- Une suite de champs définissant ses métadonnées (ID du patient, type d'images, date, etc.).

Les métadonnées sont organisées de manière hiérarchique selon le modèle suivant :

1. **Patient**
2. **Étude** : correspond à un examen pour un patient (par exemple, une IRM du cerveau réalisée le 27 janvier 2025).
3. **Série** : correspond à un sous-ensemble de données ou d'images issues de l'étude (par exemple, une séquence T1 ou T2 en IRM).
4. **Instance** : correspond à un fichier d'image individuel ou à un objet DICOM unique (par exemple, une coupe d'image IRM).

Le problème est que le standard DICOM est très permissif et adaptable au besoin de chaque établissement, par exemple pour nommer les études et les séries.

Description minimale d'un examen d'imagerie

Deux informations sont indispensables pour définir un examen d'imagerie :

- La modalité
- La date et l'heure de l'examen.

Une « modalité » fait référence aux différentes techniques pour capturer des images.

Exemples de codes DICOM des différentes modalités d'imagerie :

- CR : Radiographie numérique (Computed Radiography)
- CT : Tomodensitométrie (Scanner, Computed Tomography)
- MR : Imagerie par résonance magnétique (Magnetic Resonance)

- NM : Médecine nucléaire (Nuclear Medicine)
- US : Ultrasonographie (Ultrasound)
- XA : Angiographie par rayons X (X-Ray Angiography)
- MG : Mammographie (Mammography)
- PET : Tomographie par émission de positons (Positron Emission Tomography)

Au cours d'une même journée, un patient peut bénéficier de plusieurs examens ; pour les différencier, outre la modalité, il est nécessaire de préciser la date (YYYY – MM – DD) et l'heure d'acquisition (HH – MM).

PMSI et imagerie médicale

Quand les examens d'imagerie médicale sont réalisés pendant une hospitalisation, le code CCAM de l'examen est retrouvé dans le résumé standardisé de sortie mais de façon très inconstante ; cette donnée n'est pas nécessaire à visée de facturation car l'acte n'est pas classant.

Quand l'examen d'imagerie est réalisé en dehors d'un séjour, le code CCAM s'applique mais l'information n'est conservée que dans le logiciel de facturation, source rarement utilisée par les établissements pour alimenter OSIRIS RWD.

Il n'est donc pas recommandé d'utiliser le PMSI ou la CCAM pour structurer les informations concernant l'imagerie médicale.

Les prélèvements biologiques

Ce chapitre décrit les informations nécessaires pour utiliser les échantillons issus des prélèvements biologiques dans des projets de recherche.

Prélèvements et échantillons biologiques

Il est nécessaire de différencier prélèvements et échantillons biologiques.

Un prélèvement est réalisé à une date précise avec une topographie connue. Le prélèvement peut être du tissu tumoral, du tissu sain, du sang, de l'urine. Les prélèvements de tissu sont obtenus lors d'un acte chirurgical ou d'une biopsie.

Un échantillon biologique est un élément dérivé d'un prélèvement et il est possible d'extraire de nombreux échantillons à partir d'un seul prélèvement. Les échantillons sont extraits à des dates variables, pas toujours connues, dans différents laboratoires. L'échantillon peut être du tissu fixé, du tissu congelé, du plasma, de l'ADN, de l'ARN, etc.

Il existe donc deux niveaux : les prélèvements et les échantillons biologiques. Il est proposé, par convention, que le niveau prélèvement soit décrit dans le modèle clinique OSIRIS RWD et que le niveau échantillon soit décrit dans le modèle OSIRIS omique. Un lien logique entre les deux modèles est proposé.

Dans le cadre de projets utilisant des données de « vie réelle », il est utile de disposer des informations sur les prélèvements et les échantillons, savoir si les prélèvements sont

conservés dans l'établissement et si de nouvelles analyses sont possibles. L'information concernant la disponibilité du prélèvement n'est pas retenue dans le modèle.

Exemples de questions posées fréquemment :

- Faisabilité d'une étude clinique intégrant des données biologiques et des données omiques
- Lister les patients présentant une tumeur avec du tissu congelé en tumorothèque
- Permettre un lien entre les données de la tumorothèque et des données cliniques
- Disponibilité de données cliniques, d'images et de prélèvements pour une étude.

Les biomarqueurs et les marqueurs tumoraux

Les biomarqueurs en cancérologie se sont développés avec l'utilisation de l'immunohistochimie (IHC) et le séquençage des gènes. Ils permettent de dépister un cancer, mieux caractériser une tumeur, suivre son évolution, guider le traitement, être un outil prédictif ou pronostic. Les marqueurs sériques sont basés sur des techniques biochimiques. Leur utilisation en pratique clinique a diminué au fur de temps sauf dans certains cancers, comme les cancers de la prostate (PSA) ou les tumeurs germinales.

Nous proposons d'adopter une convention pour structurer les différents biomarqueurs car ils sont de nature très différente. Ainsi, on considère trois grands types :

- Les biomarqueurs tissulaires et moléculaires tumoraux
- Les biomarqueurs moléculaires constitutionnels
- Les marqueurs tumoraux (ou marqueurs sériques).

Les biomarqueurs tissulaires et moléculaires tumoraux

Les biomarqueurs tissulaires et moléculaires sont d'une grande valeur informative en raison de leur caractère diagnostique, théranostique ou pronostic. Pour détecter une mutation (Ex : une mutation de EGFR) ou une amplification d'un gène (Ex : gène MET), plusieurs techniques peuvent être employées et ces techniques évoluent au cours du temps (par exemple, passer d'une technique de séquençage à une technique d'IHC).

Il n'existe pas une seule manière de les quantifier ; par exemple, pour KI67 et PDL1, on mesure un pourcentage, pour l'hyper-expression de HER2 en IHC, on attend une classification en +, ++, +++, ou parfois l'information, positif (+) ou négatif (-). Lorsque l'information disponible est une valeur réelle (par exemple, le pourcentage de cellules RO +), elle devrait être conservée telle quelle car l'interprétation peut changer en fonction du seuil.

La nomenclature utilisée est LOINC. Néanmoins, comme il n'existe pas un code unique pour un biomarqueur donné et des choix sont nécessaires. Il faudra aussi assurer la maintenance d'une liste de réponse et de code LOINC déterminée et utiliser une liste de synonyme (Exemple : HER2 / *c-erb B2*).

Le référentiel doit permettre un niveau de détail conforme à ce qui est observé dans les soins courants. Ainsi, il est utile de coder les variants ou les exons dans certains cas, notamment ceux concernant les gènes mutés lorsque cette information est disponible cliniquement dans

les dossiers médicaux. A titre d'exemple, il est utile de récupérer les variants cliniquement pertinents dans la pratique, tels que EGFR Exon 18, BRAF V600E ou KRAS G12C.

Il est important de présenter les résultats avec les items suivants :

- Résultat positif (avec éventuellement une précision ou une valeur numérique)
- Résultat négatif
- Résultat non contributif.
- Information manquante

L'information manquante (car non retrouvée dans le dossier médical) ou si l'examen est non réalisé sont considérés comme identique.

En pratique,

- Pour une tumeur, plusieurs biomarqueurs tissulaires ou moléculaires sont possibles
- Prélevés sur la tumeur primaire ou lors d'un événement tumoral,
- Avec une évolution possible lors des différents événements tumoraux ce qui nécessite de connaître la date
- Détectés par l'anatomopathologie (immunohistochimie), FISH, analyse génomique par séquençage de l'ADN (exome ou panel de gènes)
- Avec des résultats exprimés différemment selon les biomarqueurs.
Exemples : Hyper-expression de Her2, RH+ (RO/RP), translocation de ALK, mutation de EGFR (détecté par IHC ou séquençage), mutation de KRAS, mutation de BRAF, réarrangements du gène RET, BRCA tumoral

Les biomarqueurs moléculaires constitutionnels

Un biomarqueur constitutionnel, également appelé biomarqueur génétique, fait référence à une caractéristique biologique intrinsèque à la personne, présente dès la naissance, généralement héritée et associé à un risque accru de développer un cancer.

En pratique,

- Un seul biomarqueur par patient
- Rattaché au patient
- Pas d'évolution au cours du temps
- Peut être mesuré dans divers échantillons biologiques (sang, salive)
- Prédispositions génétiques à certaines tumeurs (sein, ovaire)
- Exemples : mutation de BRCA1, mutation de p53

Les marqueurs tumoraux (ou marqueurs sériques)

Les marqueurs tumoraux sont des substances circulantes, souvent des protéines ou des hormones, qui sont produites par les cellules cancéreuses (ou parfois par l'organisme en réponse à la croissance d'un cancer). Ces substances sont dosées dans le sang, le LCR, la plèvre. Les dates de résultats des marqueurs tumoraux sont essentielles car elles permettent un suivi de l'évolution dans le temps.

Les marqueurs tumoraux sont principalement utilisés pour surveiller l'efficacité des traitements et ne possèdent pas de valeur diagnostique (à l'exception de marqueurs tels que le PSA ou les marqueurs sériques des tumeurs germinales).

Les principaux marqueurs tumoraux sont : PSA, ACE, CA 12-5, CA 19-9, CA 15-3, SCC, β HCG (Hormone Chorionique Gonadotrope), α -foeto protéine (AFP), Cyfra 21, Thyrocalcitonine, NSE.

En pratique,

- Plusieurs marqueurs tumoraux possibles pour un patient
- Attachés au patient et à une date
- Mesurés sur des liquides par des techniques biologiques (par prélèvement sanguin, parfois prélevés dans un épanchement, pleural ou dans le LCR)

Données complémentaires au MDS

Il est créé deux tables pour des données complémentaires non prévues dans le MDS : la table « observation » et la table « mesure ».

Ces deux tables sont alimentées à partir de données sources selon le principe d'une paire « attribut – valeur » à laquelle on ajoute l'identifiant du patient (la clé). Les mesures sont stockées sous forme « clé – attribut – valeur », l'attribut étant le concept de mesure et la valeur représentant le résultat.

Pour certaines données, il sera utile de disposer du meilleur rattachement possible de la paire « attribut – valeur », soit au patient, à la tumeur, à l'évolution tumorale ou au traitement.

La table observation

La table « observation » contient des informations cliniques attachées au patient, obtenues dans le cadre d'un interrogatoire, d'un examen ou d'une procédure.

La table « observation » est un espace flexible pour capturer des informations médicales qui ne sont pas dans le MDS et dans les autres tables du modèle.

La table « observation » est utilisée pour capturer des observations qualitatives structurées.

La table Mesure

La table « Mesure » est utilisée pour des valeurs quantitatives et standardisées (ex : durée du tabagisme, taux de glucose, dosage de la créatininémie).

La table « Mesure » contient des valeurs numériques obtenues par un interrogatoire, un examen ou des tests biologiques.

La table « Mesure » contient à la fois les attributs et les résultats de mesures telles que les tests de laboratoire, les signes vitaux, des résultats quantitatifs, etc.

LE MINIMUM DATA SET

1. Identification du patient

Les informations du MDS qui définissent un patient dans la base sont :

- Le numéro du patient dans la base OSIRIS RWD
- L'identifiant de l'établissement qui alimente la base
- L'année et le mois de naissance
- Le sexe biologique à la naissance

1.1 Numéro du patient dans la base OSIRIS RWD

Nom de l'item	Numéro du patient dans la base
Numéro de l'item	1.1
Nom de la variable	PatientId
Définition	Numéro pseudonymisé unique du patient dans la base OSIRIS RWD
Valeur attendue	Un « numéro » pseudonymisé constitué d'une suite de caractères alphanumérique
Règle	Le numéro pseudonymisé est unique pour un patient. Il ne doit contenir aucune donnée identifiante ou indirectement identifiante. Il ne doit pas comporter d'information relative au patient, par exemple ses initiales ou une date.
Justification du choix de la règle	L'identifiant permanent du patient dans un établissement ne peut être utilisé comme identifiant.
Caractère obligatoire ou optionnel	Obligatoire
Recommandations	Il est recommandé de créer une table de correspondance entre les pseudonymes dans la base de données et les données directement identifiantes.

1.2. Identifiant de l'établissement alimentant la base OSIRIS RWD

Nom de l'item	Identifiant de l'établissement alimentant la base OSIRIS RWD
Numéro de l'item	1.2
Nom de la variable	PatientOriginCenterId
Définition	Code FINESS géographique de l'établissement fournisseur des données
Nomenclature utilisée	<p>Le code FINESS (Fichier National des Etablissements Sanitaires et Sociaux) est le répertoire national de référence des personnes morales intervenant dans les domaines sanitaires, médico-social et social.</p> <p>Chaque entité inscrite au répertoire FINESS dispose d'un numéro d'identité unique et définitif composé de 9 caractères, les deux premiers chiffres désignant le département d'implantation.</p> <p>Dans le cadre des études multicentriques, cette information permet de connaître l'établissement à l'origine des données.</p> <p>C'est une donnée fixe pour un établissement qui est enregistré comme paramètre dans une table paramètre.</p>
Valeur attendue	Chaine de caractères
Règle	Le code FINESS utilisé est le code FINESS EG c'est-à-dire celui de l'établissement qui fournit la donnée en d'autres termes, la structure qui alimente la base OSIRIS RWD.
Justification du choix de la règle	Historiquement dans OSIRIS, il y avait deux codes FINESS, l'établissement fournisseur de la donnée et l'établissement qui réalise le soin. Il est décidé de ne garder que l'établissement fournisseur de la donnée car le code FINESS de la structure qui effectue le soin n'est pas toujours connu.
Obligatoire ou optionnel	Obligatoire
Recommandations	Ce code est saisi une seule fois au moment de la création de la base OSIRIS RWD. Ce code FINESS est utile dans les études multicentriques qui consolident les informations venant de plusieurs établissements de santé.

1.3 Date de naissance

Nom de l'item	Date de naissance
Numéro de l'item	1.3
Nom de la variable	PatientBirthDate

Définition	Date de naissance de la personne
Valeur attendue	YYYY - MM - DD
Règle	La source est le dossier administratif du patient.
Caractère obligatoire ou optionnel	Obligatoire

1.4 Sexe du patient

Nom de l'item	Sexe du patient
Numéro de l'item	1.4
Nom de la variable	PatientGender
Définition	Sexe biologique du (de la) patient(e)
Valeur attendue	M = sexe masculin F = sexe féminin
Règle	La source est le dossier administratif du patient
Justification du choix de la règle	Par convention, les valeurs M ou F sont préférées aux valeurs 1 ou 2.
Caractère obligatoire ou optionnel	Obligatoire

2 - Information et non-opposition du patient pour la réutilisation de ses données de santé

Les informations à recueillir sont de trois types :

- Que la personne a été informée de la réutilisation de ses données et à quelle date
- Que la personne ne s'est pas opposée à la réutilisation de ses données
- Et si la personne s'y est opposée, à quelle date.

2.1 Information du patient

Nom de l'item	Information du patient
Numéro de l'item	2.1
Nom de la variable	Information

Définition	Le patient a été informé de la réutilisation de ses données pour la recherche.
Valeur attendue	Oui/Non
Règle	Cette information doit être tracée et consignée dans le dossier du patient de l'établissement
Justification du choix de la règle (exemple)	Les personnes doivent être informées par le responsable de traitement que les données collectées lors de leur prise en charge sont versées au sein de la base de données.
Caractère obligatoire ou optionnel	Optionnel
Recommandations	<p>L'établissement a informé le patient de la réutilisation de ses données pour la recherche.</p> <p>Il est de la responsabilité de chaque établissement de tracer cette information et de la mettre à disposition.</p>
Options	<p>Le responsable de traitement peut faire valoir une exception à l'obligation d'information individuelle pour la constitution de la base de données s'il justifie dans son registre d'activité de traitement que la fourniture des informations exigerait des efforts disproportionnés.</p> <p>À ce titre, peuvent notamment être invoqués, au vu de sa situation : le nombre de personnes concernées, l'ancienneté des données, le coût et le temps de la délivrance des informations.</p>

2.2 Date d'information du patient

Nom de l'item	Date d'information
Numéro de l'item	2.2
Nom de la variable	InformationDate
Définition	Date d'information de la personne de l'enregistrement de ses données dans la base de données.
Règle	Aucune règle ne sera appliquée. Si une valeur est non disponible sur le jour (DD), le mois (MM) ou l'année (YYYY). Le champ correspondant sera laissé sans valeur.
Valeur attendue	YYYY - MM – DD
Justification du choix de la règle (exemple)	Cf en Annexe Référentiel CNIL – 8. Information des personnes
Caractère obligatoire ou	Optionnel

optionnel	
Recommandations	<p>Cette information n'est pas toujours recueillie dans les hôpitaux.</p> <p>Cette information sur la date d'information est optionnelle car secondaire par rapport au fait que le patient soit informé.</p> <p>Il se peut que l'information de la création d'un EDS par l'établissement ne se fasse pas de manière individuelle. Dans ce cas, il s'agit d'une information collective par une campagne de communication globale (livret d'accueil, site web par exemple) et la date de l'information n'est pas forcément pertinente dans ce cas.</p>

2.3 Non-opposition pour l'utilisation des données

Nom de l'item	Non-opposition pour l'utilisation des données
Numéro de l'item	2.3
Nom de la variable	NonOpposition
Définition	La personne ne s'oppose à la réutilisation de ses données de santé.
Valeur(s) attendue(s)	Oui (ne s'oppose pas) /Non
Règle	Par défaut à l'information du patient, on enregistre « Oui » si ce dernier ne manifeste pas son droit d'opposition.
Caractère obligatoire ou optionnel	Obligatoire
Recommandations	<p>L'information est souvent présente dans le système d'information des hôpitaux.</p> <p>Il est de la responsabilité de chaque établissement de tracer cette information.</p> <p>Il n'est pas nécessaire d'un consentement spécifique pour les données génétiques ; le consentement pour les données génétiques et les données cliniques sont réunis dans un seul item.</p> <p>Si les données d'un patient sont présentes dans la base de données et que le patient décide de s'opposer à la réutilisation de ces données. C'est de la responsabilité de l'établissement de tracer l'information puis de supprimer les données du patient concerné.</p>
Options	Si la personne s'oppose à la réutilisation de ses données, le patient peut ne pas figurer dans l'EDS et l'alimentation des données de cette personne sont bloquées à la source de la base.

2.4 Date d'opposition pour l'utilisation de ses données

Nom de l'item	Date d'opposition pour la réutilisation de ses données
Numéro de l'item	2.4
Nom de la variable	NonOppositionDate
Définition	Date de signature de la demande d'opposition du (de la) patient(e) pour l'utilisation de ses données cliniques et génétiques
Valeur attendue	YYYY - MM – DD
Règle	Si une valeur est non disponible sur le jour (DD), le mois (MM) ou l'année (YYYY). Le champ correspondant sera laissé sans valeur.
Justification du choix de la règle (exemple)	Cf en Annexe Référentiel CNIL – 9. Droits des personnes
Caractère obligatoire ou optionnel	Optionnel
Options	Se pose la question de l'utilité de cet item puisqu'on peut penser que si le patient s'est opposé à son enregistrement dans l'EDS il ne doit plus y figurer. Pour autant, le patient peut accepter puis signifier son refus. Ainsi, toute étude réalisée avec son consentement doit être tracée, ce qui implique de conserver la date d'opposition. Dans ce cas, il est proposé de retirer toutes les informations le concernant.

3 - Facteurs de risque

Les informations recueillies dans le modèle sont, pour chacun des 13 facteurs de risque reconnus par le CIRC, Présent (oui), absent (non), inconnu

Nom de l'item	Facteurs de risque
Numéro de l'item	3.1
Nom de l'item	PatientSmoking ; PatientAlcohol ; PatientOverweight ; PatientUnBalancedDiet, PatientSedentaryLifestyle ; PatientHormone ; PatientInfectiousDisease ; PatientIonisingRadiation ; PatientInadequatBreastFeeding ; PatientPollution ; PatientUVRadiation ; PatientOccupationalExposure ; PatientChemicalExposure
Définition	Facteurs de risque externes
Nomenclature utilisée	Les facteurs de risque sont exprimés sous forme d'information binaire, patients présentant ou non tel facteur de risque

Valeur attendue	<p>Oui / Non / Inconnu (pour chaque variable)</p> <ul style="list-style-type: none"> - Tabagisme (dont tabagisme passif) - Consommation d'alcool - Alimentation déséquilibrée - Surpoids et obésité - Activité physique insuffisante - Utilisation d'hormones exogènes - Durée insuffisante de l'allaitement infantile - Infections (ex : hépatite B et C, papillomavirus, H. Pylori) - Radiations ionisantes (radon, exposition diagnostic) - Pollution atmosphérique, particules fines - Rayonnement solaire (UV) - Expositions professionnelles (ex : amiante) - Exposition à certaines substances chimiques (ex : arsenic, benzène).
Règle	La collecte de l'information se fait par TALN à partir des dossiers textuels ou en récupérant une information structurée.
Justification du choix de la règle (exemple)	La liste proposée est réalisée avec les propositions des facteurs de risque de l'INCA basée sur la liste du CIRC.
Caractère obligatoire ou optionnel	Optionnel

4 - Statut vital du patient

Les informations à recueillir sont :

- Le statut vital (Vivant, décédé)
- La source qui permet la mise à jour du statut vital
- La date de mise à jour des données concernant le statut vital
- La date de dernière venue dans l'établissement
- La date de dernier contact (par courrier par exemple) ou date de dernière nouvelle
- La date de décès

4.1 Statut vital

Nom de l'item	Statut vital
Numéro de l'item	4.1
Nom de la variable	DeathStatus
Définition	Statut vital de la personne
Valeur attendue	Dans les études cliniques, les informations "vivant / décédé" sont codées

	<p>dans les bases de données de manière standardisée pour faciliter l'analyse statistique, assurer la conformité réglementaire et permettre l'interopérabilité entre systèmes.</p> <ul style="list-style-type: none"> - 0 = Vivant - 1 = Décédé (événement survenu)
Règle	<p>Le système récupère l'information dans le dossier administratif.</p> <p>En l'absence d'information, le patient est considéré comme vivant, parfois à tort ; il est nécessaire de vérifier qu'il n'est pas décédé (par exemple dans le fichier INSEE) pour réaliser des études sur la survie.</p> <p>Par défaut le patient est mis en statut « vivant » jusqu'à ce qu'on remonte l'information qu'il est décédé</p> <p>En utilisant les informations disponibles, avec par ordre de priorité ;</p> <ul style="list-style-type: none"> - Le dossier administratif de l'établissement, - Les données de l'INSEE - Ou par le biais du dossier médical.
Caractère obligatoire ou optionnel	Obligatoire
Recommandations	<p>Il est recommandé, si le patient n'est pas déclaré « décédé », d'interroger le fichier des personnes décédées de l'INSEE (https://www.insee.fr/fr/information/4190491) pour savoir si le patient n'est pas inscrit dans le fichier INSEE comme décédé.</p> <p>Un autre site peut être utilisé permettant d'importer des listes CSV (https://deces.matchid.io/search)</p>

4.2 Source de l'information sur le statut vital

Nom de l'item	Source de l'information sur le statut vital
Numéro de l'item	4 .2
Nom de la variable	DeathStatusSource
Définition	Information concernant le statut vital de la personne récupérée dans les bases de données de l'établissement et de l'INSEE
Valeur attendue	Donnée issue de l'INSEE Donnée structurée issu de l'établissement Donnée issue de la lecture du dossier médical
Règle	Il est conseillé de privilégier la donnée de décès si elle est structurée dans le système d'information de l'établissement, puis de rechercher l'information dans les données de l'INSEE.
Caractère obligatoire ou optionnel	Obligatoire

Recommandations	<p>Les algorithmes de recherche pour un appariement d'un patient avec la base INSEE sont différents entre les établissements. Il serait intéressant de définir une même méthodologie.</p> <p>La qualité de l'information issue de l'INSEE dépend directement de l'algorithme et des critères de filtre, plus ou moins stricts, et selon les traits utilisés (nom de naissance, prénom, commune de naissance, etc.).</p> <p>La recommandation est d'utiliser les mêmes méthodes pour les études multicentriques.</p>
-----------------	---

4.3 Date de mise à jour des données

Nom de l'item	Date de mise à jour des données
Numéro de l'item	4.3
Nom de la variable	DeathStatusUpdateDate
Définition	Date de dernière interrogation de la base INSEE et de la mise à jour de la base administrative de l'établissement
Valeur attendue	YYYY - MM – DD
Règle	L'information concernant le statut vital de la personne récupérée dans les bases de données de l'INSEE doit être complétée de la date d'interrogation de la base. Le même algorithme est mis en place pour les données issues de l'établissement
Caractère obligatoire ou optionnel	Obligatoire

4.4 Date de dernière venue

Nom de l'item	Date de dernière venue
Numéro de l'item	4.4
Nom de la variable	DeathStatusLastVisit
Définition	Date de dernière venue physique de la personne dans l'établissement
Valeur attendue	YYYY - MM – DD
Règle	<p>L'établissement s'entend par l'établissement qui alimente la base OSIRIS RWD.</p> <p>Il ne s'agit pas d'une date de rendez-vous mais bien de la venue physique dans l'établissement.</p>

	<p>Cette date est alimentée en automatique à partir du dossier de l'établissement.</p> <p>L'établissement choisit la source la plus fiable, toute celle qui définit qu'une venue physique a bien été réalisée.</p>
Justification du choix de la règle	<p>Cette date est réputée fiable car récupérée à chaque venue. Elle peut être récupérée automatiquement. Elle peut être utilisée par les statisticiens comme la date de dernière nouvelle.</p> <p>La date de dernière visite dans l'établissement est une donnée toujours disponible et fiable si le patient a bénéficié d'une consultation, une hospitalisation ou un passage en hôpital de jour.</p>
Caractère obligatoire ou optionnel	Obligatoire
Recommandations	Il n'est pas retenu le principe qu'une date de document soit considérée comme une date de dernière venue.

4.5 Date de dernier contact

Nom de l'item	Date de dernier contact
Numéro de l'item	4.5
Nom de la variable	DeathStatusLastContact
Définition	Date à laquelle l'information sur le statut vital a été mise à jour à partir des informations recueillies dans les documents externes à l'établissement (appelé également date de dernières nouvelles)
Valeur attendue	YYYY - MM - DD
Règle	En complément des autres sources d'information, il est possible de rechercher les informations sur le statut vital dans les documents du dossier médical pour alimenter la « date de dernier contact ».
Caractère obligatoire ou optionnel	Optionnel

4.6 Date de décès

Nom de l'item	Date de décès
Numéro de l'item	4.6
Nom de la variable	DeathDateDate

Définition	Date à laquelle la personne est décédée.
Valeur attendue	YYYY - MM - DD
Caractère obligatoire ou optionnel	Optionnel
Recommandations	Si elle est structurée dans l'établissement, cette date doit prévaloir car cela signifie que l'établissement avait conservé un lien avec le patient.

5 - Antécédents familiaux de cancer

5.1 Topographie de l'antécédent familial de cancer

Nom de l'item	Topographie de l'antécédent familial de cancer
Numéro de l'item	5.1
Nom de la donnée	FamilyCancerHistoryTopographyCode
Définition	Localisation de l'antécédent carcinologique d'un membre de la famille
Nomenclature utilisée	CIM-O3 : axe topographique
Valeur attendue	Chaîne de caractères (ex : C18.4)
Caractère obligatoire ou optionnel	Obligatoire

5.2 Lien de parenté avec le patient

Nom de l'item	Lien de parenté
Numéro de l'item	5.2
Nom de la variable	FamilyCancerHistoryParentage
Définition	Lien de parenté entre le patient et la personne de la même famille présentant un cancer
Nomenclature utilisée	Liste fermée
Valeur attendue	<ul style="list-style-type: none">- Mère- Père- Fils- Fille- Grand-père- Grand-mère- Petit-fils- Petite-fille- Frère- Sœur
Caractère obligatoire ou optionnel	Obligatoire
Recommandations	La recommandation est de structurer toutes les personnes ayant eu un cancer en lien de parenté du 1er ou 2ème degré avec le patient.

6 - Antécédent personnel de cancer

Les antécédents personnels de cancer sont comptabilisés comme des occurrences de cancer (Cf. chapitre 8 - Le cancer).

7 - Antécédent médical, comorbidité et événement indésirable

Les comorbidités, événements indésirables et les antécédents sont regroupées dans un seul item appelé **évènement morbide**.

La date de survenue, de découverte ou de recueil dans le dossier médical est recueillie et permettra au moment de l'analyse d'interpréter les résultats comme un antécédent, une comorbidité ou un possible effet indésirable.

On utilise une seule classification, la CIM-10, pour structurer maladies et symptômes, antécédent, comorbidité et effets indésirables.

7.1 Evènement morbide

Nom de l'item	Evènement morbide
Numéro de l'item	7.1
Nom de la variable	RelatedPathologyCode
Définition	Toute pathologie hors cancer : antécédent, comorbidité ou évènement indésirable.
Nomenclature utilisée	CIM-10
Valeur attendue	Chaine de caractères
Règle	<p>La règle est de structurer dans cet item les pathologies associées (comorbidités, antécédents médicaux) et les événements indésirables des traitements.</p> <p>C'est la date de début qui permettra de qualifier et de séparer un antécédent d'une comorbidité.</p> <p>C'est la probabilité de survenue et la date qui permettront de séparer les événements indésirables des autres pathologies.</p>
Justification du choix de la règle	Cet item permet de distinguer les pathologies associées survenues avant le diagnostic de cancer (antécédent médical du cancer) des pathologies associées survenues pendant la maladie cancéreuse (comorbidités).
Caractère obligatoire ou optionnel	Obligatoire

7.2 Date de survenue de l'évènement morbide

Nom de l'item	Date de survenue de l'évènement morbide
Numéro de l'item	7.2
Nom de la variable	RelatedPathologyDiagnosisDateDay, RelatedPathologyDiagnosisDateMonth, RelatedPathologyDiagnosisDateYear
Définition	Date à laquelle le symptôme ou la maladie sont notés ou si cette information n'est pas précisée, on propose de prendre la date du compte-rendu.
Valeur attendue	YYYY - MM - DD
Règle	La date de survenue est la date notée dans le dossier
Caractère obligatoire ou optionnel	Obligatoire

Le cancer

8 – La tumeur primitive

Un cancer est défini par :

- Une localisation tumorale (ou code topographique)
- Une anatomo-pathologie (ou morphologie)
- Une latéralité

D'autres données ont été retenues dans le cadre du MDS :

- L'occurrence du cancer, en raison des cancers multiples
- Un lieu de prise en charge
- Une date de diagnostic (anatomo-pathologique ou cytologique)
- Un bilan d'extension initial : TNM et/ou stade
- La taille de la tumeur primitive, la date de la mesure et la méthode de mesure
- Un grade histologique
- Des biomarqueurs prédictifs, pronostiques ou théranostiques (voir le chapitre biomarqueurs)
- La taille de la tumeur
- La date de la mesure
- La méthode de mesure : clinique, imagerie, pathologique
- Une date de début de traitement (voir le chapitre traitements)

8.1 Occurrence du cancer

Nom de l'item	Occurrence du cancer (numéro d'ordre)
Numéro de l'item	8.1
Nom de la variable	CancerOrder
Définition	Ordre d'apparition du cancer pour le patient, quel que soit le lieu de prise en charge
Règle	Le numéro du cancer correspond à l'ordre des cancers intervenus dans la vie de la personne. Pour les cancers pris en charge dans l'établissement, la date de diagnostic permet de discerner l'ordre. Pour les antécédents de cancer, l'année présente dans le dossier médical est prise en compte.
Valeur attendue	Entier de 1 à n
Justification du choix de la règle (exemple)	Le numéro du cancer 1, 2, 3, etc. correspond à l'ordre des cancers intervenus dans la vie des patients quel que soit le lieu de prise en charge. <i>Exemple : Une patiente vient dans l'établissement A pour un cancer du sein gauche. Elle a eu auparavant un cancer du sein droit il y a 10 ans</i>

	<i>traité dans l'établissement B.</i> <i>Le cancer du sein droit est enregistré en numéro 1.</i> <i>Le cancer du sein gauche est enregistré en numéro 2.</i>
Recommandation	Le numéro du cancer (ou occurrence) est calculé automatiquement à partir des informations sur la date du diagnostic, le code topographique et la latéralité

8.2 Prise en charge dans l'établissement

Nom de l'item	Prise en charge dans l'établissement
Numéro de l'item	8.2
Nom de la variable	CancerCareOriginCenter
Définition	Prise en charge totale ou partielle (au moins un acte) du patient dans l'établissement pour l'occurrence de cancer
Règle	<p>L'établissement dans lequel une prise en charge totale ou partielle (au moins un acte) a lieu est considéré comme un lieu de prise en charge.</p> <p>Cet item permet de différencier les antécédents personnels de cancers, qui n'ont jamais été pris en charge dans l'établissement, des cas index, pris en charge intégralement ou partiellement dans l'établissement.</p>
Valeur attendue	Y / N
Justification du choix de la règle (exemple)	<p><i>Un patient a été pris en charge intégralement pour un cancer en dehors de l'établissement sans jamais venir pour aucun acte diagnostic, thérapeutique ou une surveillance, il est considéré que la prise en charge se fait en dehors de l'établissement.</i></p> <p><i>S'il vient ensuite pour une 2ème tumeur et est pris en charge dans l'établissement, il est considéré comme pris en charge pour la 2ème tumeur, même si cette prise en charge est partielle (acte diagnostic ou thérapeutique).</i></p> <p><i>Un patient est traité par chimiothérapie dans l'établissement A ; la radiothérapie est réalisée dans l'établissement B. Il est considéré comme pris en charge dans l'établissement A et dans l'établissement B.</i></p>
Option	Pour les antécédents de cancer, la modalité « Non » est retenue.

8.3 Date de diagnostic du cancer

Nom de l'item	Date de diagnostic du cancer
Numéro de l'item	8.3
Nom de la variable	CancerDiagnosticDateDay ; ---Month ; ---Year
Définition	Date de l'examen ou de l'acte qui permet le diagnostic de certitude du cancer.
Valeur attendue	YYYY - MM – DD
Règle	<p>La date de diagnostic du cancer correspond au premier prélèvement confirmant le caractère malin, quel que soit le site du prélèvement, site de la tumeur primitive ou site d'une localisation métastatique, avant tout traitement du cancer.</p> <p>Ceci inclut l'acte chirurgical si le diagnostic de certitude est porté à la suite de cet acte. En cas de biopsie réalisée avant un acte chirurgical, c'est la date de la biopsie qui prévaut.</p> <p>Une seule date est retenue en cas de prélèvements multiples et toujours la première, sans prioriser l'histologie sur la cytologie, si la cytologie permet un diagnostic de malignité.</p> <p>Pour certaines tumeurs, le diagnostic est porté sans nécessité de prélèvement (par exemple le diagnostic repose sur l'imagerie pour certaines tumeurs cérébrales, le rétinoblastome, les cancers primitifs hépatiques).</p> <p>Dans tous les cas, c'est la date de prélèvement (ou de l'examen) qui prévaut et non la date du résultat de l'examen.</p> <p>Si une valeur est non disponible sur le jour (DD), le mois (MM) ou l'année (YYYY). Le champ correspondant sera laissé sans valeur.</p>
Caractère obligatoire ou optionnel	Obligatoire Si antécédent de cancer, seul l'année est obligatoire

8.4 Méthode de diagnostic du cancer

Nom de l'item	Méthode de diagnostic du cancer
Numéro de l'item	8.4
Nom de la variable	CancerDiagnosisMethod
Définition	Méthode d'investigation et de confirmation du diagnostic de cancer (ou base de diagnostic)

Règle	Un seul code est retenu en utilisant le code qui permet le diagnostic de certitude du cancer. Le diagnostic du cancer correspond au premier prélèvement (ou au premier examen) confirmant le caractère malin.
Nomenclature utilisée	Basée sur les règles européennes de codage dans les registres. https://encr.eu/sites/default/files/Recommandations/ENCR%20Recommandation%20BoD_Oct2022_FR.pdf
Valeurs attendues	1 = Clinique 2 = Investigation clinique (y compris imagerie et chirurgie exploratoire) 4 = Marqueurs tumoraux spécifiques 5 = Cytologie 7 = Histologie 8 = Tests cytogénétiques ou moléculaires ou biologiques 9 = Inconnu Exemple 1 : si la cytologie permet le diagnostic de certitude, c'est la date de la cytologie qui est retenue, avec le code 5 comme méthode de diagnostic du cancer. Exemple 2 : si le diagnostic d'un rétinoblastome se fait par imagerie, c'est la date de la première imagerie qui prévaut, avec le code 2
Caractère obligatoire ou optionnel	Optionnel

8.5 Diagnostic réalisé dans l'établissement (sera remplacé par l'item source de la donnée)

Nom de l'item	Diagnostic réalisé dans l'établissement
Numéro de l'item	8.5
Nom de la variable	?
Définition	On différencie le fait que le diagnostic est réalisé dans l'établissement (« in ») ou dans un autre établissement (« out »). Lorsque le diagnostic est porté dans l'établissement, l'information est en partie structurée et le compte-rendu médical provient du SIH.
Règle	Lorsque le diagnostic est porté par cytologie ou histologie, le document de référence est le CR de cytologie ou d'anatomopathologie. Il convient de différencier la source du document (document qui provient du SIH de l'ETS ou si le document est scanné). Si le prélèvement est réalisé dans un autre ETS ou « en ville » et que la

	<p>relecture (avec un CR de l'ETS) est réalisée dans l'ETS, on considère que le diagnostic est réalisé dans l'ETS.</p> <p>Si une relecture d'examen est réalisée dans l'ETS et si un CR provient du SIH, on considère que le diagnostic est réalisé dans l'ETS si le CR est antérieur avant tout traitement.</p> <p>Lorsque le diagnostic est porté par imagerie, le diagnostic est considéré comme réalisé dans l'ETS si l'imagerie a été relue avec un CR réalisé dans l'établissement.</p>
Nomenclature utilisée	In (dans l'établissement) ; Out (en dehors de l'établissement)
Caractère obligatoire ou optionnel	Optionnel

8.6 Code topographique

Nom de l'item	Code topographique du cancer
Numéro de l'item	8.6
Nom de la variable	CancerDiagnosisTopographyCode
Définition	Localisation de la tumeur primitive.
Nomenclature utilisée	CIMO-3 – axe topographique
Valeur attendue	Chaîne de caractères (ex : C18.4)
Règle	<p>Le code topographique décrit le site anatomique d'origine de la tumeur.</p> <p>Le code a toujours un préfixe de « C », suivi d'un nombre à trois chiffres qui indique le site (deux chiffres) et le sous-site (un chiffre), séparés par un point décimal.</p> <p>Exemple : C18.4 : le C18 indique que le site est le côlon et le 4 indique que le sous-site est le côlon transverse.</p>
Caractère obligatoire ou optionnel	Obligatoire

8.7 Groupe topographique

Nom de l'item	Groupe topographique du cancer
---------------	--------------------------------

Numéro de l'item	8.7
Nom de la variable	
Définition	Localisation de la tumeur primitive (regroupement)
Nomenclature utilisée	CIMO-3 – axe topographique (regroupement des codes OMS)
Valeur attendue	Chaîne de caractères
Règle	<p>Le code topographique décrit le site anatomique d'origine de la tumeur.</p> <p>Les tables permettent des regroupements automatiques, calculés</p> <p>Exemple :</p> <p>C19 (Jonction recto-sigmoïdienne) et C20 (rectum) sont regroupés en C20.9.</p>
Caractère obligatoire ou optionnel	Obligatoire

8.8 Code histologique / morphologique

Nom de l'item	Code histologique / morphologique
Numéro de l'item	8.7
Nom de la variable	CancerDiagnosisMorphologyCode
Définition	Description morphologique du cancer
Nomenclature utilisée	CIMO-3 – axe morphologie
Valeur attendue	Chaîne de caractères (ex : 8144/3)
Règle	<p>Les règles utilisées sont celles de la CIMO-3.</p> <p>Pour une même entité tumorale, on différencie les codes comportementaux (le 5ème digit) et des codes morphologiques (les 4 premiers digits).</p> <p>La règle est d'utiliser un seul code morphologique, même si la tumeur est composite ou multiple dans un même site primitif.</p>
Caractère obligatoire ou optionnel	Obligatoire
Options	Si l'établissement utilise la codification ADICAP, il existe un

	transcodage en code CIMO-3 dans la mesure où la classification ADICAP n'est pas maintenue ; elle n'est donc pas recommandée.
--	--

8.9 Groupe morphologique

Nom de l'item	Groupe morphologique du cancer
Numéro de l'item	8.9
Nom de la variable	
Définition	Description morphologique du cancer (Regroupement)
Nomenclature utilisée	CIMO-3 – axe morphologique (regroupement des codes OMS)
Valeur attendue	Chaîne de caractères
Règle	<p>Le code morphologique décrit l'antomocytopathologie du cancer (site anatomique d'origine de la tumeur ou site métastatique pour les tumeurs d'emblée métastatique.</p> <p>Les tables permettent des regroupements automatiques,</p> <p>Exemple :</p> <p>Les codes 9050 - 9055 sont regroupés en un seul code : mésothéliome</p>
Caractère obligatoire ou optionnel	Obligatoire

8.10 Latéralité de la tumeur primitive

Nom de l'item	Latéralité
Numéro de l'item	8.10
Nom de la variable	CancerDiagnosisLaterality
Définition	Latéralité de la tumeur primitive
Valeurs attendues	Droit : R Gauche : L Non applicable (par exemple pour les lésions médianes) : NA Inconnu : UN
Caractère obligatoire ou	Obligatoire

optionnel	
Recommandations	<p>La règle est de créer deux cancers lorsqu'un diagnostic de cancer est posé sur une localisation où une latéralité est possible.</p> <p>Dans tous les cas (cancers synchrones ou cancers successifs avec un traitement entre les deux), on crée un cancer d'un côté, puis un deuxième cancer en notant la latéralité, soit deux cancers.</p> <p>Le code « cancer bilatéral » n'est pas retenu car un cancer bilatéral n'a pas nécessairement les mêmes caractéristiques biologiques et le traitement peut être différent (à la différence d'un cancer localisé associé à une métastase controlatérale).</p> <p>L'information peut être « inconnu » ou « non applicable » comme pour les lymphomes.</p> <p>Une liste des cancers par localisation est fournie en annexe avec les codes possibles (R/L/NA/UN)</p> <p>Exemples :</p> <ul style="list-style-type: none"> - Un cancer de la langue : non applicable - Un cancer du sein est soit à droite, soit à gauche ; - Cas particulier des cancers de l'ovaire : la classification TNM ou FIGO permet de distinguer les tumeurs unilatérales des tumeurs bilatérales

8.11 Taille de la tumeur

Nom de l'item	Taille de la tumeur primitive au diagnostic
Numéro de l'item	8.11
Nom de la variable	TumorSize
Définition	Permet de décrire la taille de la tumeur primaire.
Valeur attendue	Entier
Règle	<p>La taille de la tumeur primitive (la valeur du plus grand axe) est mesurée par les radiologues ou les pathologistes et l'information est consignée dans le dossier médical.</p> <p>L'unité de valeur est le millimètre.</p>
Caractère obligatoire ou optionnel	Optionnel
Recommandations	<p>On ne conserve qu'une seule mesure à savoir la plus grande dimension (ou le plus grand axe) en millimètre, avant tout traitement.</p> <p>La mesure pathologique (donc la chirurgie) prime sur imagerie</p>

	En l'absence de chirurgie, c'est l'imagerie qui compte.
Option	Pour les tumeurs multifocales, notamment les cancers du sein, il est proposé de retenir la plus grande taille.

8.12 Date de la mesure de la taille

Nom de l'item	Date de la mesure de la tumeur primitive au diagnostic
Numéro de l'item	8.12
Nom de la variable	TumorSizeDate (Day, Month, Year)
Définition	Date du compte rendu qui indique la taille de la tumeur primitive
Valeur attendue	Date (YYYY - MM – DD)
Caractère obligatoire ou optionnel	Optionnel

8.13 Méthode de mesure de la taille

Nom de l'item	Méthode de mesure de la tumeur primitive au diagnostic
Numéro de l'item	8.13
Nom de la variable	TumorSizeMethod
Définition	Type de compte rendu qui indique la taille de la tumeur primitive
Valeur attendue	Code alphanumérique selon la classification SNOMED CT Examen pathologique - Macroscopique (SCT 787377000) Examen d'imagerie - IRM (SCT 113091000) - TDM (SCT 77477000) - Echographie (SCT 16310003) - Radiologie standard, (SCT 363680008)
Caractère obligatoire ou optionnel	Optionnel
Règle	Pour l'imagerie, on ne conserve qu'une seule mesure. Lorsque plusieurs mesures sont réalisées par des appareils différents, on prend en compte dans l'ordre :

	<ul style="list-style-type: none"> - IRM - TDM - Echographies - Radio standard (ex : mammographie)
--	--

8.14 Cotation du T

Nom de l'item	Cotation du T
Numéro de l'item	8.14
Nom de la variable	TNMEventTValue
Définition	La cotation du T (qui représente la taille de la tumeur) dans la classification TNM.
Nomenclature utilisée	Notre recommandation est d'utiliser un fichier qui liste l'ensemble des codes de la classification de l'UICC TNM (version en fonction de la date du CR médical)
Valeur attendue	Chaîne de caractères (liste fermée) Liste de codes de la classification TNM (présente et passées)
Règle	<p>La valeur du T est une valeur unique codée au moment du diagnostic de la tumeur (pour cT et pT) et peut différer selon le type de classification.</p> <ul style="list-style-type: none"> - Un seul cT coté par le clinicien avant tout traitement, - Un seul pT coté par le pathologiste - Un seul yT après traitement néoadjuvant.
Caractère obligatoire ou optionnel	Optionnel
Recommandations	Lorsque, pour une même tumeur, deux stades de T sont notés dans le dossier médical, on prend la valeur la plus élevée.

8.15 Cotation du N

Nom de l'item	Cotation du N
Numéro de l'item	8.15
Nom de la variable	TNMEventNValue
Définition	La cotation du N (qui représente l'atteinte éventuelle de ganglions lymphatiques régionaux) dans la classification TNM.
Nomenclature utilisée	La classification TNM (diffère selon la version) Notre recommandation est d'utiliser un fichier qui liste l'ensemble des codes de la classification de l'UICC TNM (version en fonction de la date du CR médical)

Valeur attendue	Chaîne de caractères (liste fermée) Liste de codes de la classification TNM (présente et passées)
Règle	Même règle que pour le T
Caractère obligatoire ou optionnel	Optionnel

8.16 Cotation du M

Nom de l'item	Cotation du M
Numéro de l'item	8.16
Nom de la variable	TNMEventMValue
Définition	Cotation du M qui représente l'atteinte à distance de la tumeur initiale
Valeur attendue	1 ou 0
Règle	<p>Lorsque l'information du TNM est structurée, le M du TNM est notée à 1.</p> <p>Lorsque la tumeur est d'emblée métastatique, l'information du TNM n'est pas toujours présente car la valeur du T ou du N perd d'intérêt par rapport à la notion de maladie métastatique, que ce soit pour la décision de traitement ou pour évaluer le pronostic. Dans ce cas, l'information est présente dans le dossier et doit être codée lors de l'analyse en M1.</p> <p>Il est possible de créer une variable intermédiaire structurée avec la source de la donnée ; métastase d'emblée avec comme source le TALN</p>
Caractère obligatoire ou optionnel	Obligatoire

8.17 Type de TNM

Nom de l'item	Type de TNM
Nom de la variable	TNMEventType
Numéro de l'item	8.17
Définition	Le type de la classification TNM
Nomenclature utilisée	UICC

Valeur attendue	c → clinique p → pathologique y → : après traitement néoadjuvant
Règle	Le type est indiqué en fonction de la source de la donnée ou si c'est indiqué dans le compte rendu médical.
Caractère obligatoire ou optionnel	Optionnel
Recommandations	Le TNM est codé au diagnostic. Il est invariant et unique (monovalué). Le TNM est clinique quand il est codé par le clinicien Le TNM est pathologique quand il est codé par le pathologiste quand le patient est opéré. On parle de yTNM après traitement médical et lorsque le patient est opéré.
Options	

8.18 Type de Stade

Nom de l'item	Type de Stade
Numéro de l'item	8.18
Nom de la variable	PrimaryCancerStageType
Définition	Le système de classification ayant permis de définir le stade de la tumeur.
Nomenclature utilisée	Le type de classification utilisé est rarement noté La codification dépend du type de tumeur Si l'information sur le type de classification est disponible, elle est récupérée ; dans le cas contraire, on utilise le code inconnu (UK).
Valeur attendue	Liste fermée
Règle	On prend le type (Ex : FIGO) quand il est renseigné dans le CR.
Caractère obligatoire ou optionnel	Optionnel

8.19 Stade de la tumeur

Nom de l'item	Stade de la tumeur (extension)
Numéro de l'item	8.19
Nom de la variable	PrimaryCancerStageValue
Définition	Le stade du cancer est une donnée synthétique qui mesure l'extension de la maladie cancéreuse au diagnostic et avant tout traitement
Nomenclature utilisée	Liste fermée : 0 (cancers in situ), I, II, III, IV, UK La codification en stade repose sur de multiples classifications souvent noté de 0 à 4 (de rares exceptions sont possibles).
Valeur attendue	Chaîne de caractères
Règle	Le stade n'est pas calculé à partir du TNM Il est récupéré si la donnée est structurée (Ex : RCP) En l'absence de donnée structurée, l'information peut être récupérée par TALN dans les comptes rendus médicaux. Il est possible que deux codifications du stade soient retrouvées dans le dossier. C'est le dernier stade mentionné dans le dossier avant tout traitement qui prime
Caractère obligatoire ou optionnel	Optionnel
Recommandations	Si l'information sur le stade est structurée, celle-ci prime sur les informations récupérées par TALN.

8.20 Type de grade histologique

Nom de l'item	Type de grade (histologique)
Numéro de l'item	8.20
Nom de la variable	PrimaryCancerGradeHistologicalType
Définition	Le système de classification ayant permis de définir le grade de la tumeur.
Nomenclature utilisée	Code Snomed CT

Valeur attendue	Chaîne de caractères
Caractère obligatoire ou optionnel	Optionnel
Recommandations	<p>La codification dépend du type de tumeur</p> <p>Par exemple, pour le cancer du sein, le pathologiste évalue trois paramètres : l'architecture tumorale, la forme et la taille du noyau de la cellule et l'activité mitotique. Le grade est exprimé par les lettres SBR, suivies des chiffres I, II ou III.</p> <p>La codification en grade repose sur de multiples classifications, parfois plusieurs pour la même tumeur ; par exemple, le système de classification du grade pour les cancers de la prostate est basé sur le score de Gleason ou le score ISUP.</p> <p>On attribue habituellement au grade un chiffre de 1 à 3 ou 4.</p> <p>Le grade peut aussi être exprimé en termes de « bas grade » pour les tumeurs les moins agressives et « haut grade » pour les tumeurs les plus agressives.</p> <p>Si l'information est disponible sans ambiguïté, elle est récupérée ; dans le cas contraire, elle est notée UN ».</p>

8.21 Grade histologique

Nom de l'item	Valeur du grade histologique
Numéro de l'item	8.21
Nom de la variable	PrimaryCancerGradeHistologicalValue
Définition	Le grade du cancer (ou grade histologique) est une mesure de la différenciation des cellules et de l'agressivité de la tumeur
Nomenclature utilisée	Liste fermée
Valeur attendue	<p>Si l'information est structurée dans un CR (Ex : CR de RCP ou « fiche tumeur », elle prime sur les autres documents.</p> <p>Si l'information n'est pas structurée, elle est récupérée par TALN uniquement dans les CR d'anatomopathologie ou dans les fiches de RCP.</p> <p>La donnée est parfois numérique (1, 2, 3), alpha-numérique (1A) ou en texte (différencié/indifférencié).</p>

Caractère obligatoire ou optionnel	Optionnel
Recommandations	Il est proposé, pour chaque type de tumeur (sein, prostate, etc.) une liste des systèmes de grade utilisé

9 - Evènements tumoraux

Pour chaque évènement tumoral, on peut distinguer :

- Le type de l'évènement tumoral (récidive ou rechute, métastase, transformation)
- La date de 1ère évocation de l'évènement tumoral (clinique, biologique, imagerie)
- Les investigations réalisées pour établir le diagnostic des évènements tumoraux
- La date de diagnostic lorsqu'un prélèvement cytologique ou histologique est réalisé
- Le code anatomo-pathologique (en cas de prélèvement)
- Des biomarqueurs éventuels
- La topographie des métastases et la date de découverte de chaque métastase (qui surviennent à des dates distinctes).
- La première date de traitement de l'évènement (cf. chapitre traitement)

9.1 Type de l'évènement tumoral

Nom de l'item	Type de l'évènement tumoral
Numéro de l'item	9.1
Nom de la variable	TumEventType
Définition	Les évènements tumoraux sont définis par les évènements cliniques suivants : Pour les cancers solides : <ul style="list-style-type: none"> - Récidive locale ou loco-régionale, - Métastase Pour les cancers hématologiques : <ul style="list-style-type: none"> - Rechute - Transformation (ou acutisation)
Nomenclature utilisée	Codification
Valeur attendue	1 = Récidive locale (ou loco-régionale) 2 = Métastase 3 = Rechute 4 = Transformation (ou acutisation)

Règle	Cette information étant rarement structurée à la source, elle peut l'être à partir de des compte rendus médicaux par TALN et des traitements administrés
Caractère obligatoire ou facultatif	Obligatoire

9.2 Date de première évocation de l'évènement tumoral

Nom de l'item	Date de première évocation de l'évènement tumoral
Numéro de l'item	9.2
Nom de la variable	TumEventMentionDateDay, TumEventMentionDateMonth, TumEventMentionDateYear,
Définition	Date de première évocation de l'évènement dans le dossier médical.
Valeur attendue	YYYY - MM - DD
Règle	<p>Pour la gestion des dates de début des évènements tumoraux, ces évènements sont définis par leur première occurrence dans le dossier médical.</p> <p>Celle-ci marque le commencement ou la date de l'évènement tumoral</p> <ul style="list-style-type: none"> - récurrence locale ou loco-régionale (1^{er} occurrence) - maladie métastatique (1^{er} occurrence) - rechute hématologique - Transformation <p>Si une valeur est non disponible sur le jour (DD), le mois (MM) ou l'année (YYYY). Le champ correspondant sera laissé sans valeur.</p>
Caractère obligatoire ou facultatif	Optionnel

9.3 Date de diagnostic histologique

Nom de l'item	Date de diagnostic histologique de l'évènement tumoral
Numéro de l'item	9.3
Nom de la variable	TumEventDiagnosisDateDay, TumEventDiagnosisDateMonth, TumEventDiagnosisDateYear,
Définition	Date de confirmation du diagnostic de l'évènement tumoral par un prélèvement histologique ou cytologique

Valeur attendue	YYYY - MM - DD
Règle	<p>Si le diagnostic est réalisé en dehors de l'établissement, on indique la date indiquée dans le dossier. Dans ce cas, la date du diagnostic de la récurrence ou de la métastase peut être structurée sans un compte rendu récupéré de l'extérieur, par exemple un compte rendu d'anatomopathologie.</p> <p>Si cette information n'est pas structurée à la source, elle est récupérée dans le dossier.</p> <p>Si une valeur est non disponible sur le jour (DD), le mois (MM) ou l'année (YYYY). Le champ correspondant sera laissé sans valeur.</p>
Caractère obligatoire ou facultatif	Obligatoire

9.4 Code histologique / morphologique

Nom de l'item	Code histologique / morphologique
Numéro de l'item	9.4
Nom de la variable	TumEventMorphologyCode
Définition	Description morphologique de l'évènement tumoral
Nomenclature utilisée	CIMO-3 Axe morphologie
Valeur attendue	Chaîne de caractères (ex : 8144/6)
Règle	Notre recommandation est de considérer le compte-rendu d'anatomopathologie comme référence
Caractère obligatoire ou facultatif	Optionnel
Recommandations	Quand la maladie progresse, l'histologie confirme ou permet de déterminer le traitement.

9.5 Code topographique des métastases

Nom de l'item	Code topographique des métastases
Numéro de l'item	9.5
Nom de la variable	MetastasisTopographyCode

Définition	Localisation des métastases dans l'évènement « maladie métastatique ».						
Nomenclature utilisée	CIMO-3 axe topographique						
Valeur attendue	Chaîne de caractères (ex :C18.4)						
Règle	Cette information ne concerne que les métastases de l'évènement « maladie métastatique » ; on recueille le code topographique des métastases ainsi que leur date d'apparition (ou de découverte).						
Caractère obligatoire ou facultatif	Optionnel						
Recommandations	<p>Les localisations métastatiques ne nécessitent pas la même précision que la localisation des tumeurs primitives.</p> <p>On se limite au premier niveau de topographie de la classification CIMO-3</p> <p>Exemples :</p> <table> <tr> <td>C22</td><td>Foie</td></tr> <tr> <td>C34</td><td>Poumon</td></tr> <tr> <td>C71</td><td>Cerveau</td></tr> </table>	C22	Foie	C34	Poumon	C71	Cerveau
C22	Foie						
C34	Poumon						
C71	Cerveau						

9.6 Dates de découverte des métastases par topographie

Nom de l'item	Date de découverte des métastases par topographie
Numéro de l'item	9.6
Nom de la variable	MetastasisDiscoveryDateDay, MetastasisDiscoveryDateMonth, MetastasisDiscoveryDateYear
Définition	Date de survenue des métastases rattachées à un code topographique de la métastase et à l'évènement « maladie métastatique ».
Valeur attendue	YYYY - MM - DD
Règle	<p>Cette information ne concerne que les métastases de l'évènement « maladie métastatique ». Dans cet évènement, il faut recueillir les codes topographiques des métastases ainsi que leur date d'apparition au fur et à mesure de leur survenue.</p> <p>Si une valeur est non disponible sur le jour (DD), le mois (MM) ou l'année (YYYY). Le champ correspondant sera laissé sans valeur.</p>
Caractère obligatoire ou facultatif	Optionnel

9.7 Investigations réalisées pour établir le diagnostic des événements tumoraux

Nom de l'item	Investigation réalisée pour le diagnostic des événements tumoraux (récidives, métastases, rechutes, transformations)
Numéro de l'item	9.8
Nom de la variable	TumEventDiagnosisMethod
Définition	Méthode d'investigation et de confirmation du diagnostic de l'événement tumoral
Règle	Dès qu'une investigation avec un ordre plus élevé (cf. infra) est réalisée, cette modalité prime sur les codes inférieurs (basée sur les règles européennes de codage dans les registres).
Valeur attendue	1 = Clinique 2 = Investigation clinique (y compris imagerie et chirurgie) 4 = Marqueurs tumoraux spécifiques (ex : PSA, HCG) 5 = Cytologie 7 = Histologie 8 = Tests cytogénétiques ou moléculaires ou biologiques 9 = Inconnu
Caractère obligatoire ou facultatif	Optionnel

Etat général et examen physique (taille, poids, PS)

Pour chaque item, les informations à recueillir sont :

- Un type de mesure : poids, taille, performance status
- Une valeur ; poids en kilogrammes, taille en centimètre, performance status (code selon la classification OMS)
- Une date de la mesure décomposée en année, mois, jour (YYYY, MM, DD)

10.1 Type de mesures de l'état général et de l'examen physique

Nom de l'item	Nom de la variable mesurant l'état physique et l'état général
Numéro de l'item	10.1
Nom de la variable	MeasureType
Définition	Mesures sur l'état général du patient (PS) et son examen physique (taille, poids)
Valeur attendue de	PS (Performance Status) ; Poids ; Taille

la mesure	
Caractère obligatoire ou optionnel	Optionnel

10.2 Valeurs de la mesure de l'état général et de l'examen physique

Nom de l'item	Mesures sur l'état général et l'examen physique														
Numéro de l'item	10.2														
Nom de la variable	MeasureValue														
Définition	Valeur de la mesure sur l'état général du patient (PS) et son examen physique (taille, poids)														
Valeur attendue de la mesure	<p>Chaines de caractères</p> <p>PS : codes 0,1,2,3,4</p> <p>Poids : entier en kilogramme</p> <p>Taille : entier en centimètres</p>														
Règle	<p>La recommandation est de conserver toutes les mesures avec les dates. Cette recommandation concerne le PS, le poids, la taille.</p> <p>Concernant le PS, et si les médecins codent en index de Karnofsky (IK), la recommandation est de le transformer en PS selon la classification OMS et d'utiliser la table de correspondance entre PS et IK :</p> <table border="1"> <thead> <tr> <th>Score ECOG</th><th>Index de Karnofsky</th></tr> </thead> <tbody> <tr> <td>0</td><td>100</td></tr> <tr> <td>1</td><td>80-90</td></tr> <tr> <td>2</td><td>60-70</td></tr> <tr> <td>3</td><td>40-50</td></tr> <tr> <td>4</td><td>10-30</td></tr> <tr> <td>5</td><td>0</td></tr> </tbody> </table> <p>Lorsque la donnée est disponible sous forme structurée dans une source (logiciel de prescription, dossier de soins), on priorise cette source par rapport à une source de données non structurée.</p> <p>La donnée peut être récupérée par un traitement du langage dans les documents médicaux.</p>	Score ECOG	Index de Karnofsky	0	100	1	80-90	2	60-70	3	40-50	4	10-30	5	0
Score ECOG	Index de Karnofsky														
0	100														
1	80-90														
2	60-70														
3	40-50														
4	10-30														
5	0														
Justification du choix de la règle (exemple)	Si un médecin code un PS avec une valeur ambiguë, (exemple 1-2), on conserve la valeur la plus « péjorative » c'est-à-dire le chiffre le plus grand.														
Caractère obligatoire ou	Optionnel														

optionnel	
-----------	--

10.3 Date de recueil de la mesure

Nom de l'item	Date de recueil de la mesure
Numéro de l'item	10.3
Nom de la variable	MeasureDateDay, MeasureDateMonth, MeasureDateYear
Définition	La date à laquelle la mesure est recueillie
Valeur attendue de la mesure	YYYY - MM - DD
Règle	On prend chaque mesure associée à une date. Si une valeur est non disponible sur le jour (DD), le mois (MM) ou l'année (YYYY). Le champ correspondant sera laissé sans valeur.
Caractère obligatoire ou optionnel	Optionnel

Les traitements du cancer

Structuration de la chirurgie et de l'imagerie interventionnelle

Variable	Valeur attendue	Règle
Date de l'acte	YYYY - MM - DD	Date de l'acte dans le PMSI de l'établissement ou Date présente dans le CR opératoire ou le CR d'imagerie interventionnelle
Type d'acte	Chapitre CCAM (8-digit)	Chapitre de l'acte réalisé en CCAM ou Codage à partir du libellé repéré dans le CR opératoire ou le CR d'imagerie interventionnelle
Code précis de l'acte	Code de l'acte CCAM	Code de l'acte lorsque l'acte est réalisé (
Acte réalisé dans l'établissement	Y/N	Indication du lieu de réalisation de l'acte : Y – Réalisé dans l'établissement N – Réalisé en dehors de l'établissement
Source du document	Code interne	Structuré à la source TALN d'un document de l'établissement TALN d'un document extérieur

11.1 Date de l'acte Nom de l'item	Date de l'acte de chirurgie ou d'imagerie interventionnelle
Numéro de l'item	11.1
Nom de la variable	SurgeryDateDay ; SurgeryDateMonth ; SurgeryDateYear
Définition	Date de l'acte de chirurgie ou d'imagerie interventionnelle : date de l'acte dans le PMSI de l'établissement ou date présente dans le CR opératoire ou le CR d'imagerie interventionnelle
Valeur attendue	YYYY - MM - DD
Règle	<p>Si l'acte est réalisé dans l'établissement, la date est récupérée à partir des données du résumé de sortie standardisé du PMSI (la date de la chirurgie est précisée dans le résumé d'unité médicale).</p> <p>Si l'acte est réalisé en dehors de l'établissement, c'est la date du compte-rendu opératoire ou du compte-rendu d'imagerie qui prévaut. Si une valeur est non disponible sur le jour (DD), le mois (MM) ou l'année (YYYY), le champ correspondant sera laissé sans valeur.</p>
Caractère obligatoire ou optionnel	Obligatoire

11.2 Type d'acte (regroupement)

Nom de l'item	Type d'acte de chirurgie ou d'imagerie interventionnelle
Numéro de l'item	11.2
Nom de la variable	SurgeryType
Définition	Description de l'acte chirurgical ou d'imagerie interventionnelle
Nomenclature utilisée	CCAM – code hiérarchique sur 8 digits
Valeur attendue	900 codes environ constituent les codes hiérarchies
Justification du choix de la règle (exemple)	<p>Pour les actes réalisés en interne, on transcode le code CCAM précis vers le code de la hiérarchie. Il s'agit donc d'une variable dérivée du code CCAM.</p> <p>Pour les actes réalisés à l'extérieur, on propose de structurer l'information directement vers le code de la hiérarchie.</p>
Caractère obligatoire ou optionnel	Optionnel
Recommandations	Pour l'imagerie à visée interventionnelle qui est codée également dans les codes CCAM, il est recommandé de la coder dans la même table que la chirurgie.

11.3 Acte chirurgical ou d'imagerie interventionnelle

Nom de l'item	Acte réalisé dans l'établissement
Numéro de l'item	11.3
Nom de la variable	SurgeryProcedure
Définition	Description précise de l'acte chirurgical ou de l'acte d'imagerie interventionnelle réalisé dans l'établissement
Nomenclature utilisée	CCAM (environ 7000 codes)
Règle	On conserve les actes chirurgicaux non thérapeutiques (chirurgies reconstructrices ou prophylactiques)
Justification du choix de la règle (exemple)	Les actes réalisés en interne sont codés en CCAM à la source La CCAM est la liste des actes médicaux techniques, utilisant des codes, commune aux secteurs privé et public, qui permet une description précise de l'activité chirurgicale et de l'imagerie interventionnelle
Caractère obligatoire ou optionnel	Obligatoire
Recommandations	Un acte chirurgical ou d'imagerie interventionnelle est parfois décrit avec plusieurs codes CCAM. On n'agrège pas les codes et on aura une ligne par code CCAM avec une date. Plusieurs actes CCAM peuvent figurer pour une même date de chirurgie du CRO. Cela consiste à créer plusieurs lignes avec les mêmes dates.

11.4 Acte réalisé dans l'établissement

Nom de l'item	Acte réalisé dans l'établissement
Numéro de l'item	11.4
Nom de la variable	SurgeryOriginCenter
Définition	Indicateur du lieu de réalisation de l'acte
Valeur attendue	« Y » / « N » pour Oui / Non
Caractère obligatoire ou optionnel	Obligatoire
Remarques	Lorsque l'acte est réalisé dans l'établissement, la source de données est le PMSI de l'établissement. Sinon, l'acte de chirurgie ou d'imagerie interventionnelle est repérée par NLP dans, les comptes rendus opératoire ou les compte rendus d'imagerie interventionnelle.

11.5 Source de l'information

Nom de l'item	Source de l'information
Numéro de l'item	11.5
Nom de la variable	
Définition	Indicateur de la source d'information pour le codage des actes de chirurgie et d'imagerie interventionnelle
Valeur attendue	1 : Structuré dans le DPI et progiciels de soins 2 : Autre source structurée interne à l'établissement (Bases internes) 3 : Source structurée externe à l'établissement 4 : TALN d'un document texte ou PDF 5 : TALN d'un document image océrisé
Caractère obligatoire ou optionnel	Optionnel
Remarques	Lorsque l'acte est réalisé dans l'établissement, la source de données pour la chirurgie sera le PMSI de l'établissement. Sinon, l'acte de chirurgie ou l'acte d'imagerie interventionnelle est repérée par NLP dans des CR. On différencie les comptes rendus provenant de l'établissement des comptes rendus provenant de l'extérieur.

Structuration des traitements systémiques

Variable	Valeur attendue	Règle
Code la molécule	Code ATC	Toutes les données sont converties en code ATC
Nom de la molécule	Texte	Concerne tous les médicaments (avec ou sans codes ATC).
Date d'administration	YYYY – MM - DD	Date d'administration de toutes les molécules injectées ou Première date de prescription des thérapies orales
Lieu d'administration	Y/N	Indication du lieu d'administration des molécules injectables : Y – Réalisé dans l'établissement N – Réalisé en dehors de l'établissement
Source de la donnée	Code interne	Structuré TALN d'un document de l'établissement TALN d'un document extérieur

Accès précoce	Y/N	Concerne les médicaments prescrits avant l'autorisation de mise sur le marché (AMM) et la décision de remboursement
Essai clinique	Y/N	Concerne les patients qui sont inclus dans un essai clinique pour la thérapeutique en cours

12.1 Code de la molécule

Nom de l'item	Code de la molécule administrée
Numéro de l'item	12.1
Nom de la variable	DrugCode
Définition	Le code de la molécule correspond à un code ATC qui renvoie à la Dénomination commune internationale (DCI).
Nomenclature utilisée	Code ATC. Le Système de classification anatomique, thérapeutique et chimique (en anglais : Anatomical Therapeutic Chemical Classification System ou ATC) est publié par le Collaborating Centre for Drug Statistics Methodology de l'Organisation mondiale de la santé (OMS).
Règle	La classification ATC est préféré aux autres classifications nationales (Vidal, Base Claude Bernard, CIP, UCD).
Valeur attendue	Exemples : <ul style="list-style-type: none"> - L01XC18 (pembrolizumab) - L01CD01 (paclitaxel) - L01XY02 : association fixe de pertuzumab et trastuzumab
Caractère obligatoire ou optionnel	Optionnel

12.2 Nom de la molécule administrée

Nom de l'item	Nom de la molécule administrée
Numéro de l'item	12.2
Nom de la variable	DrugName
Définition	Label correspondant au code ATC. Si hors ATC (cas des molécules en phase d'essais cliniques non commercialisée et qui n'ont pas encore de code ATC).
Nomenclature utilisée	ATC si possible sinon Dénomination commune internationale (DCI).
Valeur attendue	Chaine de caractère
Règle	Quand la molécule administrée ne fait pas partie de la classification ATC et qu'elle est enregistrée dans le logiciel de prescription type CHIMIO, alors on la prend telle quelle.

Caractère obligatoire ou optionnel	Obligatoire
Recommandations	Les médicaments en accès précoces peuvent être inclus dans cette variable.

12.3 Traitement médical administré dans l'établissement

Nom de l'item	Traitement administré dans l'établissement
Numéro de l'item	12.3
Nom de la variable	DrugOriginCenter
Définition	Indicateur du lieu d'administration du traitement médical
Valeur attendue	Yes / No
Caractère obligatoire ou optionnel	Obligatoire

12.4 Source de l'information

Nom de l'item	Source de l'information des médicaments prescrits ou administrés
Numéro de l'item	12.4
Nom de la variable	
Définition	Indicateur de la source d'information pour le codage des médicaments
Valeur attendue	1 : Structuré dans le DPI et progiciels de soins 2 : Autre source structurée interne à l'établissement (Bases internes) 3 : Source structurée externe à l'établissement 4 : TALN d'un document texte ou PDF 5 : TALN d'un document image océrisé
Caractère obligatoire ou optionnel	Optionnel
Remarques	Lorsque le médicament est prescrit ou administré dans l'établissement, la source de données sera l'outil de prescription de l'établissement. Sinon, le médicament est repéré par NLP dans les comptes rendus médicaux On différencie les comptes rendus provenant de l'établissement des comptes rendus provenant de l'extérieur.

12.5 Date d'administration du traitement médical

Nom de l'item	Date
Numéro de l'item	12.5
Nom de la variable	DrugDateYear ; DrugDateMonth ; DrugDateDay
Définition	Date d'administration ou de prescription de la molécule
Valeur attendue	YYYY - MM - DD Si une valeur est non disponible sur le jour (DD), le mois (MM) ou l'année (YYYY). Le champ correspondant sera laissé sans valeur.
Caractère obligatoire ou optionnel	Optionnel
Recommandations	<p>Date d'administration pour les chimiothérapies et immunothérapies : toutes les dates pour chaque administration sont recueillies</p> <p>Date de prescription pour les traitements oraux (thérapies ciblées, hormonothérapie) ; seule la première date de prescription est récupérée</p> <p>Une information peut apparaître en doublon, à la fois dans la source structurée et dans le dossier médical.</p> <p>Si deux médicaments sont identiques, avec un intervalle de date inférieur à un seuil, la source privilégiée est la date issue de la source structurée.</p> <p>Un intervalle de 14 jours est la borne recommandée, dans la mesure peu de protocoles ont un intervalle entre deux cycles inférieurs à 21 jours.</p>

12.6 Essai clinique

Nom de l'item	Essai clinique
Numéro de l'item	12.6
Nom de la variable	ClinicalTrial
Définition	Le traitement s'inscrit dans le contexte d'un essai clinique
Valeur attendue	Y / N
Caractère obligatoire ou optionnel	Optionnel
Recommandation	L'information est recueillie que l'essai ait lieu dans l'établissement ou en dehors de l'établissement. L'information peut être structurée à la source ou récupérée dans le texte par TALN.

12.7 Médicament en accès précoce

Nom de l'item	Accès précoce
---------------	---------------

Numéro de l'item	12.7
Nom de la variable	EarlyAccess
Définition	Le traitement s'inscrit dans le contexte d'un accès précoce
Valeur attendue	Y / N
Caractère obligatoire ou optionnel	Optionnel

13 – Radiothérapie

Note : les variables présentes dans ce chapitre seront amenées à évoluer pour correspondre aux variables minimales permettant d'identifier une observation dans le modèle OSIRIS - RADIOTHERAPIE.

Structuration de la radiothérapie

Variable	Valeur attendue	Règle
Type de radiothérapie	Codes CCAM	Acte CCAM issu du RSS séance OU Codage à partir du libellé repéré dans le CR de fin de traitement ou les comptes rendus médicaux
Date de début de la radiothérapie	YYYY - MM - DD	Date de début pour l'ensemble du traitement
Topographies des organes irradiés	CIMO-3	Données multivaluées tenant compte de l'ensemble des organes irradiées
Radiothérapie réalisée dans l'établissement	Y/N	Indication du lieu de réalisation de la radiothérapie : Y – Réalisé dans l'établissement N – Réalisé en dehors de l'établissement
Source de la donnée	Code interne	Structuré TALN d'un document de l'établissement TALN d'un document extérieur
Lien vers le MDS OSIRIS RT		

13.1 Type de radiothérapie

Nom de l'item	Type de traitement de radiothérapie
Numéro de l'item	13.1

Nom de la variable	
Définition	Description de la technique de radiothérapie
Nomenclature utilisée	CCAM
Caractère obligatoire ou optionnel	Obligatoire
Recommandations	<p>La radiothérapie réalisée en ambulatoire est codée en « RSS séances » pour le PMSI.</p> <p>La radiothérapie peut comporter plusieurs actes possibles pour un traitement.</p> <p>Le regroupement est réalisé à l'analyse.</p> <p>Un traitement de radiothérapie peut être administré en une seule séance (radiochirurgie) soit une suite de séance de radiothérapie.</p>

13.2 Date de début de la radiothérapie

Nom de l'item	Date de début de la radiothérapie
Numéro de l'item	13.2
Nom de la variable	
Définition	Date de début du traitement de radiothérapie pour l'ensemble du traitement réalisé
Valeur attendue	<p>YYYY - MM - DD</p> <p>Si une valeur est non disponible sur le jour (DD), le mois (MM) ou l'année (YYYY). Le champ correspondant sera laissé sans valeur.</p>
Recommandations	<p>On ne considère que la date de début de traitement pour l'ensemble du traitement (« course » en anglais).</p> <p>Exemple : si un traitement comporte une radiothérapie du sein qui commence à une date x, et que pendant le traitement une radiothérapie des ganglions axillaires est réalisée quelques jours plus tard, on ne garde que la date de début.</p> <p>En cas d'interruption de traitement pour effets indésirables, on garde toujours la date du début</p> <p>Cependant, si une radiothérapie est débutée avec un intervalle supérieur à x jours, on considère qu'il existe deux traitements avec deux dates de début de traitement. On propose que x soit fixé à 30 jours.</p>
Caractère obligatoire ou optionnel	Optionnel

13.3 Topographie des organes irradiés

Nom de l'item	Topographie des organes irradiés
Numéro de l'item	13.3
Nom de la variable	
Définition	Ensemble des sites topographies où les traitements de radiothérapie ont été délivrés.
Nomenclature utilisée	CIM-10 axe topographique
Caractère obligatoire ou optionnel	Optionnel
Recommandations	<p>L'information est à récupérer en priorité dans le logiciel de radiothérapie ou dans le PMSI</p> <p>Le choix dépend de l'établissement en fonction du niveau de confiance dans le codage à la source.</p> <p>En l'absence d'information structurée de qualité, l'information est récupérée dans les compte rendus médicaux ou dans le compte rendu de fin de radiothérapie</p>
Options	Un code pour la radiothérapie du corps entier est nécessaire

13.4 Traitement de radiothérapie réalisé dans l'établissement

Nom de l'item	Traitement de radiothérapie réalisé dans l'établissement
Numéro de l'item	13.4
Nom de la variable	
Définition	Indicateur du lieu de réalisation de la radiothérapie
Nomenclature utilisée	Booléen
Valeur attendue	Yes / No
Caractère obligatoire ou optionnel	Optionnel
Recommandations	Le système peut se baser sur l'information du PMSI c'est-à-dire sur la source de données. Si le traitement est réalisé dans l'établissement, il

	<p>existe un RSS dit « séance ».</p> <p>En l'absence de données issues du PMSI, la radiothérapie peut être considérée en dehors de l'établissement.</p> <p>Un traitement de radiothérapie débuté est toujours poursuivi dans le même établissement.</p>
--	---

13.5 Source de l'information

Nom de l'item	Source de l'information
Numéro de l'item	13.5
Nom de la variable	
Définition	Indicateur de la source d'information pour le codage des actes de radiothérapie
Valeur attendue	<p>1 : Structuré dans le DPI et progiciels de soins</p> <p>2 : Autre source structurée interne à l'établissement (Bases internes)</p> <p>3 : Source structurée externe à l'établissement</p> <p>4 : TALN d'un document texte ou PDF</p> <p>5 : TALN d'un document image océrisé</p>
Caractère obligatoire ou optionnel	Optionnel
Remarques	<p>Lorsque la radiothérapie est réalisée dans l'établissement, la source de données sera le PMSI de l'établissement.</p> <p>Sinon, la radiothérapie est repérée par NLP dans les CR médicaux.</p> <p>On différencie les comptes rendus provenant de l'établissement des comptes rendus provenant de l'extérieur.</p>

13.6 Lien vers le set de données radiothérapie V1.0

Nom de l'item	Lien vers le set de données radiothérapie V1.0
Numéro de l'item	13.6
Définition	<p>Lien avec le DICOM RT du patient</p> <p>Dans le cas où le traitement est réalisé dans l'établissement, il est recommandé d'alimenter la source OSIRIS RADIOTHERAPIE pour récupérer de façon automatique les données au format DICOM RT</p>
Règle	On récupère dans le dossier DICOM RT du patient ce qui a été délivré avec la date de début du traitement.

Caractère obligatoire ou optionnel	Optionnel
------------------------------------	-----------

14. Réponse, résistance ou progression

14.1 Progression

Nom de l'item	Progression
Numéro de l'item	14.1
Nom de la variable	
Définition	Le terme "progression" désigne l'évolution d'un cancer lorsque la maladie évolue défavorablement malgré le traitement : augmentation de volume de la tumeur ou des métastases, apparition de nouvelles lésions ou de métastases
Valeur attendue	Progression = Y
Règle	L'information étant rarement structurée à la source, elle est structurée par TALN. On notera qu'il existe plusieurs « synonymes » utilisés par les médecins qui signifient qu'il existe une résistance à un traitement
Caractère obligatoire ou optionnel	Obligatoire

14.2 Date de la progression

Nom de l'item	Date de la progression
Numéro de l'item	14.2
Nom de la variable	
Définition	Date qui permet de déterminer que la maladie évolue défavorablement malgré le traitement
Valeur attendue	YYYY - MM – DD
Règle	La problématique est de déterminer par convention la date de progression

	<p>Il serait logique de proposer de retenir comme date de progression la date qui permet de définir que l'item progression a été codé à oui (Y)</p> <p>Dans la mesure où l'information recueillie par TALN est difficilement reproductible, on propose de définir comme date de progression la date de début de la ligne suivante si un traitement est administré</p> <p>En l'absence de traitement, par exemple si le traitement est suspendu ou définitivement arrêté, on propose de garder la date du premier compte-rendu qui évoque la notion de progression ou de résistance au traitement.</p>
Caractère obligatoire ou optionnel	Obligatoire

15 - Imagerie et médecine nucléaire

Structuration de l'imagerie médicale

Variable	Valeur attendue	Règle
Date de l'examen	YYYY - MM - DD HH - MM	Date et heure de l'examen d'imagerie
Modalités d'imagerie	Codification DICOM	Code DICOM issu du PACS Ou Codage à partir du libellé repéré dans le CR d'imagerie
Topographies des organes étudiés	Codification propre	Codage à partir de la données DICOM dans le champ « study » ou À partir du compte-rendu d'imagerie.
Imagerie réalisée dans l'établissement	Y/N	Indication du lieu de réalisation de la radiothérapie : Y – Réalisé dans l'établissement N – Réalisé en dehors de l'établissement
Source de la donnée	Code interne	Structuré TALN d'un document de l'établissement TALN d'un document extérieur
Lien vers le MDS OSIRIS imagerie		Study_ID + Date + ID_patient

15.1 Date de l'examen d'imagerie

Nom de l'item	Date de réalisation ou d'acquisition de l'examen d'imagerie
Numéro de l'item	15.1
Nom de la variable	
Définition	Date et heure de réalisation ou d'acquisition de l'examen d'imagerie
Valeur attendue	Un champ date (YYYY - MM – DD) et un champ heure (HH - MM)
Règle	<p>La date de l'acte d'imagerie est la date de la réalisation de l'acte.</p> <p>Si la date ne peut être récupérée à partir du PACS, c'est la date du compte-rendu d'imagerie qui prévaut/</p> <ul style="list-style-type: none">- YYYY - MM - DD et HH - MM car nécessaire pour identifier une suite d'images si l'information est récupérée à partir des données du PACS- YYYY - MM - DD si la source est le compte rendu <p>Si une valeur est non disponible sur le jour (DD), le mois (MM) ou l'année (YYYY). Le champ correspondant sera laissé sans valeur.</p>
Caractère obligatoire ou facultatif	Obligatoire
Recommandations	<p>Si la source n'est pas le PACS ou si l'examen d'imagerie a été fait dans un autre établissement ou un cabinet privé, la date de l'examen est récupérée dans le compte-rendu d'imagerie ou récupéré de l'établissement extérieur à l'établissement.</p> <p>La recommandation est de récupérer toutes les dates de tous les examens d'imagerie du patient.</p> <p>La recommandation est d'extraire la date des données DICOM de l'imagerie correspondante.</p>

15.2 Modalité utilisée pour l'examen

Nom de l'item	Modalité à l'origine de l'examen d'imagerie
Numéro de l'item	15.2
Nom de la variable	
Définition	Type d'équipement utilisé permettant l'acquisition des images
Nomenclature utilisée	Codification DICOM

Valeur attendue	Liste de l'ensemble des modalités utilisées
Caractère obligatoire ou facultatif	Obligatoire
Recommandations	La recommandation est l'utilisation du DICOM pour automatiser le remplissage de cet item en récupérant l'information depuis le tag DICOM correspondant.

15.3 Topographie des organes examinés

Nom de l'item	Topographie des organes examinés
Numéro de l'item	15.3
Nom de la variable	
Définition	Ensemble des sites topographies examinés
Nomenclature utilisée	Propre à OSIRIS RWD
Caractère obligatoire ou optionnel	Optionnel
Recommandations	Données tenant compte de l'ensemble des organes étudiées Codage à partir de la données DICOM dans le champ « study » ou À partir du compte-rendu d'imagerie.

15.4 Imagerie réalisée dans l'établissement

Nom de l'item	Imagerie réalisée dans l'établissement
Numéro de l'item	15.4
Nom de la variable	
Définition	Indicateur du lieu de réalisation de l'imagerie médicale
Nomenclature utilisée	Booléen
Valeur attendue	Yes / No
Caractère obligatoire ou optionnel	Optionnel

Recommandations	L'information est issue des fichiers DICOM où figure le nom de l'établissement
-----------------	--

15.5 Source de l'information

Nom de l'item	Source de l'information
Numéro de l'item	15.5
Nom de la variable	
Définition	Indicateur de la source d'information pour le codage des actes d'imagerie
Valeur attendue	1 : Structuré dans le DPI et progiciels de soins 2 : Autre source structurée interne à l'établissement (Bases internes) 3 : Source structurée externe à l'établissement 4 : TALN d'un document texte ou PDF 5 : TALN d'un document image océrisé
Caractère obligatoire ou optionnel	Optionnel
Remarques	Lorsque l'examen d'imagerie est réalisé dans l'établissement, la source de données sera le PACS de l'établissement. On différencie les comptes rendus provenant de l'établissement des comptes rendus provenant de l'extérieur.

15.6 Lien vers le set de données du modèle OSIRIS imagerie

Nom de l'item	Lien vers le set de données imagerie
Numéro de l'item	15.6
Définition	Numéro unique permettant d'identifier l'étude associé à une date et au patient
Valeur attendue	Study_ID + Date + ID_patient
Règle	Nous récupérons le numéro unique du PACS lorsque l'image est réalisée dans l'établissement ou lorsqu'elle est acquise à l'extérieur et intégrée au PACS. Dans ce dernier cas, les informations disponibles sont identiques à celles d'une image capturée dans l'établissement.
Caractère	Optionnel

obligatoire ou facultatif	
---------------------------	--

16 - Les prélèvements biologiques

Les six informations permettant de décrire les prélèvements biologiques

1. Identifiant (ou numéro) du prélèvement
2. Lieu de conservation du prélèvement
3. Date du prélèvement
4. Type du prélèvement (pièce opératoire, biopsie, etc.)
5. Nature du prélèvement (tumoral, tissu sain)
6. Topographie du prélèvement (ex : sein, métastase pulmonaire)

16.1 Identifiant du prélèvement

Nom de l'item	Identifiant du prélèvement biologique
Numéro de l'item	16.1
Nom de la variable	
Définition	Identifiant du prélèvement dans l'établissement
Valeur attendue	Chaîne de caractères
Caractère obligatoire ou optionnel	Obligatoire
Recommandations	<p>Seuls les prélèvements pris en charge dans l'établissement où le patient est suivi sont structurés dans la base OSIRIS RWD, ce qui inclut les biopsies, les prélèvements réalisés lors d'un acte chirurgical, une prise de sang, un prélèvement de liquide (urine, liquide pleural, ascite, liquide céphalo-rachidien).</p> <p>La pseudonymisation de l'identifiant est réalisé en amont de l'alimentation dans la base OSIRIS</p>
Options	Les prélèvements qui viendraient d'un établissement extérieur peuvent être enregistrés et répertoriés avec un numéro interne à l'établissement lorsqu'une relecture histologique ou lorsqu'une analyse génomique sur échantillon sont réalisées.

16.2 Lieu de conservation du prélèvement

Nom de l'item	Lieu de conservation du prélèvement
---------------	-------------------------------------

Numéro de l'item	16.2
Nom de la variable	
Définition	Lieu de conservation du prélèvement (dans l'établissement / en dehors)
Nomenclature utilisée	Liste fermée
Valeur attendue	<ul style="list-style-type: none"> - Dans l'établissement - Hors de l'établissement
Règle	Il est important de faire une distinction entre les prélèvements conservés dans l'établissement et ceux conservés en dehors pour déterminer si de nouvelles analyses sur échantillon à partir des prélèvements sont possibles, notamment pour les études de faisabilité.
Caractère obligatoire ou optionnel	Optionnel

16.3 Date du prélèvement

Nom de l'item	Date du prélèvement
Numéro de l'item	16.3
Nom de la variable	
Définition	Date de prélèvement (de la biopsie ou de l'acte chirurgical, de la prise de sang)
Valeur attendue	YYYY - MM - DD
Règle	Si une valeur est non disponible sur le jour (DD), le mois (MM) ou l'année (YYYY). Le champ correspondant sera laissé sans valeur.
Caractère obligatoire ou facultatif	Obligatoire

16.4 Type du prélèvement

Nom de l'item	Type du prélèvement
Numéro de l'item	16.4
Nom de la variable	
Définition	Information indiquant le type de prélèvement

Nomenclature utilisée	Liste fermée utilisant la classification UMLS (Unified Medical Language System)
Valeur attendue	<ul style="list-style-type: none"> • Pièce opératoire • Biopsie • Prélèvement cellulaire • Liquide (comprenant le sang, la plèvre, le liquide d'ascite, le LCR) • Inconnu
Caractère obligatoire ou optionnel	Obligatoire

16.5 Nature du prélèvement

Nom de l'item	Nature du prélèvement
Numéro de l'item	16.5
Nom de la variable	
Définition	Information indiquant la nature du prélèvement
Nomenclature utilisée	Liste fermée utilisant la classification UMLS (Unified Medical Language System)
Valeur attendue	<ul style="list-style-type: none"> • Tissu tumoral • Tissu sain
Caractère obligatoire ou optionnel	Optionnel

16.6 Topographie du prélèvement

Nom de l'item	Topographie du prélèvement
Numéro de l'item	16.6
Nom de la variable	
Définition	Localisation anatomique du prélèvement selon la Classification Internationale des Maladies pour l'Oncologie.
Nomenclature utilisée	CIMO-3
Valeur attendue	<p>Il est suffisant de se limiter au premier niveau de topographie de la classification CIMO-3.</p> <p>Exemples :</p>

	C22 Foie C34 Poumon
Caractère obligatoire ou optionnel	Optionnel

17. Les biomarqueurs et les marqueurs tumoraux

Les informations à recueillir sont les suivantes :

1. Type de biomarqueurs (tumoral, constitutionnel, marqueur tumoral)
2. Nom du biomarqueur tissulaire ou moléculaire
3. Technique utilisée pour l'analyse du biomarqueur (ex : IHC, séquençage)
4. Statut du biomarqueur (ex : positif ; négatif ; altération du gène ; pas d'altération du gène, résultat non contributif)
5. Information de précision sur le biomarqueur (ex : exon 18) en texte libre
6. Information de précision en numérique (par exemple % ou valeur numérique du marqueur sérique)
7. Unité utilisée
8. Date du résultat (ou du prélèvement)
9. Nature du prélèvement (prélèvement sanguin, épanchement, pleural, LCR, salive).

17.1. Type du biomarqueur (tissulaire ou moléculaire ou sérique)

Nom de l'item	Type du biomarqueur (tissulaire ou moléculaire ou sérique)
Numéro de l'item	17.1
Nom de la variable	BiomarkerType
Définition	Classement des biomarqueurs selon leur type et leur nature
Nomenclature utilisée	<ul style="list-style-type: none"> - Tumoral, - Constitutionnel, - Marqueur tumoral
Caractère obligatoire ou optionnel	Optionnel

17.2 Nom du biomarqueur

Nom de l'item	Nom du biomarqueur tissulaire ou moléculaire ou marqueur tumoral
---------------	--

Numéro de l'item	17.2
Nom de la variable	BiomarkerName
Définition	Nom du biomarqueur
Nomenclature utilisée	LOINC
Caractère obligatoire ou optionnel	Obligatoire
Recommandations	Le référentiel doit permettre un niveau de détail conforme à ce qui est observé dans les soins courants.

17.3 Technique utilisée pour l'analyse du biomarqueur tissulaire ou moléculaire

Nom de l'item	Type d'analyse
Numéro de l'item	17.3
Nom de la variable	BiomarkerAnalysis
Définition	Type d'analyse utilisée pour la détection du biomarqueur moléculaire ou tissulaire
Nomenclature utilisée	LOINC
Valeur attendue	IHC FISH CISH
Caractère obligatoire ou optionnel	Optionnel

17.4 Statut du biomarqueur

Nom de l'item	Statut du biomarqueur
Numéro de l'item	17.4
Nom de la variable	BiomarkerValueTxt
Définition	Valeur d'expression du biomarqueur
Valeur attendue	<ul style="list-style-type: none"> - Résultat positif (par exemple pour une mutation de gène) - Résultat négatif (pas de mutation du gène) - Résultat non contributif.

	- Information manquante (ou examen on réalisé=
Caractère obligatoire ou optionnel	Optionnel
Recommandations	Il est important de présenter les résultats positifs mais également les résultats négatifs ou non contributifs.

17.5 Précision du biomarqueur

Nom de l'item	Précisions du biomarqueur
Numéro de l'item	17.5
Nom de la variable	BiomarkerPrecision
Définition	Information de précision sur les résultats du biomarqueur
Nomenclature utilisée	LOINC
Caractère obligatoire ou optionnel	Optionnel
Recommandations	<p>Il est utile de coder les variants ou les exons dans certains cas, notamment ceux concernant les gènes mutés lorsque cette information est disponible cliniquement dans les dossiers médicaux.</p> <p>A titre d'exemple, il est utile de récupérer les variants cliniquement pertinents dans la pratique, tels que BRAF V600E ou KRAS G12C.</p>

17.6 Résultat numérique de l'analyse du marqueur tumoral

Nom de l'item	Valeur numérique de l'analyse
Numéro de l'item	17.6
Nom de la variable	BiomarkerValueNum
Définition	Valeur d'expression du marqueur tumoral
Valeur attendue	Valeur numérique du marqueur tumoral.
Caractère obligatoire ou optionnel	Optionnel

17.7 Unité

Nom de l'item	Unité
Numéro de l'item	17.7
Nom de la variable	BiomarkerUnit
Définition	Unité utilisée pour les résultats du biomarqueur
Caractère obligatoire ou optionnel	Optionnel

17.8 Date du prélèvement

Nom de l'item	Date du prélèvement
Numéro de l'item	17.8
Nom de la variable	?
Définition	Date du résultat pour les biomarqueurs moléculaires et tissulaires
Valeur attendue	YYYY - MM - DD
Règle	<p>Les dates des résultats des marqueurs tumoraux sont essentielles car elles permettent un suivi de l'évolution dans le temps.</p> <p>Lorsque la source biologie est disponible, la date de prélèvement est privilégiée.</p> <p>Cependant, lorsqu'une analyse est effectuée en dehors de l'établissement, les dates sont extraites des dossiers médicaux à l'aide de techniques de traitement du langage naturel (TALN), privilégiant la date du compte rendu.</p>
Justification du choix de la règle (exemple)	<p>Lorsque l'analyse est effectuée en dehors de l'établissement, la date du compte rendu est toujours privilégiée</p> <p>Dans certains cas, l'analyse moléculaire peut être réalisée plusieurs mois, voire années, après la date de la biopsie. La décision du clinicien concernant le début d'une thérapie spécifique est basée sur les résultats de l'analyse, et donc sur cette date.</p> <p>Ainsi, pour aligner la date sur la décision clinique, la date du compte rendu est utilisée comme référence.</p>
Caractère obligatoire ou facultatif	Optionnel

17.9 Nature du prélèvement

Nom de l'item	Nature du prélèvement pour le dosage du biomarqueur
Numéro de l'item	17.9
Définition	Nature du prélèvement pour le dosage du marqueur tumoral, constitutionnel ou sérique
Valeur attendue	Sérum, LCR, liquide pleural, salive
Caractère obligatoire ou optionnel	Optionnel

18. Données complémentaires : la table observation

Il est créé deux tables pour des données complémentaires non prévues dans le MDS : la table « observation » et la table « mesure ».

Les champs de la table observation

Variable	Valeur attendue	Règle et exemples
Type d'observation	Code selon une nomenclature	(Exemples : type de tabagisme, sevré ou non)
Nomenclature utilisée	Code de la nomenclature	SNOMED, LOINC, etc.
Date de l'observation	YYYY - MM - DD	
Valeur de l'observation	Donnée structurée	Exemples : - Tabagisme actif / passif - Tabagisme sevré / non sevré
Concept auquel l'information est rattachée	Patient Tumeur Evolution tumorale Traitement.	Par défaut, l'information est rattachée au patient
Source de la donnée	Code interne	Structuré TALN d'un document de l'établissement TALN d'un document extérieur

18.1 Type d'observation

Nom de l'item	Type d'observation ou nom de la variable
Numéro de l'item	18.1
Nom de la variable	
Définition	Information qualitative ne faisant pas parti du MDS
Valeur attendue	Donnée alphanumérique
Recommandations	<p>Dans la table observation, chaque établissement définit en fonction de ses projets l'observation à récupérer. Dans les projets multicentriques, il est judicieux que le nom attribué au type d'observation soit identique.</p> <p>Par exemple, si une étude s'intéresse au tabagisme, la seule information disponible dans le MDS est tabagisme oui/non.</p> <p>Il serait utile dans une table observation de différencier les items suivants qui sont autant de type d'information différente :</p> <ul style="list-style-type: none">- Tabagisme actif / tabagisme passif- Tabagisme sévère / tabagisme non sévère

18.2 Nomenclature utilisée

Nom de l'item	Nomenclature utilisée pour structurer l'observation
Numéro de l'item	18.2
Nom de la variable	
Définition	Type d'information selon une nomenclature
Valeur attendue	Donnée structurée (SNOMED, LOINC, CCAM, CIMO-3, nomenclature propre à l'étude, etc.)
Recommandations	On privilégie les nomenclatures internationales

18.3 Date de l'observation

Nom de l'item	Date de l'observation
---------------	-----------------------

Numéro de l'item	18.3
Nom de la variable	
Valeur attendue	YYYY - MM - DD

18.4 Valeur de l'observation

Nom de l'item	Valeur de l'observation
Numéro de l'item	18.4
Nom de la variable	
Définition	Valeur de la variable correspondant au type d'observation à une date donnée
Valeur attendue	Donnée structurée

18.5 Concept de rattachement de l'observation

Nom de l'item	Concept de rattachement de l'observation
Numéro de l'item	18.5
Nom de la variable	
Définition	Rattachement de la variable correspondant au type d'observation à une table du MDS
Valeur attendue	Donnée structurée faisant référence à une des tables du modèle
Recommandations	<p>Par défaut, le type d'observation et sa valeur sont rattachés au patient.</p> <p>Pour certaines données, il est possible de disposer du meilleur rattachement possible de la valeur, par exemple à une tumeur, à une évolution tumorale ou à un traitement.</p>

18.6 Source de la donnée

Nom de l'item	Source de la donnée
Numéro de l'item	18.6
Nom de la variable	

Définition	Indicateur de la source d'information pour l'alimentation de la donnée de l'observation
Valeur attendue	1 : Structuré dans le DPI et progiciels de soins 2 : Autre source structurée interne à l'établissement (Bases internes) 3 : Source structurée externe à l'établissement 4 : TALN d'un document texte ou PDF 5 : TALN d'un document image océrisé
Caractère obligatoire ou optionnel	Optionnel

19. Données complémentaires : la table mesure

Les champs de la table Mesure

Variable	Valeur attendue	Règle et exemples
Type de mesure (attribut)	Code selon une nomenclature	Créatininémie
Date de la mesure	YYYY – MM – DD	
Valeur de la mesure	Nombre	Exemple : 110
Unité de mesure	Code	Micromoles/litre
Concept auquel l'information est rattachée	Patient Tumeur Evolution tumorale Traitement.	Par défaut, l'information est rattachée au patient
Source de la donnée	Code interne	Structuré TALN d'un document de l'établissement TALN d'un document extérieur

19.1 Type de mesure

Nom de l'item	Type de mesure ou nom de la variable
Numéro de l'item	19.1
Nom de la variable	
Définition	Type de mesure ne faisant pas parti du MDS
Valeur attendue	Donnée alphanumérique

Recommandations	<p>Dans la table mesure, chaque établissement définit en fonction de ses projets l'observation à récupérer. Dans les projets multicentriques, il est judicieux que le nom attribué au type de la mesure soit identique.</p> <p>Par exemple, si une étude s'intéresse à l'insuffisance rénale comme événement indésirable, et que certaines variables ne sont pas récupérées, il est créé une variable dans la table mesure</p> <ul style="list-style-type: none"> - Protéinurie des 24 h
-----------------	---

19.2 Unité de la mesure

Nom de l'item	Unité de la mesure
Numéro de l'item	19.2
Nom de la variable	
Nomenclature utilisée	Les unités de mesure en biologie peuvent être codées avec UMLS
Valeur attendue	Exemple Unité de mesure : Milligramme par décilitre (mg/dl) : C0025384
Recommandations	

19.3 Date de la mesure

Nom de l'item	Date de la mesure
Numéro de l'item	19.3
Nom de la variable	
Valeur attendue	YYYY - MM - DD

19.4 Valeur de la mesure

Nom de l'item	Valeur de l'observation
Numéro de l'item	19.4
Nom de la variable	
Définition	Valeur de la variable correspondant au type de mesure à une date donnée
Valeur attendue	Donnée structurée

19.5 Concept de rattachement de la mesure

Nom de l'item	Concept de rattachement de la mesure
Numéro de l'item	19.5
Nom de la variable	
Définition	Rattachement de la variable correspondant au type de mesure à une des tables du MDS
Valeur attendue	Donnée structurée faisant référence à une des tables du modèle
Recommandations	<p>Par défaut, le type de mesure et sa valeur sont rattachés au patient.</p> <p>Pour certaines données, il est possible de disposer du meilleur rattachement possible de la valeur, par exemple à une tumeur, à une évolution tumorale ou à un traitement.</p>

19.6 Source de la donnée

Nom de l'item	Source de la donnée
Numéro de l'item	19.6
Nom de la variable	
Définition	Indicateur de la source d'information pour l'alimentation de la mesure
Valeur attendue	<p>1 : Structuré dans le DPI et progiciels de soins</p> <p>2 : Autre source structurée interne à l'établissement (Bases internes)</p> <p>3 : Source structurée externe à l'établissement</p> <p>4 : TALN d'un document texte ou PDF</p> <p>5 : TALN d'un document image océrisé</p>
Caractère obligatoire ou optionnel	Optionnel

PARTICIPANTS AU GROUPE DE TRAVAIL (V1)

1. Alain Livartowski (Médecin Oncologue, Unicancer, coordonnateur)
2. Olivier Carsenti (PSCC, coordinateur-adjoint)
3. Emmanuel Desandes (Institut de cancérologie de Lorraine)
4. Toufik Belguendouz Bendahmane (Médecin Oncologue)
5. Alexandre Yazigi (Chirurgien Oncologue CEO, Gimli)
6. Ali Bellamine (Médecin Santé Publique Data scientist, Elsan)
7. Aurélie Bertaut (Médecin - Centre Georges Francois Leclerc)
8. Mathieu Robain (Directeur scientifique, Unicancer)
9. Vincent Le Texier (Chef de Projet Données de santé, Fondation Synergie Lyon Cancer)
10. Julie Perlberg Samson (Médecin DIM, Institut Curie)
11. Francois Lion (Gustave Roussy)
12. Laetitia Chanas, (Direction des data, Institut curie)
13. Marc Deloger (Responsable service bio-informatique, Gustave Roussy)
14. Myriam Kossai (Médecin anatomo-pathologiste, CLCC Clermont Unicancer)
15. Elodie Pleyner (Centre Lyon Bérard - Unicancer)
16. Yec'han Laizet (Institut Bergonié)
17. Guillaume Danton (Data Scientist, Centre Eugène Marquis)
18. Julien Huyard (Data Manager, Unicancer)
19. Samantha Padeloup (Elsan)
20. Raphaël Sigogne (Responsable Data, Paris Saclay Cancer Cluster)
21. Hugo Crochet (Directeur des Systèmes d'Information, Centre Léon Bérard)
22. Pierre Heudel (Médecin Oncologue, Centre Lyon Bérard)
23. Francois Bocquet (Institut de Cancérologie de l'Ouest Nantes-Angers)

PARTICIPANTS AU COMITE DE RELECTURE (V2)

1. Alain Livartowski (Médecin Oncologue, PSCC, coordonnateur)
2. Emmanuel Desandes (Institut de cancérologie de Lorraine)
3. Aurélie Bertaut (Médecin - Centre Georges Francois Leclerc)
4. Julie Perlberg Samson (Médecin DIM, Institut Curie)
5. Laetitia Chanas, (Direction des data, Institut curie)
6. Julien Huyard (Data Manager, Unicancer)
7. Mohammed Hurera (Attaché de recherche clinique, Institut Curie)
8. Emilie BENSADON (Cheffe de projet, Unicancer)
9. Federica MARTINA (Unicancer)

Glossaire

AINS : anti-inflammatoire non stéroïdien

CIM-10 : classification internationale des maladies – 10^{ème} révision

CIMO-3 : Classification internationales des maladies pour l'oncologie – 3^{ème} édition

CNIL : Commission Nationale de l'informatique et des libertés

DICOM : Digital imaging and communications in medicine

EDS : entrepôt de données de santé

ETS : Etablissement de santé

INCa : Institut National du cancer

INSEE : Institut national de la statistique et des études économiques

LCR : liquide céphalo-rachidien

MDS : minimum data set

MTEV : maladie thrombo-embolique veineuse

PACS : Picture archiving and communication system

PMSI : Programme de médicalisation des systèmes d'information

RGPD : règlement général pour la protection des données

RSA : résumé de sortie anonymisé

RSS : résumé standardisé de sortie

SIH : système d'information hospitalier

TALN : traitement automatique du langage naturel

UMLS : Unified Medical Language System

ANNEXES

Le codage des cancers

La CIM-10 (Classification Internationale des Maladies, 10ème révision) et la CIMO-3 (Classification Internationale des Maladies pour l'Oncologie, 3ème édition) sont utilisées pour classer les tumeurs mais elles ont des finalités différentes.

CIM-10 (Classification Internationale des Maladies - 10ème révision)

La CIM-10 est une classification internationale des maladies et des problèmes de santé. Elle est utilisée pour coder une large gamme de diagnostics, les cancers mais également les autres maladies, les affections, les symptômes et autres états de santé.

Les codes CIM-10 sont alphanumériques, composés d'une lettre suivie de chiffres (par exemple, C34.1 pour "cancer du lobe supérieur du poumon").

La CIM-10 est utilisée pour les statistiques sanitaires, en épidémiologie et pour le remboursement des soins de santé. Elle est utilisée en France pour le codage des résumés de sortie standardisés dans le cadre du PMSI. Un des avantages est que cette donnée est structurée, qualifiée, systématique (sauf pour les consultations externes), partagée par tous les systèmes d'information et toutes les structures (publiques ou privées) en France. Un point particulier : la CIM-10 différencie tumeurs primitives et métastases (codées en C77, C78, C79). La CIM-10 est aussi utilisée par les médecins pour le codage des localisations tumorales (fiche RCP, fiche « tumeur », etc).

Il est tentant d'utiliser cette information structurée à la source mais il existe de nombreuses limites à la CIM-10 notamment pour les maladies hématologiques (leucémies, lymphomes), les sarcomes et les tumeurs cutanées (mélanomes).

Un autre inconvénient de la CIM10 est de mélanger des codes topographiques et des codes morphologiques, par exemple pour les lymphomes C81 à C96 avec la difficulté par exemple de coder un lymphome localisé du sein. Autre limite dans la mesure où la CIM-10 tient compte du comportement évolutif de la tumeur car on trouve des codes topographiques différents pour les cancers, les cancers in situ (D00-D09), les tumeurs bénignes, les tumeurs à évolution imprévisible.

CIMO-3 (Classification Internationale des Maladies pour l'Oncologie – 3ème édition)

La CIMO-3 est spécifiquement utilisée pour coder les cancers. Elle comprend deux axes : l'axe topographique (qui décrit le site anatomique) et l'axe morphologique (qui décrit le type histologique de la tumeur).

La CIMO-3 est principalement utilisée par les registres du cancer pour coder les caractéristiques des tumeurs, leur site anatomique (topographie) et leur type histologique (morphologies).

Les pathologistes utilisent prioritairement les codes CIMO-3 (ou ADICAP) dans les comptes rendus d'anatomo-cytopathologie, sachant qu'il existe une correspondance ADICAP/CIMO-3.

CIMO-3 - axe topographique

Les codes topographiques en CIMO-3 sont alphanumériques, commençant par la lettre "C" suivie de deux chiffres pour représenter la localisation principale de la tumeur (par exemple, C34.1 pour "lobe supérieur du poumon"). Ces codes sont le plus souvent similaires à ceux de la CIM-10, avec des niveaux de détails adaptés à l'oncologie ; les codes topographiques de la CIMO-3 n'intègrent pas les informations sur la morphologie.

CIMO-3 - axe morphologique

Le code morphologique de la CIMO-3 est un système de codification utilisé pour classer et décrire les caractéristiques morphologiques des tumeurs, en particulier les tumeurs cancéreuses. Elle est souvent utilisée par les anatomo-cytopathologistes.

Le code morphologique de la CIMO-3 se compose de cinq chiffres (ou cinq digits) :

- Les 4 premiers digits indiquent le type histologique de la tumeur, c'est-à-dire la structure ou la composition cellulaire de la tumeur (par exemple, adénocarcinome, sarcome, etc.).
- Le 5^{ème} digit est une barre oblique suivie d'un chiffre, qui désigne le comportement biologique de la tumeur et son évolutivité :
 - /0 : Tumeur bénigne.
 - /1 : Tumeur de malignité incertaine ou à potentiel malin.
 - /2 : Carcinome in situ (tumeur non invasive).
 - /3 : Tumeur maligne, primaire.
 - /6 : Tumeur maligne, métastatique.
 - /9 : Tumeur maligne, incertaine si primaire ou métastatique.
- Exemple de code : 8140/3 : "adénocarcinome, tumeur maligne" où 8140 représente un adénocarcinome et /3 indique que c'est une tumeur maligne primaire.
- Pour le 5^{ème} digit : les « /3 », « /6 » et « /9 » représentent des tumeurs malignes mais il est recommandé de les recoder tous en /3, comme cela est fait dans les registres (ces codes sont utilisés par les pathologistes).

Les groupes par localisation et morphologie pour le codage des cancers

Tableau 1: Groupes des tumeurs malignes considérées comme histologiquement "différents" dans le cadre de la définition des tumeurs multiples (adaptation de Berg, 1996)

Groupes	
<i>Carcinomes</i>	
1. Carcinomes épidermoïdes et carcinomes à cellules transitionnelles	8051 - 8084, 8120 - 8131
2. Carcinomes basocellulaires	8090 - 8110
3. Adénocarcinomes	8140 - 8149, 8160 - 8162, 8190 - 8221, 8260 - 8337, 8350 - 8551, 8570 - 8576, 8940 - 8941
4. Autres carcinomes spécifiques	8030 - 8046, 8150 - 8157, 8170 - 8180, 8230 - 8255, 8340 - 8347, 8560 - 8562, 8580 - 8671
5. Carcinomes non spécifiés (SAI)	8010 - 8015, 8020 - 8022, 8050
<i>Sarcomes</i>	
6. Sarcomes et tumeurs des parties m	8680 - 8713, 8800 - 8921, 8990 - 8991, 9040 - 9044, 9120 - 9125, 9130 - 9136, 9141 - 9252, 9370 - 9373, 9540 - 9582
<i>Mésothéliomes</i>	
7. Mésothéliome	9050 - 9055
<i>Tumeurs du système hématopoïétique et du tissu lymphoïde</i>	
8. Leucémies myéloïdes	9840, 9861 - 9931, 9945 - 9946, 9950, 9961 - 9964, 9980 - 9987
9. Lymphomes matures à cellules B	9670 - 9699, 9728, 9731 - 9734, 9761 - 9767, 9769, 9823 - 9826, 9833, 9836, 9940
10. Lymphomes matures à cellules T e	9700 - 9719, 9729, 9768, 9827 - 9831, 9834, 9837, 9948
11. Lymphome hodgkinien	9650 - 9667
12. Tumeurs à mastocytes	9740 - 9742
13. Tumeurs à Histiocytes et cellules lymphoïdes accessoires	9750 - 9758
14. Types non spécifiés	9590 - 9591, 9596, 9727, 9760, 9800 - 9801, 9805, 9820, 9832, 9835, 9860, 9960, 9970,

	9975, 9989
15. <i>Sarcome de Kaposi</i>	9140
Autres types de cancer spécifiés	
16. <i>Autres types de cancer spécifiés</i>	8720 – 8790, 8930 – 8936, 8950 – 8983, 9000 – 9030, 9060 – 9110, 9260 – 9365, 9380 – 9539,
Types de cancer non spécifiés	
17. <i>Types de cancer non spécifiés</i>	8000 – 8005

Tableau 2: Groupes de codes topographiques considérés comme un seul site pour la définition des tumeurs primitives multiples

Code topo. CIM-O-2/3	Libellé	Si diagnostic métachrone ² , coder le premier site. Si diagnostic synchrone, coder les codes ci-dessous
C01	Base de la langue	C02.9
C02	Autres localisations et localisations non spécifiées de la langue	
C00	Lèvre	C06.9
C03	Gencives	
C04	Plancher de la bouche	
C05	Palais	
C06	Autres localisations et localisations non spécifiées de la bouche	
C09	Amygdale	C14.0
C10	Oropharynx	
C12	Sinus piriforme	
C13	Hypopharynx	
C14	Autres localisations et localisations mal définies des lèvres, de la cavité buccale et du pharynx	
C19	Jonction recto-sigmoïdienne	C20.9
C20	Rectum	
C23	Vésicule biliaire	C24.9
C24	Autres localisations et localisations non spécifiées des voies biliaires	
C33	Trachée	C34.9
C34	Bronches et poumons	
C40	Os, articulations et cartilage articulaire des membres	C41.9

² Tumeurs synchrones : le délai de diagnostic entre les tumeurs est inférieur à 6 mois,
Tumeurs métachrones : le délai est supérieur ou égal à 6 mois .

C41	Os, articulations et cartilage articulaire de localisations autres et non spécifiés	
C65	Bassinet (du rein)	C68.9
C66	Urètre	
C67	Vessie	
C68	Autres localisations et localisations mal définies des organes urinaires	

Les méthodes de diagnostic du cancer

Code	Description	Critères
1	Clinique	Diagnostic effectué avant le décès, mais sans aucun des éléments suivants (codes 2-8).
2	Investigations cliniques	Toutes les techniques de diagnostic, y compris la radiographie, l'endoscopie, l'imagerie, l'échographie, la chirurgie exploratoire (comme la laparotomie) sans diagnostic tissulaire.
4	Marqueurs tumoraux spécifiques	Y compris les marqueurs biochimiques et/ou immunologiques qui sont spécifiques d'un site tumoral.
5	Cytologie	Examen de cellules provenant d'un site primitif ou secondaire, y compris les fluides aspirés par endoscopie ou par aiguille ; comprend également l'examen microscopique de sang périphérique et de ponction de moelle osseuse, l'immunophénotypage par cytométrie de flux et la biopsie liquide en l'absence d'anatomopathologie.
7	Histologie	Examen histologique du tissu provenant de la tumeur (primitive ou métastatique), quel que soit le mode d'obtention, y compris toutes les techniques de coupe et les biopsies de moelle osseuse
8	Tests cytogénétiques ou moléculaires	Détection d'anomalies génétiques spécifiques à la tumeur ou de modifications génétiques dans la tumeur, y compris des techniques telles que le caryotype, la FISH (hybridation in situ en fluorescence), la PCR (réaction en chaîne par polymérase), le séquençage de l'ADN.
9	Inconnu	

Les changements de traitement

Ce texte a pour but de définir un **changement de traitement** et le différencier d'une **ligne de traitement** dans le modèle OSIRIS RWD.

La notion de ligne de traitement est difficile à définir et il n'existe pas de consensus absolu. La définition d'une ligne de traitement, (1ère ligne, 2ème ligne) peut correspondre à un changement de médicament, dépend du protocole administré qui correspond à une séquence de traitements définis a priori ou du changement d'un des médicaments.

Quelques exemples de problèmes rencontrés :

- Un protocole associant trois molécules peut être administré en association comme par exemple le protocole (épirubicine – cyclophosphamide - docétaxel) pendant 6 cycles ou en séquentiel selon le schéma (E + C) pendant 3 cycles puis D pendant trois cycles. Il s'agit d'une seule ligne de traitement alors qu'on observe un changement de médicaments.
- Une association de deux molécules (cisplatine + vinorelbine) transformée en (carboplatine + vinorelbine) correspond à un changement de médicaments mais considérée par certains comme la même ligne de traitement car le cisplatine et le carboplatine appartiennent à la même classe pharmacologique.
- Un passage de l'association (cisplatine + vinorelbine) en (cisplatine + gemcitabine) est considéré comme un changement de traitement et un changement de ligne. A l'inverse, si le protocole prévoit d'emblée le passage de l'un à l'autre, il s'agirait d'une seule ligne.
- Le passage d'une association A + B + C à une association A + B ne sera pas considérée ni comme changement de ligne ni comme changement de traitement.

Il existe donc deux notions et il est nécessaire d'avoir un consensus afin de permettre une interopérabilité et une façon identique de définir chaque item pour l'alimentation automatique et pour l'analyse.

La recommandation est de définir par consensus tous les changements de traitement et de laisser au moment de l'analyse la définition du calcul des lignes.

Les changements de traitement concernent les chimiothérapies cytotoxiques, les traitements par voie orale (hormonothérapie, thérapie ciblée orale type TKI), les anticorps monoclonaux, l'immunothérapie.

Il n'est pas défini de ligne de traitement pour les chimiothérapies adjuvantes, néo-adjuvantes et les associations chimiothérapie – radiothérapie.

Ligne de traitement

Une ligne de traitement, également appelée "ligne de chimiothérapie" ou "protocole de traitement", désigne un ensemble de médicaments administrés à un patient atteint de cancer pendant une période donnée. Une ligne de traitement contient plusieurs cycles et un cycle correspondant à un médicament ou une association de médicaments, administré à intervalle régulier.

La notion de ligne de traitement ne s'applique que pour les patients en phase métastatique.

On propose que le **changement de ligne** et le calcul du **nombre de lignes** soit une variable calculée au moment de l'analyse en fonction des protocoles et des localisations tumorales.

Changement de traitement

Le **changement de traitement** est défini par la modification ou la substitution d'un ou de plusieurs médicaments. La date du changement de traitement est la date à laquelle le ou les traitements ont été modifiés ou substitués.

On propose que le **changement de traitement** soit une variable calculée de manière identique dans la base avec des règles communes établies par consensus, sans tenir compte du caractère protocolaire de ce changement de traitement, ni des localisations.

Principes généraux

- On repère de façon automatique les changements de molécules administrées.
- La source structurée est privilégiée sur les documents non structurés.
- La source de l'information est une donnée nécessaire.

La source des données

Trois sources peuvent être utilisées :

- Les logiciels de prescription avec une information structurée
- Les comptes rendus médicaux de l'établissement avec une information en texte libre notamment pour les thérapeutiques orales ou les chimiothérapies systémiques administrées en dehors de l'établissement.
- Les documents récupérés de l'extérieur de l'hôpital et scannés dans le dossier.

La source de données à privilégier est celle issue des outils de prescription de chimiothérapie. On peut ainsi disposer de toutes les molécules administrées avec les dates d'administration (et non les traitements simplement prescrits).

- Molécules en DCI (par exemple docétaxel) ou par le nom de la spécialité pharmaceutiques (TAXOTERE®), en ne tenant compte que des doses réellement administrées (et non des doses prescrites ou prévues).
- Date d'administration (si administration en continu sur plusieurs jours, date de début)

A titre d'exemple, la source CHIMIO® dispose uniquement des molécules injectées ou qui nécessitent une reconstitution et cela ne concerne que les patients traités dans l'établissement :

- Les traitement IV uniquement (sauf exception) c'est-à-dire la chimiothérapie systémique, l'immunothérapie, les anticorps monoclonaux
- On ne récupère pas les « traitements associés » (acide folinique, corticoïdes...)

Les comptes rendus médicaux

Certains traitements ne se trouvent pas dans la source structurée :

- Patients non traités dans l'établissement
- Traitements oraux : hormonothérapie, chimiothérapie orale, thérapie ciblée

Dans ce cas, il est nécessaire de reconstituer le traitement pour structurer l'information à partir des dossiers médicaux.

Recommandations

Ces recommandations concernent la structuration automatique de l'item « changement de traitement »

Cas particulier des traitements de maintenance

Dans certains cas, après une rémission ou une stabilisation de la maladie, un patient peut être placé sous traitement de maintenance, qui consiste à poursuivre un des médicaments administrés dans la ligne précédente pour maintenir la rémission ou contrôler la maladie sur une période prolongée. Certains appellent ce traitement, chimiothérapie de continuation.

Exemple : le traitement standard des adénocarcinomes pulmonaires associe le cisplatine et le pémétrexed (4 cycles) suivi de l'utilisation du pémétrexed en continuation tant que le traitement est efficace, c'est-à-dire tant qu'il n'y a pas de progression de la maladie, et tant que le traitement est bien supporté. La poursuite du pémétrexed seul ne constitue pas un changement de traitement.

Chimiothérapie comportant A + B puis A seul : pas de changement de traitement

Cas particulier du switch de certaines molécules

Dans certains cas, un des médicaments présente des effets indésirables qui obligent à l'arrêt ; il peut alors être remplacé par un médicament de la même classe mais avec un profil de toxicité différent. Cette situation peut se voir mais avec un médicament appartenant à une autre classe avec un mode d'action différent.

Exemple : le traitement standard des carcinomes épidermoïdes bronchiques associe le cisplatine et la vinorelbine (4 cycles). Lorsqu'il apparaît une toxicité rénale, le cisplatine peut être remplacé par la carboplatine. Lorsqu'il apparaît une toxicité digestive à type d'occlusion, la vinorelbine peut être remplacée par la gemcitabine.

Un débat agite les oncologues médicaux ; pour certains, il s'agit bien d'une nouvelle ligne, pour d'autres, d'une simple adaptation.

Chimiothérapie comportant A + B puis A + C

On propose de considérer qu'il s'agit d'un changement de traitement et de laisser au moment de l'analyse la décision quant au changement de ligne ou non.

Cas de l'arrêt d'un des médicaments lors d'une association

Dans certains cas, pour des protocoles à 3 médicaments, un des médicaments a une toxicité qui oblige à l'arrêt ; il n'est pas remplacé mais définitivement arrêté. [OBJ]

Chimiothérapie comportant A + B + C puis A + B : pas de changement de traitement

Cas d'une diminution de doses

Lors d'un traitement, quel que soit la molécule administrée, il est parfois diminué les doses d'un ou des médicaments administrés en association. Il ne s'agit pas d'un changement de traitement mais d'une simple adaptation.

Cas de l'augmentation de l'intervalle de temps entre les cycles

Si les médicaments administrés ne changent pas, il ne s'agit pas d'un changement de traitement, quel que soit l'intervalle de temps entre les cycles y compris si le délai est très long. En effet, lorsque le délai est long et que le protocole est identique, on parle de « re-challenge » et non une nouvelle ligne. Le délai de 6 mois est en général la convention qui est adoptée.

Cas des protocoles qui alternent des associations différentes

Le nom du protocole n'est pas utilisable car le paramétrage dépend du centre et il est impossible, dans un protocole, de détecter des changements de ligne qui n'en sont pas. C'est au moment de l'analyse qu'il est décidé si le traitement est protocolaire ou s'il s'agit d'un changement de traitement et/ou d'un changement de ligne.

Hormonothérapies et thérapies ciblées

Lorsqu'un patient est sous hormonothérapie ou thérapie ciblée orale, on compte les lignes pour l'hormonothérapie ou les thérapies ciblées orales de la même façon.

Exemple 1 : une patiente sous tamoxifène en 1^{ère} ligne progresse ; on passe une anti-aromatase ; il s'agit d'un changement de traitement.

Exemple 2 : dans la suite, si on constate une progression, on passe à une chimiothérapie, il s'agit d'un changement de traitement.

Topographies des actes d'imagerie en oncologie

Rachis

- Rachis
- Sacrum

Membres

- Membre supérieur
- Membre inférieur

Tête et cou

- Crâne
- Cou
- Panoramique dentaire
- Glandes salivaires
- Glande thyroïde
- Glandes parathyroïdes

Vaisseaux

- Veine cave inférieure et ses affluents
- Artères iliaques et des membres inférieurs
- Veines des membres supérieurs
- Veines des membres inférieurs et des veines iliaques

Thorax

- Poumon et médiastin
- Grill costal et paroi thoracique
- Sternum
- Cœur

Abdomen

- Région oeso-gastro-duodénale
- Estomac
- Grêle
- Côlon
- Foie
- Rein et uretère

Pelvis

- Vessie
- Testicules
- Prostate
- Utérus (col et corps)
- Ovaires et trompes

Tissu mou et peau

Sein

Région thoraco-abdominale

Région thoraco-abdomino-pelvienne

Exploration du corps entier par TDM ou par TEP-Scanner

Exploration osseuse par scintigraphie