

# Contrôle de qualité des registres du cancer basés sur la population

Tous les registres du cancer doivent pouvoir donner des indications objectives sur la qualité des données recueillies. Les méthodes disponibles ont été décrites dans un des premiers Rapports techniques du CIRC (Parkin et coll., 1994) et mises à jour dans des articles publiés en 2009 (Parkin et Bray, 2009 ; Bray et Parkin, 2009). Ces papiers décrivent quatre dimensions définissant la qualité des registres : la comparabilité, la validité, la ponctualité et l'exhaustivité.

## 1. Comparabilité

Pour pouvoir comparer les statistiques générées pour différentes populations et à différentes périodes, il est nécessaire de standardiser les pratiques relatives à la classification et au codage des nouveaux cas, et d'uniformiser les définitions de l'inci-

dence, notamment pour l'enregistrement et la notification des cancers primitifs multiples survenant chez un même individu. La *Classification internationale des maladies pour l'oncologie*, CIM-O, publiée par l'OMS, fournit les normes pour la classification et le codage des cancers, notamment pour le codage de la topographie (localisation de la tumeur dans l'organisme), de la morphologie (aspect microscopique de la tumeur), du comportement (si la tumeur est maligne, bénigne, ou in situ) et du stade (extension de la tumeur). La CIM-O-3 publie en outre un système de codage normalisé pour la base de diagnostic ainsi que les règles du CIRC pour le codage des cancers primitifs multiples. Comme la cancérogenèse est un processus qui peut parfois prendre des décennies, la définition de la date d'incidence est arbitraire ; c'est pourquoi

il est particulièrement important de suivre les normes en vigueur. Les règles à suivre pour la définition de la date d'incidence sont données par le Réseau européen des registres du cancer (<http://www.encreu/images/docs/recommandations/recommandations.pdf>).

Dans les PRFI, certains obstacles objectifs peuvent empêcher de suivre les normes internationales. Par exemple, l'absence de couverture par les laboratoires d'anatomopathologie ou la difficulté d'accès au diagnostic va réduire le pourcentage de cas confirmés sur des bases morphologiques et différer la date d'incidence telle que définie par le Réseau européen des registres du cancer, qui demande d'utiliser en priorité la date de la première confirmation histologique ou cytologique du cancer comme date d'incidence.

## 2. Validité

Il est possible d'améliorer nettement l'exactitude des données enregistrées en effectuant des vérifications de la cohérence des informations au moment de la saisie des données ; de telles vérifications sont incorporés dans le logiciel CanReg (voir Annexe 1). La plupart des registres vont aussi, de façon formelle ou informelle, vérifier la précision du travail de l'équipe en réalisant une sorte de nouvelle extraction des données (en retournant à la source ou à plusieurs sources pour vérifier l'exactitude de l'enregistrement) ou des exercices de recodage ; ces vérifications seront, si besoin, suivies d'actions visant à corriger les erreurs mises en évidence.

La plupart des registres vont s'attacher à trois statistiques qui ont une incidence sur l'exactitude des données recodées. Ce sont :

- la proportion (ou pourcentage) de cas pour lesquels il manque des données
- le pourcentage de cas dont le diagnostic a été vérifié sur des bases morphologiques (VM%)
- le pourcentage de cas pour lesquels la seule information disponible provient du certificat de décès (CDS%).

### 2.1 Proportion (ou pourcentage) de cas pour lesquels il manque des données

La proportion de cas pour lesquels les valeurs de différents éléments tels que l'âge ou le stade ne sont pas connues, représente également un indicateur de la qualité des données. Un élément important à évaluer ici est la proportion de cas dont la localisation primitive est incertaine (LPI%). La catégorie correspondant aux localisations primitives inconnues (code C80.9) selon la CIM-O

doit également inclure les localisations mal définies.

Les données sur certains éléments peuvent être très difficiles à obtenir dans les PRFI. Cela peut concerner, par exemple, le numéro d'identification personnel, ce qui entraîne des procédures de recouplement astreignantes pour un niveau de précision moindre. Beaucoup de PRFI ont le même problème d'absence ou de mauvaise qualité des données de mortalité. Cela pose de nombreux problèmes pour les registres du cancer, comme le sous-enregistrement du fait de l'absence de cas enregistrés sur la base du certificat de décès seulement (CDS) contribuant à l'incidence, et l'impossibilité de calculer les indicateurs normalisés de qualité des données (à part le pourcentage de cas dont le diagnostic a été vérifié sur des bases morphologiques [VM%]). Le seul aperçu de l'exhaustivité de l'enregistrement du cancer en l'absence de données de mortalité provient des méthodes de détermination des cas par croisement de sources indépendantes ou des méthodes de capture–recapture (décrites plus loin).

### 2.2 Pourcentage de cas dont le diagnostic a été vérifié sur des bases morphologiques (VM%)

La vérification morphologique se réfère aux cas dont le diagnostic repose sur l'histologie ou la cytologie. Ce type de diagnostic est considéré comme une sorte de « gold standard », tandis que le diagnostic fondé sur d'autres critères suscite des doutes quant à sa précision (même si l'on peut se demander si la cytologie exfoliative est vraiment toujours plus précise que l'IRM ou la TDM). On considère qu'un pourcentage élevé de diagnostics vérifiés morphologiquement (VM%) signe l'exactitude du diagnostic, tandis qu'un

VM% faible jette des doutes sur la validité des données.

Les vérifications effectuées lors de la préparation de *Cancer Incidence in Five Continents* (CI5) comprennent une comparaison formelle des VM% (par sexe, pour les principaux sites de cancer) avec un « standard », fondé sur les valeurs observées dans la même région 5 ans auparavant. L'Annexe 2 fournit des tableaux avec les valeurs « standard » de certains indicateurs sélectionnés de la qualité des données, dont le VM% par pays ou par région, et utilisés lors du processus de publication des volumes de CI5. Un VM% plus faible qu'attendu suscite des doutes sur la validité des données enregistrées, mais ce ne sont généralement pas les registres du cancer qui peuvent influencer l'existence de services d'anatomopathologie ou leur utilisation dans leur zone géographique. Dans les PRFI, c'est même l'inverse – un VM% relativement élevé – qui est préoccupant. Il est plus facile de recueillir les données des services de pathologie que d'aller à la pêche aux informations dans les services cliniques ou dans des archives hospitalières mal organisées. Une proportion élevée de cas diagnostiqués via le service d'anatomopathologie peut très bien être le signe de failles dans la recherche des cas et, par conséquent, d'un enregistrement incomplet. Pire, l'enregistrement sera non seulement incomplet mais biaisé, avec un déficit des cancers pour lesquels la biopsie n'est pas facile (par exemple cancers du poumon, du foie, du cerveau et du pancréas).

### 2.3 Pourcentage de cas enregistrés sur la base du seul certificat de décès (CDS%)

Les « cas CDS » sont les cas de cancer enregistrés sur la base du

seul certificat de décès, et pour lesquels on n'a pas pu retrouver d'autres informations. Comme décrit précédemment, la nature des certificats de décès dans les PRFI est très variable, allant des extraits des registres d'état civil aux certificats délivrés par la morgue des hôpitaux. Mais presque toujours, l'exactitude des informations portant sur le diagnostic est discutable, la personne rédigeant le certificat pouvant n'avoir eu que peu de contacts avec le patient avant son décès et avoir été mal informée sur la façon de rapporter la cause du décès. Aussi, si l'on n'arrive pas à trouver le dossier médical des personnes qui, apparemment, sont décédées de cancer (ou avec un cancer), on peut raisonnablement soupçonner que le diagnostic était tout simplement faux. Néanmoins, la pratique des registres demande que ces cas soient inclus, mais quand ils représentent une large proportion des cas, la validité des données est douteuse.

Il est difficile d'établir des critères objectifs d'acceptabilité du CDS% – cela dépend des conditions locales, par exemple de la disponibilité des certificats de décès, de la possibilité de faire des recoupements avec la base de données du registre, de la qualité des déclarations des causes de décès et de la facilité à retrouver les cas.

## 2.4 Cohérence interne

Il convient de vérifier et de corriger les données des dossiers nouvellement soumis pour vérifier la validité des éléments, la cohérence interne et la cohérence entre les dossiers, avant de les relier à la base de données centrale. Ces vérifications et corrections doivent aussi s'appliquer à la base de données du registre chaque fois qu'un changement a été effectué.

## 3. Ponctualité

Il est souvent nécessaire de communiquer rapidement les données des registres du cancer. Mais pour les registres du cancer (et leurs clients), il convient de trouver un compromis entre la ponctualité dans la communication des données et leur niveau d'exhaustivité. La ponctualité dépend de la rapidité avec laquelle le registre peut collecter, traiter et rapporter des données suffisamment complètes et exactes. Dans certains pays, comme le Royaume-Uni, la saisie électronique des données a accéléré le processus d'enregistrement. Certains réseaux de registres, tels que SEER et l'Association nord-américaine des registres centraux du cancer (*North American Association of Central Cancer Registries*), demandent à leurs membres de rapporter leurs données dans les 22 à 24 mois qui suivent la clôture de l'année du diagnostic. Certains registres utilisent des méthodes de modélisation du délai pour estimer le sous-dénombrement au moment du rapport, ou des projections à court terme pour les estimations concernant l'année en cours.

## 4. Exhaustivité

Parkin et Bray (2009) distinguent :

- les méthodes qualitatives (ou semi-quantitatives) qui donnent une indication sur le degré d'exhaustivité par rapport aux autres registres, ou au cours du temps, et
- les méthodes quantitatives, qui permettent une évaluation numérique de l'exhaustivité de l'enregistrement de tous les cas éligibles.

### 4.1 Méthodes semi-quantitatives

Parmi les méthodes semi-quantitatives, la possibilité qu'un VM% rela-

tivement élevé puisse représenter un recueil des données incomplet a déjà été mentionnée.

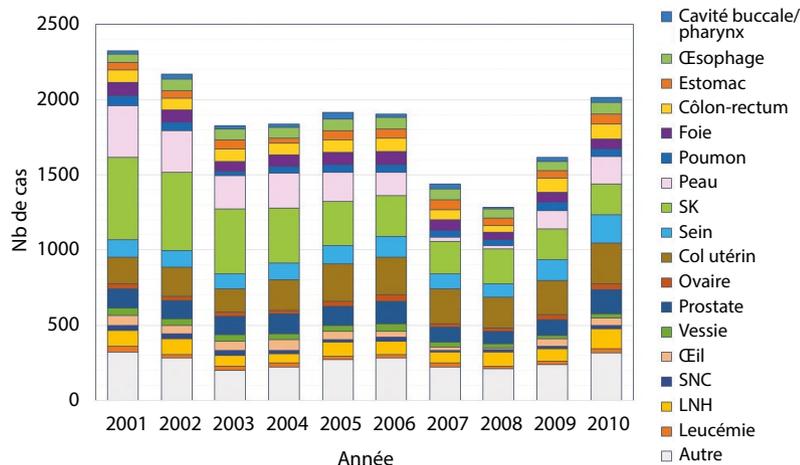
Un cas de cancer donné peut être identifié à partir de différentes sources (hôpitaux, laboratoires ou certificats de décès), et l'on considère généralement que, en recourant à un grand nombre de sources différentes pour chaque cas, il est relativement rare de se retrouver avec zéro source (c'est-à-dire dans la situation où le cas n'a été trouvé dans aucune des sources). Les autres indicateurs largement utilisés sont :

- le rapport mortalité sur incidence
- la stabilité de l'incidence dans le temps
- la comparaison des taux d'incidence avec ceux d'autres populations (similaires).

#### 4.1.1 Rapport mortalité sur incidence

Le rapport mortalité sur incidence (M:I) est un indicateur important largement utilisé – par exemple dans CI5 – pour identifier un manque d'exhaustivité éventuel. Il s'agit de comparer le nombre de décès imputables à un cancer particulier, obtenus à partir d'une source indépendante du registre (généralement les statistiques démographiques), et le nombre de nouveaux cas de ce cancer enregistrés pendant la même période. L'application de cette méthode nécessite toutefois des données de mortalité de bonne qualité (spécialement en ce qui concerne l'enregistrement précis de la cause de décès), de façon que M:I se situe autour de : 1 – probabilité de survie (à 5 ans). Cela permet d'établir des normes objectives pour M:I, applicables aux régions où la survie est probablement plus ou moins la même (voir Annexe 2). Cette méthode ne peut pas être utilisée quand il n'existe pas d'enregistrement exhaustif des décès, ou quand la cause de décès est

**Fig. 5.1.** Nombre de nouveaux cas de cancer par localisation dans un registre du cancer, 2001–2010. LNH, lymphome non hodgkinien ; SK, sarcome de Kaposi ; SNC, système nerveux central.



absente ou inexacte – situation que l'on retrouve dans presque tous les pays d'Afrique, et dans beaucoup de pays d'Asie.

#### 4.1.2 Stabilité de l'incidence avec le temps

Il est facile pour le registre de vérifier rapidement le nombre de cas enregistrés chaque année. En l'absence de changements notables dans la population, on peut ainsi identifier rapidement les failles possibles dans l'identification des cas.

La Fig. 5.1 en fournit un exemple. On observe un déficit évident de cas pour les années 2007, 2008 et 2009, et bien que cela concerne pratiquement tous les sites de cancer, ce déficit est particulièrement marqué pour les cancers de la peau.

#### 4.1.3 Comparaison des taux d'incidence entre différentes populations (similaires)

Bien entendu, les profils d'incidence des populations ne sont pas tous pareils ; l'observation de ces différences est l'un des objectifs des registres du cancer. Il convient néanmoins de comparer les résultats avec

ceux des registres couvrant une population similaire (par sa situation géographique ou par sa composition ethnique) – à condition que les données des autres registres soient elles-mêmes de bonne qualité – pour rechercher les différences. On peut s'attendre à certaines variations, ou on peut les expliquer sur la base de l'exposition à des facteurs de risque connus, mais une différence systématique (nombreux taux plus faibles que prévu) peut laisser soupçonner un sous-enregistrement.

Cette méthode est utilisée par les responsables de *Cancer Incidence in Five Continents*, qui comparent les résultats de chaque registre avec ceux d'un groupe de registres du même pays (ou de la même région géographique) (Annexe 2).

### 4.2 Méthodes quantitatives

Trois méthodes sont disponibles pour obtenir une évaluation quantitative du degré d'exhaustivité des registres :

- vérification des cas de manière indépendante
- méthodes de capture–recapture
- méthodes basées sur les certificats de décès.

#### 4.2.1 Vérification des cas de manière indépendante

La comparaison de la base de données du registre avec des listes de cas de cancer établies indépendamment et non utilisées par le registre est une méthode particulièrement utile et objective pour évaluer l'exhaustivité. En croisant la base de données du registre et la liste indépendante de cas, on pourra estimer le nombre de cas de « manqués » par le registre. La proportion de patients éligibles figurant déjà dans le registre donne une estimation directe et quantitative de l'exhaustivité.

L'existence de telles listes de patients atteints de cancer dans la zone couverte par le registre – provenant par exemple d'études ou d'enquêtes – offre une occasion d'évaluer l'exhaustivité du registre qu'il convient de ne pas laisser passer.

#### 4.2.2 Méthodes de capture–recapture

Ces méthodes exploitent le fait que les mêmes cas de cancer peuvent être notifiés aux registres par de multiples sources. Généralement les sources d'information sont groupées dans trois catégories plus ou moins indépendantes les unes des autres : hôpitaux, laboratoires (d'anatomopathologie) et certificats de décès. L'idée de base est que, si l'on sait combien de cas sont notifiés par une source, par deux sources ou par les trois sources, il est possible d'estimer combien sont notifiés par aucune (c'est-à-dire combien ont été manqués). Dans la pratique, l'analyse de l'exhaustivité par capture–recapture demande de faire des croisements de fichiers (de façon que les cas identifiés par chacune des multiples sources soient correctement classifiés). Ce n'est pas un problème pour les utilisateurs de CanReg, où les sources d'information sont

rassemblées pour chaque cas de cancer. CanReg étant structuré pour les croisements de fichiers, cette sorte d'analyse devrait être particulièrement simple.

#### 4.2.3 Méthodes basées sur les certificats de décès

Les méthodes basées sur les certificats de décès dépendent de la disponibilité de certificats de relativement bonne qualité, mentionnant la cause de décès (de façon complète et exacte) dans la zone couverte par le registre du cancer ; ces méthodes ne seront pas facilement applicables dans de nombreuses situations rencontrées dans les PRFI. Les deux autres méthodes, en revanche, peuvent être appliquées facilement.

### 5. Indicateurs de la qualité des données de survie

L'estimation de la survie demande, à la différence des données d'inci-

dence, des informations de haute qualité sur le suivi des patients atteints de cancer. C'est possible quand les registres disposent de données sur toutes les causes de mortalité et si des procédures de recoupement (basées idéalement sur le numéro d'identification unique) sont en place. Comme les registres d'état civil sont souvent inexistantes, non fiables ou inaccessibles, de nombreux registres des PRFI ont recours aux méthodes actives de suivi. Les indicateurs utilisés pour évaluer la qualité des données de survie sont la fréquence des cas enregistrés sur la base du seul certificat de décès, et la fréquence des cas exclus de l'analyse du fait de l'absence de suivi (Swaminathan et coll., 2011). L'absence de suivi est responsable de biais même dans les registres des pays à revenu élevé. Même une faible sous-estimation des décès peut entraîner une surestimation de la survie à long terme (Brenner et Hakulinen, 2009). Dans les PRFI

où les systèmes de statistiques sur la santé fonctionnent mal en routine et où les données de mortalité sont souvent inexistantes, les estimations de survie venant des registres de population sont souvent les seules à donner une idée de l'état de la prise en charge du cancer dans le pays.

Comme noté par Skeet dans *Cancer Registration: Principles and Methods* (Skeet, 1991), « tous les registres devraient être capables de citer des mesures objectives [de vérification] plutôt que de compter sur les idées reçues et les vœux pieux. » C'est un bon conseil, qui n'est pas toujours suivi. Avant de rapporter les résultats du registre, il convient d'évaluer leur qualité, d'autant plus que leur objectif est presque toujours de pouvoir comparer les taux et les risques de cancer entre différentes populations, différents sous-groupes ou différentes périodes ; il ne faudrait pas que les différences observées soient dues à des artefacts liés au processus d'enregistrement.

## Points clés

- Tous les registres du cancer basés sur la population devraient pouvoir fournir des indications objectives sur la qualité des données qu'ils ont collectées.
- Les méthodes disponibles sont décrites et mises à jour, et couvrent quatre dimensions de la qualité : comparabilité, validité, ponctualité et exhaustivité.