

Chapitre 9. Méthodes statistiques pour les registres

**Centre International de Recherche sur le Cancer
P. Boyle et D.M. Parkin**

150 cours Albert Thomas, 69372 Lyon Cédex 08, France

L'objectif de ce chapitre n'est pas de remplacer les ouvrages statistiques de référence. Son objectif est seulement d'aider ceux qui sont impliqués dans l'enregistrement des cancers à comprendre les calculs nécessaires à la présentation de leurs données. Pour les registres reposant sur une base de population, il s'agit essentiellement des taux d'incidence. Les méthodes nécessaires à l'utilisation de ces taux dans des études comparatives (par exemple comparaison des taux d'incidence provenant de différentes périodes de temps ou de différentes zones géographiques) sont aussi décrites mais quand les taux d'incidence ne peuvent être calculés, les résultats du registre doivent être présentés sous forme de proportions, et les méthodes correspondantes sont aussi présentées ici.

PREMIÈRE PARTIE : MÉTHODES POUR L'ÉTUDE DE L'INCIDENCE

Définitions

Le taux d'incidence

Le problème majeur des registres de cancers reposant sur une base de population est le calcul des taux d'incidence des cancers et leur utilisation pour étudier le risque pour un individu résidant dans la zone géographique du registre d'avoir un cancer comparativement à un individu d'une autre zone géographique, ou de comparer différents sous-groupes de la population à l'intérieur de la zone géographique du registre (voir chapitre 3).

L'incidence exprime le nombre de nouveaux cas de cancers qui apparaissent dans une population définie d'individus non malades, et le

taux d'incidence représente le nombre de tels événements dans une période de temps définie. Ainsi :

$$\text{Taux d'incidence} = \frac{\text{Nbre de nouveaux cas de la maladie}}{\text{population à risque}} \text{ dans une période de temps.}$$

Cette mesure donne une estimation directe de la probabilité ou du risque de maladie. Elle est d'une importance fondamentale dans les études épidémiologiques.

Comme les taux d'incidence se rapportent à une période de temps, il est nécessaire de définir la date exacte du début de la prise en compte des nouveaux cas de la maladie. Pour les registres de cancers il s'agit de la date d'incidence (Chapitre 6, item 16). Bien que cette date ne corresponde pas au début de la maladie cancéreuse, elle est plus pertinente et surtout plus facile à définir que d'autres (par exemple la date de début des symptômes, la date d'entrée à l'hôpital, ou la date de traitement).

Période d'observation

Le vrai risque instantané de la maladie est donné par le taux d'incidence pour une période de temps infiniment courte, le taux "instantané" ou "force de morbidité". Avec des périodes de temps plus longues, la population à risque devient moins clairement définie (à cause des naissances, des décès et des migrations), et le taux lui même peut être variable avec le temps. En pratique, le cancer dans les populations humaines est un événement relativement rare, et pour l'étudier, des populations importantes doivent être observées sur une période de plusieurs années. Les taux d'incidence sont par convention exprimés en taux annuels (c'est à dire par an), et quand les données sont collectées sur plusieurs années le dénominateur est converti en une estimation du nombre de personnes-années exposées.

Population à risque

Les études de cohorte portent le plus souvent sur de relatives petites populations d'individus dans lesquelles on étudie la présence ou l'absence d'un facteur de risque. De façon inévitable, des individus du groupe sont perdus de vue (à cause des décès, des migrations, d'impossibilité à les suivre), et souvent de nouveaux individus vont s'ajouter à cette cohorte.

De ce fait, les individus sont observés et soumis au risque d'avoir la maladie pendant des périodes de temps variables. Le dénominateur du taux d'incidence est alors calculé en faisant la somme pour chaque individu du nombre de personnes-années pour lequel il contribue.

Les registres de cancers font habituellement le calcul des taux d'incidence pour des populations entières, et le dénominateur pour de tels taux ne peut provenir de la connaissance de la contribution de chaque individu à la population à risque. Celui-ci est donc généralement estimé par la population au milieu de l'année (ou la moyenne de la population du début et de la fin de l'année ou de la période), obtenue par un recensement. La variance de l'estimation du taux d'incidence est déterminée par le nombre de cas utilisés dans le numérateur du taux ; pour cette raison, il est habituel de cumuler plusieurs années d'observation, et de calculer le taux moyen annuel. Le dénominateur est alors estimé en personnes-années, idéalement en additionnant les estimations de la population du milieu de l'année pour chacune des années considérées. Quand celles-ci ne sont pas disponibles, une solution moins performante consiste à utiliser les effectifs de la population provenant d'un ou deux points pendant la période de temps considérée pour estimer le nombre de personnes-années à prendre en compte. Cette approximation est raisonnable s'il n'existe pas de changements rapides ou irréguliers dans la structure de la population. Des exemples, illustrant des estimations du nombre de personnes-années observées en fonction des données de population dont on peut disposer sont montrés dans le tableau 1. Par convention les taux d'incidence des cancers sont exprimés en cas pour 100 000 personnes-années, permettant d'éviter l'utilisation de petites décimales. Pour les cancers de l'enfance, le taux est souvent exprimé par million.

Tableau 1. Calcul des personnes-années à risque utilisant différentes données de population.
Données pour les hommes entre 45 et 49 ans. Ecosse, 1980-1984

Années	1. Chaque année ^a	2. Point intermédiaire ^b	3. Points irréguliers ^c
1980	140.800	—	—
1981	142.700	—	142.700
1982	140.600	140.600	—
1983	141.200	—	141.200
1984	141.500	—	—

^a Méthode 1. Personnes-années = 140.800 + 142.700 + 140.600 + 141.200 + 141.500 = 706.800

^b Méthode 2. Personnes-années = 140.600 X 5 = 703.000

^c Méthode 3. Diminution de la population, entre l'année 2 et l'année 4 = 1500 ; diminution annuelle = 1500/2 = 750 ; personnes-années = (142.700 + 750) + 142.700 + (142.700 - 750) + 141.200 + (141.200 - 750) = 709.750

Quand les estimations de la population sont utilisées pour avoir une approximation du nombre de personnes-années à risque, le dénominateur du taux prendra en compte quelques personnes qui ne sont pas vraiment à risque. Heureusement, pour l'étude des taux d'incidence de cancers

spécifiques, ceci induit peu de différence, le nombre de personnes dans la population qui sont vivantes et qui présentent un cancer d'une localisation spécifique étant relativement petit. Cependant, si une partie substantielle de la population est de façon certaine non à risque d'avoir la maladie, elle devrait être exclue du dénominateur. Un exemple évident est d'exclure le sexe opposé du dénominateur des taux de cancers spécifiques à un sexe, et les taux d'incidence des cancers de l'utérus sont mieux calculés si l'on ne prend en compte que les femmes ayant un utérus (une relativement grande proportion de femmes d'âge moyen ont eu une hystérectomie). Ceci est spécialement vrai si des comparaisons doivent être faites pour différentes périodes de temps ou différents lieux où la fréquence de l'hystérectomie peu varier (Lyon et Gardner, 1977; Parkin et coll 1985a).

Calcul des taux

De nombreux indices ont été développés pour exprimer l'occurrence d'une maladie dans une communauté. Ils ont été clairement présentés par Inskip et ses collaborateurs (Inskip et coll, 1983) et d'autres auteurs ont aussi abordé ce sujet (Armitage, 1971 ; Armitage et Berry, 1987 ; Breslow et Day, 1980, 1987 ; Doll et Cook, 1967 ; Fleiss, 1981 ; MacMahon et Pugh, 1970). Ce chapitre s'intéressera particulièrement aux méthodes qui sont les plus appropriées à l'enregistrement des cancers et sera illustré d'exemples. Les exemples utilisent les données d'incidence du cancer du poumon chez les hommes en Ecosse. Même si on a essayé d'introduire le plus possible d'étapes intermédiaires dans les calculs, il n'a pas toujours été possible de les introduire toutes. Aussi, le recalcul de quelques uns des stades intermédiaires peut donner des résultats légèrement différents à cause du degré de précision utilisé dans les calculs et des arrondis. Ainsi le lecteur qui reprendrait tous les calculs, obtiendrait le même résultat final mais devrait s'attendre à quelques imprécisions mineures dans les résultats intermédiaires présentés dans ce texte.

Taux brut et taux spécifique

Supposons qu'il y ait A groupes d'âge pour lesquels le nombre de cas et le nombre correspondant de personnes-années à risque peut être estimé. Habituellement le nombre de groupe est 18 ($A = 18$) et les classes d'âge utilisées sont 0-4, 5-9, 10-14, 15-19 ... 80-84 et 85 et plus (85+). Cependant, des changements de classifications sont souvent utilisés, par exemple en séparant les enfants âgés de moins d'un an (0) et ceux qui sont âgés de 1 à 4 ans (1-4) ou en limitant les classes d'âges à 75 ans, c'est à dire en ayant des classes d'âges jusqu'à 70-74 et 75 +. Appelons r_i le nombre de

cas qui sont apparus dans la i ème classe d'âge. Si tous les cas sont d'âge connu, alors le nombre total de cas R est :

$$R = \sum_{i=1}^A r_i = r_1 + r_2 + r_3 + \dots + r_A \quad (9.1)$$

De la même façon, on appelle n_i le nombre de personnes-années observé dans la i ème classe d'âge pendant la même période de temps pendant laquelle les cas ont été comptabilisés, le nombre total de personnes-années observées N est :

$$N = \sum_{i=1}^A n_i = n_1 + n_2 + n_3 + \dots + n_A \quad (9.2)$$

Le taux brut (B) pour 100 000 (tous âges) peut être aisément calculé en divisant le nombre total de cas (R) par le nombre total de personnes-années exposées (N) et en multipliant le résultat par 100 000.

$$\text{Taux brut} = B = \frac{R}{N} \times 100\,000 \quad (9.3)$$

C'est à dire, quand l'âge est connu pour tous les cas,

$$B = \frac{\sum_{i=1}^A r_i}{\sum_{i=1}^A n_i} \times 100\,000 \quad (9.4)$$

Le taux spécifique pour la classe d'âge i , que nous appellerons a_i , peut aussi être calculé simplement comme un taux pour 100 000 en divisant le nombre de cas dans la classe d'âge (r_i) par le nombre correspondant de personnes-années exposées (n_i) et multipliant le résultat par 100 000. Ainsi,

$$a_i = \frac{r_i}{n_i} \times 100\,000 \quad (9.5)$$

Standardisation sur l'âge - généralités

Un des problèmes rencontré le plus fréquemment en épidémiologie des cancers est la comparaison des taux d'incidence d'un cancer particulier entre deux populations différentes ou pour la même population à deux périodes différentes. La comparaison de simples taux bruts peut fréquemment donner une fausse image de la réalité à cause de différences dans la structure d'âge des populations à comparer. Si une population est en moyenne plus jeune qu'une autre, si les taux spécifiques par l'âge sont les mêmes dans les deux populations, il y aura plus de cas dans la population la plus vieille que dans la plus jeune. Le tableau 2 montre comment les taux spécifiques augmentent très rapidement avec l'âge.

Exemple 1. Calcul des taux bruts et des taux spécifiques par âge

Le tableau 2 présente les données permettant de calculer l'incidence du cancer de la trachée, des bronches et des poumons (numéro 162 de la CIM-9 chez l'homme en Ecosse de 1980 à 1984).

Tableau 2. Incidence du cancer du poumon chez l'homme en Ecosse de 1980 à 1984

N° d'ordre	Groupe d'âge	Nombre de cas	Personne/années observées	Taux spécifique par âge pour 100.000
(i)		(r_i)	(n_i)	(r_i/n_i)
1	0-4	0	827 400	0,00
2	5-9	0	856 500	0,00
3	10-14	0	1 061 500	0,00
4	15-19	0	1 157 400	0,00
5	20-24	4	1 074 900	0,37
6	25-29	3	917 700	0,33
7	30-34	29	890 300	3,26
8	35-39	61	816 000	7,48
9	40-44	153	724 400	21,12
10	45-49	376	706 800	53,20
11	50-54	902	703 800	128,16
12	55-59	1 819	691 200	263,17
13	60-64	2 581	610 900	422,49
14	65-69	3 071	511 800	600,04
15	70-74	3 322	425 600	780,55
16	75-79	2 452	266 800	919,04
17	80-84	1 202	122 500	981,22
18	85+	429	54 700	784,28
		16 404	12 420.200	

Les taux spécifiques par âge sont calculés en appliquant la formule (9.5). Par exemple, pour le groupe d'âge 40-44 ans ($i=9$)

$$\begin{aligned}
 a_9 &= \frac{r_9}{n_9} \times 100.000 \\
 &= \frac{153}{724.400} \times 100.000 = 21,1
 \end{aligned}$$

Donc, pour le groupe d'âge 40-44 ans, le taux d'incidence moyen annuel spécifique par âge est de 21,1 pour 100.000. Les autres taux spécifiques présentés dans le tableau sont calculés de la même façon.

Le taux brut, B, est calculé par la formule (9.3) R étant le nombre total de cas (soit 16.404) et N le nombre total de personne-années observées (soit 12.420.200).

$$B = \frac{16.404}{12.420.200} \times 100.000 = 132,1$$

Le taux d'incidence moyen annuel tous âges confondus du cancer du poumon chez les hommes en Ecosse est donc de 132,1 pour 100.000 dans la période 1980-84.

Ainsi, si l'on compare les niveaux de cancer entre deux régions, ou si l'on mesure l'évolution du cancer au cours du temps dans la même région, il est important de prendre en compte les différences de la structure d'âge des populations. C'est ce que fait la standardisation sur l'âge. *Mais on n'insistera jamais assez sur le fait que la seule solution correcte pour comparer des taux d'incidence entre des populations ayant des pyramides des âges très différents consiste à limiter les comparaisons aux taux spécifiques par âge.* (Doll et Smith, 1982). Une mesure résumée telle que celle produite par la technique d'ajustement sur l'âge ne peut remplacer l'examen des taux spécifiques par âge. Cependant, il est très utile, particulièrement lorsqu'on compare de nombreux taux d'incidence, de disposer d'une mesure résumée telle que ce taux standardisé sur l'âge.

Deux méthodes de standardisation sur l'âge sont largement utilisées: la méthode directe et la méthode indirecte. La méthode directe est d'abord décrite car elle est beaucoup plus facile à interpréter que la méthode indirecte (pour une discussion complète voir Rothman, 1986), et est généralement préférée quand cela est possible (pour plus d'informations voir Breslow et Day, 1987, pages 72-75).

Tableau 3. Population Mondiale de Référence

N° d'ordre (i)	Groupe d'âge	Population (p_i)
1	0-4	12 000
2	5-9	10 000
3	10-14	9 000
4	15-19	9 000
5	20-24	8 000
6	25-29	8 000
7	30-34	6 000
8	35-39	6 000
9	40-44	6 000
10	45-49	6 000
11	50-54	5 000
12	55-59	4 000
13	60-64	4 000
14	65-69	3 000
15	70-74	2 000
16	75-79	1 000
17	80-84	500
18	85+	500
		100 000

Source : Doll et coll., 1966

Standardisation sur l'âge - méthode directe

Le taux standardisé sur l'âge est le taux théorique qui aurait été obtenu si les taux spécifiques observés pour chaque âge étaient appliqués à une population standard : cette population est communément appelée la population de référence.

Les effectifs de chaque classe d'âge de la population de référence sont connus comme des pondérations utilisées dans le processus de standardisation. Plusieurs pondérations possibles w_i peuvent être utilisées. L'utilisation de ces différentes pondérations (c'est-à-dire l'utilisation de différentes populations de référence) donnera des résultats différents pour le taux standardisé. La population de référence mondiale (voir tableau 3), modification par Doll et coll de celle proposée par Segi (1960), utilisée dans la série d'ouvrages "Cancer Incidence in Five Continents", est la plus fréquemment utilisée. Sa large diffusion facilite grandement la comparaison du niveau de cancer entre les différentes régions du monde.

Si p_i est l'effectif de la i ème classe d'âge de la population de référence, et si, comme ci-dessus $i = 1, 2, \dots, A$, si a_i représente le taux spécifique pour l'âge dans la classe d'âge i , le taux standardisé pour l'âge (TSA) est :

$$\text{TSA} = \frac{\sum_{i=1}^A a_i p_i}{\sum_{i=1}^A p_i} \quad (9.6)$$

Une base de données peut contenir des cas de cancers ayant un âge inconnu. Ceci signifie que l'équation (9.1) n'est plus valide, puisque le nombre total de cas (R) est plus grand que la somme des cas dans les groupes d'âge individuels ($\sum r_i$), si bien que le TSA, obtenu à partir des taux spécifiques par âge (équation 9.5), sera une sous estimation de la vraie valeur.

Doll et Smith (1982) ont proposé d'appliquer une correction, en multipliant le TSA (calculé comme dans l'équation 9.6) par :

$$\frac{R}{\sum_{i=1}^A r_i}$$

L'utilisation de cet ajustement implique que la distribution par âge des cas dont l'âge est inconnu, est la même que pour les cas dont l'âge est connu. Bien que cette hypothèse ne soit pas souvent vérifiée, car c'est souvent chez les personnes âgées que l'âge n'est pas enregistré, l'effet n'est pas très important tant que la proportion de cas d'âge inconnu reste petite ($< 5\%$).

Exemple 2. Calcul des Taux standardisés par âge, méthode directe

Le tableau 4 montre les taux spécifiques par âge chez les hommes d'Ecosse calculés plus haut (tableau 2). Ces taux spécifiques par âge (a_i) sont multipliés par les effectifs des différents groupes d'âge de la Population Mondiale de Référence (p_i) ce qui donne $a_i \times p_i$, la somme des $a_i p_i$ fait 9.902.410 soit,

$$\sum_{i=1}^A a_i p_i = 9.062.410$$

Tableau 4. Calcul des taux d'incidence standardisés sur l'âge par la méthode directe. Cancer du poumon chez l'homme en Ecosse période 1980-84.

N° d'ordre	Groupe d'âge	Taux spécifique par âge pour 100.000	Population Mondiale de Référence	
(i)		(a_i)	(p_i)	(a_i/p_i)
1	0-4	0,00	12 000	0
2	5-9	0,00	10 000	0
3	10-14	0,00	9 000	0
4	15-19	0,00	9 000	0
5	20-24	0,37	8 000	2 960
6	25-29	0,33	8 000	2 640
7	30-34	3,26	6 000	19 560
8	35-39	7,48	6 000	44 880
9	40-44	21,12	6 000	126 720
10	45-49	53,20	6 000	319 200
11	50-54	128,16	5 000	640 800
12	55-59	263,17	4 000	1 052 680
13	60-64	422,49	4 000	1 689 960
14	65-69	600,04	3 000	1 800 120
15	70-74	780,55	2 000	1 561 100
16	75-79	919,04	1 000	919 040
17	80-84	981,22	500	490 610
18	85+	784,28	500	392 140
			100.000	9.062.410

Les effectifs de la population de référence utilisés sont ceux de la Population Mondiale de Référence dont la somme fait 100.000 par convention soit,

$$\sum_{i=1}^A p_i = 100.000$$

Pour le cancer du poumon, le taux d'incidence annuel moyen standardisé sur l'âge (TSA) pour 100.000 hommes en Ecosse dans la période 1980-84 se calcule ainsi :

$$TSA = \frac{\sum_{i=1}^A a_i p_i}{\sum_{i=1}^A p_i} = \frac{9.062.410}{100.00} = 90,62410$$

C'est à dire 90,6 pour 100.000 et par an (les unités des TSA pour 100.000 et par an sont celles des taux spécifiques par âge, a_i utilisés dans les calculs)

Taux tronqués

Doll et Cook (1967) ont proposé le calcul des taux sur la base d'un âge tronqué (35-64 ans), surtout à cause des problèmes posés par les personnes âgées chez qui le diagnostic et l'enregistrement des cancers peut être beaucoup moins fiable. Plusieurs auteurs continuent à présenter des données en utilisant des taux tronqués, bien qu'on ne soit pas sûr que l'exactitude gagnée compense l'augmentation de la complexité des calculs et de l'interprétation, et le gaspillage des données collectées inutilement. En effet, le calcul limite nettement l'intérêt d'une partie des données contenues dans le tableau 4.

Le taux tronqué standardisé par l'âge (TTSA) peut être écrit de la façon suivante :

$$\text{TTSA} = \frac{\sum_{i=8}^{13} a_i p_i}{\sum_{i=8}^{13} p_i} \quad (9.7)$$

Il est clair que l'expression (9.7) est un cas particulier de l'expression (9.6) avec une sommation débutant à la classe d'âge 8 (correspondant à 35-39 ans) et se terminant avec la classe d'âge 13 (correspondant à 60-64 ans). De la même façon, pour les comparaisons des taux d'incidence dans l'enfance, la classe d'âge tronquée 0-14 a été utilisée, avec la partie correspondante de la population de référence (Parkin et coll, 1988).

Ecart-type des taux standardisés - méthode directe

Un taux d'incidence standardisé calculé à partir de données réelles est considéré, en théorie statistique, comme une estimation d'une vraie valeur d'un paramètre (qui aurait été connu seulement si le nombre d'observations était infiniment grand). Il est donc habituel de présenter avec ce taux une mesure de sa précision : son écart type.

L'écart-type peut aussi être utilisé pour calculer l'intervalle de confiance du taux, qui est plus facile à interpréter. L'intervalle de confiance à 95% représente une étendue de valeurs à l'intérieur de laquelle on a 95% de chance que se trouve la vraie valeur du taux d'incidence (cela signifie que seulement 5 estimations sur 100 auront un intervalle de confiance qui n'inclura pas la vraie valeur). On peut d'autre part présenter un intervalle de confiance à 99% qui aura une étendue plus large que l'intervalle à 95% car il présentera un plus grand degré de certitude.

De façon générale, l'intervalle de confiance d'un taux standardisé à $[100(1 - \alpha)]$ % d'un taux standardisé (TSA), avec un écart-type e.t.(TSA) peut s'écrire :

Exemple 3. Calcul des taux tronqués standardisés sur l'âge par la méthode directe**Tableau 5. Calcul du taux d'incidence tronqué (35-64 ans) standardisé sur l'âge - Cancer du poumon chez les hommes en Ecosse, période 1980-84**

N° d'ordre	Groupe d'âge	Taux spécifique par âge pour 100.000	Population Mondiale de Référence	
(i)		(a _i)	(p _i)	(a _i × p _i)
1	0-4			
2	5-9			
3	10-14			
4	15-19			
5	20-24			
6	25-29			
7	30-34			
8	35-39	7,48	6 000	44.880
9	40-44	21,12	6.000	126.720
10	45-49	53,20	6.000	319.200
11	50-54	128,16	5.000	640.800
12	55-59	263,17	4.000	1.052.680
13	60-64	422,49	4.000	1.689.960
14	65-69			
15	70-74			
16	75-79			
17	80-84			
18	85+			
			31.000	3.874.240

Dans cet exemple

$$\sum_{i=8}^{13} p_i = 31\ 000$$

et

$$\sum_{i=8}^{13} a_i p_i = 3\ 874\ 240$$

Il faut rappeler que l'on ne prend en compte que les populations et les taux correspondant aux tranches d'âges étudiées, donc :

$$TTSA = \frac{\sum_{i=8}^{13} a_i p_i}{\sum_{i=8}^{13} p_i} = \frac{3\ 874\ 240}{31\ 000} = 124,97548$$

Ce qui correspond pour les cancers du poumon à un taux d'incidence annuel moyen tronqué (35-64) standardisé sur l'âge chez les hommes en Ecosse durant la période 1980-84 de 125 pour 100.000.

$$\text{TSA} \pm Z_{\alpha/2} \times (\text{e.t.}(\text{TSA})) \quad (9.8)$$

où $Z_{\alpha/2}$ est la déviation de la loi normale centrée réduite (voir Armitage et Berry, 1987 pour discussion des principes généraux). Par exemple, l'intervalle de confiance peut être calculé en choisissant $Z_{\alpha/2} = 1,96$, le 97,5 ième percentile de la distribution normale. Pour un intervalle de confiance à 99%, $Z_{\alpha/2}$ est égal à 2,58.

Il existe 2 méthodes pour calculer un écart-type d'un taux directement ajusté sur l'âge : l'approximation binômiale et l'approximation de Poisson, qui sont illustrées ci-dessous. Elles donnent des résultats identiques et chacune peut être utilisée.

Le taux d'incidence standardisé (TSA) peut être calculé à partir de la formule (9.6).

On démontre que la variance du TSA est :

$$\text{Var (TSA)} = \frac{\sum_{i=1}^A [a_i p_i^2 (100\,000 - a_i) / n_i]}{\sum_{i=1}^A p_i} \quad (9.9)$$

L'écart-type du TSA (e.t.(TSA)) peut être calculé simplement :

$$\text{e.t.}(\text{TSA}) = \sqrt{\text{Var (TSA)}} \quad (9.10)$$

L'intervalle de confiance à 95% du TSA calculé dans l'exemple 2 est donné par la formule 9.8 :

$$\begin{aligned} \text{TSA} \pm Z_{\alpha/2} \times (\text{e.t.}(\text{TSA})) &= 90.62 \pm 1.96 \times 0.73 \\ &= 89.19 \text{ à } 92.05 \end{aligned}$$

On peut obtenir une autre expression, comme l'ont souligné Armitage et Berry (1987), si les a_i sont petits (comme c'est généralement le cas) en faisant une approximation de Poisson pour la variance binômiale des a_i . On obtient alors une expression de la variance du taux standardisé (Var (TSA)) :

$$\text{Var (TSA)} = \frac{\sum_{i=1}^A (a_i p_i^2 \times 100\,000 / n_i)}{\left(\sum_{i=1}^A p_i \right)^2} \quad (9.11)$$

et l'écart-type du taux standardisé (e.t. (TSA)) est la racine carrée de la variance, comme précédemment (formule 9.10).

Comparaison de deux taux standardisés calculés selon la méthode directe

Il est souvent intéressant d'étudier le rapport de taux standardisés de différents groupes de population, par exemple de deux régions différentes, ou de groupes ethniques différents, ou de différentes périodes de temps. Le rapport de deux taux standardisés par la méthode directe, TSA_1 / TSA_2 , est appelé le ratio des taux standardisés (RTS), et représente le risque relatif d'avoir la maladie dans la population 1 par rapport à la population 2. Il est habituel de calculer aussi la signification statistique de ce ratio de taux standardisés (pour savoir si le ratio observé est significativement différent de 1). Plusieurs méthodes sont disponibles pour calculer l'intervalle de confiance exact de ratio de taux standardisés (Breslow et day, 1987, page 64 ; Rothman, 1986 ; Checkoway et coll, 1989) ; une approximation peut être obtenue avec la formule suivante (Smith, 1987) :

$$(TSA_1 / TSA_2)^{1 \pm (Z_{\alpha/2} / X)}$$

$$\text{où } X = \frac{(TSA_1 - TSA_2)}{\sqrt{(e.t.(TSA_1)^2 + e.t.(TSA_2)^2)}} \quad (9.12)$$

et $Z_{\alpha/2} = 1.96$ (au niveau 95%)
ou $Z_{\alpha/2} = 2.58$ (au niveau 99%)

Si 1 est compris dans cet intervalle, les taux standardisés TSA_1 et TSA_2 ne sont pas significativement différents (au degré de signification de 5% si $Z_{\alpha/2} = 1,96$ a été utilisé, ou au degré de signification de 1% si $Z_{\alpha/2} = 2,58$ a été utilisé).

Si l'on veut comparer des taux standardisés de plusieurs sous-populations, un moyen logique de procéder est de comparer le taux standardisé de chaque sous-population avec celui de la population prise dans son ensemble, au lieu de comparer toutes les paires possibles. Par exemple, pour l'atlas d'incidence des cancers en Ecosse, Kemp et coll (1985) ont obtenu un numérateur et un dénominateur pour 56 districts d'Ecosse sur une période de 6 ans (1975-1980).

Pour chacune des localisations cancéreuses et séparément pour chaque sexe, une moyenne annuelle du taux d'incidence standardisé pour 100 000 personnes-années a été calculée avec la méthode directe en utilisant la population mondiale de référence (comme décrite ci-dessus). L'écart-type a été calculé pour chaque district et pour l'Ecosse toute entière de la même façon, permettant un résumé des comparaisons statistiques. Pour éviter de comparer les districts à population élevée (par exemple Glasgow avec 17% de la population totale d'Ecosse) avec le taux de l'Ecosse, qui est lui même affecté par leur contribution, le taux de chaque district a été comparé avec le taux du reste de l'Ecosse (par exemple Glasgow avec l'ensemble de l'Ecosse moins Glasgow). La méthode de

Exemple 4. Calcul de l'écart type d'un taux standardisé sur l'âge (approximation binomiale)

Le tableau 6 présente les données nécessaires au calcul de l'écart type du taux d'incidence standardisé sur l'âge pour le cancer du poumon chez l'homme en Ecosse en 1980-84.

Tableau 6. Calcul de l'écart type du taux d'incidence annuel moyen (tous âges confondus) standardisé sur l'âge pour 100.000 par la méthode binomiale - Cancer du poumon chez l'homme en Ecosse durant la période 1980-84.

Groupe d'âge	Taux spécifique par âge pour 100.000	Population Mondiale de Référence	Personne-années	$a_i p_i^2 (100\ 000 - a_i)$
	(a_i)	(p_i)	(n_i)	p_i
0-4	0,00	12 000	827 400	0
5-9	0,00	10 000	856 500	0
10-14	0,00	9 000	1 061 500	0
15-19	0,00	9 000	1 157 400	0
20-24	0,37	8 000	1 074 900	2 202 988
25-29	0,33	8 000	917 700	2 301 398
30-34	3,26	8 000	890 300	13 181 644
35-39	7,48	6 000	816 000	32 997 532
40-44	21,12	6 000	724 400	104 936 424
45-49	53,20	6 000	706 800	270 823 584
50-54	128,16	5 000	703 800	454 659 552
55-59	263,17	4 000	691 200	607 586 624
60-64	422,49	4 000	610 900	1 101 862 784
65-69	600,04	3 000	511 800	1 048 838 592
70-74	780,55	2 000	425 600	727 873 536
75-79	919,04	1 000	266 800	341 301 952
80-84	981,22	500	122 500	198 284 096
85+	784,28	500	54 700	355 634 848

$$\sum_{i=1}^{18} \left[a_i p_i^2 (100\ 000 - a_i) / n_i \right] = 5\ 262\ 486\ 016$$

et

$$\left(\sum_{i=1}^{18} p_i \right)^2 = 10\ 000\ 000\ 000$$

En appliquant la formule (9.9), on obtient :

$$\begin{aligned} \text{Var (TSA)} &= \frac{5\ 262\ 486\ 016}{10\ 000\ 000\ 000} \\ &= 0,526249 \end{aligned}$$

Puis avec la formule (9.10), e.t. (TSA) = 0,73

Donc l'écart type du taux d'incidence moyen annuel standardisé sur l'âge du cancer du poumon en Ecosse entre 1980-84 est de 0,73.

Exemple 5. Calcul de l'erreur type d'un taux d'incidence standardisé sur l'âge (approximation de poisson)

Le tableau 7 présente les données nécessaires au calcul de l'écart type du taux d'incidence standardisé sur l'âge utilisant cette deuxième méthode.

Tableau 7. Ecart type du taux d'incidence standardisé sur l'âge (approximation de poisson)

Groupe d'âge	Taux spécifique par âge pour 100.000	Population Mondiale de Référence	Personnes-années	$\frac{a_i p_i^2 \times 100\ 000}{n_i}$
	(a_i)	(p_i)	(n_i)	
0-4	0,00	12 000	827 400	0
5-9	0,00	10 000	856 500	0
10-14	0,00	9000	1 061 500	0
15-19	0,00	9000	1 157 400	0
20-24	0,37	8000	1 074 900	2 202 986
25-29	0,33	8000	917 700	2 301 406
30-34	3,26	8000	890 300	13 182 074
35-39	7,48	6000	816 000	33 000 000
40-44	21,12	6000	724 400	104 958 592
45-49	53,20	6000	706 800	270 967 744
50-54	128,16	5000	703 800	455 242 976
55-59	263,17	4000	691 200	609 189 824
60-64	422,49	4000	610 900	1 106 537 856
65-69	600,04	3000	511 800	1 055 169 984
70-74	780,55	2000	425 600	733 599 616
75-79	919,04	1000	266 800	344 467 776
80-84	981,22	500	122 500	200 248 976
85+	784,28	500	54 700	358 446 048

$$\sum_{i=1}^{18} [a_i p_i^2 \times 100\ 000 / n_i] = 5\ 289\ 515\ 520$$

et

$$\left(\sum_{i=1}^{18} p_i \right)^2 = 10\ 000\ 000\ 000$$

En utilisant la formule (9.11) on obtient :

$$\begin{aligned} \text{Var (TSA)} &= \frac{5\ 289\ 515\ 520}{10\ 000\ 000\ 000} \\ &= 0,52895 \end{aligned}$$

Puis avec la formule (9.10)

$$\text{e.t. (TSA)} = 0,73$$

On observe dans cet exemple que le résultat est le même, à la centésimale près, que celui obtenu par la méthode précédente, ce qui met en évidence la similitude des deux approches.

comparaison était celle décrite ci-dessus pour les taux standardisés par la méthode directe et les ratios ont été classés en 5 catégories comme : significativement élevé à 1% (++) ; (2) significativement élevé à 5% (+) ; (3) non significativement élevé ou bas ; (4) significativement bas à 5 % (-) ; ou (5) significativement bas à 1% (--).

Exemple 6. Calcul de l'intervalle de confiance du rapport de deux taux standardisés

Nous avons vu que le taux d'incidence standardisé sur l'âge du cancer du poumon chez l'homme en Ecosse était de 90,6 avec une erreur type de 0,73 pour la période 1980-84. En 1960-64 il était de 68,3 avec une erreur de 0,67.

Le rapport de ces taux d'incidence $SAR_1/SAR_2 = 90,6/68,3 = 1,3265$

Pour calculer l'intervalle de confiance de ce rapport des taux standardisés d'incidence, on utilise la formule (9-12) :

$$X = \frac{TSA_1 - TSA_2}{\sqrt{(e.t.(TSA_1))^2 + e.t.(TSA_2)^2}}$$

$$= \frac{90,6 - 68,3}{\sqrt{(0,73)^2 + (0,67)^2}} = \frac{22,3}{\sqrt{0,9818}} = 22,51$$

Pour avoir l'intervalle de confiance à 95% nous prenons $Z_{\alpha/2} = 1,96$; avec lequel on obtient

$$\text{Limite inférieure} = (TSA_1/TSA_2)^{1-(Z_{\alpha/2}/X)} = (1,3265)^{1-(1,96/22,51)} = 1,29$$

$$\text{Limite supérieure} = (TSA_1/TSA_2)^{1+(Z_{\alpha/2}/X)} = (1,3265)^{1+(1,96/22,51)} = 1,36$$

Si les taux avaient été les mêmes pour les 2 périodes, le rapport (TSA_1/TSA_2) aurait été de 1. Comme l'intervalle de confiance ne comprend pas cette valeur [1,29 - 1,36], on peut conclure que les taux sont significativement différents avec un risque d'erreur inférieur à 5%.

Le tableau 8 présente les taux d'incidence du cancer du poumon provenant de l'atlas d'Ecosse (Kemp et coll, 1985). Chez les hommes, le plus haut taux rapporté provenait du district 33 (Glasgow City : 130,6 pour 100 000, écart-type 2,01) qui était significativement différent au degré de 1% du taux du reste de l'Ecosse. La banlieue d'Inverclyde (109,9 ; 5,35) présentait également un taux élevé statistiquement significatif à 1%, de même que la ville d'Edinburgh (103,2 ; 2,32). Il est important de noter l'effet de la taille de la population sur le degré de signification statistique. Bien que la ville d'Edinburgh soit classée seulement 7ème en terme de taux d'incidence des cancers du poumon chez l'homme, grâce à la taille de sa population, elle était un des trois districts du groupe où la signification statistique était la plus forte.

Le même phénomène s'observe pour les femmes avec la ville de Glasgow (33,3 ; 0,90) qui a le plus haut taux. Cependant, le deuxième taux le plus élevé provenait de Badenoch (31,8 ; 9,36), qui n'est pas statistiquement différent du reste de l'Ecosse, à cause de la petite taille de la population.

Test de tendance des taux standardisés

Il est possible d'étendre la notion de test des différences entre des paires de taux standardisés (voir ci-dessus) à une notion plus générale. En effet, quelquefois des taux standardisés ordonnés sur une quelconque échelle sont disponibles. Les classes de cette échelle peuvent correspondre au degré d'exposition, à un facteur étiologique ou simplement au temps.

Tableau 8. Taux d'incidence du cancer du poumon dans différents districts d'Ecosse 1975-80

District		Hommes				Femmes			
N°	Nom	Cas	TSA	e.t.	Ordre	Cas	TSA	e.t.	Ordre
7	Badenoch	21	55,0 ⁻⁻⁻	12,58	47	14	31,8	9,36	2
21	Edinburgh	2.087	103,2 ⁺⁺	2,32	7	734	25,9 ⁺⁺	1,05	8
24	Tweeddale	49	73,6	11,03	28	28	29,1	6,24	3
33	Glasgow	4.579	130,6 ⁺⁺	2,01	1	1.802	33,3 ⁺⁺	0,90	1
37	Cumbernauld	145	109,1	9,25	3	36	21,8	3,58	18
45	Inverclyde	438	109,9 ⁺⁺	5,35	2	137	27,5	2,46	5
54	Orkney	36	40,2 ⁻⁻⁻	6,97	56	12	13,6 ⁻	4,07	48
55	Shetland	39	46,1	7,70	53	7	5,8 ⁻⁻⁻	2,33	56
Total		19.239	91,4	0,67		6.136	23,1	0,31	

TSA, Taux standardisé sur l'âge pour 100.000 (méthode directe, Population Mondiale de Référence)

E.T. Ecart type

++, significativement supérieur au reste de l'Ecosse, $p < 0,01$

--, significativement inférieur au reste de l'Ecosse, $p < 0,01$

-, significativement inférieur au reste de l'Ecosse, $p < 0,05$

Source : Kemp et coll, 1985

Des exemples simples sont les taux standardisés provenant de différentes périodes dans le temps ou de différentes classes socio-économiques. On pourrait alors classer ces taux standardisés provenant de différentes zones géographiques (provinces par exemple) en fonction de la pluviométrie moyenne, de l'altitude ou du degré de pollution atmosphérique.

Dans ces circonstances, l'investigateur est intéressé non seulement par la comparaison des paires de taux standardisés, mais veut aussi savoir si les taux d'incidence suivent une quelconque tendance liée aux catégories d'exposition. La méthode la plus simple pour représenter une tendance linéaire est de tracer une ligne droite représentant l'équation de la régression.

Les taux d'incidence standardisés annuels du cancer du poumon chez l'homme en Ecosse entre les années 1960 et 1970 seront utilisés comme exemple. Pour estimer les tendances temporelles, l'année actuelle peut être utilisée pour ordonner les taux ; cependant, pour simplifier les calculs, on peut soustraire 1959 à chacune des années, si bien que 1960 devient 1, 1961 devient 2, ... et 1970 devient 11. Les mêmes résultats seront obtenus que l'on utilise l'une ou l'autre des valeurs.

Dans une régression simple¹, il y a deux types de variables : la variable explicative (dans ce cas, l'année notée x) et la variable à expliquer (dans ce cas, le taux standardisé noté y) ; l'équation de la régression linéaire peut être écrite :

$$y = a + bx \quad (9.13)$$

ou y = taux d'incidence standardisé du cancer du poumon,

x = numéro de l'année (année de départ : 1959)

a = "ordonnée à l'origine"

b = pente de la droite de régression

Les formules pour a et b , et les écarts-types correspondants proviennent de Bland (1987). Par exemple,

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

peut être écrite :

$$b = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \quad (9.14)$$

ou n = nombre de paires d'observations

et $\bar{y} = \sum y / n$ et $\bar{x} = \sum x / n$

L'écart-type de la pente b est :

$$\text{e.t.}(b) = \sqrt{\frac{1}{n-2} \left\{ \sum (y_i - \bar{y})^2 - b^2 \sum (x_i - \bar{x})^2 \right\} \over \sum (x_i - \bar{x})^2} \quad (9.15)$$

¹ il est souvent plus correct d'utiliser une régression pondérée où chaque point ne contribue pas de la même façon à l'ajustement de la droite de régression. On utilise souvent comme pondération $p_i = 1/\text{Var}(y_i)$.

(voir Armitage et Berry 1987).

L'ordonnée à l'origine (a) peut être calculée :

$$a = \bar{y} - b\bar{x} \quad (9.16)$$

La pente calculée (b) représente l'augmentation moyenne du taux d'incidence standardisé avec l'augmentation d'une unité de la variable explicative, c'est-à-dire, dans cet exemple, l'augmentation moyenne d'une année sur l'autre. L'écart-type de la pente (e.t. (b)) peut être utilisé pour calculer l'intervalle de confiance de la pente, de manière analogue à celle qui a été utilisée dans la formule 9.8.

Un test peut être fait pour savoir si la pente est significativement différente de 1, en calculant le ratio de la pente à son écart-type (b/e.t.(b)) qui suivra une distribution de t avec n-2 degrés de liberté (voir Armitage et Berry, 1987).

La standardisation par l'âge - méthode indirecte

Une autre méthode de standardisation fréquemment utilisée est la standardisation indirecte. Il s'agit d'une comparaison entre un nombre de cas observés et attendus. Le nombre de cas attendus est calculé en appliquant les taux par âges théoriques (a_i) à la population étudiée :

$$\sum_{i=1}^A e_i = \sum_{i=1}^A a_i n_i / 100\ 000 \quad (9.17)$$

ou un e_i , le nombre de cas attendus dans la classe d'âge i, est le produit du "taux théorique" et du nombre de personnes dans la classe d'âge i dans la population étudiée.

Le ratio standardisé (M) peut être calculé en comparant le nombre de cas observé ($\sum r_i$) avec celui attendu.

$$M = \frac{\sum_{i=1}^A r_i}{\sum_{i=1}^A e_i} = \frac{\sum_{i=1}^A r_i}{\sum_{i=1}^A a_i n_i / 100\ 000} \quad (9.18)$$

Il est généralement exprimé sous forme de pourcentage en multipliant par 100. S'il est appliqué aux données d'incidence, il est communément connu comme étant le RSI (Ratio Standardisé d'Incidence) ; s'il est appliqué aux données de mortalité, il est connu comme étant le RSM (Ratio Standardisé de Mortalité).

Exemple 7. Calcul du changement annuel moyen du taux d'incidence standardisé sur l'âge pour le cancer du poumon chez l'homme en Ecosse et test de la tendance évolutive.

Entre 1960 et 1970, les taux d'incidence annuels tous âges confondus du cancer du poumon chez l'homme en Ecosse étaient respectivement de 77,05 ; 81,78 ; 87,78 ; 89,05 ; 85,68 ; 87,04 ; 89,97 ; 100,50 ; 104,85 ; 104,77 et 107,57 pour 100.000.

Dans cet exemple l'année de référence est 1959. En d'autres termes, 1960 devient 1, 1961 devient 2 et ainsi pour chaque années jusqu'à 1970 qui devient 11.

$$\begin{aligned} n &= 11 \\ \sum x_i &= 66 \\ \sum x_i^2 &= 506 \\ \sum y_i &= 1016,0 \\ \sum y_i^2 &= 94913,0 \\ \sum x_i y_i &= 6419,2 \end{aligned}$$

Avec la formule (9.14)

$$\begin{aligned} b &= \frac{6419,2 - ((66 \times 1016)/11)}{506 - ((66 \times 66)/11)} \\ &= \frac{323,2}{110} \\ &= 2,938 \end{aligned}$$

et

$$\begin{aligned} \bar{x} &= \sum x_i / n = 6 \\ \bar{y} &= \sum y_i / n = 92,36 \end{aligned}$$

A partir de l'expression (9.15) on obtient

$$\text{e.t.}(b) = 0,351$$

L'intervalle de confiance à 95% de la pente (b) se calcule comme suit :

$$\begin{aligned} \text{limite inférieure} &= 2,938 - (1,96 \times 0,351) \\ &= (2,938 - 0,688) \\ &= 2,250 \\ \text{limite supérieure} &= 2,938 + (1,96 \times 0,351) \\ &= 3,626 \end{aligned}$$

L'intervalle de confiance à 95% de la pente b est [2,22 ; 3,36]. Pour vérifier que la pente diffère bien de zéro on calcule la quantité

$$\frac{b}{\text{e.t.}(b)} = \frac{2,938}{0,351} = 8,375$$

Quantité, que l'on compare aux valeurs de la table de t avec (n - 2) degrés de liberté. Ici la pente est significativement différente de 0 avec un niveau de significativité à 1% ($p < 0,01$), ce qui est très significatif.

On peut conclure qu'il existe une augmentation significative de l'incidence du cancer du poumon en Ecosse entre 1960 et 1970 et que cette augmentation des taux standardisés est en moyenne annuellement de 2,9 cas/100 000.

Exemple 8. Calcul du rapport standardisé d'incidence par la méthode indirecte

Le tableau 9 présente les éléments nécessaires au calcul du RSI du cancer du poumon en Ecosse en 1980-84 en utilisant comme taux de référence les taux de 1960-64.

Tableau 9. Calcul du rapport standardisé d'incidence pour le cancer du poumon chez les hommes en Ecosse en 1980-84. Les taux de 1960-64 sont utilisés comme référence

Groupe d'âge	Taux pour 100.000 en 1960-64	Personnes - années d'observation	Nombre de cas attendus	Nombre de cas observés
	(a _i)	(n _i)	(e _i =a _i n _i /100 000)	(r _i)
0-4	0,00	827 400	0,00	0
5-9	0,00	827 400	0,00	0
10-14	0,00	856 500	0,00	0
15-19	0,00	1 061 500	1,04	0
20-24	1,12	1 157 400	12,04	4
25-29	1,70	1 074 900	15,60	3
30-34	4,91	917 700	43,71	29
35-39	16,25	890 300	132,60	61
40-44	29,38	816 000	212,83	153
45-49	79,92	724 400	564,87	376
50-54	151,07	703 800	1063,23	902
55-59	269,58	703 800	1863,34	1819
60-64	391,41	610 900	2391,12	2581
65-69	459,74	511 800	2352,95	3071
70-74	400,46	425 600	1704,36	3322
75-79	285,21	266 800	760,94	2452
80-84	207,49	122 500	254,18	1202
85+	100,84	54 700	55,16	429
			11 427,97	16 404

$$\sum e_i = \sum_{i=1}^A a_i n_i / 100\,000 = 11\,427,97$$

et

$$\sum r_i = 16\,404$$

En appliquant la formule (9.18) le ratio standardisé d'incidence, $M \times 100$, pour la période 1980-84 est :

$$\frac{16\,404}{11\,427,97} \times 100 = 144$$

En d'autres termes la fréquence du cancer du poumon a augmenté de 44% de 1960-64 à 1980-84 compte tenu des différences de structure d'âge de la population.

Ecart-type du ratio standardisé

Le ratio standardisé (M) est calculé par la formule 9.18 et sa variance, Var (M), est :

$$\text{Var}(M) = \frac{\sum_{i=1}^A r_i}{\left(\sum_{i=1}^A a_i n_i / 100\,000 \right)^2} \quad (9.19)$$

et l'écart-type de ce ratio indirect, e.t.(M), est la racine carrée de la variance, comme précédemment (formule 9.10).

Exemple 9. Calcul de l'erreur type du rapport d'incidence standardisé par la méthode indirecte

Le tableau 9 présente les éléments nécessaires au calcul de l'erreur type du rapport d'incidence standardisé (RSI) du cancer du poumon chez les hommes en Ecosse en 1980-84 relativement à la période 1960-64.

Si nous écrivons le RSI sous la forme $M \times 100$,

$$\text{Var}(\text{RSI}) = \text{Var}(M \times 100) = 10\,000 \text{ Var}(M)$$

et en appliquant la formule (9.19)

$$\begin{aligned} \text{Var}(\text{RSI}) &= 10\,000 \frac{\sum_{i=1}^A r_i}{\left(\sum_{i=1}^A a_i n_i \right)^2} \\ &= \frac{10\,000 \times 16\,404}{(11\,427,97)^2} \\ &= 1,2561 \end{aligned}$$

et

$$\text{s.e.}(\text{RSI}) = \sqrt{\text{Var}(\text{RSI})} = 1,12$$

Aussi, dans cet exemple le RSI est 144 et son écart type est de 1,12. En conséquence l'intervalle de confiance à 95% du RSI est

$$\text{RSI} \pm (Z_{\alpha/2} \times (\text{s.e.}(\text{RSI}))) = 144 \pm (1,96 \times 1,12)$$

c'est à dire, (141,8 ; 146,2)

On peut conclure qu'il existe une réelle augmentation de l'incidence du cancer du poumon chez les hommes en Ecosse entre 1960 et 1970, avec une augmentation moyenne d'environ 2,9 cas pour 100.000 chaque année.

$$e.t.(M) = \frac{\sqrt{\sum_{i=1}^A r_i}}{\sum_{i=1}^A a_i n_i / 100\ 000} \quad (9.20)$$

Vandenbroucke (1982) a proposé une méthode simplifiée de calcul de l'intervalle de confiance à $(100(1-\alpha))\%$ du ratio standardisé, en procédant en deux étapes. Les limites inférieures et supérieures pour le nombre d'évènements observés sont d'abord calculées :

$$\begin{aligned} \text{Limite inférieure} &= \left[\sqrt{\text{evenements observés}} - (Z_{\alpha/2} \times 0,5) \right]^2 \\ \text{Limite supérieure} &= \left[\sqrt{\text{evenements observés}} + (Z_{\alpha/2} \times 0,5) \right]^2 \end{aligned}$$

La division de ces limites pour le nombre observé par le nombre attendu d'évènements amène à une approximation de l'intervalle de confiance à 95% (ou 99%) pour le RSI.

La limite inférieure du RSI

$$\begin{aligned} &= \frac{\left[\sqrt{\text{evenements observés}} - (Z_{\alpha/2} \times 0,5) \right]^2}{\text{evenements attendus}} \\ &= \frac{\left\{ \sqrt{\sum_{i=1}^A r_i} - (Z_{\alpha/2} \times 0,5) \right\}^2}{\sum_{i=1}^A a_i n_i / 100\ 000} \end{aligned} \quad (9.21)$$

La limite supérieure du RSI

$$= \frac{\left\{ \sqrt{\sum_{i=1}^A r_i} + (Z_{\alpha/2} \times 0,5) \right\}^2}{\sum_{i=1}^A a_i n_i / 100\ 000} \quad (9.22)$$

Test de la différence entre le ratio standardisé et la valeur attendue

Il peut être obtenu simplement en calculant l'intervalle de confiance approprié, si bien que l'on peut voir si la valeur de 100 est incluse ou exclue de l'intervalle.

Exemple 10. Calcul approché de l'intervalle de confiance à 95% du rapport d'incidence standardisé

Les éléments nécessaires au calcul sont donnés au tableau 9.

Nombre d'événements observés = 16 404

Nombre d'événements attendus = 11 427,97

Pour une intervalle de confiance à 95%, $Z_{\alpha/2}$ est 1,96

$$\text{limite inférieure} = \frac{[\sqrt{16\,404} - (1,96 \times 0,5)]^2}{11\,427,97} = 1,411353$$

$$\text{limite supérieure} = \frac{[\sqrt{16\,404} + (1,96 \times 0,5)]^2}{11\,427,97} = 1,45747$$

Compte tenu du fait que le RSI s'exprime sous forme de pourcentage, ses limites approchées sont 141,4 et 145,7.

Ces limites sont très proches de celles obtenues dans l'exemple 9 qui sont plus précises mais aussi beaucoup moins faciles à calculer.

Exemple 11. Test de la signification des rapports standardisés d'incidence par la méthode indirecte.

Dans l'exemple 9, nous avons observé que le RSI des années 1980-84 en Ecosse était de 144 avec une erreur type de 1,12 et en conséquence, un intervalle de confiance à 95% de 141,8 à 146,2, qui ne comprend pas 100.

De même, l'intervalle de confiance à 99% de ce RSI était $144 \pm (2,58 \times 1,12)$; soit de 141,1 à 146,9 qui ne comprend pas 100 lui non plus.

On peut donc conclure que le taux d'incidence des cancers du poumon observé en 1980-84 était significativement plus élevé que celui de 1960-64 avec un risque de 1% ($p < 0,01$).

Il faut noter qu'avec la standardisation indirecte, les pondérations qui sont utilisées dans le processus de standardisation sont des effectifs spécifiques par âge dans le sous-groupe étudié. Ainsi, si les RSI sont calculés pour beaucoup de sous-groupes de population (par exemple différentes régions, différents groupes ethniques) avec différentes structures de populations, les différents RSI ne sont relatifs qu'à la population de référence (comme dans l'exemple 11) et ne peuvent pas être comparées aux autres. Ainsi, si le RSI pour les cancers du poumon chez les hommes en Ecosse entre 1970 et 1974 est estimé à 1,22 (ou 122 en pourcentage) en utilisant les taux d'incidence de 1960 à 1964 comme référence, on ne peut pas en déduire que le risque relatif entre 1980 et 1984 par rapport à 1970-74 est $142/122$, c'est à dire : 1,18.

Taux cumulé et risque cumulé

Day (1987) a proposé le taux cumulé comme un autre taux d'incidence standardisé. Dans le volume IV de la série "Cancer Incidence in Five Continents", cette mesure remplace les calculs faits sur les populations européennes et africaines de référence (Waterhouse et coll, 1982).

Le risque cumulé est le risque pour un individu de développer le cancer étudié jusqu'à un certain âge, s'il n'y avait pas d'autres causes de décès. Il est essentiel de préciser la période d'âge sur laquelle le risque est cumulé : habituellement c'est de 0 à 74 ans, représentant l'ensemble de la vie. Pour les cancers de l'enfance, la période de 0 à 14 ans peut être utilisée. Le taux cumulé est la somme, pour chaque année d'âge, des taux d'incidence spécifiques, en prenant de la naissance à l'âge de 74 ans pour le taux 0-74. Il s'agit d'un taux standardisé par la méthode directe avec une population ayant le même effectif pour chaque groupe d'âge. Il peut être considéré comme une approximation du risque cumulé.

On rappelle que a_i est le taux d'incidence spécifique dans la i ème classe d'âge qui couvre t_i année. En d'autres termes, si les classes d'âges utilisées sont 0,1-4, 5-9 ... alors t_1 sera 1, t_2 sera 4, t_3 sera 5, ect... Le taux cumulé peut s'écrire :

$$\text{Taux cumulé} = \sum_{i=1}^A a_i t_i \quad (9.23)$$

où la sommation se fait jusqu'à la classe d'âge A. Si le calcul des taux spécifiques a été fait en utilisant des classes d'âge de 5 ans, pour le taux cumulé de 0 à 74 ans, A = 15 et :

$$\text{Taux cumulé (0 - 74)} = \sum_{i=1}^{15} 5a_i$$

Habituellement, cette quantité est exprimée sous forme de pourcentage plutôt que pour 100 000.

Day (1987) a démontré que le risque cumulé était :

$$\text{Risque cumulé} = 100 \times [1 - \exp (-\text{Taux cumulé}/100)] \quad (9.24)$$

Exemple 12. Calcul du taux cumulé et du risque cumulé

Le tableau 10 présente les éléments nécessaires au calcul du taux cumulé de cancers du poumon chez les hommes en Ecosse de 1980-84. Dans cet exemple on utilise des groupes d'âges d'amplitude égale, toujours 5 ans.

Tableau 10. Calcul du taux cumulé et du risque cumulé (0-74) de cancers du poumon chez les hommes en Ecosse de 1980-84.

Groupe d'âge	Taux spécifiques par âge pour 100 000	Amplitude du groupe d'âge	(Taux spécifiques par âge) X (amplitude du groupe)
	(a _i)	(t _i)	(a _i t _i)
0-4	0,00	5	0
5-9	0,00	5	0
10-14	0,00	5	0
15-19	0,00	5	0
20-24	0,37	5	1,85
25-29	0,33	5	1,65
30-34	3,26	5	16,30
35-39	7,48	5	37,40
40-44	21,12	5	105,60
45-49	53,20	5	266,00
50-54	128,16	5	640,80
55-59	263,17	5	1315,85
60-64	422,49	5	2112,45
65-69	600,04	5	3000,20
70-74	780,55	5	3902,75
75-79	-	-	-
80-84	-	-	-
85+	-	-	-
			<hr/> 11 400,85

$$\text{Taux cumulés} = \sum_{i=1}^{15} a_i t_i = 11\,400,85$$

Le taux cumulé (0-74) est égal à : 11 400,9 pour 100.000, ou 11,4%

Le risque cumulé (0-74) est égal à :

$$100 \times [1 - \exp(-11,4/100)]$$

$$= 10,8\%$$

Ainsi, en l'absence d'une autre cause de décès un homme présente en Ecosse un risque estimé à 10,8% de développer un cancer du poumon avant 75 ans.

Exemple 13. Calcul de l'écart type d'un taux cumulé

Le tableau 11 présente les éléments nécessaires pour calculer l'écart type d'un taux cumulé.

Tableau 11. Calcul de l'erreur type du taux cumulé (0-74) pour le cancer du poumon chez l'homme en Ecosse, 1980-84

Groupe d'âge	Taux spécifique par âge pour 100.000	Amplitude du groupe d'âge	Personnes-années	
	(a _i)	(t _i)	(n _i)	a _i t _i ² /n _i
0-4	0,00	5	827 400	0
5-9	0,00	5	856 500	0
10-14	0,00	5	1 061 500	0
15-19	0,00	5	1 157 400	0
20-24	0,37	5	1 074 900	0,00001
25-29	0,33	5	917 700	0,00001
30-34	3,26	5	890 300	0,00009
35-39	7,48	5	816 000	0,00023
40-44	21,12	5	724 400	0,00073
45-49	53,20	5	706 800	0,00188
50-54	128,16	5	703 800	0,00455
55-59	263,17	5	691 200	0,00952
60-64	422,49	5	610 900	0,01729
65-69	600,04	5	511 800	0,02931
70-74	780,55	5	425 600	0,04584
				<hr/>
				0,10947

$$\text{Var (Taux cum.)} = \sum_{i=1}^{15} a_i t_i^2 / n_i = 0,10947 \text{ pour } 100\ 000$$

$$= 0,00010947\%$$

L'écart type du taux cumulé (0-74), e.t. (taux cum.), s'obtient par extraction de la racine carrée de cette valeur

$$\text{e.t. (Taux cum.)} = \sqrt{0,00010947} \text{ pour } \sqrt{100}$$

$$= 0,105\%$$

Donc le taux cumulé (0-74) du cancer du poumon chez l'homme en Ecosse est de 11,4% avec un écart type de 0,1%

L'erreur-type du taux cumulé

La variance et l'erreur-type du taux cumulé peuvent être obtenues à partir des formules de la variance et de l'erreur-type du taux ajusté par la méthode directe (9.10 et 9.11) en utilisant les pondérations appropriées (c'est-à-dire la taille des intervalles d'âge, t_i) et l'approximation de Poisson :

$$\text{Var (taux cumulé)} = \sum_{i=1}^A (a_i t_i^2 / n_i) \quad (9.25)$$

et l'écart-type du taux cumulé, e.t.(taux cum.) est :

$$\text{e.t. (taux cum.)} = \sqrt{\sum_{i=1}^A (a_i t_i^2 / n_i)} \quad (9.26)$$

L'intervalle de confiance à 95% du taux cumulé est facilement obtenu grâce à l'équation 9.8 :

$$11.4 \pm (1.96 \times 0.105)$$

$$\text{soit } [11.6, 11.2]$$

DEUXIÈME PARTIE : LES MÉTHODES PROPORTIONNELLES

La fréquence relative

Si la taille de la population d'où sont issus les cas enregistrés est inconnue, il est impossible de calculer des taux d'incidence. Dans ce cas on peut étudier la distribution des cas en observant la proportion de différents types de cancer. La procédure habituelle consiste à calculer le pourcentage (ou fréquence relative) de chaque cancer par rapport au total :

$$\text{fréquence relative} = \frac{R}{T} \quad (9.27)$$

ou R = nombre de cas de cancer qui nous intéressent dans le groupe étudié ;
 T = nombre de cas de cancer (toutes localisations confondues) dans le groupe étudié.

Il est aussi possible de calculer le ratio de fréquence (Doll, 1968) où chaque cancer est exprimé comme une proportion des autres cancers plutôt qu'une proportion du total :

$$\text{fréquence rapport} = \frac{R}{T - R} \quad (9.28)$$

Ceci peut avoir des avantages dans certaines circonstances (par exemple si l'on étudie un cancer qui représente une grande proportion du nombre total de cancers), mais ce calcul présente aussi des inconvénients, que l'on ne développera pas ici.

Des comparaisons de fréquence relative peuvent être faites entre les registres ou à l'intérieur des registres, par exemple, entre différentes zones géographiques, différents groupes ethniques ou différentes périodes de temps. L'utilisation des fréquences relatives pose des problèmes car cette comparaison est souvent interprétée comme une vraie différence de risque entre deux sous-groupes alors que cette information ne peut provenir que de la comparaison de deux taux d'incidence. Le rapport des fréquences n'est égal au risque relatif que lorsque les taux d'incidence globaux (pour l'ensemble des cancers) sont égaux.

Dans l'exemple exposé dans la figure 1, le rapport entre les taux d'incidence (rapport des taux) du cancer du foie à Cali et à Singapour, qui ont les mêmes taux globaux d'incidence, est 6,9. Le rapport des fréquences est alors une bonne approximation (7,3). En revanche, bien que le rapport des taux (risque relatif) du cancer du foie à Singapour et Dakar soit de 1,0, le rapport entre les deux proportions est de 0,3. Car le taux d'incidence global à Dakar (37,0 pour 100 000) est beaucoup plus faible que celui de Singapour (126,2 pour 100 000), les cancers autres que le foie y étant beaucoup moins fréquents.

Un problème analogue apparaît dans la comparaison des fréquences entre hommes et femmes. Dans pratiquement toutes les études, l'incidence des cancers spécifiquement féminins (sein, utérus, ovaire) est considérablement plus grande que celle des cancers spécifiquement masculins (prostate, testicule, pénis). Cependant, comme le pourcentage total doit toujours être égal à 100 dans les comparaisons de la fréquence relative, la fréquence des cancers qui sont communs aux deux sexes semblera toujours plus basse chez les femmes.

Dans l'exemple montré, le risque de cancer de l'estomac chez les hommes par rapport aux femmes à Cali est de 1,49 si l'on compare les taux d'incidence (figure 2). Cependant, le ratio des fréquences relatives est de 2,03 à cause des cancers liés au sexe qui représentent environ la moitié des cancers chez les femmes, alors qu'il ne compte que pour seulement 12% chez les hommes.

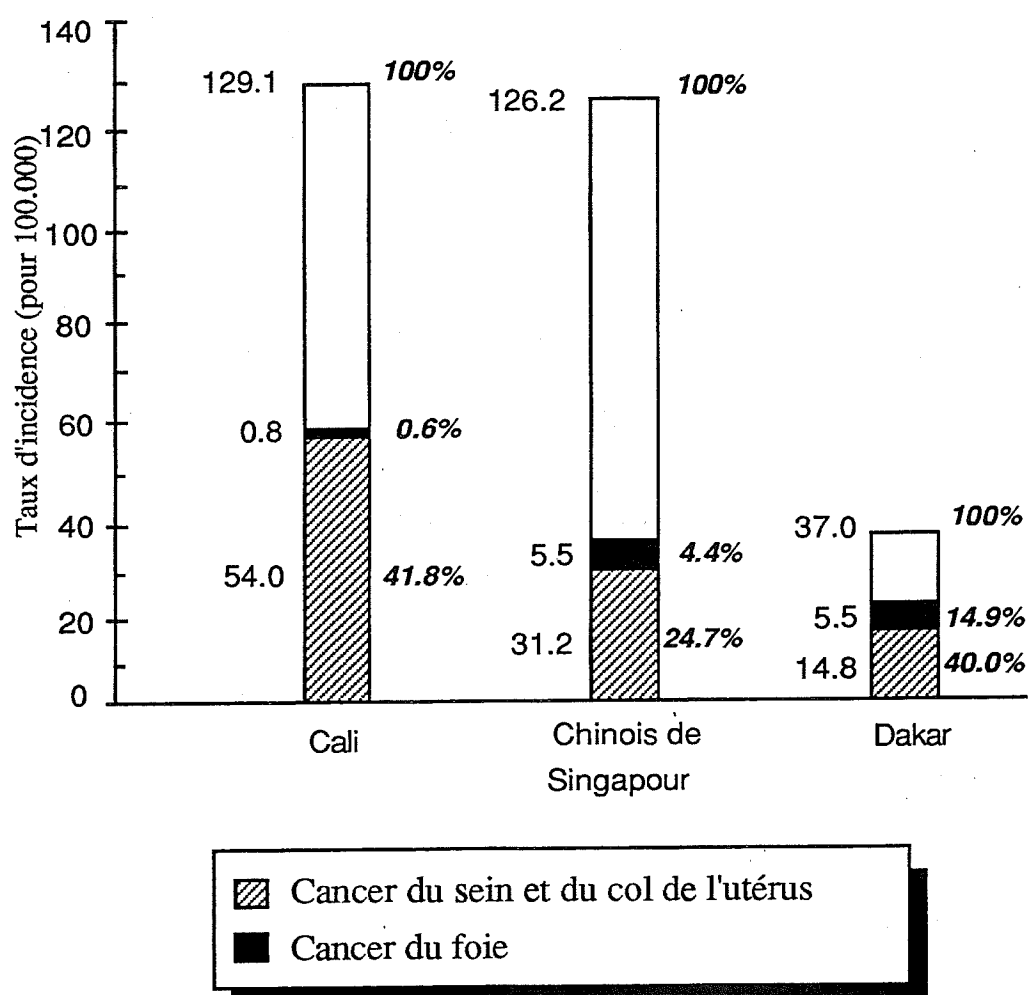


Figure 1. Taux d'incidence (pour 100.000) et fréquences en pourcentage des cancers féminins de 3 registres

Cancer du sein et du col (CIM 174,180) ; cancer du foie (CIM 155). Pour le cancer du foie, le rapport des taux d'incidence "chinois de Singapour/Cali" = $5,5/0,8=6,9$; "chinois de Singapour/Dakar" = $5,5/5,5=1,0$; le rapport des pourcentages "chinois de Singapour/Cali" = $4,4/0,6=7,3$; "chinois de Singapour/Dakar" = $4,4/14,9=0,3$.

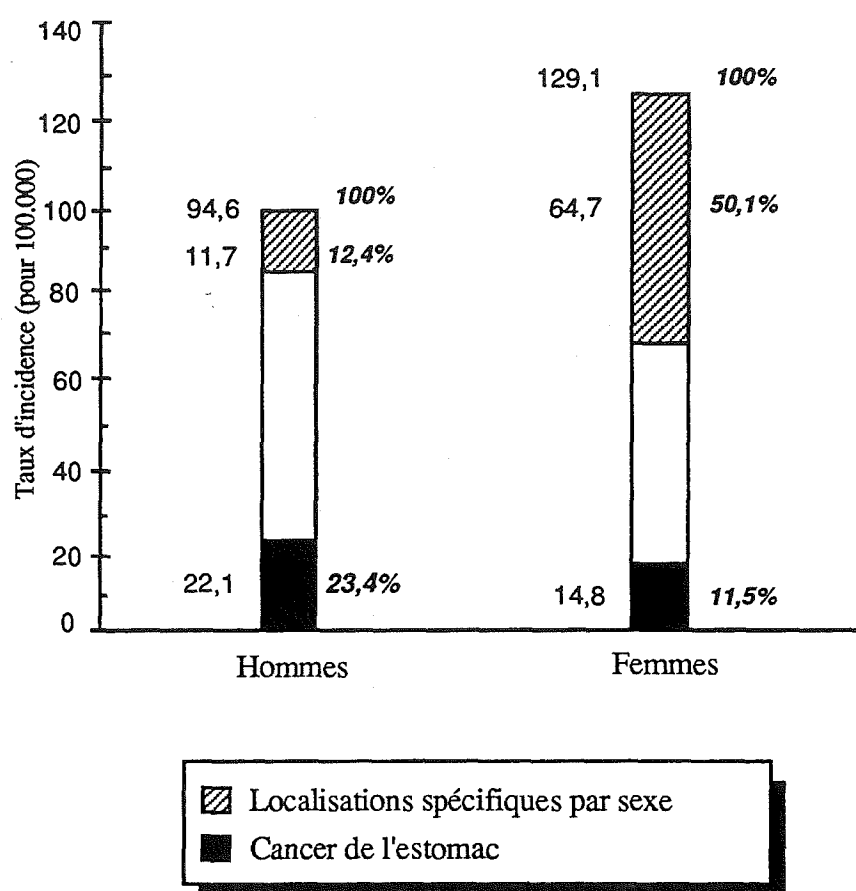


Figure 2. Taux d'incidence (pour 100.000) et fréquences en pourcentage du cancer de l'estomac et de cancers spécifiques du sexe masculin et féminin, Cali, Colombie, 1972-76

Cancers spécifiques du sexe féminin : CIM 174-183 ; Cancers spécifiques de sexe masculin : CIM 185-187. Cancer de l'estomac, CIM.151. Sex ratio du cancer de l'estomac : rapport des taux d'incidence, M/F = 22,1/14,8 = 1,49. Ratio des pourcentages, M/F = 23,4/11,5 = 2,03. Ratio des pourcentages après exclusion des cancers spécifiques du sexe, M/F = 26,7/23,0 = 1,16.

Les comparaisons de fréquences relatives faites séparément pour chaque sexe, donnent bien sûr les mêmes résultats que les comparaisons des taux d'incidence.

Une solution du problème posé par la comparaison des fréquences relatives entre différents centres où l'occurrence de certaines tumeurs fréquentes est hautement variable, consiste à calculer des fréquences résiduelles, c'est-à-dire le pourcentage d'un cancer particulier après avoir enlevé les tumeurs ayant les taux les plus variables d'une série à l'autre. Cette procédure peut être utile pour comparer les séries où les différences dans les taux d'incidences globaux sont essentiellement dues à quelques tumeurs très variables (c'est ce qui a été fait par exemple pour comparer des séries africaines par Cook et Burkitt (1971)).

Cependant, ceci complique quelque peu l'interprétation, et les résultats pourront être moins clairs qu'en utilisant la simple fréquence relative. Ainsi, dans l'exemple de la figure 2, le fait d'enlever les localisations liées au sexe du dénominateur signifie que l'incidence totale devient plus élevée chez les hommes que chez les femmes, si bien que le ratio des fréquences résiduelles pour le cancer de l'estomac (1,16) devient une sous-estimation plutôt qu'une sur-estimation du vrai risque relatif (1,49).

Dans l'exemple déjà présenté dans la figure 1, les cancers du col de l'utérus et du sein représentent 40% des cancers à Dakar mais seulement 24,7% à Singapour. Si ces tumeurs variables sont exclues du dénominateur, les fréquences résiduelles du cancer du foie sont 5,8% (4,4/100 - 24,7) à Singapour et 24,8% (14,9/100 - 40,0) à Dakar. L'estimation du risque relatif obtenue en comparant ces fréquences résiduelles est de 0,23 (5,8/24,8), qui est plus éloignée de la vraie valeur (1,0) que l'estimation obtenue en comparant les pourcentages bruts (0,30).

Standardisation sur l'âge

Comme dans le cas des comparaisons des taux d'incidence, la comparaison des proportions est rendue plus complexe quand il existe des différences dans la structure d'âge des populations à comparer.

La fréquence relative des différents types de cancers varie considérablement avec l'âge ; par exemple, certaines tumeurs, comme la leucémie aigüe sont plus fréquentes dans l'enfance alors que d'autres (cancers respiratoires et digestifs) sont très rares et représentent une grande proportion des cancers chez les personnes âgées. La proportion des différents cancers dans une série est donc fortement influencée par la structure d'âge. Une forme de standardisation par l'âge est donc nécessaire si l'on veut faire des comparaisons.

Deux méthodes ont été utilisées pour la standardisation par l'âge, le rapport standardisé de cancer (ASCAR), qui est l'équivalent en fréquence proportionnelle de la standardisation directe sur l'âge (Tuyns, 1968), et le rapport d'incidence proportionnelle standardisée (RIPS ou RIP), qui est une standardisation indirecte. Des deux, le RIP est nettement supérieur, l'ASCAR étant seulement valable si les bases de données proviennent de sources complètement différentes et quand il n'y a pas de référence évidente pour la comparaison.

Le rapport standardisé de cancer (ASCAR)

L'ASCAR est une standardisation directe qui nécessite de choisir une répartition des cancers par âge qui servira de référence sur laquelle la série

à comparer sera standardisée. Le choix est arbitraire, mais si la référence est identique à la distribution d'âge de tous les cancers dans les séries à comparer, l'ASCAR sera relativement proche de la fréquence relative brute. Les proportions servant de référence pour comparer les fréquences de cancers dans différents pays en voie de développement (Parkin, 1986) sont données dans le tableau 12.

L'ASCAR est calculé de la façon suivante :

$$\text{ASCAR} = \sum_{i=1}^A (r_i/t_i) p_i \quad (9.29)$$

ou r_i = nombre de cancers qui nous intéressent dans le groupe étudié dans la classe d'âge i
 t_i = nombre de cas de cancer toutes localisations confondues dans le groupe étudié et dans la classe d'âge i
 p_i = la proportion de référence pour la classe d'âge i

Tableau 12. Distribution des âges utilisée comme référence dans les pays en voie de développement

Groupe d'âge	%
0-14	5
15-24	5
25-34	5
35-44	10
45-54	20
55-64	25
65-74	20
75+	10
Tous	100

Source : Parkin, 1986.

L'ASCAR est interprété comme étant la fréquence relative d'un cancer qui aurait été observée si les proportions spécifiques par âge avaient été observées dans une population ayant une distribution par âge de l'ensemble des cancers identiques à la distribution de référence. On doit souligner que les problèmes rencontrés lorsque l'on fait des comparaisons entre des bases de données ayant des taux d'incidence globaux différents, persistent et ne sont pas corrigés par la standardisation.

Exemple 14. Calcul du rapport standardisé de cancer (ASCAR)

Le tableau 13 présente les éléments nécessaires au calcul de l'ASCAR du cancer nasopharyngé chez l'homme à Tunis. Avec la formule (9.29), $ASCAR = 10,98$, que nous pouvons comparer avec la fréquence relative calculée sans ajustement sur l'âge qui est

$$\frac{344}{3.073} \times 100 = 11,19\%$$

Tableau 13. Calcul du rapport standardisé de cancer (ASCAR) pour le cancer du nasopharynx chez les hommes de Tunis 1976-80

Groupe d'âge	Nombre de cas		Proportion de cancer du nasopharynx	Proportion référence %	Pourcentage attendu %
	Nasopharynx (r_i)	Tous cancers (t_i)	(r_i/t_i)	(p_i)	(r_i/t_i) p_i
0-14	16	257	0,062	5	0,31
15-24	37	239	0,155	5	0,78
25-34	22	132	0,167	5	0,84
35-44	60	292	0,205	10	2,05
45-54	88	612	0,144	20	2,88
55-64	76	744	0,102	25	2,55
65-74	40	619	0,065	20	1,30
75+	5	178	0,028	10	0,28
	344	3073	0,112	100	10,98

Source : Parkin, 1986

Les problèmes statistiques de comparaison des scores ASCAR n'ont pas été développés et il n'existe pas de formule disponible pour calculer l'écart-type.

Le rapport d'incidence proportionnelle (RIP)

Le rapport d'incidence proportionnelle est la méthode de choix pour comparer des bases de données lorsque l'on dispose d'une répartition standard de proportion par âge pour chaque type de cancer (comme la standardisation indirecte nécessite des taux d'incidence spécifique de référence). Le cas de figure habituel est représenté par un registre qui veut comparer différentes sous-classes de cas à l'intérieur de celui-ci (définies par exemple par le lieu de résidence, le groupe ethnique, la catégorie professionnelle...). Dans ce cas une référence est donnée par les proportions spécifiques par âge de chaque type de cancer pour l'ensemble du registre

(une référence extérieure est préférable car l'ensemble du registre inclura aussi les sous-groupes étudiés). En pratique si aucun sous-groupe ne représente un grand pourcentage (30% ou plus) du total, ceci est relativement peu important.

Dans le rapport d'incidence proportionnelle, le nombre de cas attendus dans le groupe étudié pour un cancer spécifique est calculé, et le RIP est le rapport des cas observés sur les cas attendus (exactement comme le RSI) et il est habituellement exprimé sous forme d'un pourcentage.

Le nombre de cas attendus pour un cancer particulier est obtenu en multipliant le nombre total de cancers dans chaque groupe d'âge faisant partie des données à étudier par les proportions spécifiques par âge et par type de cancer de la référence. On peut représenter symboliquement :

$$\text{RIP} = (R/E) \times 100 \quad (9.30)$$

$$E = \sum_{i=1}^A t_i (r_i^* / t_i^*) \quad (9.31)$$

où R = nombre de cas observés de la localisation qui nous intéresse dans le groupe étudié

E = nombre de cas attendus de la localisation qui nous intéresse dans le groupe étudié

r_i^* = nombre de cas de cancers qui nous intéressent dans le groupe d'âge i dans la population de référence

t_i^* = nombre de cas de cancers tous sites confondus dans le groupe d'âge i dans la population de référence

t_i = nombre de cancers toutes localisations confondues dans le groupe d'âge i dans le groupe étudié.

Breslow et Day (1987) ont donné une formule pour l'écart-type du log RIP :

$$\text{e.t.}(\log \text{RIP}) = \frac{\left[\sum_{i=1}^A r_i (t_i - r_i) / t_i \right]^{1/2}}{R} \quad (9.32)$$

Exemple 15. Calcul du rapport d'incidence proportionnelle

Les informations figurant dans le tableau 14 permettent de calculer le RIP du cancer du foie dans une région de Thaïlande, en utilisant comme référence les distributions spécifiques par âge de ce cancer pour l'ensemble de la Thaïlande.

Tableau 14. Données permettant de calculer le RIP pour le cancer du foie dans une région de Thaïlande

Age	Thaïlande			Region 4		
	Cancer du foie	Tous cancers	Proportion	Cancer du foie	Tous cancers	Cancer du foie attendus
	(r_i^*)	(t_i^*)	(r_i^*/t_i^*)	(v_i)	(t_i)	$t_i(r_i^*/t_i^*)$
0-4	2	210	0,010	0	9	0,090
5-9	1	143	0,007	0	5	0,035
10-14	4	145	0,027	0	4	0,108
15-19	7	230	0,030	1	12	0,360
20-24	23	265	0,087	2	23	2,001
25-29	50	368	0,136	11	37	5,032
30-34	120	492	0,244	22	57	13,908
35-39	169	685	0,247	31	84	20,748
40-44	314	1 077	0,292	52	123	35,916
45-49	383	1 540	0,249	107	213	53,037
50-54	470	2 155	0,218	95	220	47,960
55-59	388	2 093	0,185	66	182	33,670
60-64	323	2 161	0,150	74	174	26,100
65-69	230	1 910	0,120	41	152	18,240
70-74	148	1 631	0,091	27	90	8,190
75-79	69	980	0,070	12	35	2,450
80-84	21	426	0,049	4	15	0,735
85+	5	172	0,029	0	8	0,232
	2 727	16 683		545	1443	268,812

Source : Scrivatanabul et coll., 1988

En utilisant la formule (9.31)

$$E = \sum_{i=1}^A t_i (r_i^*/t_i^*) = 268,812$$

En utilisant la formule (9.30)

$$\text{RIP} = (R/E) \times 100 = 545/268,812 \times 100 = 203\%$$

où r_i = nombre de cas de cancers qui nous intéressent dans le groupe d'âge i dans le groupe étudié.

Une formule simple peut être utilisée pour avoir une bonne approximation de la formule 9.32 tant que la proportion de cas due à la cause qui nous intéresse est relativement petite :

$$\text{e.t.}(\log \text{RIP}) = 1/\sqrt{R} \quad (9.33)$$

A partir des données du tableau 14, en utilisant la formule (9.32), l'écart-type peut être calculé :

$$\text{e.t.}(\log \text{RIP}) = \frac{\sqrt{325,03}}{545} = 0,033$$

et en utilisant l'approximation de la formule (9.33)

$$\text{e.t.}(\log \text{RIP}) = \frac{1}{\sqrt{545}} = 0,043$$

Breslow et Day (1987) n'ont pas recommandé l'utilisation de procédures statistiques spéciales pour le RIP ; les questions de signification statistique et de différence observée pouvant être évaluées avec l'intervalle de confiance.

Pour obtenir l'intervalle de confiance à 95% d'un RIP à 2,03 (exemple 15), et en utilisant l'écart-type $\text{e.t.}(\log \text{RIP})$ calculé par la formule (9.32).

$$\begin{aligned} \text{RIP} &= 2,03 \\ \log \text{RIP} &= 0,708 \\ \text{intervalle de confiance 95\% de } \log \text{RIP} &= 0,708 \pm (1,96 \times 0,033) \\ &= 0,643 ; 0,773 \\ \text{intervalle de confiance 95\% de RIP} &= 1,90 ; 2,17 \end{aligned}$$

Relation entre le RIP et le RSI

Comme le calcul du RIP ne nécessite pas d'information sur la population à risque, un RIP élevé ne signifie pas forcément que le risque de la maladie est élevé, mais simplement qu'il y a une plus grande proportion de cas de cette pathologie que dans la population de référence.

Les relations entre le RIP et le RSI ont été étudiées de façon empirique par plusieurs groupes (Decouflé et coll, 1980 ; Kupper et coll 1978 ; McDowall, 1983 ; Roman et coll, 1984).

En pratique, on peut démontrer que pour chaque groupe étudié,

$$\text{RIP} = \frac{\text{RSI}}{\text{RSI (tous cancers)}}$$

Le ratio RSI/RSI (tous cancers) est appelé le RSI relatif. Ainsi, un RSI relatif plus grand que 100 suggèrera que le taux d'incidence spécifique à la cause étudiée dans la population de l'étude est plus grand que ce qui aurait été attendu sur la base du taux d'incidence global pour tous les cancers. La conséquence de ceci est que le RIP peut être plus grand que 100 tandis que le RSI sera plus petit, ou vice versa.

Le tableau 15 montre un exemple tiré du registre des cancers d'Israël (Steinitz et coll, 1989). Dans cet exemple, les hommes d'origine asiatique ont un taux plus bas d'incidence de cancer toutes localisations confondues que la population de référence (ici «tous les hommes Juifs»), donnant un RSI (tous cancers) de 77%. Ils ont également un RSI plus bas pour le cancer du poumon que «tous les hommes Juifs» (86%). Cependant, comme le cancer du poumon est proportionnellement plus important chez les «hommes asiatiques» que chez «les hommes Juifs» dans leur ensemble, le RIP dépasse 100.

Tableau 15. Relation entre RIP et RSI. Incidence du cancer chez les Juifs d'Israël ; hommes nés en Asie comparés à l'ensemble des hommes Juifs

Cause	cas observés	RSI (%)	RIP (%)	Relatif RSI (%)
Tous cancers	6771	77	100	100
Oesophage	114	105	139	136
Estomac	693	76	100	99
Lèvre	125	110	140	143
Langue	1062	86	112	112