

LA COHORTE CONSTANCES

CALCULS DE PUISSANCE

JUILLET 2010

Épidémiologie des déterminants professionnels et sociaux de la santé
Centre de recherche en Épidémiologie et Santé des Populations
Inserm Unité 1018 - Université de Versailles Saint-Quentin

16 av Paul Vaillant-Couturier, F-94800 Villejuif
Tel (33) 1 77 74 74 00 – fax (33) 1 77 74 74 03

Objectif

Il est prévu d'inclure 200 000 sujets sur cinq ans dans *CONSTANCES*. Comme cela est indiqué dans le protocole scientifique, ce chiffre est quelque peu arbitraire dans la mesure où il n'est pas possible de faire, dans le contexte d'une cohorte « généraliste » visant notamment à constituer une plateforme de recherche pour des questions très diversifiées, de véritables calculs de nombre de sujets nécessaires comme cela est le cas lorsqu'il s'agit de tester une hypothèse précise. Le nombre de 200 000 a donc été le fruit d'un compromis entre une puissance *a priori* globalement satisfaisante et les aspects opérationnels, logistiques et budgétaires du projet.

Il a semblé cependant utile de procéder à des simulations de puissance en calculant des nombres de sujets nécessaires sous certaines hypothèses, afin d'évaluer les possibilités d'études que permettra une cohorte de 200 000 sujets adultes. Ces simulations ont été faites dans un contexte analytique (estimation des associations entre exposition à des facteurs de risque et des pathologies), mais elles permettent également de juger de la précision qui pourra être obtenue dans un contexte descriptif.

Méthode

Afin d'illustrer les aspects liés à l'effectif des sujets de la cohorte en termes de puissance des études qui pourront être réalisées, on a tout d'abord estimé le nombre d'événements de santé qui se produiront à terme de 5, 10 et 15 ans de suivi après l'inclusion (mortalité et incidence) pour certaines pathologies d'intérêt pour lesquelles on dispose de sources de données nationales ; pour l'incidence des cancers, on a procédé à ces estimations par localisations. Les simulations sont faites à partir de 200 000 sujets inclus, dont la répartition est proportionnelle aux classes d'âge population française (recensement 1999).

Les calculs de nombre de sujets nécessaires sont ensuite effectués pour obtenir une puissance de 80 % pour un risque alpha de première espèce égal à 5 % ; ils prennent en compte le fait que les personnes sont recrutées sur 5ans. Les calculs ont été effectués par le logiciel Stata, à l'aide du programme ART¹. Les résultats sont présentés sous forme de nombres de sujets nécessaires (exposés+non exposés) selon :

- la fréquence de l'exposition (E=1 %,2.5 %,5 %,10 %).
- le risque relatif (ou Hazard Ratio : HR=1.5, 2, 3, 4, 5, 10)
- la durée de l'étude après la fin de la période d'inclusion (D=1, 5, 10).
- l'incidence annuelle dans le groupe non exposé. (I=1 pour 100 000 personnes ; 10 pour 100 000 personnes ; et 50 pour 100 000 personnes)

Pour les combinaisons représentées en rouge, la puissance sera insuffisante pour un effectif de 200 000 personnes dans *CONSTANCES* ; en orange la puissance sera peut être suffisante (selon que les analyses seront effectuées séparément chez les hommes et chez les femmes.) ; en vert et bleu, la puissance sera suffisante et de petits nombres de sujets nécessaires permettront de faire des tests dans des sous populations et d'étudier des interactions.

¹ Analysis of Resources for Trials, Royston P, Barthel F, Babiker A (MRC Clinical Trials Unit, London NW1 DA,UK).

Remarques

Certains aspects liés au fait que la cohorte vieillira au cours du temps n'ont pas été pris en compte dans les calculs :

- La fréquence de l'exposition est considérée comme fixée, alors qu'il est vraisemblable que les conditions d'exposition des personnes évolueront au cours du temps.
- De même, l'incidence annuelle est considérée comme fixée, alors qu'elle va très certainement augmenter au cours du suivi pour la majorité des problèmes de santé puisque les sujets seront plus âgés : ainsi, pour la maladie d'Alzheimer, il est clair que l'incidence annuelle dans la cohorte augmentera au cours du suivi : elle sera d'environ 50 pour 100 000 personnes durant les 5 premières années, puis de 100 pour 100 000 personnes pour les années 6 à 10 et de 200 pour 100 000 pour les années 11 à 15. Par contre, il semble que pour certains cancers, l'incidence annuelle dans toute la cohorte augmente peu : par exemple, pour les cancers Lèvre-Bouche-Pharynx, l'incidence annuelle dans la cohorte est estimée à environ 60 pour 100 000 personnes pour les années 1 à 5, 68 pour 100 000 pour les années 6 à 10 et 72 pour 100 000 pour les années 11 à 15. Pour la maladie d'Alzheimer, il ne semblait pas nécessaire de tenir compte de cette augmentation, car l'incidence est suffisamment élevée pour que la puissance soit satisfaisante ; pour la plupart des cancers l'augmentation est faible pour une durée de suivi de 10 ans après la fin de l'inclusion.
- De plus, des décès vont se produire au cours du suivi, et le nombre de personnes à risque va diminuer au cours du suivi ; il s'ensuit que les nombres de sujets nécessaires fournis dans le tableau ci-dessous sont légèrement sous-estimés.

Résultats

1 - Estimation du nombre de cas attendus

Nombres attendus de décès toutes causes

(Taux de décès : CépiDc 1999)

	5 ans			10 ans			15 ans		
	Hommes	Femmes	Total	Hommes	Femmes	Total	Hommes	Femmes	Total
Toutes Causes	4 131	2 133	6 264	9 727	5 502	15 229	16 983	10 736	27 719
Cancer	1 632	852	2 483	3 731	1 998	5 729	6 250	3 445	9 695
App. Circulat.	960	487	1 446	2 414	1 422	3 836	4 467	3 096	7 563
Suicide	157	60	217	321	124	445	486	189	675
Tout Traumatismes	458	186	644	956	430	1 386	1 506	767	2 273

Nombres attendus de pathologies incidentes

(Référence : Incidence et prévalence de différentes maladies - ARME Pharmacovigilance - 2007)

Pathologies	5 ans			10 ans			15 ans		
	Hommes	Femmes	Total	Hommes	Femmes	Total	Hommes	Femmes	Total
Alzheimer ¹	265	240	505	793	1 007	1 800	1 548	2 469	4 018
Cardiopathies ischémiques (35-64 ans) ²	681	138	819	1 418	290	1 708	2 178	452	2 630

¹Source : Cohorte Paquid

²Source : Registres Francim

Nombres attendus de cancers incidents par localisation

(Référence : Évolution incidence et mortalité par cancer en France de 1978 à 2000, Réseau Francim, Août 2003)

Localisation (CIM9)	5 ans			10 ans			15 ans		
	Hommes	Femmes	Total	Hommes	Femmes	Total	Hommes	Femmes	Total
	3162	2220	5381	7036	4855	11892	11444	7823	19267
Lèvre-Bouche-Pharynx	306	47	353	644	103	747	1005	165	1170
Oesophage	88	16	104	191	35	226	304	59	363
Estomac	79	33	112	183	79	263	309	141	450
Colon-Rectum	357	251	608	817	590	1407	1360	1011	2370
Foie	106	16	121	232	36	268	373	60	433
Pancréas	54	34	88	121	78	199	197	133	330
Larynx	89	7	96	189	16	205	298	26	324
Poumon	502	91	593	1093	199	1292	1743	321	2064
Plèvre	13	4	17	30	9	38	48	14	62
Mélanome	66	86	152	139	180	319	216	278	494
Sein	0	900	900	0	1922	1922	0	3031	3031
Col Utérus	0	75	75	0	155	155	0	236	236
Corps Utérus	0	100	100	0	220	220	0	357	357
Ovaires	0	91	91	0	196	196	0	312	312
Prostate	689	0	689	1627	0	1627	2755	0	2755
Vessie	166	22	188	380	55	435	633	98	730
Rein	113	55	168	245	123	367	390	198	589
Syst. Nerveux	56	52	109	118	110	228	181	171	352
Thyroïde	19	66	85	39	134	173	58	201	259
Tissus Lymphoïdes	109	75	183	235	167	403	375	275	650
Hodgkin	15	12	27	29	22	51	42	31	73
Myélome Mult.	51	26	77	112	61	173	181	103	284
Leucémie	66	42	108	145	94	239	234	154	387

2 – Estimation de nombres de sujets nécessaires

Fréquence Exposition	Incidence annuelle chez les non exposés								
	1/100 000			10/100 000			50/100 000		
	Durée de suivi après la fin de l'inclusion			Durée de suivi après la fin de l'inclusion			Durée de suivi après la fin de l'inclusion		
HR	1 an	5 ans	10 ans	1 an	5 ans	10 ans	1 an	5 ans	10 ans
1.5	100 10 ⁶	48 10 ⁶	28 10 ⁶	10 10 ⁶	5 10 ⁶	3 10 ⁶	2 10 ⁶	967 10 ³	581 10 ³
2	28 10 ⁶	13 10 ⁶	8 10 ⁶	3 10 ⁶	10 ⁶	804 10 ³	575 10 ³	270 10 ³	162 10 ³
3	8 10 ⁶	4 10 ⁶	2.3 10 ⁶	845 10 ³	394 10 ³	237 10 ³	169 10 ³	79 10 ³	48 10 ³
4	4 10 ⁶	1.9 10 ⁶	1.2 10 ⁶	427 10 ³	199 10 ³	120 10 ³	85 10 ³	40 10 ³	24 10 ³
5	2.7 10 ⁶	1.2 10 ⁶	746 10 ³	266 10 ³	124 10 ³	75 10 ³	53 10 ³	25 10 ³	15 10 ³
10	752 10 ³	351 10 ³	211 10 ³	75 10 ³	35 10 ³	21 10 ³	15 10 ³	7159	4348
Fréquence Exposition	Incidence annuelle chez les non exposés								
	1/100 000			10/100 000			50/100 000		
	Durée de suivi après la fin de l'inclusion			Durée de suivi après la fin de l'inclusion			Durée de suivi après la fin de l'inclusion		
HR	1 an	5 ans	10 ans	1 an	5 ans	10 ans	1 an	5 ans	10 ans
1.5	42 10 ⁶	19 10 ⁶	12 10 ⁶	4 10 ⁶	2 10 ⁶	10 ⁶	843 10 ³	394 10 ³	237 10 ³
2	12 10 ⁶	5 10 ⁶	3 10 ⁶	1 10 ⁶	546 10 ³	328 10 ³	234 10 ³	109 10 ³	66 10 ³
3	3.5 10 ⁶	1.6 10 ⁶	966 10 ³	345 10 ³	161 10 ³	97 10 ³	69 10 ³	32 10 ³	19 10 ³
4	1.7 10 ⁶	812 10 ³	487 10 ³	174 10 ³	81 10 ³	49 10 ³	35 10 ³	16 10 ³	9866
5	10 ⁶	507 10 ³	304 10 ³	108 10 ³	51 10 ³	30 10 ³	22 10 ³	10 10 ³	6174
10	303 10 ³	141 10 ³	85 10 ³	30 10 ³	14 10 ³	8530	6115	2880	1749
Fréquence Exposition	Incidence annuelle chez les non exposés								
	1/100 000			10/100 000			50/100 000		
	Durée de suivi après la fin de l'inclusion			Durée de suivi après la fin de l'inclusion			Durée de suivi après la fin de l'inclusion		
HR	1 an	5 ans	10 ans	1 an	5 ans	10 ans	1 an	5 ans	10 ans
1.5	21 10 ⁶	10 10 ⁶	6 10 ⁶	2 10 ⁶	10 ⁶	607 10 ³	434 10 ³	203 10 ³	121 10 ³
2	6 10 ⁶	2.8 10 ⁶	1.7 10 ⁶	604 10 ³	280 10 ³	169 10 ³	121 10 ³	57 10 ³	34 10 ³
3	1.8 10 ⁶	833 10 ³	500 10 ³	179 10 ³	83 10 ³	50 10 ³	36 10 ³	17 10 ³	10 10 ³
4	900 10 ³	420 10 ³	252 10 ³	90 10 ³	42 10 ³	25 10 ³	1810 ³	8469	5105
5	562 10 ³	262 10 ³	157 10 ³	56 10 ³	26 10 ³	16 10 ³	10 10 ³	5289	3192
10	155 10 ³	72 10 ³	43 10 ³	15 10 ³	7265	4370	3132	1475	895
Fréquence Exposition	Incidence annuelle chez les non exposés								
	1/100 000			10/100 000			50/100 000		
	Durée de suivi après la fin de l'inclusion			Durée de suivi après la fin de l'inclusion			Durée de suivi après la fin de l'inclusion		
HR	1 an	5 ans	10 ans	1 an	5 ans	10 ans	1 an	5 ans	10 ans
1.5	11 10 ⁶	5 10 ⁶	3.2 10 ⁶	10 ⁶	538 10 ³	323 10 ³	231 10 ³	108 10 ³	65 10 ³
2	3.2 10 ⁶	1.5 10 ⁶	905 10 ³	323 10 ³	151 10 ³	90 10 ³	65 10 ³	30 10 ³	18 10 ³
3	960 10 ³	448 10 ³	269 10 ³	96 10 ³	45 10 ³	27 10 ³	19 10 ³	9005	5421
4	485 10 ³	227 10 ³	136 10 ³	49 10 ³	23 10 ³	14 10 ³	9749	4565	2751
5	303 10 ³	142 10 ³	85 10 ³	30 10 ³	14 10 ³	8515	6092	2855	1722
10	84 10 ³	39 10 ³	24 10 ³	8425	3938	2368	1697	798	484

Plus de 200 000 personnes

Entre 25 000 et 100 000 personnes

Entre 100 000 et 200 000 personnes

Moins de 25 000 personnes

Conclusion

On constate que dans de nombreuses situations communes, lorsque les pathologies et les expositions à des facteurs de risque ont une fréquence suffisamment importante dans la population, il sera possible de procéder, parfois très rapidement après l'inclusion, à des études dotées d'une puissance satisfaisante pour des facteurs de risque ayant un pouvoir pathogène élevé, ainsi qu'à des estimations d'incidence et de prévalence avec une précision correcte.

Par contre, pour des pathologies et des prévalences d'exposition peu fréquentes, et pour des risques relatifs faibles, même un effectif de 200 000 sujets ne sera pas suffisant pour tester des hypothèses spécifiques, particulièrement lorsqu'il s'agira de prendre en compte des interactions avec divers facteurs. De même, les estimations d'incidence et de prévalence ne seront pas très précises pour les pathologies et les expositions les moins fréquentes.

Au total, on peut dire qu'avec un effectif de 200 000 sujets, *CONSTANCES* permettra la réalisation de nombreux travaux dans des conditions favorables, mais que pour d'autres études, il sera nécessaire de mettre des données en commun avec d'autres cohortes. C'est une des raisons des partenariats qui ont été entrepris avec d'autres équipes, françaises ou internationales.

On peut rappeler que ce problème de puissance suffisante est largement compris dans d'autres pays qui développent actuellement des projets de cohortes en population générale aux objectifs voisins de ceux de *CONSTANCES* ; ainsi, en Allemagne, la *Cohorte Helmholtz* doit inclure 200 000 personnes, en Suède le projet *Lifegene* concerne 500 000 sujets², et au Canada se met actuellement en place la cohorte Cartagene qui comprendra 300 000 personnes (il est intéressant de souligner, que comme dans le cas de *CONSTANCES*, ce projet se fait en collaboration avec la Régie de l'assurance maladie du Québec-RAMQ, qui a constitué une unité appelée «le Service Cartagene de la RAMQ» : Service CaG-RAMQ)³.

² <http://lifegene.ki.se/>

³ <http://www.cartagene.qc.ca/>