

# COLLECTE ET TRAITEMENT DES DONNÉES STATISTIQUES DE RECHERCHE ET DÉVELOPPEMENT DES ENTREPRISES EN BELGIQUE

*Roger Kalenga-Mpala,*

*Politique scientifique fédérale – Belgique, Service des indicateurs de R&D*

## Introduction

Les activités de recherche et développement (R&D) des entreprises en Belgique présentent des caractéristiques importantes suivantes.

- Les travaux de R&D en entreprises peuvent être exécutés soit de manière permanente, reconduites d'une année en année, soit de manière occasionnelle. En outre, elles peuvent être effectuées en interne et /ou sous-traitées auprès des autres institutions de recherche telles que les autres entreprises, les universités, les centres de recherche publics, etc.

- Les activités de R&D des entreprises en Belgique sont fortement concentrées auprès d'un nombre très limité d'entre elles, essentiellement des grandes entreprises de quelques secteurs de technologie de pointe. Malgré qu'on puisse aussi identifier des petites entreprises très dynamiques dans le domaine de la recherche. En effet, plus deux tiers des dépenses intra-muros de R&D de l'ensemble des entreprises sont réalisés par les 100 plus importantes entreprises en matière de R&D.

- Enfin, les évolutions récentes fluctuantes des dépenses afférentes aux travaux de R&D et du personnel affecté à ces travaux dans les entreprises font qu'il est de plus en plus difficile de trouver des variables auxiliaires économiques fort corrélées avec les variables de R&D.

Ces caractéristiques font que la mesure des données des dépenses et du personnel de R&D des entreprises en Belgique via la méthode classique de sondage aléatoire, même stratifiée à plusieurs dimensions notamment liées aux secteurs d'activités et aux catégories de taille de l'entreprise, va nécessiter un échantillon très large si l'on veut tenir compte de la non homogénéité des travaux de R&D dans les entreprises. Ce peut se traduire par des conséquences importantes sur les coûts exorbitants de réalisation des enquêtes et le faible taux de participation. Par contre, lorsque l'échantillon est particulièrement réduit, les estimations risquent d'être moins fiables, en raison de l'hétérogénéité des activités de R&D dans les entreprises et aussi des facteurs d'extrapolation utilisés. En outre, la méthode de l'échantillonnage ne tient pas compte des travaux antérieurs de R&D des firmes. Par ailleurs, la réalisation d'un recensement exhaustif de toutes les entreprises qui exécutent la R&D peut s'avérer très difficile à mettre en oeuvre, car il est extrêmement difficile de tenir des registres entièrement et constamment actualisés des entreprises exécutant la R&D, en particulier la R&D occasionnelle, à partir des diverses sources d'information.

Pour mesurer correctement et régulièrement les données de R&D des entreprises en Belgique tant au niveau de l'ensemble de la population des entreprises qu'au niveau désagrégé, c'est à dire des secteurs d'activités et catégories de taille des entreprises, une des méthodes consiste à combiner systématiquement lors des enquêtes spécialisées de R&D un inventaire des entreprises dont il est connu ou supposé qu'elles exécutent en interne la R&D permanente ou occasionnelle et/ou sous-traitent la R&D à un sondage afin de recueillir des informations sur la R&D exécutée par des entreprises qui ne font partie des registres répertoriant les entreprises qui exécutent des travaux de R&D. La constitution de l'échantillon doit naturellement reposer sur un registre des activités économiques de bonne qualité. La méthode qui combine à la fois l'inventaire et le sondage présente l'avantage d'aboutir à une estimation de la R&D de la population cible qui comporte beaucoup moins d'incertitudes que la méthode de l'échantillonnage simple. En outre, les données estimées sont bien comparables dans le temps car les travaux antérieurs de R&D d'une entreprise de l'inventaire sont pris en compte. Enfin, cette méthode convient mieux pour les petits pays, comme la Belgique, contrairement aux grands pays dans lesquels les coûts peuvent s'avérer très élevés pour la mettre en pratique.

Ce papier décrit brièvement quelques indications méthodologiques concernant la réalisation d'enquêtes sur la R&D des entreprises en Belgique. Ces indications concernent notamment le cadre et la portée des enquêtes sur les activités de R&D des entreprises en Belgique, l'identification de la population cible pour les enquêtes R&D, l'organisation et la conduite de ces enquêtes ainsi que le traitement statistique des données de R&D des entreprises y compris le traitement de la non réponse.

## 1. Cadre et portée des enquêtes sur la R&D des entreprises

Le Manuel de Frascati de l'OCDE<sup>1</sup> définit les activités de R&D comme celles qui englobent les travaux de création entrepris de façon systématique en vue d'accroître la somme des connaissances, y compris la connaissance de l'homme, de la culture et de la société, ainsi que l'utilisation de cette somme de connaissances pour de nouvelles applications. Même si les entreprises ne sont pas les seules à effectuer la R&D, elles demeurent le plus étroitement liées à la création de nouveaux produits et procédés de production ainsi qu'à l'amélioration significative des techniques existantes. La R&D est alors un moteur essentiel et indispensable de l'innovation industrielle et de la croissance économique, et donc du bien-être. Le terme R&D recouvre trois activités: la recherche fondamentale, la recherche appliquée et le développement expérimental.

L'objectif principal des enquêtes spécialisées sur la R&D des entreprises en Belgique est de permettre tant au niveau national que régional de faire l'inventaire du potentiel scientifique et de mesurer ainsi toutes les ressources financières et humaines consacrées aux activités de R&D par l'ensemble des entreprises et également au niveau plus détaillé des secteurs d'activité et de catégories de taille des entreprises. En Belgique, les enquêtes de R&D des entreprises et celles des institutions non marchandes, autres exécutants des travaux de R&D, ont été instituées par un arrêté royal et réglées par un accord de coopération<sup>2</sup> entre l'autorité fédérale et les pouvoirs publics fédérés .

Par ailleurs, les statistiques de R&D sont utiles et demandées de plus en plus par beaucoup d'utilisateurs du monde politique et scientifique. L'intensité de 3% des dépenses de R&D par rapport au produit intérieur brut constitue un des objectifs importants dans le contexte du processus de Lisbonne et son évolution est bien suivie dans le cadre de la stratégie européenne de la croissance et de l'emploi.

A des fins statistiques, les enquêtes sur la R&D des entreprises mesurent sur base annuelle deux principaux intrants: les dépenses engagées au titre de la R&D et le personnel employé à ces travaux. Les dépenses de R&D sont appréhendées en distinguant les dépenses de R&D intra-muros et les dépenses de R&D extra-muros. Les premières comprennent toutes les dépenses afférentes aux travaux de R&D exécutés dans l'entreprise, y compris les coûts liés au personnel de R&D et les secondes couvrent les dépenses afférentes à la R&D effectuée en dehors de l'entreprise. Quant au personnel affecté à la R&D, il est exprimé en équivalents temps plein (ETP) ou en nombre d'années-personnes affectées à la R&D et en personnes physiques.

Par ailleurs, pour une meilleure comparabilité internationale des données statistiques de R&D, le Manuel de Frascati de l'OCDE donne à ce sujet quelques directives et recommandations méthodologiques fondées sur les pratiques exemplaires et les expériences des enquêtes de R&D dans les pays membres.

Les enquêtes sur la R&D des entreprises en Belgique sont conjointement menées par la Politique scientifique fédérale<sup>3</sup> et les régions, entités fédérées responsables entre autre de la politique de recherche et de l'innovation des entreprises, tous les deux ans et portent sur les données annuelles de la période précédant l'année de la tenue de l'enquête. La dernière enquête a été tenue en 2008 et se réfère aux données annuelles de 2006 et de 2007 ainsi qu'aux prévisions de 2008. Les unités statistiques dans les enquêtes de R&D des entreprises comprennent, d'une part, toutes les firmes, organismes et institutions dont l'activité première est la production marchande de biens ou de services

<sup>1</sup> OCDE, 2002, "Manuel de Frascati: «Méthode type proposée pour les enquêtes sur la recherche et le développement expérimental (R&D)»", OCDE, Paris.

<sup>2</sup> Adaptation de l'Accord de coopération du 12 juillet 1994 portant exécution de l'article 6bis, §2, point 6, de la loi spéciale du 8 août 1980 de réformes institutionnelles, inséré par la loi du 8 août 1988 et modifié par la loi du 16 juillet 1993, du 16 avril 2006.

<sup>3</sup> Administration publique fédérale chargée de la préparation, de l'exécution et de l'évaluation de la politique scientifique et de ses prolongements au niveau fédéral de la Belgique.

(autres que d'enseignement supérieur) en vue de leur vente au public à un prix qui correspond à la réalité économique et, d'autre part, les institutions privées sans but lucratif principalement au service de ces entreprises. L'essentiel des unités statistiques est donc constitué par des entreprises privées (sociétés et quasi sociétés) distribuant ou non des bénéfiques. Mais parmi les unités statistiques, on trouve également des entreprises publiques, et des institutions privées sans but lucratif, producteurs marchands de biens et de services autres que d'enseignement supérieur.

## 2. Identification de la population cible

Dans les enquêtes de R&D des entreprises en Belgique, pour déterminer la population cible, on réalise d'une part un inventaire des entreprises dont il est connu ou supposé qu'elles exécutent de la R&D permanente ou occasionnelle en leur sein (R&D intra-muros) et/ou sous-traitent la R&D auprès des autres exécutants<sup>4</sup> de R&D (R&D extra-muros) et d'autre part un sondage sur le reste de la population des entreprises afin de mesurer la R&D non prise en compte par les entreprises répertoriées.

### 2.1. Inventaire des entreprises de R&D

Toutes les entreprises exécutant de la R&D, qu'il s'agisse d'une activité permanente ou occasionnelle et même celles qui font sous-traiter leurs activités de R&D auprès des autres exécutants sont systématiquement répertoriées. L'inventaire ne permet pas seulement de recenser les entreprises actives en R&D mais aussi de distinguer si possible parmi elles, celles qui font la R&D permanente ou occasionnelle ou encore celles qui ne réalisent que la R&D extra-muros. Pour cette dernière raison et celle consistant à disposer d'un registre le plus exhaustif et le plus actualisé possible des entreprises qui exécutent la R&D, on s'appuie sur une série des listes d'entreprises et sources d'informations suivantes :

- les listes d'entreprises faisant état d'activités de R&D lors des enquêtes précédentes sur la R&D, des enquêtes européennes sur l'innovation ou d'autres récentes enquêtes d'entreprises qui collecte les informations sur la R&D, telles que les enquêtes sur la biotechnologie, les enquêtes structurelles de l'institution belge des statistiques, etc;
- les listes d'entreprises bénéficiant de subventions des pouvoirs publics régionaux et de contrats publics de R&D;
- les listes d'entreprises demandant des réductions d'impôts au titre de leurs activités de R&D et/ou bénéficiant des mesures fiscales en faveur de la R&D et de l'innovation;
- les répertoires des laboratoires de R&D et des entreprises des sites ou parcs scientifiques, des listes des membres des certaines fédérations patronales de haute technologie;
- les listes de petites entreprises travaillant en étroite collaboration avec les universités (spin offs) ; listes des entreprises ayant obtenu des brevets auprès de l'office national;
- les autres sources d'information y compris les médias, etc.

En outre, la constitution d'un registre exhaustif des entreprises exécutant la R&D permet de réduire quelque peu l'importance de mettre sur pied d'un échantillon afin de mesurer la R&D résiduelle du reste de la population des entreprises. En Belgique, après plusieurs vagues successives d'enquêtes sur la R&D des entreprises, la part dans le total des dépenses de R&D intra-muros du reste de la population approché via l'échantillon se situe en dessous de 9% depuis déjà quelques années. En d'autres termes, plus de 90% du total des dépenses de R&D intra-muros de l'ensemble des entreprises sont obtenus à partir des données de R&D des entreprises répertoriées. En ce qui concerne les chiffres du personnel total de R&D en équivalents temps plein (ETP), la part des entreprises répertoriées est légèrement en dessous de 90% (tableau 1).

**Tableau 1 : Répartition de la R&D en fonction de la classification des entreprises lors des enquêtes antérieures récentes**

	2000	2001	2002	2003	2004	2005	2006(p)
<b>Dépenses de R&amp;D intra-muros (%)</b>							
Entreprises répertoriées pour la R&D	91,0	91,0	91,4	91,3	91,3	91,2	91,2

<sup>4</sup> Les autres exécutants de R&D de R&D visés dans ce cadre sont les autres entreprises, les universités et autres institutions de l'enseignement supérieur, les centres de recherche publique, les institutions sans but lucratif qu'ils soient situés sur le territoire national ou à l'étranger.

Population entreprises non répertoriées	9,0	9,0	8,6	8,7	8,7	8,8	8,8
Total	100,0	100,0	100,0	100,0	100,0	100,0	100,0
<b>Personnel total de R&amp;D en ETP (%)</b>							
Entreprises répertoriées pour la R&D	86,1	85,5	86,2	86,4	87,6	87,8	87,9
Population entreprises non répertoriées	13,9	14,5	13,8	13,6	12,4	12,2	12,1
Total	100,0	100,0	100,0	100,0	100,0	100,0	100,0

Source : CFS/STAT, Politique scientifique fédérale, 2007.

Pour revenir aux entreprises répertoriées, lors de la dernière enquête R&D des entreprises organisées en 2008, l'inventaire, à travers les différentes sources exploitées, a permis de répertorier 3890 entreprises en Belgique dont il est connu ou supposé qu'elles exécutent de la R&D permanente ou occasionnelle en leur sein et/ou sous-traitent la R&D auprès des autres exécutants (tableau 2). Près de la moitié de ces entreprises sont opérationnelles dans les industries manufacturières malgré que cette branche d'activité économique représente moins de 30% dans le total de la valeur ajoutée de la Belgique.

**Tableau 2 : Inventaire pour l'enquête 2008 : répartition par nature de R&D, secteur d'activités et catégorie de taille d'entreprises**

Entreprises connues ou supposées pour		R&D permanente				R&D occasionnelle ou R&D exclusivement sous-traitée			
Secteur d'activités	Nace 2.0	< 50	50-249	≥ 250	Total	< 50	50-249	≥ 250	Total
Agriculture, chasse et pêche	01-03	12	4		16	11	1		12
Industries extractives	05-09	1	1		2	5	6	1	12
Industrie manufacturière	10-32	447	411	229	1087	657	267	48	972
<i>Industries de haute technologie</i>		241	173	95	509	236	67	16	319
<i>Industries de moyenne technologie</i>		98	119	54	271	227	93	15	335
<i>Industries de faible technologie</i>		108	119	80	307	194	107	17	318
Équipements publics, construction	33-43	29	17	10	56	103	30	13	146
Commerce, hôtels et restaurants	45-47, 55-56	46	20	4	70	144	27	11	182
Transport, entreposage et poste	49-53	8	9	9	26	32	22	10	64
Editions, productions images, son	58-60	8	1	4	13	28	7	3	38
Télécommunications	61	1		5	6	10	1	2	13
Activités informatiques	62-63	190	42	10	242	159	15	1	175
Activités de services financiers	64-66	7	5	15	27	44	7	10	61
Architecture, ingénierie, contrôle	71	126	17	5	148	143	11	4	158
Services de R&D	72	42	15	3	60	35	1		36
Autres services aux entreprises	73-74, 77-82	42	9	11	62	144	28	5	177
Autres services	75, 84-99	4	1	1	6	17	5	1	23
<b>Total général</b>		<b>963</b>	<b>552</b>	<b>306</b>	<b>1821</b>	<b>1532</b>	<b>428</b>	<b>109</b>	<b>2069</b>

Source : CFS/STAT, Politique scientifique fédérale, 2009.

## 2.2. Population des entreprises non répertoriées et sondage

Il est très difficile de parvenir par l'inventaire à réaliser un recensement exhaustif de toutes les entreprises exécutant la R&D et à tenir entièrement actualisé le registre de ces entreprises notamment celles qui exécutent des travaux de R&D de manière occasionnelle. C'est pourquoi, l'inventaire est complété par un sondage sur le reste de la population des entreprises non répertoriées. Le sondage réalisé sur le reste de la population permet non seulement à contribuer à la consolidation de l'estimation du total des dépenses de R&D et du personnel de R&D de l'ensemble des entreprises, mais il permet aussi de corriger l'éventuelle sous-estimation de la R&D des secteurs d'activités et des catégories de taille des entreprises moins intensives en R&D. Ces secteurs et catégories sont susceptibles d'être moins représentées dans le registre des entreprises répertoriées pour leurs activités de R&D. Ce sont souvent les cas des secteurs de services et de catégories des petites et moyennes entreprises. Quand on sait qu'il est difficile d'identifier la R&D de manière catégorique dans les secteurs des services que dans ceux de l'industrie manufacturière parce qu'elle ne revêt pas nécessairement un caractère «spécialisé». Un autre élément important à prendre en considération dans la constitution d'un échantillon de la population des autres entreprises qui ne font pas partie des entreprises dont il est connu ou supposé qu'elles exécutent de la R&D est le coût. La taille de l'échantillon est donc très limitée. Enfin, la dimension territoriale ou régionale entre aussi en compte pour réaliser le sondage. Ce qui signifie que l'échantillon est constitué au niveau des chacune des

trois régions de la Belgique. L'échantillon global de la Belgique n'est autre que le résultat de l'ensemble des trois échantillons régionaux.

La population cible pour l'échantillon est composée de l'ensemble des autres entreprises qui ne font pas partie des entreprises, dont il est connu ou supposé qu'elles exécutent de la R&D, qui sont déjà en principe répertoriées par l'inventaire. Le sondage est guidé par le souci de prendre en compte des sous-populations d'entreprises relativement à faible dispersion et que l'on veut voir être représentées au sein de l'échantillon. L'échantillon constitué au niveau des régions est donc stratifié en fonction des secteurs d'activités et catégories de taille des entreprises. En outre, compte tenu de la taille de l'échantillon souhaitée très limitée, la stratification est certainement une des meilleures manières d'introduire une information auxiliaire, qui tient compte à la fois des secteurs d'activités et de catégories de taille des entreprises, dans le sondage afin d'améliorer la précision des estimateurs.

En ce qui concerne les catégories de taille, les entreprises de moins de 10 employés ne font pas en principe partie de la population cible à l'exception des quelques secteurs des services de haute technologie<sup>5</sup>. Selon le secteur d'activités, la population cible comprend généralement toutes les entreprises des branches d'activités économiques qui sont présentées dans le tableau 3. Dans la branche des industries manufacturières, un découpage des secteurs en fonction de leur niveau de technologie est aussi envisagé. Les branches d'activités des services réputées pour être très peu ou non intensives en R&D ne sont pas souvent prises en compte.

**Tableau 3 : Branches d'activités des entreprises faisant partie de la population cible**

Description des branches d'activités	Code Nace 2.0
Activités de soutien à l'agriculture et traitement primaire des récoltes	016
Industries extractives	05-09
Industries manufacturières	10-32
Réparation et installation de machines et d'équipements	33
Équipements publics, construction	35-39, 41
Commerce de gros	46
Transports, entreposage et stockage, activités de poste et de courrier	49-53
Editions, productions cinématographiques, vidéo, son et audiovisuels	58-60
Télécommunications	61
Activités informatiques	62-63
Activités de services financiers	64-66
Activités d'architecture et d'ingénierie, de contrôle et analyses techniques	71
Services de R&D	72
Autres services aux entreprises	73-74, 77-82

Source : OCDE, Manuel de Frascati (2002) et CFS/STAT, Politique scientifique fédérale (2009).

Pour constituer l'échantillon, on effectue, indépendamment les uns des autres, des tirages aléatoires sans remise dans chacune des sous-populations ou strates constituées des branches d'activités et de catégories de taille des entreprises non répertoriées. Le plan de sondage dans les différentes strates est établi non seulement en fonction de la limitation de la taille de l'échantillon global mais aussi des résultats de R&D dans les strates correspondantes des enquêtes antérieures les plus récentes. Idéalement, l'objectif visé d'utiliser une allocation optimale afin d'obtenir des bons niveaux de précision pour toutes les données de R&D estimées.

Un recensement est souvent prévu dans les sous-populations concernant les catégories des entreprises de 250 employés et plus, étant donné le nombre très limité d'entreprises dans ces strates et aussi la forte probabilité de rencontrer des entreprises actives en R&D dans ces catégories. Il en est de même dans certaines strates des entreprises de moins de 250 employés des secteurs de haute technologie. En outre, dans les strates des entreprises de moins de 250 employés dans lesquelles le sondage est prévu, les entreprises qui ont déclaré ne pas avoir des activités de R&D aux enquêtes antérieures les plus récentes ne sont pas sélectionnées dans l'échantillon.

**Tableau 4 : Population cible et échantillon des entreprises non répertoriées pour l'enquête 2008 : répartition par secteur d'activités et catégorie de taille d'entreprises**

	Population cible	Echantillon

<sup>5</sup> Services de la programmation informatique, activités d'architecture et d'ingénierie, activités de contrôle et analyses techniques et services de R&D.

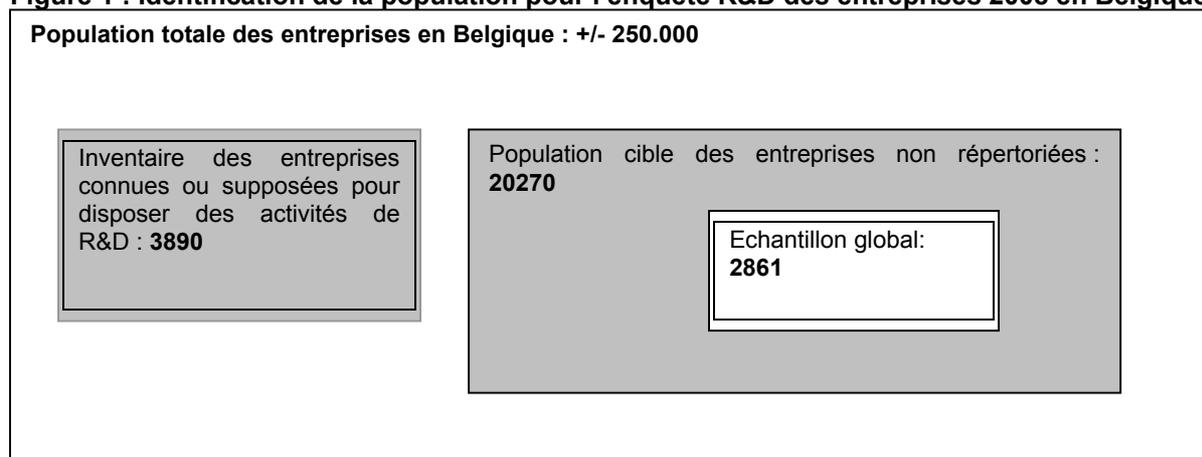
Secteur d'activités	Nace 2.0	< 50	50-249	≥ 250	Total	< 50	50-249	≥ 250	Total
Serv. à l'agriculture et ind. extractives	016, 05-09	83	4	0	87	26	4	0	30
Industrie manufacturière	10-32	3771	667	80	4518	611	177	80	868
<i>Industries de haute technologie</i>		493	121	39	653	115	45	39	199
<i>Industries de moyenne technologie</i>		1436	242	23	1701	214	75	23	312
<i>Industries de faible technologie</i>		1842	304	18	2164	282	57	18	357
Équipements publics, construction	35-43	3446	465	43	3954	271	58	43	372
Commerce, réparation et installation	45-46, 33	3858	437	36	4331	352	90	36	478
Transport, entreposage et poste	49-53	1990	346	37	2373	199	66	37	302
Editions, productions images, son	58-60	156	37	5	198	24	23	5	52
Télécommunications	61	34	22	4	60	13	14	4	31
Activités informatiques	62-63	700	67	15	782	101	32	15	148
Activités des services financiers	64-66	453	88	25	566	46	25	25	96
Architecture, ingénierie, contrôle	71	1076	30	3	1109	118	21	3	142
Services de R&D	72	50	0	0	50	50	0	0	50
Autres services aux entreprises	73-74, 77-82	1741	362	86	2189	144	45	86	275
Santé et activités vétérinaires	75, 86	47	6	0	53	11	6	0	17
<b>Total général</b>		<b>17405</b>	<b>2531</b>	<b>334</b>	<b>20270</b>	<b>1966</b>	<b>561</b>	<b>334</b>	<b>2861</b>

Source : CFS/STAT, Politique scientifique fédérale, 2009.

Le tableau 4 donne un aperçu de la population cible des entreprises non répertoriées pour leurs activités de R&D et de l'échantillon constitué au niveau de la Belgique pour la dernière enquête R&D des entreprises (2008). Les secteurs d'activités et les catégories de taille des entreprises non répertoriées par l'inventaire et susceptibles de faire partie de la population cible sont représentés dans l'échantillon global de la Belgique. Rappelons aussi que cet échantillon global de la Belgique est constitué des résultats des échantillons représentant les trois régions du pays.

Concernant la population d'enquête de R&D 2008, comme on peut s'apercevoir avec les tableaux 2 et 4, la dernière enquête de R&D des entreprises en Belgique a porté donc sur un total de 6571 entreprises parmi lesquelles 3890 entreprises répertoriées dont il est connu ou supposé qu'elles exécutent de la R&D. Le reste d'entreprises, c'est à dire 2861, sont donc sélectionnées par un sondage réalisé dans toutes les autres entreprises des branches d'activités et de catégories de taille d'au moins 10 employés dans la plupart des cas, les entreprises qui peuvent avoir des activités de R&D mais qui ne font pas partie des entreprises déjà répertoriées. La figure 1 donne une idée avec les informations de la dernière enquête concernant la détermination de la population d'enquête de R&D des entreprises en Belgique à partir du registre officiel belge des employeurs de l'ONSS<sup>6</sup> adapté pour les besoins des enquêtes de R&D des entreprises.

**Figure 1 : Identification de la population pour l'enquête R&D des entreprises 2008 en Belgique**



Source : ONSS, CFS/STAT, Politique scientifique fédérale, 2009.

### 3. Organisation et conduite des enquêtes

<sup>6</sup> Office national de sécurité sociale.

Les enquêtes sur la R&D des entreprises en Belgique sont organisées tous les deux ans et portent sur les données annuelles de la période précédant l'année de la tenue de l'enquête. La dernière enquête a été tenue en 2008 et se réfère aux données annuelles effectives de 2006 et de 2007 ainsi qu'aux prévisions de 2008. Le questionnaire utilisé pour l'enquête comprend un nombre minimal de questions fondamentales sur l'activité de R&D, afin de permettre la production de statistiques harmonisées et comparables à transmettre aux organisations internationales telles que l'OCDE et Eurostat. En outre, par souci de disposer des informations indispensables pour la Politique scientifique en Belgique tant au niveau fédéral que régional, un nombre très limité d'autres questions simplifiées peuvent aussi être ajoutées au questionnaire. Dans cette section, nous abordons successivement du questionnaire de l'enquête, de la conduite de l'enquête et de la participation des entreprises à celle-ci.

### **3.1. Questionnaire**

L'élaboration du questionnaire de l'enquête de R&D des entreprises poursuit entre autres des objectifs suivants: pertinence, utilité et clarté des questions pour les répondants. Le questionnaire est aussi présenté de façon structurée et logique et il est accompagné par des directives, des définitions<sup>7</sup>, de notes explicatives et d'instructions claires. Compte tenu de la charge de travail qu'il engendre pour les répondants, le questionnaire doit être, en principe, aussi simple, aussi court que possible.

Deux types de questionnaires sont généralement proposés aux entreprises sélectionnées pour l'enquête. Le questionnaire destiné aux entreprises de l'échantillon est très fortement simplifié par rapport au questionnaire principal conçu pour les entreprises dont il est connu ou supposé qu'elles exécutent de la R&D permanente ou occasionnelle en leur sein et/ou sous-traitent la R&D auprès des autres exécutants.

#### **3.1.1. Questionnaire principal**

Le questionnaire principal destiné aux entreprises répertoriées pour leurs activités de R&D se rapporte aux données signalétiques, administratives et économiques (chiffre d'affaires et effectifs totaux) de l'entreprise dans un premier module. Dans le deuxième et principal module, il est question des caractéristiques des efforts de R&D, des chiffres des dépenses et du personnel de R&D avec leurs sous-agrégats ou répartitions fonctionnelles. Enfin dans le reste du questionnaire, une série d'autres questions essentiellement fermées portent sur des informations concernant la distribution sectorielle et territoriale de la R&D de la firme en Belgique, sur des renseignements au sujet de la relation entre la R&D et l'innovation ainsi que sur les autres informations concernant notamment la coopération technologique, la R&D des entreprises faisant partie du groupe sur la technologie et l'accroissement des connaissances.

#### **3.1.2. Questionnaire simplifié**

Concernant le questionnaire simplifié destiné aux entreprises de l'échantillon, seules les données essentielles signalétiques, administratives et économiques ainsi que les données agrégées sur les efforts de R&D peuvent être rapportées. On peut recourir au formulaire simplifié pour faire une dernière et ultime relance de l'enquête à toutes les entreprises non répondantes y compris celles répertoriées pour leur R&D.

### **3.2. Modes de collecte des données**

Après avoir déterminé la population des entreprises pour laquelle on a besoin d'informations concernant leurs activités de R&D et identifié ses unités pour l'enquête, il faut définir une technique de collecte des données en tenant compte des coûts, des délais notamment celui de publication et de livraison des statistiques.

L'information visée par les enquêtes de R&D des entreprises est collectée par l'auto-dénombrement, c'est à dire l'information requise est fournie par l'entreprise qui complète elle-même le questionnaire. La R&D est une activité non seulement complexe, mais aussi inextricablement liée à un certain nombre d'autres activités. La plupart des données sur les activités de R&D des entreprises en fonction des critères du Manuel de Frascati ne sont pas directement disponibles en entreprise. De plus, les travaux de R&D d'une entreprise peuvent ne pas apparaître clairement dans son organisation interne et dans ses comptes financiers. Les répondants sont le plus souvent amenés à procéder à des

---

<sup>7</sup> La plupart des définitions proposées dans le questionnaire sont tirées du Manuel de Frascati de l'OCDE (2002).

estimations. En outre, plusieurs responsables dans l'entreprise peuvent être concernés pour répondre au même questionnaire des enquêtes de R&D. C'est pourquoi un délai relativement suffisant est accordé aux entreprises. La période de collecte des données dure généralement six mois<sup>8</sup>, au cours de laquelle des relances d'enquêtes par voie postale et par téléphone sont régulièrement organisées.

En plus de la version classique papier du questionnaire, les entreprises peuvent également répondre par voie électronique via les sites Internet des institutions fédérale et régionales organisant l'enquête. Le questionnaire électronique présente l'avantage de pouvoir être rempli en autant de sessions qu'on souhaite. Ce procédé de collecte d'enquête semble être aujourd'hui de plus en plus le choix d'une partie non négligeable des entreprises. Environ plus de 35% des réponses aux enquêtes sont collectées par voie électronique.

### **3.3. Simulation de la coopération avec les répondants et leur participation aux enquêtes**

Pour améliorer le taux de réponse, généralement bas pour les enquêtes non obligatoires pratiquant le mode de collecte par l'auto-dénombrément, une série de mesures ci-dessous sont mises en oeuvre. Un bon contact est indispensablement établi avec les répondants, en particulier les entreprises les plus importantes en matière de R&D. Lorsque le répondant est identifié, il faut essayer de trouver la personne la mieux à même de remplir le questionnaire et de susciter son intérêt pour l'enquête. Dans les enquêtes de R&D des entreprises, celle-ci fait souvent partie, soit du service de comptabilité ou financier ou du service des ressources humaines, soit du département chargé des activités de R&D. La coopération entre les différents responsables d'entreprises répondants à l'enquête est privilégiée. Le questionnaire est généralement envoyé à la personne qui a répondu lors de la dernière enquête en ce qui concerne les entreprises répertoriées. Sinon et pour les autres entreprises, celui-ci est adressé au directeur général ou l'administrateur délégué de l'entreprise. Les contacts personnels avec les répondants qui ont besoin de conseils ou qui ont fourni des réponses non satisfaisantes sont souvent organisés.

Le nombre minimal de relances ou rappels d'enquête, avec à chaque fois un questionnaire papier envoyé par poste, est généralement de deux. Mais toutefois, les non répondants appartenant aux entreprises répertoriées pour leurs travaux de R&D sont en plus régulièrement rappelés par téléphone. Des efforts supplémentaires sont consentis pour s'assurer de la forte participation aux enquêtes des plus importantes entreprises en matière de R&D et des programmes de suivi pour chaque enquête peuvent être mis en oeuvre pour ces entreprises. La dernière relance par courrier à tous les non répondants quelle que soit la classification s'effectue exclusivement avec le questionnaire simplifié initialement destiné aux entreprises non répertoriées sélectionnées pour l'enquête.

Pour s'assurer également de la coopération des personnes qui répondent à l'enquête, nous avons le devoir de les aider à prendre conscience des utilisations potentielles des données et d'être attentif à leurs besoins éventuels en matière de statistiques de R&D même si l'on sait que la plupart sont rarement des utilisateurs de statistiques de R&D. L'engagement est également pris auprès des répondants afin de respecter la confidentialité des données collectées et s'assurer par la suite que tous les utilisateurs sont conscients des préoccupations des répondants. Seules des données agrégées sous forme d'indicateurs, de tableaux et de graphes sont publiées et communiquées aux instances régionales, nationales et internationales. Il est donc important de montrer à quoi servent les données pour encourager la coopération. Outre, les besoins des politiques scientifiques tant au niveau fédéral que régional conduisant notamment à développer les mesures spécifiques à encourager et soutenir la recherche et l'innovation, on adresse aussi aux répondants les publications ou les condensés d'indicateurs de R&D.

Le tableau 5 donne un aperçu concernant la participation des entreprises lors de la dernière enquête de R&D en 2008. Le taux de réponse global non pondéré est estimé à 42,23 %. La participation aux enquêtes de R&D des entreprises reste stable depuis plus d'une décennie et se situe en moyenne entre 40% et 45%. Compte tenu du caractère fort concentré des activités de R&D du secteur privé auprès d'un nombre très limité d'entreprises en Belgique, le taux de réponse aux enquêtes peut mieux s'apprécier en analysant notamment la participation des entreprises les plus importantes en matière

---

<sup>8</sup> Le début de l'enquête est souvent prévu au plus tard le mois d'avril de l'année au cours de laquelle l'enquête est tenue et sa clôture a généralement lieu à la fin du mois d'octobre de la même année.

de R&D. Lorsqu'on considère le groupe de 159 entreprises les plus importantes en matière de R&D, le taux de réponse non pondéré atteint près de 75%.

**Tableau 5 : Taux de réponse<sup>9</sup> à l'enquête R&D de 2008 par classification des entreprises**

	Nombre d'unités avec réponses	Nombre total d'unités	Taux de réponse en %
Entreprises répertoriées pour R&D permanente <i>dont les 159 plus importantes en R&amp;D</i>	939 119	1821 159	51,57 74,84
Entreprises répertoriées pour R&D occasionnelle ou R&D exclusivement sous-traitée	651	2069	31,46
Population des entreprises non répertoriées (échantillon)	1261	2861	44,08
<b>Total</b>	<b>2851</b>	<b>6751</b>	<b>42,23</b>

Source : CFS/STAT, Politique scientifique fédérale, 2009.

Note : les 159 entreprises les plus importantes en matière de R&D exécutent en généralement plus de 70% du total des dépenses de R&D de l'ensemble des entreprises.

La participation des entreprises répertoriées pour leurs travaux de R&D occasionnelle ou exclusivement sous-traitée est en moyenne faible par rapport à celle des autres catégories de classification des entreprises. Puisque les efforts supplémentaires de collecte et les programmes de suivi d'enquête sont moins orientés vers ces entreprises. En outre, la participation de ces entreprises aux enquêtes est aussi irrégulière que leur caractère non continu des travaux de R&D. Beaucoup d'entre elles ne souhaitent pas répondre quand elles n'ont pas de projets de R&D.

Le taux de réponse non pondéré des entreprises de l'échantillon de la population des entreprises non répertoriées est estimé à 44,08%. Conformément aux recommandations méthodologiques d'Eurostat concernant les enquêtes sur l'innovation<sup>10</sup> des entreprises, lorsque le taux de réponse ne dépasse pas 70% une analyse des non répondants est effectuée. On peut procéder de la même façon dans le cadre des enquêtes de R&D en ce qui concerne les non répondants de l'échantillon de la population des entreprises non répertoriées. L'analyse des non réponses dans un échantillon représentatif d'unités n'ayant pas répondu consiste à organiser une simple enquête<sup>11</sup> s'appuyant sur un échantillon aléatoire simple d'au moins 10% des non répondants excepté les entreprises non actives. Le but de cette analyse est de vérifier si les non répondants ont un comportement différent que les répondants à l'enquête de départ. En d'autre terme, il s'agit de vérifier l'hypothèse selon que la proportion des entreprises actives en R&D est identique parmi les non répondants et les répondants l'enquête de départ. Le taux de réponse à cette enquête doit être très élevé. En plus, cette enquête est réalisée dans la mesure du possible en fonction des différentes strates c'est à des secteurs d'activités et des catégories de taille des entreprises.

A la dernière enquête sur la R&D des entreprises tenue en 2008, l'analyse des non répondants de l'échantillon de la population des entreprises non répertoriées a été seulement organisée dans une des trois régions et a concerné 129 entreprises non répondantes à l'enquête de départ. Le taux de réponse a été de 87%. Les conclusions de l'analyse ont débouché à confirmer l'hypothèse selon laquelle le comportement est identique entre les non répondants et les répondants de l'enquête de départ. Par conséquent, il n'existe pas de différence statistiquement significative (voir les résultats du test dans l'encadré 1).

**Encadré 1 : Test d'égalité entre la proportion sur la R&D des répondants et celle des non répondants de l'échantillon d'une des trois régions de la Belgique**

Soient  $P_r = 0,138$ : la proportion des entreprises actives en R&D parmi les répondants de l'échantillon  
 $P_{nr} = 0,194$ : la proportion des entreprises actives en R&D parmi les non répondants sélectionnés  
 $n_r = 168$ : nombre de répondants de l'échantillon

<sup>9</sup> Le taux de réponse est non pondéré et est établi en fonction des unités répondantes aux questions effectives sur la R&D dans le questionnaire de l'enquête.

<sup>10</sup> A la demande d'Eurostat, les enquêtes harmonisées européennes sur l'innovation des entreprises (CIS) sont régulièrement menées dans les pays membres de l'Union européenne en vue de dresser un panorama des activités d'innovation des firmes.

<sup>11</sup> L'enquête simple est souvent téléphonique et porte sur des questions simples qui visent à savoir si l'entreprise dispose des activités de R&D et dans l'affirmative quels effectifs approximatifs sont alloués à ces activités.

$N = 420$  : nombre total d'entreprises de l'échantillon

$r_e = 40,0$  : taux de réponse de l'échantillon

$n_{nr} = 112$  : nombre considéré de non répondants sélectionnés

Hypothèses  $H_0 : P_r = P_{nr}$  ou  $P_r - P_{nr} = 0$

$H_1 : P_r \neq P_{nr}$

Test statistique : Normale Z,

$$Z = \frac{(\bar{P}_r - \bar{P}_{nr})}{\sqrt{S^2(\bar{P}_r) + S^2(\bar{P}_{nr})}} = -1,517548$$

$S^2(\bar{P}_r)$  : Variance liée à  $P_r$ ,

$$S^2(\bar{P}_r) = \left( \frac{\bar{P}_r(1-\bar{P}_r)}{n_r} \right) = 0,00071$$

$S^2(\bar{P}_{nr})$  : Variance liée à  $P_{nr}$ ,

$$S^2(\bar{P}_{nr}) = \left( \frac{\bar{P}_{nr}(1-\bar{P}_{nr})}{n_{nr}} \right) \left( 1 - \frac{n_{nr}}{N(1-r_e)} \right) = 0,00079$$

Décision :  $-1,96 < Z = -1,517548 < 1,96$ , l'hypothèse  $H_0$  est acceptée et donc il n'existe pas de différence significative entre les deux proportions à un niveau de significativité de 5%.

## 4. Traitement statistique

Cette section concernant le traitement statistique des données de R&D des entreprises collectées par les enquêtes s'articule essentiellement autour des points suivants : le contrôle de cohérence des données, les méthodes d'estimation et l'évaluation de la qualité des données produites.

### 4.1. Contrôle de cohérence des données de R&D

Comme déjà souligné dans les sections précédentes, les données de R&D collectées conformément au Manuel de Frascati de l'OCDE sont rarement disponibles dans les entreprises voire dans leurs comptes financiers. Les répondants aux enquêtes sont amenés à les estimer eux-mêmes sur base des définitions formelles des variables de R&D, des notes explicatives et autres directives contenues dans le questionnaire de l'enquête. Pour s'assurer de la qualité et de la cohérence des données collectées, une série de tests sont mis en oeuvre. Ces tests ont été également établis dans l'objectif de préserver d'une part les relations de pertinence entre les différentes principales variables de R&D (dépenses, personnel, chercheurs) et d'autre part les relations entre les variables de R&D et certaines variables économiques telles que le chiffre d'affaires et les effectifs totaux. La majeure partie de ces tests sur la cohérence des données sont intégrés au questionnaire électronique de sorte à signaler les éventuelles données incohérentes lors de l'envoi électronique des données. Notons que le questionnaire électronique sert également de formulaire pour l'encodage des réponses des enquêtes en version papier.

Les données jugées incohérentes sont dans un premier temps vérifiées auprès des répondants dans le cadre de la coopération établie avec eux. En cas de non coopération du répondant, ces données sont traitées en exploitant notamment d'autres sources d'informations (enquêtes de R&D antérieures, autres enquêtes, comptes financiers, sites internet, etc.). Enfin de compte, les données incohérentes non corrigées, à défaut de collaboration des répondants ou par manque de disponibilité d'informations des autres sources, peuvent être automatiquement traitées par des procédures informatisées développées à ce sujet.

Parallèlement au contrôle de cohérence et de qualité des données collectées aux enquêtes de R&D, les tests mis au point permettent aussi de relever les réponses incomplètes ou manquantes. Une partie de réponses incomplètes ou manquantes peuvent être éventuellement complétées par les procédures automatiques de correction qui recourent notamment aux autres renseignements de l'entreprise fournis dans le cadre de la même enquête.

**Tableau 6 : Aperçu des réponses partiellement incomplètes et/ou incohérentes non encore traitées automatiquement à l'enquête R&D de 2008 par classification des entreprises**

	Nombre d'unités avec réponses incomplètes et /ou incohérentes	Nombre d'unités avec réponses	Taux en %
Entreprises répertoriées pour R&D permanente	377	939	40,15
Entreprises répertoriées pour R&D occasionnelle ou R&D exclusivement sous-traitée	146	651	22,43
Population des entreprises non répertoriées (échantillon)	115	1261	9,12
<b>Total</b>	<b>638</b>	<b>2851</b>	<b>22,38</b>

Source : CFS/STAT, Politique scientifique fédérale, 2009.

Dans le tableau 6, le taux des réponses partiellement incomplètes et/ou incohérentes non encore automatiquement traitées des entreprises répondantes répertoriées pour leurs travaux de R&D permanente est évidemment élevé. Puisque ces entreprises doivent répondre à l'entière du questionnaire principal contrairement aux autres entreprises qui font de la R&D occasionnelle ou qui n'ont aucune activité de R&D ou encore qui sont interrogées par un questionnaire simplifié notamment les entreprises de l'échantillon de la population non répertoriées. Les non-réponses partielles sont donc plus fréquentes pour des entreprises qui exécutent les travaux de R&D permanente et concernent essentiellement les questions sur les différentes répartitions fonctionnelles des dépenses et du personnel de R&D<sup>12</sup>.

## 4.2. Estimation des données de R&D

Les enquêtes de R&D des entreprises en Belgique visent à estimer le total pour l'ensemble des entreprises et au niveau des secteurs d'activités et des catégories de taille des entreprises des principales variables de R&D suivantes :

- dépenses de R&D intra-muros (Intexp),
- dépenses de R&D extra-muros (Extexp),
- personnel total de R&D en ETP (Tpersfte) et en personnes physiques (Tperspu) et
- nombre de chercheurs en ETP (Tresfte) et en personnes physiques (Trespu).

En outre, il est aussi question d'estimer des sous-agrégats ou des répartitions institutionnelles de ces principales variables de R&D. D'autres variables liées aux activités en rapport avec la R&D sont aussi traitées lors des enquêtes de R&D.

La compilation des statistiques de ces variables de R&D fait intervenir différentes méthodes d'estimation. Ces méthodes sont établies en fonction de la nature des travaux de R&D et de la classification des entreprises à savoir : les entreprises répertoriées pour la R&D permanente, les entreprises répertoriées pour R&D occasionnelle ou avec R&D exclusivement sous-traitée et les entreprises de l'échantillon de la population des entreprises non répertoriées.

Soit  $Y_t$  le total d'une des principales variables de R&D pour l'ensemble des entreprises à la période t.

En fait, le total d'une des principales variables de R&D constitue la somme des totaux de chaque classification des entreprises de R&D :

$$Y_t = Y_{p_t} + Y_{o_t} + Y_{s_t}$$

Avec respectivement  $Y_{p_t}$ ,  $Y_{o_t}$  et  $Y_{s_t}$  le total d'une des principales variables de R&D des entreprises répertoriées pour la R&D permanente, le total des entreprises répertoriées pour R&D occasionnelle ou avec R&D exclusivement sous-traitée et le total de la population des entreprises non répertoriées obtenu par un sondage.

<sup>12</sup> Notez que dans les enquêtes de R&D, les répartitions fonctionnelles des dépenses de R&D intra-muros visent notamment les sources de financement, les types de recherche ou les différentes sortes de coûts et celles du personnel de R&D concernent la part des femmes, les professions ou les niveaux de qualification du personnel de R&D ou des chercheurs.

#### 4.2.1. Estimation du total de R&D des entreprises répertoriées pour la R&D permanente

L'estimation du total d'une des principales variables de R&D pour les entreprises répertoriées pour leur R&D permanente se fonde en principe sur les réponses aux enquêtes de R&D. Mais dans la pratique, les réponses aux enquêtes de R&D sont souvent incomplètes. On distingue les non-réponses complètes et les non-réponses partielles. Une non-réponse complète signifie que l'entreprise ne répond pas du tout à l'enquête, c'est à dire les données de toutes les variables principales de R&D pour cette entreprise sont complètement manquantes. En cas de non-réponse partielle, les données d'au moins une des variables principales de R&D sont disponibles à partir de l'enquête. Pour remédier aux non-réponses partielles ou complètes, une série de méthodes dites «méthodes d'imputation» ont été développées et elles visent donc à estimer les non-réponses à partir d'informations complémentaires de la même enquête ou des enquêtes antérieures les plus récentes<sup>13</sup>.

Dans le cas de non-réponse partielle, les données de R&D partiellement disponibles de l'entreprise dans le cadre de la même enquête sont utilisées avec un ensemble des ratios qui mettent en rapport les différentes variables de R&D de entreprise de la même enquête ou des enquêtes antérieures les plus récentes. Ces ratios sont notamment établis en fonction des relations entre les principales variables de R&D. Les principaux ratios au niveau de l'entreprise (i) servant à imputer les données manquantes de R&D dans le cas de non-réponse partielle sont les suivants :

$$\frac{Int\ exp_{it}}{Tpersfte_{it}}, \frac{Ext\ exp_{it}}{Int\ exp_{it}}, \frac{Tpersfte_{it}}{Tperspu_{it}}, \frac{Tresfte_{it}}{Trespu_{it}}, \frac{Tresfte_{it}}{Tpersfte_{it}}.$$

L'information partielle sur au moins une des variables principales de R&D de l'entreprise à l'enquête et la connaissance d'un de ces ratios soit à partir de la même enquête soit à partir des enquêtes antérieures les plus récentes permettent ainsi d'estimer les données manquantes des autres principales variables de R&D. Cette méthode part de l'hypothèse que les rapports entre les principales variables de R&D d'une entreprise restent quasiment identiques au cours des années successives d'une période généralement estimées à quatre ans.

En ce qui concerne la non-réponse complète, les données antérieures récentes disponibles sur les principales variables de R&D de la firme sont utilisées pour estimer les données manquantes correspondantes pour la période sous revue. Les données antérieures sont évidemment adaptées en fonction de l'évolution récente de R&D de l'entreprise et de l'évolution de la R&D du secteur d'activités et de la catégorie de taille auxquels l'entreprise est associée. Lorsqu'on ne dispose pas de données antérieures récentes sur les principales variables de R&D de la firme, on peut utiliser les ratios moyens de la cellule, c'est à dire du secteur d'activités et de la catégorie de taille de l'entreprise, pour estimer les données manquantes de R&D des entreprises. Parmi les ratios énumérés ci-dessous, on peut aussi ajouter d'autres ratios concernant les données des dépenses de R&D intra-muros par rapport au chiffre d'affaires (Turn) et du personnel total de R&D en ETP par rapport aux effectifs totaux en ETP (Empfte). On sait que la R&D étant une variable métrique en partie corrélée à certaines variables économiques telles que le chiffre d'affaires et l'emploi.

L'ensemble des ratios considérés au niveau d'une cellule (j) sont alors les suivants :

$$\frac{Int\ exp_{jt}}{Tpersfte_{jt}}, \frac{Ext\ exp_{jt}}{Int\ exp_{jt}}, \frac{Tpersfte_{jt}}{Tperspu_{jt}}, \frac{Tresfte_{jt}}{Trespu_{jt}}, \frac{Tresfte_{jt}}{Tpersfte_{jt}}, \frac{Int\ exp_{jt}}{Turn_{jt}}, \frac{Tpersfte_{jt}}{Empfte_{jt}}.$$

Les ratios d'une cellule sont des rapports des moyennes pondérées respectives calculées avec les données des entreprises répondantes répertoriées pour la R&D permanente de la cellule. Lors du calcul des moyennes, les données aberrantes de la cellule sont exclues<sup>14</sup>.

<sup>13</sup> En pratique, seules les données des deux dernières enquêtes sont considérées comme des enquêtes antérieures les plus récentes. Etant donné que les enquêtes sont organisées tous les deux et se réfèrent aux données annuelles des deux précédentes années, ce sont donc les données de quatre années récentes qui peuvent être prises en compte.

<sup>14</sup> En réalité, toutes les observations qui sont au delà de l'intervalle interquartile augmenté (1,5 fois l'intervalle interquartile) de la cellule sont écartées dans le calcul de la moyenne. Un minimum de plus ou moins 10

$$\frac{\sum_i Int\ exp_{it}}{\sum_i Tpersfte_{it}}, \frac{\sum_i Ext\ exp_{it}}{\sum_i Int\ exp_{it}}, \frac{\sum_i Tpersfte_{it}}{\sum_i Tperspu_{it}}, \frac{\sum_i Tresfte_{it}}{\sum_i Trespu_{it}}, \frac{\sum_i Tresfte_{it}}{\sum_i Tpersfte_{it}}, \frac{\sum_i Int\ exp_{it}}{\sum_i Turn_{it}},$$

$$\frac{\sum_i Tpersfte_{it}}{\sum_i Empfte_{it}}.$$

Le total d'une des principales variables de R&D concernant les entreprises répertoriées pour leur R&D permanente au cours d'une période donnée ( $Yp_t$ ) est donc la somme des données fournies à l'enquête par les répondants et des estimations ou imputations effectuées pour les non-réponses partielles ou complètes.

$$Yp_t = \sum_{i=1}^{n_{pr}} y_{it} + \sum_{i=1}^{n_{pnr}} y_{it}$$

$n_{pr}$  et  $n_{pnr}$  se réfèrent respectivement au nombre d'unités répondantes et au nombre d'unités non répondantes parmi les entreprises répertoriées pour leur R&D permanente.

#### 4.2.2. Estimation du total de R&D des entreprises répertoriées pour la R&D occasionnelle ou avec R&D exclusivement sous-traitée

L'estimation du total d'une des principales variables de R&D pour cette catégorie d'entreprises se fonde évidemment sur les réponses à l'enquête et les estimations des données partiellement ou complètement manquantes. La différence importante par rapport aux entreprises répertoriées pour leur R&D permanente réside dans l'estimation des non-réponses complètes. Les méthodes d'imputation pour les non-réponses complètes dans ce cas ne prennent pas en compte des données antérieures de R&D de la firme. Puisque les entreprises qui n'ont que des activités de R&D de façon occasionnelle ou des activités de R&D extra-muros, peuvent s'engager dans un projet une année et ne plus exécuter de R&D l'année suivante. Les données manquantes des principales variables de R&D sont donc estimées en imputant les moyennes des résultats des entreprises répondantes au niveau de la cellule. On remplace donc la valeur manquante par une moyenne calculée à partir des résultats des unités répondantes de la cellule.

Le total de R&D des entreprises répertoriées pour la R&D occasionnelle ou avec R&D exclusivement sous-traitée est donné par la formule suivante :

$$Yo_t = \sum_{i=1}^{n_{or}} y_{it} + \sum_{i=1}^{n_{onrp}} y_{it} + \sum_{j=1}^J \left( \frac{n_j^{onr}}{n_j^{or}} \left( \sum_{i=1}^{n_{jor}} y_{it} + \sum_{i=1}^{n_{jonrp}} y_{it} \right) \right)$$

Les deux premiers termes sommes de la formule concernent respectivement les données des répondants et les estimations des répondants partiels. Le reste de la formule représente la partie calculée pour les non répondants complets.

$n_j^{or}$ ,  $n_j^{onr}$  constituent respectivement le nombre d'unités répondantes et partiellement répondantes et le nombre d'unités totalement non répondantes d'une cellule (j). Pour procéder aux estimations des entreprises partiellement répondantes, on recourt notamment aux ratios des cellules des entreprises répertoriées pour leur R&D permanente.

---

observations est aussi requis pour calculer les moyennes pondérées de la cellule. A défaut les cellules peuvent être fusionnées avec celles qui leur sont proches, souvent avec des cellules des secteurs de même niveau de technologie.

### 4.2.3. Estimation du total de R&D de la population des entreprises non répertoriées

L'estimation du total des principales variables de R&D de la population des entreprises non répertoriées s'effectue via les résultats du sondage. Les résultats de R&D des entreprises répondantes de l'échantillon à l'enquête sont extrapolés à l'aide de diverses méthodes pour correspondre à l'ensemble de la population des entreprises non répertoriées. En d'autres termes, les résultats des principales variables de R&D des entreprises répondantes de l'échantillon sont pondérés en fonction de l'échantillonnage et de la non-réponse afin de produire des résultats valides de R&D de la population cible des entreprises non répertoriées. Les facteurs d'extrapolation utilisés au niveau des strates ou cellules sont des rapports entre le nombre total d'unités de la strate et le nombre d'unités répondantes du sous-échantillon représentant cette strate. En principe, on doit assumer que chaque unité dans une strate a la même probabilité d'inclusion. Mais en pratique, on peut procéder à la correction des poids de certaines unités répondantes jugées fort intensives en R&D<sup>15</sup>. Pour ces dernières, les poids appliqués correspondent à ceux des entreprises répertoriées pour leur R&D permanente, c'est à dire de valeur égale à 1. En cas de correction, le nombre total d'unités et le nombre d'unités répondantes dans une strate sont également adaptés.

Les unités répondantes sont constituées par des unités totalement répondantes, des unités partiellement répondantes à l'enquête ainsi que par des unités non répondantes de l'échantillon mais ayant déclaré aux enquêtes antérieures les plus récentes ne pas exécuter les travaux de R&D. Les données de R&D partiellement manquantes sont estimées notamment avec l'aide des ratios des cellules des entreprises répertoriées pour leur R&D permanente.

Le total d'une principale variable de R&D de la population des entreprises non répertoriées est estimé par la formule suivante :

$$Y_{S_t} = \sum_{h=1}^H \left( \frac{N_h}{n_h^{SR}} \sum_{i=1}^{n_h^{SR}} y_{it} \right) + \sum_{i=1}^{n^{SRP}} y_{it}$$

La première partie de la formule est obtenue par extrapolation au niveau des strates (h) des résultats de R&D des entreprises répondantes excepté éventuellement des résultats de certaines unités répondantes fort intensives en R&D repris en deuxième partie de la formule.

$N_h$  et  $n_h^{SR}$  représentent respectivement le nombre total adapté d'unités et le nombre adapté d'unités répondantes dans une strate. Et  $n^{SRP}$  constitue le nombre d'unités répondantes considérées fort intensives en R&D.

### 4.2.4. Résultats du total de R&D de l'ensemble des entreprises

Pour terminer cette section, nous présentons successivement les résultats obtenus pour les dépenses de R&D intra-muros et le personnel total de R&D en ETP de l'ensemble des entreprises au cours des dernières années dans les tableaux 7 et 8. Ces résultats sont répartis par classification des entreprises de R&D et une distinction est faite selon que les données soient des réponses à l'enquête ou des estimations. En analysant rapidement les résultats pour 2004 et 2005, on peut s'apercevoir que près de 75% du total des dépenses de R&D intra-muros sont estimés par les réponses de l'enquête et pour le total du personnel total de R&D en ETP les réponses à l'enquête représentent plus de 65%. On peut noter avec satisfaction que les estimations des non-réponses complètes sans données antérieures des entreprises répertoriées pour la R&D permanente ne représentent plus ou moins que 1% du total général pour les deux variables principales de R&D.

**Tableau 7: Dépenses de R&D intra-muros des entreprises en Belgique par classification des entreprises de R&D (en millions d'euros)**

	2000	2001	2002	2003	2004	2005	2006(p)
<b>Entreprises répertoriées pour la R&amp;D permanente</b>							

<sup>15</sup> Les critères d'intensité de R&D sont explicitement définis. Généralement, toutes les entreprises dont les dépenses de R&D intra-muros dépassent 1 million d'euros sont concernées.

-Réponses	2301,3	2539,7	2327,5	2301,3	2457,8	2495,4	2436,7
-Estimations des non-réponses partielles	58,4	36,2	13,3	15,3	5,7	7,6	60,1
-Estimations des non-réponses complètes avec données antérieures	656,0	686,3	711,4	657,8	662,2	641,5	775,2
-Estimations des non-réponses complètes sans données antérieures	37,2	38,1	51,8	51,9	29,9	28,7	30,2
<b>-Total</b>	<b>3052,9</b>	<b>3300,3</b>	<b>3103,9</b>	<b>3026,3</b>	<b>3155,5</b>	<b>3173,2</b>	<b>3302,2</b>
<b>Entreprises répertoriées pour la R&amp;D occasionnelle</b>							
-Réponses	131,7	169,8	142,7	159,5	204,8	214,6	230,6
-Estimations des non-réponses complètes	82,3	99,1	101,2	109,9	48,6	54,2	56,0
<b>-Total</b>	<b>214,0</b>	<b>268,9</b>	<b>243,9</b>	<b>269,4</b>	<b>253,4</b>	<b>268,8</b>	<b>286,6</b>
<b>Population des entreprises non répertoriées</b>							
-Echantillon réalisé	82,2	95,3	76,5	80,9	100,8	101,4	105,3
<b>-Total</b>	<b>321,8</b>	<b>351,9</b>	<b>314,5</b>	<b>312,2</b>	<b>322,9</b>	<b>333,6</b>	<b>345,6</b>
<b>Population totale des entreprises</b>							
<b>-Total général</b>	<b>3588,6</b>	<b>3921,1</b>	<b>3662,3</b>	<b>3607,9</b>	<b>3731,8</b>	<b>3775,6</b>	<b>3934,4</b>

Source : CFS/STAT, Politique scientifique fédérale, 2009.

**Tableau 8: Personnel total de R&D des entreprises en Belgique par classification des entreprises de R&D (en ETP)**

	2000	2001	2002	2003	2004	2005	2006(p)
<b>Entreprises répertoriées pour la R&amp;D permanente</b>							
-Réponses	17409	18597	17461	17033	17156	17938	15303
-Estimations des non-réponses partielles	1618	1284	1165	1022	561	583	2446
-Estimations des non-réponses complètes avec données antérieures	6957	7004	5919	5919	6276	6069	7266
-Estimations des non-réponses complètes sans données antérieures	459	458	571	568	303	282	283
<b>-Total</b>	<b>26444</b>	<b>27343</b>	<b>25116</b>	<b>24541</b>	<b>24296</b>	<b>24872</b>	<b>25298</b>
<b>Entreprises répertoriées pour la R&amp;D occasionnelle</b>							
-Réponses	1439	1859	1258	1439	1988	2151	2246
-Estimations des non-réponses complètes	943	1155	951	1133	657	733	754
<b>-Total</b>	<b>2382</b>	<b>3014</b>	<b>2209</b>	<b>2572</b>	<b>2645</b>	<b>2883</b>	<b>3000</b>
<b>Population des entreprises non répertoriées</b>							
-Echantillon réalisé	917	1099	814	850	1019	1018	1041
<b>-Total</b>	<b>4667</b>	<b>5132</b>	<b>4360</b>	<b>4262</b>	<b>3800</b>	<b>3858</b>	<b>3908</b>
<b>Population totale des entreprises</b>							
<b>-Total général</b>	<b>33493</b>	<b>35490</b>	<b>31686</b>	<b>31375</b>	<b>30741</b>	<b>31613</b>	<b>32206</b>

Source : CFS/STAT, Politique scientifique fédérale, 2009.

### 4.3. Evaluation de la qualité des données

Dans le cadre des enquêtes R&D des entreprises, il est désormais prévu à l'initiative d'Eurostat de faire un rapport afin d'évaluer la qualité des données. L'évaluation de la qualité statistique est basée sur une série de critères ou indicateurs établis à ce sujet par Eurostat. Ces critères portent notamment sur la pertinence, la précision, la régularité, la disponibilité, la comparabilité, la cohérence ainsi que sur les coûts de production des statistiques et les charges déclaratives des répondants. Un bon nombre de ces éléments de qualité ont été déjà abordés, même sommairement, dans les sections précédentes. D'autres aspects évaluant la qualité des données tels que les coûts et les charges ne vont pas être abordés. Dans la suite nous nous focalisons principalement sur la précision des données produites à partir des enquêtes de R&D. Le but poursuivi dans cette sous-section est d'évaluer la précision de manière qualitative et non l'évaluer quantitativement par des coefficients de variations comme dans les enquêtes se fondant uniquement sur un sondage.

La précision dénote la validité des estimations par rapport aux valeurs exactes. Très souvent, les données statistiques diffèrent des valeurs exactes à cause notamment des variabilités et des biais qui sont créés tout au long du processus de collecte et de production des données. Il existe, dans une enquête, de nombreuses sources d'erreur. Les principales sont : l'erreur de couverture, l'erreur d'échantillonnage, l'erreur de mesure, l'erreur due à la non-réponse.

#### 4.3.1. Erreur de couverture

L'erreur de couverture est due au fait que certains individus ou unités statistiques dans la population ont une probabilité nulle de figurer dans l'échantillon. Cela peut survenir si la population définie par la base de sondage ne coïncide pas avec la population que l'on désire réellement étudier. Dans les enquêtes de R&D des entreprises en Belgique qui combine systématiquement un inventaire des entreprises dont il est connu ou supposé exécuter les travaux de R&D à un sondage sur le reste de la population, cette erreur est très limitée. En outre, la constitution de l'échantillon du reste de la population s'appuie sur un registre officiel de bonne qualité de l'ONSS.

#### 4.3.2. Erreur de d'échantillonnage

Ce type d'erreur affecte davantage les enquêtes dont la détermination de la population d'enquête s'effectue intégralement ou essentiellement par un sondage. Il est naturellement dû au fait qu'une partie de la population totale est seulement sélectionnée de manière aléatoire pour être considérée lors de l'enquête. On peut évaluer l'importance de l'erreur d'échantillonnage associée à un estimateur en calculant le biais et la variance. Dans les enquêtes de R&D des entreprises en Belgique, la population des entreprises non répertoriées, pour laquelle un échantillon est constitué, représente moins de 10 % du total de la R&D de l'ensemble des entreprises. En outre, le recours à l'échantillon permet aussi d'améliorer le champ couvert par l'inventaire des entreprises dont il est connu ou supposé exécuter les travaux de R&D. Par conséquent, les éventuelles incertitudes liées à l'échantillonnage ont des effets très limités dans l'estimation du total de R&D de l'ensemble des entreprises.

#### 4.3.3. Erreur de mesure

Elle provient essentiellement des inexactitudes des réponses enregistrées. Le questionnaire utilisé pour collecter les données peut notamment conduire à l'enregistrement des réponses inexactes ou incohérentes. L'effet induit par l'enquêteur sur les réponses des enquêtés et l'erreur due au répondant sont aussi les autres causes de l'erreur de mesure. Dans les enquêtes sur la R&D des entreprises, l'incompréhension des questions, la perception des concepts de R&D, l'estimation des données de R&D selon le Manuel de Frascati par les répondants conduisent parfois à des réponses inexactes ou incohérentes. Une série de mesures sont mises en oeuvre pour réduire ou limiter ce type d'erreur. Ces mesures vont de la présentation des questions dans le questionnaire avec des notes explicatives, des directives et des exemples de cas de figure au contrôle automatisé sur la qualité et la cohérence des données collectées en passant par la simulation de la coopération avec les répondants.

#### 4.3.4. Erreur due à la non-réponse

L'erreur due à la non-réponse est présente si l'on ne peut pas réaliser l'enquête auprès de toutes les unités sélectionnées. Les causes principales de la non-réponse sont le refus de répondre, la non localisation de l'unité à interroger, l'inaptitude à répondre ou l'invalidité de la réponse. Les raisons de la non-réponse due au refus concernent notamment les difficultés à répondre aux questions jugées plus spécifiques, voire complexes et détaillées. Presque tous les répondants procèdent à des estimations car les données de R&D selon le Manuel de Frascati sont peu disponibles au sein de l'entreprise. Le questionnaire paraît aussi long et nécessite beaucoup de ressources et du temps. Le traitement de la non-réponse dans les enquêtes sur la R&D des entreprises a été suffisamment abordé au point ci-dessus relatif à l'estimation des données de R&D. Rappelons qu'environ 75% du total des dépenses de R&D intra-muros sont estimés par les réponses à l'enquête.

## Conclusion

Ce papier vient de décrire brièvement quelques indications méthodologiques qui sont considérées dans la réalisation des enquêtes sur la R&D des entreprises en Belgique. En effet, les données statistiques de R&D des entreprises sont collectées et traitées dans le cadre des enquêtes spécialisées et régulières et qui sont inspirées par le Manuel de Frascati de l'OCDE pour une meilleure comparabilité internationale des données. En outre, l'objectif de ces enquêtes est de permettre en Belgique tant au niveau national que régional de faire l'inventaire du potentiel scientifique et de mesurer ainsi toutes les ressources financières et humaines consacrées aux activités de R&D.

La détermination de la population d'enquête sur la R&D des entreprises se fonde en grande partie sur un inventaire des entreprises dont il est connu ou supposé qu'elles exécutent en interne la R&D permanente ou occasionnelle et/ou sous-traitent leur R&D. Cependant, l'inventaire est complété par un sondage stratifié afin de recueillir les informations de R&D exécutée par des autres entreprises non répertoriées. L'inventaire est régulièrement tenu à jour grâce à plusieurs sources dont les enquêtes sur la R&D elles-mêmes. La méthode qui combine à la fois l'inventaire et le sondage présente l'avantage d'aboutir à une estimation du total de la R&D de la population cible qui comporte beaucoup moins d'incertitudes que la méthode de l'échantillonnage classique. L'erreur de l'échantillonnage est donc très limitée. En outre, en prenant en compte les données des travaux antérieurs de R&D d'une firme de l'inventaire qui ne répond pas à l'enquête en vue d'imputer les données manquantes, cela contribue également à rendre les données estimées de R&D de l'ensemble des entreprises beaucoup plus comparables dans le temps. Enfin, cette méthode convient mieux pour les petits pays, comme la Belgique, contrairement aux grands pays dans lesquels les coûts peuvent s'avérer très élevés pour la mettre en pratique.

La compilation des données statistiques sur la R&D des entreprises, et en particulier le traitement de la non-réponse, fait intervenir différentes méthodes d'estimation selon que les entreprises soient répertoriées ou non, ou encore exécutent la R&D permanente, la R&D occasionnelle ou sous-traitent exclusivement leur R&D. Pour traiter les non-réponses complètes et partielles des entreprises répertoriées pour la R&D permanente, les données antérieures sur la R&D sont utilisées pour être imputées aux données manquantes. Les données antérieures sont tout de même adaptées en fonction d'une série d'informations complémentaires fournies notamment dans le cadre de la même enquête. Quant aux non-réponses complètes des entreprises répertoriées pour la R&D occasionnelle, elles sont exclusivement estimées par les renseignements de la même enquête, en imputant les moyennes des résultats des entreprises répondantes au niveau de la cellule. L'estimation de la R&D des autres entreprises non répertoriées s'effectue en extrapolant les résultats des entreprises répondantes de l'échantillon. Ces méthodes d'estimation permettent ainsi d'obtenir le total des principales variables de R&D de l'ensemble des entreprises et au niveau des secteurs d'activité et des catégories de taille des entreprises. L'erreur due à la non réponse sur l'estimation du total des principales variables de R&D de l'ensemble des entreprises peut s'avérer très faible car environ 75% du montant total estimé sont composés par les réponses à l'enquête.

Enfin, des efforts peuvent continuer à être déployés en vue de limiter l'erreur de mesure concernant les travaux de R&D des entreprises. Puisque dans les enquêtes sur la R&D des entreprises, l'incompréhension des questions, la perception des concepts de R&D, l'estimation des données de R&D selon les critères du Manuel de Frascati par les répondants conduisent parfois à des réponses inexactes ou incohérentes. Notez pour terminer que les données estimées pour la R&D des entreprises ne constituent qu'une indication quantitative d'un ensemble très complexe et hétérogène d'activités et de structures d'entreprises. Des éléments qualitatifs peuvent y être ajoutés en vue d'une meilleure utilisation, analyse ou interprétation de ces données.

## Bibliographie

- [1] OCDE, « Manuel de Frascati: Méthode type proposée pour les enquêtes sur la recherche et le développement expérimental (R&D) », *OCDE*, 2002, Paris.
- [2] Cincera M., Kalenga-Mpala R., Veugelers R. and al., « Activities of the Business Sector in Flanders: Result of the R&D Surveys in the Context of the 3% Target », *IWT Studies*, 2004, n° 46, Brussels.
- [3] Ardilly P., « Les techniques de sondage (2e édition) », *Editions Technip*, 2006, Paris.
- [4] Capron H., Cincera M., Hollant A., Meeusen W., Suetens S., Teirlinck P., « ANBERD – Harmonisation of Belgian Industrial R&D Expenditures Statistic – Final Report », *Belgian Department for Scientific, Technological and Cultural Affairs (OSTC)*, 1999, Brussels.
- [5] Delhaussé P., Kalenga-Mpala R., « Quality Report for R&D statistics in the Business Enterprise Sector in Belgium », *Eurostat document for the Federal Science Policy Office*, 2007.
- [6] Vermandele C., « Cours des Méthodes de Sondage et d'Enquêtes (STAT-D-307) », *Université libre de Bruxelles*, 2007-2008.
- [7] Eurostat, « The Fourth Community Innovation Survey (CIS 4): Methodological recommendations », 2004.
- [8] Politique scientifique fédérale and Steunpunt O&O-Indicatoren, « Estimation of missing data in the R&D survey », 2008.