

Méthodologie statistique

M2017/01

**Exploitation de l'enquête
expérimentale
Qualité de vie au travail**

Tiaray Razafindranovona (DMCSI)

Document de travail



Institut National de la Statistique et des Études Économiques

M 2017/01

**Exploitation de l'enquête expérimentale
Qualité de vie au travail**

Tiaray Razafindranovona (DMCSI)

Ce document regroupe différents travaux autour de l'expérimentation Qualité de vie au travail et de sa comparaison avec l'enquête Conditions de travail. Je remercie Élisabeth Agalva, Amélie Mauroux et Thomas Coutrot pour leurs contributions à l'exploitation de cette expérimentation.

Je remercie également Gaël de Peretti, Olivier Sautory et Sylvie Lagarde pour leur relecture attentive et leurs commentaires constructifs. Je reste seul responsable des erreurs ou omissions qui pourraient rester.

Exploitation de l'enquête expérimentale Qualité de la vie au travail

Tiaray Razafindranovona *

Résumé

Afin d'accumuler de la connaissance et de bien cerner les difficultés méthodologiques que pose la collecte par Internet ou multimode auprès des ménages, l'Insee a programmé sur plusieurs années un plan d'expérimentations. L'enquête expérimentale (web/papier) "Qualité de la vie au travail (QVT)" est l'une de ces expérimentations, réalisée en 2013, en parallèle de l'enquête "Conditions de travail (CT)". L'objectif est de comparer les réponses à ces deux enquêtes.

Après contrôle des différences de sélection entre les deux enquêtes, certaines variables présentent néanmoins toujours des écarts importants qui peuvent relever de la sélection non contrôlée ou de la mesure. La substitution d'une enquête en face à face par une enquête auto-administrée sur ce type de sujet paraît ainsi délicate. De plus, les traitements correctifs par calages devraient être finement différenciés selon la nature des variables d'intérêt ce qui rend difficile la fourniture d'un jeu de pondérations unique.

Mots-clés : enquête multimode, effets de mode, effets de sélection, effets de mesure, conditions de travail, méthodologie d'enquête, calage, score de bien-être, lettre-avis.

Abstract

Insee decided to launch a series of experimental households surveys to increase knowledge in the field of internet data collection and to identify methodological issues. One of these experimental surveys (web and paper), entitled "Qualité de la vie au travail (Quality of life at work)" took place in 2013, in parallel of the face-to-face survey "Conditions de travail (Working conditions)". The aim of the experiment is the comparison of responses of these two surveys. After controlling selection, there are still important differences on some variables because of uncontrolled selection or measurement effect. Thus, replacing a face-to-face survey by an auto-administered survey on this kind of topic could be sensitive. Besides, corrections by calibrations have to be finely differentiated which could make difficult to provide operational and unique sampling weights.

Keywords : mixed-mode survey, mode effects, selection effects, measurement effects, working conditions, survey methodology, calibration, well-being index, notification letter.

* INSEE, Département des méthodes statistiques, tiaray.razafindranovona@insee.fr

Table des matières

1	Présentation et objectifs de l'expérimentation Qualité de la vie au travail	5
1.1	Les enquêtes Conditions de travail	5
1.2	L'enquête Conditions de travail 2013	5
1.3	Le plan d'expérimentations d'enquêtes multimode à l'Insee	6
1.3.1	Protocole standard des expérimentations	6
1.3.2	Enquêtes expérimentales de l'Insee	7
1.4	L'enquête expérimentale Qualité de la vie au travail	7
1.5	Objectifs de l'expérimentation QVT	7
2	Bilan sur la collecte de l'expérimentation QVT	8
2.1	Déroulement de la collecte et taux de réponse	8
2.1.1	Retour d'expérience	8
2.1.2	Taux de réponse et comparaison avec CT	9
2.1.3	Web vs Web+papier, quels taux de réponse ?	9
2.1.4	Ressenti des répondants sur le questionnement	10
2.1.5	Caractéristiques des répondants	11
2.2	Non-réponse partielle, qualité de réponse	11
2.2.1	Non-réponse partielle	11
2.2.2	Question ouverte	12
2.2.3	Qualité des réponses : réponses multiples non prévues	13
2.3	Affectation aléatoire d'un modèle de lettre-avis	13
2.3.1	Les deux modèles de lettre-avis	14
2.3.2	Taux de réponse	14
2.3.3	Caractéristiques des répondants	14
2.3.4	Comparaison des réponses selon la lettre reçue	15
3	Correction de la non-réponse totale, apurements et imputations	17
3.1	Correction de la non-réponse totale	17
3.1.1	Première phase de correction : recevoir ou non la lettre-avis	18
3.1.2	Deuxième phase de correction : répondre ou non à l'enquête	18
3.2	Apurement et imputations	20
4	Comparaison entre CT et QVT : des effets de mode ?	21
4.1	Problématique, éléments théoriques et pratiques sur la mesure des effets de mode	21
4.1.1	Problématique	21
4.1.2	Éléments théoriques sur la mesure des effets de mode	22
4.1.3	Intérêt et limites des comparaisons naïves	22

4.1.4	Le contrôle de la sélection par le calage sur marges (après correction de la non-réponse préalable)	22
4.2	Le calage sur marges de QVT pour contrôler la sélection	23
4.2.1	Utilisation de marges sociodémographiques communes à CT	23
4.2.2	Correction de la déformation	23
4.2.3	Utilisation du score de bien-être	24
4.2.4	Une marge externe supplémentaire : la santé ressentie	26
4.3	Résultats des comparaisons entre CT et QVT	27
4.3.1	Comparaisons détaillées par modules	28
A	Quelques éléments sur la partie de l'expérimentation QVT réalisée à La Réunion	43
B	Internet et papier : des différences entre ces modes auto-administrés ?	45
B.1	En théorie : pas trop de différences entre modes auto-administrés	45
B.2	Contrôle de la sélection par kernel matching sur score de propension	45
B.3	Résultats	49
B.3.1	Modélisation du score de propension	49
B.3.2	Exemple pour illustrer la correction du biais de sélection	49
B.3.3	Utilisation différente du spectre des réponses proposées	50
B.3.4	Score de bien-être	51
C	Comparaisons sur indicateurs synthétiques	53
C.1	Indicateurs synthétiques issus de CT	53
C.2	Construction des indicateurs pour comparaison CT/QVT	54
C.3	Résultats	56

Introduction

Depuis 1978, des enquêtes sur les Conditions de travail (CT) sont organisées et exploitées par la Dares¹ : il s'agit de cerner par une description concrète le travail tel qu'il est vécu. Ces enquêtes sont réalisées en face à face : un enquêteur se déplace au domicile de la personne en emploi pour l'interroger sur ses conditions de travail. Les avantages d'une collecte d'information par l'intermédiation d'enquêteurs pour de telles enquêtes sont indéniables : par leurs efforts, lors de la phase d'approche, de contact ainsi que pendant le questionnement, les enquêteurs parviennent à obtenir une participation large et des réponses globalement de qualité. Mais l'interrogation en face à face n'est pas toujours le mode optimal de collecte, notamment dans le cas de questions sensibles ou qui peuvent induire un biais de désirabilité sociale. C'est en partie pour cela qu'une partie du questionnaire de l'édition 2013 est auto-administrée via un protocole sous casque. Qu'en serait-il si un questionnaire totalement auto-administré, moins long mais traitant globalement des mêmes problématiques, était proposé aux enquêtés ? Une expérimentation d'enquête auto-administrée, appelée Qualité de la vie au travail (QVT), a ainsi été réalisée entre fin 2012 et début 2013.

La première question qui se pose est celle du taux de réponse obtenu. En effet, les taux de réponse sont généralement beaucoup plus faibles pour les protocoles auto-administrés que pour les enquêtes traditionnelles (Razafindranovona, 2015) mais la situation est peut-être légèrement différente lorsque la cible est une population de personnes en emploi. L'objectif principal de l'exploitation sera ensuite de comparer les réponses obtenues dans l'expérimentation QVT avec celles de CT 2013 et d'essayer d'interpréter les différences constatées. La question classique consiste à se demander si les écarts relèvent plutôt de l'effet de mode lié à la mesure ou de la sélection non contrôlée. Les interprétations ne seront d'ailleurs pas toujours les mêmes selon la nature des variables : certaines questions sont assez factuelles et permettent d'obtenir une description concrète des conditions de travail alors que d'autres sont plus subjectives..

1. Direction de l'animation de la recherche, des études et des statistiques, Service statistique du Ministère du travail, de l'emploi, de la formation professionnelle et du dialogue social.

Chapitre 1

Présentation et objectifs de l'expérimentation Qualité de la vie au travail

L'expérimentation Qualité de la vie au travail (QVT) est une enquête auto-administrée dont le questionnement est largement inspiré de l'enquête Conditions de travail (CT). Elle fait partie du programme d'expérimentations de l'Insee qui vise à accumuler de la connaissance sur les enquêtes ménages réalisées via des protocoles qui offrent la possibilité de répondre par internet. Le contexte général des expérimentations d'enquêtes multimode à l'Insee sera présenté avant de préciser les objectifs et le protocole de l'enquête expérimentale Qualité de la vie au travail.

1.1 Les enquêtes Conditions de travail

Les enquêtes sur les Conditions de travail sont organisées et exploitées par la Dares et collectées par l'Insee depuis 1978. Ces enquêtes visent à cerner, par une description concrète, le travail vécu tel qu'il est perçu par la personne en emploi (Bué *et al.*, 2008). Le questionnaire aborde les horaires et l'organisation du temps de travail, l'organisation et les rythmes de travail, les risques, les pénibilités et leur prévention, les contraintes psychosociales, les relations avec le public, la violence au travail, etc. L'enquête est effectuée à domicile auprès d'un échantillon représentatif de la population de 15 ans ou plus exerçant un emploi.

1.2 L'enquête Conditions de travail 2013

L'enquête Conditions de travail 2013, sixième édition de l'enquête, s'inscrit dans la continuité des enquêtes précédentes avec quelques innovations :

- une extension de l'échantillon afin de couvrir les champs de la fonction publique et le secteur hospitalier (à but lucratif ou non) ;
- une couverture de quatre départements d'Outre-mer ;
- un approfondissement de certaines problématiques via un auto-questionnaire (sous casque) sur l'exposition aux risques psychosociaux et un questionnaire postal auprès des employeurs.

L'enquête Conditions de travail s'inscrit dans un système global d'observation statistique des conditions de travail et des expositions aux risques professionnels, qui s'appuie sur des enquêtes Conditions de travail et des enquêtes Risques psychosociaux (première édition en 2016) en alternance tous les trois ans, avec un suivi des individus en panel.

Ce dispositif est complété par les enquêtes Sumer (Surveillance médicale des risques professionnels), qui sont collectées via les médecins du travail et les médecins de prévention, et fournissent grâce à l'expertise des médecins, une image plus précise des expositions aux risques physiques, chimiques et biologiques.

1.3 Le plan d'expérimentations d'enquêtes multimode à l'Insee

L'utilisation d'Internet comme mode de recueil des données privilégié ou complémentaire est une solution envisagée à plus ou moins long terme par les Instituts nationaux de statistique pour répondre à la demande toujours plus exigeante en termes de qualité et de diversité des enquêtes auprès des ménages, dans un contexte général de restriction budgétaire (de Peretti et Razafindranovona, 2014). En particulier, le recours à la collecte multimode pourrait apparaître comme la solution théorique à mettre en oeuvre que l'on se place dans le cadre de l'erreur d'enquête totale (Groves et Lyberg, 2010) ou dans celui, plus général, de la qualité totale (Lyberg, 2012).

Cependant, si la collecte par Internet est un mode peu coûteux, elle pose des problèmes méthodologiques non négligeables : couverture, auto-sélection ou biais de sélection, non-réponse et les difficultés de sa correction, *satisficing*, etc. Aussi, avant de développer ou généraliser l'utilisation du multimode, l'Insee s'est lancé dans une vaste opération d'expérimentations afin d'étudier ces différentes questions méthodologiques. En particulier, la nécessité d'expérimenter spécifiquement pour chaque type d'enquête auprès des ménages s'est imposée car les résultats de la littérature ne sont pas toujours facilement généralisables.

1.3.1 Protocole standard des expérimentations

Le protocole "standard" utilisé pour les enquêtes expérimentales auto-administrées à l'Insee est le suivant :

- envoi d'une lettre-avis avec les données de connexion à la version en ligne du questionnaire ;
- 1ère relance (3 semaines après) avec envoi d'une lettre contenant les données de connexion, un questionnaire papier ainsi qu'une enveloppe pré-affranchie ;
- 2ème relance (3 semaines après) avec les données de connexion.

La possibilité de répondre sur papier est offerte afin de ne pas exclure d'emblée de l'enquête une partie de la population, d'autant plus quand la réponse est obligatoire, comme dans l'enquête principale en face à face. En revanche, cette possibilité n'est offerte que dans un deuxième temps afin de favoriser la réponse par internet sans que cela n'ait d'incidence négative sur le taux de réponse global.

1.3.2 Enquêtes expérimentales de l'Insee

Une première expérimentation de type multimode incluant une collecte par internet a été réalisée en 2010 et portait sur le thème du logement (Amiel et Denoyelle, 2012). Depuis 2013, plusieurs expérimentations de collecte multimode sur la base du protocole décrit plus haut ont été réalisées ou sont programmées : Vols, violences et sécurité (2013), Logement (2014), Patrimoine (2015), etc.

1.4 L'enquête expérimentale Qualité de la vie au travail

L'enquête expérimentale Qualité de la vie au travail reprend globalement les mêmes thématiques que l'enquête en face à face Conditions de travail en s'appuyant pour l'essentiel sur les mêmes questions. À quelques nouvelles questions près, il s'agit donc d'une version courte du questionnaire de CT, transposé en auto-administré.

L'échantillon de 40 000 individus (37 000 en France métropolitaine et 3 000 à La Réunion) est tiré dans les fichiers fiscaux ; seules les personnes qui déclarent des revenus du travail peuvent être sélectionnés. Pour la moitié de l'échantillon, le protocole mis en oeuvre est le protocole "standard" décrit plus haut avec réponse par internet dans un premier temps, puis par internet ou par retour de questionnaire papier à partir de la relance. Pour l'autre moitié de l'échantillon, seule la réponse par internet est possible tout au long de la période de collecte.

En outre, deux modèles de lettre-avis ont été testés : un premier modèle insiste particulièrement sur les risques psychosociaux et les difficultés au travail alors que l'autre modèle est plus neutre.

1.5 Objectifs de l'expérimentation QVT

L'enquête expérimentale QVT permettra de comparer :

- les taux de réponse entre un protocole avec un seul mode de collecte (internet) et un protocole multimode (internet et papier) ;
- les taux de réponse selon le type de lettre-avis envoyée ;
- les réponses avec celles de CT et de déterminer l'amplitude des effets de mode.

Un objectif complémentaire de l'expérimentation est de tester certaines questions pour l'enquête Risques psychosociaux 2015-2016 et d'exploiter les questions ouvertes qui pourraient déboucher sur des propositions de nouvelles questions ou de modalités pour cette enquête.

Chapitre 2

Bilan sur la collecte de l'expérimentation QVT

Cette partie décrit, dans un premier temps, le déroulement de la collecte. Elle fournit des éléments sur les taux de réponse obtenus en fonction des protocoles expérimentés. De premières exploitations brutes autour du questionnaire sont ensuite présentées avant d'analyser les réponses collectées en termes de qualité. Enfin, une exploitation de l'affectation aléatoire d'un modèle de lettre-avis parmi deux est réalisée.

Dans cette partie, comme tout le reste du document, l'analyse s'appuie uniquement sur le champ métropolitain de l'expérimentation QVT. Quelques éléments sur la collecte de l'expérimentation à la Réunion sont fournis en annexe A.

2.1 Déroulement de la collecte et taux de réponse

2.1.1 Retour d'expérience

La collecte s'est déroulée entre début novembre 2012 et mi-janvier 2013.

Pendant cette période de collecte, environ 500 appels téléphoniques ont été reçus au pôle qui était chargé de l'assistance (pour une durée globale de 35 heures de communication) et 150 messages ont été envoyés sur la boîte fonctionnelle dédiée.

Pour la moitié de l'échantillon concernée, le non-envoi de questionnaires papier est difficile à justifier et à expliquer aux enquêtés qui veulent participer mais n'ont pas internet. Si l'opération a un caractère expérimental, cela n'est pas transparent pour l'enquêté : pour lui le caractère obligatoire de la réponse peut poser problème si le seul mode de réponse proposé est internet.

Certains enquêtés se sont sentis harcelés par les relances : un bon tempo doit être mis en oeuvre pour que les relances n'éveillent pas ce type de ressenti tout en préservant leur efficacité.

2.1.2 Taux de réponse et comparaison avec CT

En fin de collecte, le nombre global de questionnaires collectés pour cette opération est de 15 200, ce qui correspond à un taux de collecte de 41 %. Pour environ 12 % des individus contactés, les courriers reviennent en NPAI¹. En excluant les NPAI, le taux de réponse des individus pour lesquels un contact, via la réception de la lettre-avis, a pu être établi s'élève ainsi à 47 %. Ces taux sont largement plus faibles que pour les enquêtes "traditionnelles" auprès des ménages mais sont plus élevés que les taux généralement constatés pour les opérations auto-administrées, en partie en raison des spécificités de la population ciblée. À titre de comparaison, les taux globaux de réponse sont donnés dans le tableau 2.1 pour l'enquête CT et l'enquête QVT (sur le champ métropolitain). Ce tableau indique quelques spécificités inhérentes aux protocoles. Ainsi, certaines personnes, affectées au protocole n'offrant qu'internet comme mode de réponse, indiquent ne pas pouvoir répondre à QVT car elles ne disposent pas des moyens matériels nécessaires (ordinateur et/ou accès à internet). En revanche la proportion de personnes qui refusent de répondre à QVT est très faible car pour un tel protocole, le refus est forcément explicite et nécessite une action particulière de l'enquêté.

	CT 2013	QVT 2013
Taux de hors champ ² (CT) ou de NPAI (QVT)	9,9 %	11,8 %
Taux de non-réponse	20,3 %	47,0 %
<i>dont Refus</i>	<i>7,7 %</i>	<i>0,8 %</i>
<i>dont Non équipé</i>		<i>2,1 %</i>
Taux de collecte	71,8 %	41,2 %

TABLE 2.1 – Comparaison des taux de réponse de CT et QVT

2.1.3 Web vs Web+papier, quels taux de réponse ?

Dans cette expérimentation, les individus sélectionnés sont répartis aléatoirement selon deux configurations de protocole : pour la moitié des individus, la réponse n'est possible que sur internet, pour l'autre la possibilité de répondre par papier est offerte à partir de la première relance. L'objectif est ainsi d'observer les taux de réponse obtenus dans ces deux configurations : le fait d'offrir la possibilité de répondre en retournant un questionnaire papier augmente-t-il significativement le taux de réponse global ?

Entre le début de la collecte et l'envoi de la première relance, les deux configurations sont tout à fait identiques : il n'est donc pas surprenant que les chiffres présentés dans le tableau 2.2 soient très proches pour les deux configurations. Au bout de 3 semaines de collecte, pour l'une comme pour l'autre configuration, le taux de réponse est d'environ 17 %, les réponses étant exclusivement (ou presque³) fournies par internet.

1. N'habite pas à l'adresse indiquée.

2. Logements vacants ou non identifiés, individus impossibles à joindre, absents de longue durée, etc.

3. Les quelques réponses papier obtenues proviennent de demandes spécifiques d'enquêtés qu'il a été jugé raisonnable d'honorer même si elles déviaient du protocole prévu.

La différence essentielle entre les deux configurations survient à la première relance : pour une partie de l'échantillon, la lettre de relance est accompagnée d'un questionnaire papier accompagné d'une enveloppe pré-affranchie pour le retour. Ainsi, pour les individus relevant de ce protocole, environ 1 600 questionnaires papier sont retournés avant la deuxième relance (table 2.2). Notons que pour cette configuration, lors de la période située entre la première et la deuxième relance, la répartition des réponses reçues est de 40 % sur internet et 60 % par papier.

Offrir un mode de réponse supplémentaire ne peut être *a priori* que bénéfique du point de vue du taux de réponse : l'effet est néanmoins somme toute assez limité puisque le taux de réponse atteint 32 % pour la configuration "internet+papier" contre 29 % pour la configuration "internet seul".

Lors de la deuxième relance, le questionnaire papier n'est pas joint au courrier pour les individus relevant de la configuration "internet+papier". Sur la période allant de la deuxième relance à la fin de la collecte, pour cette configuration, la proportion de retours par papier est un peu moins importante qu'à la période précédente mais reste quand même plus importante que celle des réponses par internet (55 % contre 45 %, table 2.3). Ainsi, à la fin de la collecte, le taux de réponse atteint pour la configuration "internet+papier" est de 45 %, avec une répartition entre internet et papier, de l'ordre de deux tiers / un tiers (table 2.4). Et si l'on exclut les individus pour lesquels le courrier est revenu avec la mention "N'habite pas à l'adresse indiquée", le taux atteint est alors de 50 %. Ces taux sont ainsi supérieurs à ceux de la configuration "internet seul" qui sont de 38 % (NPAI inclus) et de 43 % (NPAI exclus). Une partie de l'écart de taux de réponse entre ces deux configurations s'explique pour des questions d'équipement et d'accès à internet : dans la configuration "internet seul", 4 % des individus enquêtés précisent qu'ils ne vont pas répondre à l'enquête car ils ne disposent pas des moyens matériels pour y répondre.

Le taux de réponse obtenu est donc légèrement supérieur (de l'ordre de 7 points) lorsque l'on offre la possibilité supplémentaire de répondre par papier : ce différentiel peut sembler à première vue relativement faible mais n'est néanmoins pas négligeable compte tenu de la faiblesse des taux considérés. Ainsi ce paramètre peut être intégré dans l'arbitrage entre ces deux types de protocoles qui doit également prendre en compte le coût, la complexité, les délais, la qualité (contenu des réponses, représentativité) ou encore le ressenti de l'utilisateur (qui peut être décontenancé si internet est le seul mode proposé).

	Répondants internet	Répondants papier	NPAI	Pas de réponse	Total
Protocole : internet+papier	3 192	3	1 162	14 143	18 500
Protocole : internet seul	3 152	0	1 244	14 104	18 500

TABLE 2.2 – Retours avant la première relance

2.1.4 Ressenti des répondants sur le questionnement

Quelques questions posées en fin d'interrogation portent sur le ressenti des répondants en ce qui concerne la longueur, la difficulté et la confidentialité du questionnaire. Si 64 % des

	Répondants internet	Répondants papier	NPAI	Pas de réponse	Total
Protocole : internet+papier	4 251	1 591	1 868	10 790	18 500
Protocole : internet seul	5 269	0	1 988	11 243	18 500

TABLE 2.3 – Retours avant la deuxième relance

	Répondants internet	Répondants papier	NPAI	Pas de réponse	Total
Protocole : internet+papier	5 329	2 917	2 122	8 132	18 500
Protocole : internet seul	7 000	2	2 236	9 262	18 500

TABLE 2.4 – Retours à la fin de la collecte

répondants jugent le questionnaire comme étant de longueur raisonnable, la proportion de répondants qui jugent le questionnaire comme trop long est loin d'être négligeable puisqu'elle est d'environ 34 %⁴. Cette proportion de personnes qui trouvent le questionnaire trop long atteint même 42 % chez ceux qui répondent par papier. Pour environ un quart des répondants, les questions sont faciles à comprendre, pour 71 %, les questions sont jugées "normales"⁵.

2.1.5 Caractéristiques des répondants

L'exploitation des variables de sondage nous permet d'observer la relative déformation de l'échantillon tiré initialement suite à la phase de réponse à l'enquête. Cela sera ultérieurement utilisé lors de la correction de la non-réponse totale, mais nous pouvons d'ores et déjà relever quelques points.

Rappelons tout d'abord que les individus sélectionnés dans les fichiers fiscaux sont des individus qui ont déclaré des revenus du travail. Les répondants à QVT sont en moyenne plus âgés que dans l'échantillon et cela est particulièrement marqué pour les répondants sur papier, dont l'âge moyen est de 47,1 ans contre 44,4 ans dans l'échantillon. Pour ce qui est des revenus, si les répondants papier ont des revenus proches de ceux de l'échantillon, les répondants internautes ont des revenus médians bien supérieurs. Enfin, nous pouvons noter la plus grande proportion dans la population de répondants de femmes et de propriétaires de leur logement que dans l'échantillon initial.

Et pour donner quelques brefs éléments de comparaison, dans le volet actifs occupés⁶ de l'enquête Conditions de travail sur le champ métropolitain, la proportion de femmes est de 46 % et l'âge moyen de 41,6 ans.

2.2 Non-réponse partielle, qualité de réponse

2.2.1 Non-réponse partielle

Alors que sur papier, hormis le fait de proposer un questionnaire le plus clair et lisible possible, il n'existe guère de moyens de motiver l'enquêté à répondre à une question, sur internet plusieurs options sont possibles. Des contrôles bloquants peuvent être implémentés afin

4. Et moins de 2 % qui, au contraire, le trouvent trop court.

5. La modalité "difficiles à comprendre" est choisie par environ 4 % des répondants.

6. Il s'agit de personnes en emploi.

	Echantillon	Ech hors NPAI	QVT internet	QVT papier
Age (moyenne)	44,4	45,4	44,7	47,1
Femmes (proportion)	50 %	50 %	52 %	55 %
Propriétaires (proportion)	58 %	64 %	69 %	67 %
Revenus (euros annuels, médiane)	18 500	18 900	21 100	18 200

TABLE 2.5 – Caractéristiques des répondants

Pondérations brutes issues du tirage

de forcer la réponse pour les questions les plus importantes : ils doivent néanmoins être utilisés avec parcimonie car le répondant peut choisir d'abandonner définitivement si ces sollicitations l'exaspèrent. Pour limiter la non-réponse partielle, une possibilité est de ne proposer la modalité de non-réponse que dans un deuxième temps, lorsque l'enquêté tente de sauter la question. C'est ce choix qui a été fait pour le questionnaire en ligne de QVT.

Ainsi, la non-réponse partielle est quasiment inexistante sur les questionnaires en ligne alors qu'une proportion non négligeable de personnes sautent les questions sur papier (table 2.6). Ainsi la non-réponse existe dès les premières questions de l'expérimentation qui portent sur la situation personnelle de l'individu. Si la non-réponse est du même ordre, entre internet et papier, pour la question sur l'année de naissance (aux alentours de 0,5 %), elle est généralement largement supérieure sur papier pour les autres questions : ainsi alors que seuls 0,1 % des individus ne disent pas s'ils sont en couple ou non sur internet, cette proportion monte à 2,4 % sur papier.

	NR QVT web (%)	NR QVT papier (%)
Situation		
Année de naissance	0,4	0,5
Sexe	0,0	0,3
Pays de naissance	0,2	4,2
En couple	0,1	2,4
Enfant	0,1	1,5
Diplôme	0,2	1,2
Situation vis-à-vis du travail	0,0	3,0

TABLE 2.6 – Non-réponse partielle dans les questionnaires web et papier de QVT

2.2.2 Question ouverte

L'un des objectifs de l'expérimentation est d'exploiter une question ouverte en vue de la conception du questionnaire de l'enquête Risques psychosociaux (RPS). L'enquête RPS concrétise les recommandations du Collège d'expertise sur le suivi statistique des risques psychosociaux au travail. Cette enquête s'articule avec l'enquête Conditions de travail (CT) : tous les trois ans, en alternance, aura lieu l'une ou l'autre de ces enquêtes. L'interrogation se fera en panel pendant 9 ans au minimum.

La question ouverte porte sur les choses que l'individu doit faire mais désapprouve. Plus précisément, lorsque l'individu déclare qu'il doit faire, dans le cadre de son travail, des choses qu'il désapprouve, il lui est demandé de citer des exemples de ces situations. Cette question ouverte a été exploitée par la Dares, ce qui a permis d'obtenir une typologie de ces situations, typologie que va ensuite reprendre le questionnement de l'enquête RPS. Ainsi, on peut distinguer les situations où la personne :

- ne peut pas faire du bon travail, doit sacrifier la qualité ;
- doit faire des choses inutiles ou dégradantes ;
- doit mentir à des clients, des usagers, des collègues ;
- doit prendre des risques pour sa santé ;
- doit prendre des risques pour la santé des usagers, des clients, des collègues ;
- est amenée à traiter injustement ou à favoriser des personnes (refus de droits, passe-droits...);
- est amenée à exercer une contrainte sur des personnes ;
- doit faire des choses qui ne sont pas de son ressort.

2.2.3 Qualité des réponses : réponses multiples non prévues

Lors de la phase de saisie des questionnaires papier, une difficulté est remontée : certains répondants ont choisi simultanément plusieurs modalités pour répondre à certaines questions alors que dans ces cas précis, la réponse multiple n'était pas prévue. Nous reviendrons sur ce point lors de la partie dédiée aux imputations, mais pour donner un exemple plus parlant, pour la question suivante, 70 personnes choisissent au moins deux réponses dont 50 personnes qui prennent les modalités 1 et 3.

Q49 Quand, au cours de votre travail, il se produit quelque chose d'anormal, est-ce que...	
1. La plupart du temps, vous réglez personnellement l'incident.....	☐ 1
2. Vous réglez personnellement l'incident mais dans des cas bien précis, prévus d'avance	☐ 2
3. Vous faites généralement appel à d'autres (un supérieur, un collègue, un service spécialisé).....	☐ 3

FIGURE 2.1 – Exemple de question suscitant des réponses multiples

2.3 Affectation aléatoire d'un modèle de lettre-avis

L'un des objectifs de l'expérimentation est de savoir quel type d'impact peut avoir la rédaction de la lettre-avis envoyée aux enquêtés sur le taux de réponse, les caractéristiques des répondants et la nature de leurs réponses. Ainsi, deux modèles de lettre-avis ont été rédigés et le protocole de l'expérimentation prévoit une affectation aléatoire de l'un de ces deux modèles aux enquêtés.

2.3.1 Les deux modèles de lettre-avis

Alors que le modèle A de lettre-avis est plutôt neutre, le modèle B insiste sur les risques psycho-sociaux qui sont associés au travail.

Modèle A : "L'Institut national de la statistique et des études économiques (Insee) réalise une étude sur la qualité de la vie au travail en France. Cette enquête permet d'obtenir une description concrète du travail et de son organisation, selon divers angles : horaires, rythmes, ambiance au travail, ..."

Modèle B : "L'Institut national de la statistique et des études économiques (Insee) réalise une étude sur la qualité de vie et les risques psychosociaux au travail en France. Cette enquête permet d'obtenir une description du travail, de son organisation et des risques psychosociaux qui lui sont parfois associés (stress, manque de reconnaissance, insécurité, mal être au travail ...)."

2.3.2 Taux de réponse

Les taux de réponse globaux pour les deux sous-échantillons déterminés par le type de lettre envoyée sont très proches : 41,3 % pour la lettre A contre 41,1 % pour la lettre B. Dans le détail, quelques petites différences existent : dans le protocole qui prévoit une réponse possible soit par internet, soit par papier, les enquêtés affectés au modèle A répondent un peu plus sur internet alors que ceux affectés au modèle B répondent un peu plus sur papier.

Taux de réponse	Lettre A	Lettre B
Protocole I+P Internet	29,7 %	27,9 %
Protocole I+P Papier	15,2 %	16,4 %
Protocole I+P Ensemble	44,9 %	44,3 %
Protocole I Internet	37,8 %	37,9 %
Ensemble QVT	41,3 %	41,1 %

TABLE 2.7 – Taux de réponse selon le modèle de lettre

2.3.3 Caractéristiques des répondants

Du point de vue des caractéristiques des répondants, il y a également peu de différences selon le type de lettre reçue (table 2.8). Au vu de tests du χ^2 réalisés, l'indépendance est manifeste entre le type de lettre reçue et l'âge, le sexe, la nationalité, le diplôme ou encore le groupe social⁷.

7. Les p-values des tests du χ^2 entre le type de lettre reçue et l'âge, le sexe, la nationalité, le diplôme et le groupe social sont respectivement de 0,26 , 0,28 , 0,49 , 0,90 et 0,76.

	Lettre A	Lettre B
Age		
Moins de 30 ans (%)	9,1	8,6
30-34 ans (%)	12,0	11,4
35-39 ans (%)	13,9	13,9
40-49 ans (%)	32,9	31,8
50-59 ans (%)	28,1	29,7
60 ans et plus (%)	4,1	4,5
Sexe		
Femme (%)	52,4	53,3
Homme (%)	47,6	46,7
Nationalité		
Français (%)	96,7	96,9
Etranger (%)	3,3	3,1
Diplôme		
Aucun diplôme (%)	7,5	7,1
CEP, Brevet (%)	5,7	5,8
BEP, CAP (%)	21,5	21,9
Baccalauréat (%)	18,5	18,7
Supérieur (%)	46,8	46,6
CS		
Agriculteurs exploitants (%)	0,4	0,3
Artisans, commerçants et chefs d'entreprise (%)	3,9	3,6
Cadres et professions intellectuelles supérieures (%)	24,5	24,0
Professions intermédiaires (%)	29,5	30,2
Employés (%)	27,1	26,9
Ouvriers (%)	14,6	15,0

TABLE 2.8 – Caractéristiques des répondants selon la lettre-avis reçue

2.3.4 Comparaison des réponses selon la lettre reçue

Lorsque les réponses recueillies sont comparées selon le modèle de lettre reçue, les différences sont globalement assez minimes. En déroulant le questionnaire et en observant les p-values des test du χ^2 entre le modèle de lettre et les différentes variables, quelques écarts (non pondérés) peuvent néanmoins être observés.

Tout d'abord, les réponses à la question sur la santé ressentie sont légèrement différentes. En effet, la proportion de personnes à se déclarer en bonne ou en très bonne santé est inférieure d'environ 2 points (58,5 % contre 60,3 %) pour ceux qui reçoivent le modèle du type B. Ces personnes déclarent aussi plus fréquemment qu'elles sont en contact avec des gens en situation de détresse (45,1 % contre 43,6 %). Celles qui reçoivent le modèle B sont également plus pessimistes quant à leurs relations avec les autres au travail : elles sont ainsi plus souvent pas d'accord ou pas du tout d'accord avec les affirmations "Mon supérieur prête attention à ce que je dis" (22,0 % contre 20,6 %), "Les collègues avec qui je travaille m'aident à mener mes

tâches à bien" (12,6 % contre 11,1 %) et "Vu tous mes efforts, je reçois le respect et l'estime que mérite mon travail" (38,1 % contre 35,5 %). Elles déclarent aussi plus souvent qu'elles doivent toujours éviter de donner leur avis (6,9 % contre 5,7 %). Enfin, elles reportent plus souvent des agressions physiques ou sexuelles de la part du public (2,4 % contre 1,8 %) ainsi que des agressions verbales de la part de collègues ou supérieurs (18,3 % contre 15,9 %).

Ainsi, si les écarts sont globalement limités, il existe néanmoins quelques questions pour lesquelles le modèle de lettre-avis a probablement eu un certain impact. La question est de savoir si le modèle de lettre a plutôt sélectionné des personnes différentes, et plus précisément des personnes qui connaissent des difficultés plus importantes dans leur travail pour celles qui reçoivent le modèle B ou si la lettre place la personne dans un contexte différent qui fait qu'à caractéristiques équivalentes les réponses reportées ne sont pas similaires. Au vu des écarts qui interpellent quelque peu sur les agressions qui sont des questions assez factuelles, l'hypothèse d'une sélection différenciée est peut-être à privilégier.

Chapitre 3

Correction de la non-réponse totale, apurements et imputations

Cette partie décrit les premiers traitements post-collecte réalisés pour l'expérimentation Qualité de la vie au travail. Pour tenir compte des mécanismes de sélection des répondants, une correction de la non-réponse totale est réalisée en deux étapes (réception du courrier de l'Insee puis participation à l'enquête) à partir des variables de la base de sondage. Quelques brefs éléments sont également fournis sur l'apurement et les imputations réalisées. Ces étapes interviennent avant la comparaison et le rapprochement avec l'enquête Conditions de travail par contrôle spécifique de la sélection, qui seront présentés dans le chapitre suivant.

3.1 Correction de la non-réponse totale

Les éléments fournis dans la partie précédente sur les caractéristiques des répondants indiquent une nécessité de corriger la déformation de l'échantillon dès lors que l'on souhaite inférer sur notre population d'intérêt. Cette étape de correction de la non-réponse totale est indispensable pour corriger la déformation de l'échantillon en utilisant les données de la base de sondage dont on dispose. Cette correction constitue également un préalable nécessaire pour comparer *in fine* les estimations issues de QVT avec celles de CT, dont les données ont également été redressées.

On choisit d'effectuer la correction de la non-réponse totale en deux phases :

- une première phase modélise le fait de recevoir ou non le courrier invitant à répondre à l'enquête ;
- une deuxième phase modélise le fait, une fois le courrier reçu, de répondre ou non à l'enquête.

Cette manière de procéder permet de bien distinguer ces deux mécanismes qui ne mettent pas forcément en jeu les mêmes phénomènes. En effet, ce sont plutôt des facteurs liés à la mobilité / stabilité résidentielle qui jouent pour la phase de réception du courrier alors que ce sont des déterminants plus classiques de la non-réponse (revenus par exemple) qui importent lors de la deuxième phase.

3.1.1 Première phase de correction : recevoir ou non la lettre-avis

Après une analyse préliminaire, à partir des variables de la base de sondage dont on dispose, on retient comme variables, pour modéliser le fait de recevoir ou non la lettre-avis :

- l'âge ;
- le statut d'occupation (locataire ou propriétaire) ;
- le type de logement (appartement ou maison) ;
- le fait d'habiter en région PACA (un nombre assez conséquent de courriers sont revenus en NPAI pour cette région).

On effectue une régression logistique avec comme variable à expliquer le fait d'avoir reçu la lettre-avis. Les odds ratios de cette régression sont présentés dans le tableau 3.1.

	Odds ratios
Tranche d'âge	
Moins de 30 ans vs 40-49 ans	0,35
30-34 ans vs 40-49 ans	0,50
35-39 ans vs 40-49 ans	0,75
50-59 ans vs 40-49 ans	1,57
60 ans et plus vs 40-49 ans	1,71
Statut d'occupation	
Locataire vs Propriétaire	0,23
Type de logement	
Appartement vs Maison	0,66
Région de résidence : PACA	
Oui vs Non	0,61

TABLE 3.1 – Modélisation de la réception de la lettre-avis

Une fois cette régression effectuée, nous récupérons les probabilités estimées $\hat{p}_{1k}$ à partir de ce modèle de réception du courrier. L'utilisation directe des probabilités estimées pouvant poser des problèmes de robustesse, nous constituons H ¹ classes d'individus à partir de ces estimations et nous calculons les proportions observées de réception de courrier au sein de ces classes : pour la classe h , cette proportion observée est $\hat{p}_1(h)$. Un individu de l'échantillon, ayant comme poids initial w_k et appartenant à la classe h se voit attribuer le poids $w_{k(h)}$ à l'issue de cette première phase, avec :

$$w_{k(h)} = w_k \frac{1}{\hat{p}_{1k}(h)}$$

3.1.2 Deuxième phase de correction : répondre ou non à l'enquête

Pour la deuxième phase de correction, notre analyse préliminaire consiste à observer les fréquences de réponses parmi ceux ayant reçu la lettre-avis. Il en ressort que les variables que nous allons retenir pour l'analyse sont :

1. On prend ici $H = 10$.

- le sexe ;
- le revenu d'activité déclaré (quintiles) ;
- le fait d'être âgé de plus de 60 ans ;
- le statut d'occupation (locataire ou propriétaire).

On effectue la régression logistique sur le champ des individus ayant reçu la lettre-avis et avec comme variable à expliquer le fait de répondre à l'enquête, que cela soit par internet ou en retournant le questionnaire papier. Les odds ratios de l'estimation sont donnés dans le tableau 3.2.

	Odds ratios
Sexe	
Femme vs Homme	1,41
Revenus	
Quintile 1 vs Quintile 3	0,71
Quintile 2 vs Quintile 3	0,85
Quintile 4 vs Quintile 3	1,26
Quintile 5 vs Quintile 3	1,51
Statut d'occupation	
Locataire vs Propriétaire	0,74
60 ans et plus	
Oui vs Non	0,81

TABLE 3.2 – Modélisation de la phase de réponse à l'enquête

On déroule le même type de traitements qu'à la phase précédente, et ayant construit L^2 classes d'individus lors de cette étape, le poids à l'issue de la correction de la non-réponse totale est au final, pour un individu répondant à l'enquête, appartenant aux classes h et l :

$$w_{k(h,l)} = w_k \frac{1}{\hat{p}_1(h)} \frac{1}{\hat{p}_2(l)}$$

Ainsi, comme l'estimateur

$$\hat{Y} = \sum_{k \in s} \frac{y_k}{\pi_k p_k}$$

(avec p_k la probabilité de réponse de l'individu k) est sans biais pour Y , Y est estimé par

$$\hat{Y}_{CNR} = \sum_{k \in s} \frac{y_k w_k}{\hat{p}_k}$$

(asymptotiquement sans biais), avec

$$\hat{p}_k = \hat{p}_1(h) \hat{p}_2(l)$$

2. Pour cette phase aussi, nous avons choisi $L=10$.

3.2 Apurement et imputations

Les principaux traitements généraux de la phase d'apurement ont consisté à supprimer les questionnaires trop lacunaires et regarder la cohérence entre informations de la base de sondage et réponses au questionnaire avec correction si nécessaire.

Le traitement post collecte des questionnaires papier a été en revanche plus fastidieux. Une saisie brute de toute l'information reportée avait été demandée afin de faciliter le travail de saisie et de laisser le choix de la méthode d'imputation lors de l'exploitation. Ainsi, les traitements spécifiques aux questionnaires papier ont consisté à :

- corriger certaines valeurs numériques qui ne correspondaient pas parfaitement au format attendu (par exemple, environ 150 corrections de ce type sur le temps de travail déclaré) ;
- corriger ou imputer certaines valeurs en tenant compte des remarques écrites par les répondants sur le questionnaire (une centaine de corrections manuelles ont été effectuées) ;
- imputer une réponse lorsque plusieurs modalités étaient choisies alors qu'une seule était attendue : dans certains cas (une cinquantaine), le choix s'imposait assez facilement entre les modalités donc l'imputation était déterministe, mais dans le cas général (un millier de cas à traiter), un tirage uniforme est effectué entre les différentes modalités.

Chapitre 4

Comparaison entre CT et QVT : des effets de mode ?

L'objet principal de la comparaison des données provenant de modes de collecte différents est de savoir si les réponses des enquêtés diffèrent ou non selon les modes. Par définition, les effets de mode sont les différences dans les réponses obtenues selon les modes de collecte. Si on peut les appréhender de manière brute, les effets de mode se réfèrent plus spécifiquement à l'erreur de mesure : toutes choses égales par ailleurs, donc à populations "comparables", quelles sont les différences liées aux différents instruments de mesure ? On cherche ainsi à savoir si les personnes répondent de la même manière lorsqu'elles sont confrontées aux différents modes de collecte.

4.1 Problématique, éléments théoriques et pratiques sur la mesure des effets de mode

4.1.1 Problématique

Ici la question qui nous intéresse particulièrement est de savoir si les individus répondent de la même manière sur leurs conditions de travail lorsqu'ils sont interrogés en face à face, via l'intermédiation d'un enquêteur, et lorsque le questionnaire est auto-administré, sur internet ou sur papier. Ce type de comparaison se justifie tout particulièrement car, en termes d'effets de mode, la dichotomie majeure entre les modes de collecte est celle qui distingue les modes auto-administrés de ceux où l'intermédiation se fait via un enquêteur.

Aussi, ces analyses pourront être déclinées assez finement car les processus cognitifs en jeu ne sont pas les mêmes selon la nature des questions posées : par exemple, alors que certaines questions visent à décrire concrètement les conditions de travail, d'autres relèvent plus de la question subjective.

L'analyse va confronter d'un côté les données de l'expérimentation Qualité de la vie au travail et de l'autre, les données de l'enquête Conditions de travail. Le regroupement des observations issues du web et du papier de QVT s'explique par le caractère auto-administré de ces modes, présentés souvent comme similaires dans la littérature. Quelques éléments de comparaison entre réponses sur internet et réponses sur papier sont néanmoins présentés en annexe B.

4.1.2 Éléments théoriques sur la mesure des effets de mode

Pour analyser les données collectées via différents modes, une première manière de procéder est de comparer directement les estimations de paramètres d'intérêt, par exemple les moyennes d'une variable quantitative ou les proportions pour une variable qualitative. Cette comparaison directe permet de mesurer l'ampleur des effets de mode au sens le plus global du terme mais ne permet pas de quantifier ce qui nous intéresse le plus, à savoir les effets de mode liés à la mesure. Le souci de ce type de comparaison directe est que ce ne sont pas forcément les mêmes types de personnes qui répondent via les différents modes de collecte étudiés : les différences constatées peuvent être largement induites par la sélection des individus. Ce genre d'analyse brute ne permet pas ainsi de distinguer l'erreur de mesure des erreurs liées à la sélection, puisque les différents effets sont confondus.

Pour essayer de distinguer les effets de mode liés à la mesure des effets de sélection, une solution est de rendre les différents répondants "comparables" en contrôlant par un ensemble de caractéristiques observables Z , pas ou très peu sensibles au mode de collecte.

Schématiquement, les différences brutes constatées dans les réponses constituent l'effet de mode global et les différences qui demeurent après un contrôle par des caractéristiques observables reflèteraient, sous certaines hypothèses, les effets de mode liés à la mesure, donc purgés des effets de sélection (Razafindranovona, 2015).

4.1.3 Intérêt et limites des comparaisons naïves

L'objectif majeur de l'expérimentation est de comparer les estimations issues de QVT avec celles obtenues dans l'enquête "principale" CT. Si l'objectif final est d'obtenir des résultats les plus robustes possibles en termes d'effet de mode, donc en contrôlant au mieux ce qui relève d'autres phénomènes (comme par exemple la sélection), il n'est néanmoins pas inintéressant de procéder à de premières comparaisons à partir des données brutes de QVT. Dans quelle mesure les estimations à partir de ces données brutes sont-elles proches ou éloignées de celles de CT ? Dans le cas où l'on constaterait de très fortes différences sur des questions bien spécifiques, il est probable que ces écarts persistent y compris après un contrôle de la sélection (que nous allons réaliser plus tard) et que l'on soit bien en présence de ce qu'on peut appeler, au sens large, un effet de mode. On fournira ainsi, pour information, les estimations brutes issues de QVT avant toute correction, dans les tableaux synthétiques de comparaisons des paramètres d'intérêt.

4.1.4 Le contrôle de la sélection par le calage sur marges (après correction de la non-réponse préalable)

Le contrôle de la sélection est appelé *back-door model* dans certains papiers de la littérature sur les effets de mode (Vannieuwenhuyze *et al.*, 2014). On peut par exemple utiliser les techniques économétriques qui visent à évaluer l'effet d'un traitement, comme les modèles de régression ou les méthodes de matching. Une autre possibilité est d'essayer de rendre comparables les populations par des techniques de calage, ce qui s'inscrit davantage dans l'optique traditionnelle du traitement aval des données d'enquête avec fourniture, *in fine*, d'un jeu de

pondérations.

Pour comparer CT et QVT, le choix d'un contrôle de la sélection par calage sur marges a été fait. Non seulement car, comme évoqué, cela s'inscrit bien dans un processus général de repondération mais aussi parce que les variables d'intérêt sont nombreuses¹ ce qui rend assez périlleux l'évaluation d'effets via des modélisations potentiellement spécifiques pour chacune des variables en question. Il ne s'agit pas ici à proprement parler d'une évaluation d'effet d'un traitement reposant sur un calage comme cela est fait dans la méthode de l'*entropy balancing* (Hainmueller, 2012), mais plutôt d'un traitement classique de données post-collecte sans *a priori* sur les variables d'intérêt à étudier.

4.2 Le calage sur marges de QVT pour contrôler la sélection

4.2.1 Utilisation de marges sociodémographiques communes à CT

Dorénavant, le champ est restreint à celui des personnes en emploi.

Les marges, calculées à partir des effectifs de CT, sont :

- le sexe ;
- l'âge ;
- le diplôme ;
- la nationalité ;
- la PCS (niveau 1 ou groupe social).

Il s'agit de variables qui ont pour la plupart été utilisées lors du redressement de CT, soit pour le calage, soit lors du traitement de la non-réponse totale. Le calage sur ces variables semble licite dans la mesure où les effets de mode ne devraient pas être particulièrement marqués pour de telles questions. De plus, il s'agit de variables qui sont *a priori* corrélées aux variables d'intérêt donc potentiellement pertinentes pour corriger au moins une partie des différences liées à la sélection.

L'ajustement à partir de ces variables de calage permet ainsi de contrôler une partie des différences entre les deux populations liées à la sélection. En effet, le caractère auto-administré du protocole de QVT peut écarter certaines sous-populations, comme, par exemple, les personnes les moins à l'aise avec l'écrit. Le calage à partir de ces marges, conjugué à l'effet de la correction de la non-réponse totale, devrait ainsi permettre de rendre les populations de CT et QVT "comparables", en termes de principales caractéristiques sociodémographiques ou encore de niveau général de littératie.

4.2.2 Correction de la déformation

Notons d'ailleurs que si le traitement de la non-réponse totale a ses objectifs propres de correction de la déformation de l'échantillon initial, il participe également dans une certaine mesure au contrôle de la sélection. En effet, il peut rapprocher, en jouant sur les pondérations,

1. Toutes les variables de l'enquête sont potentiellement variables d'intérêt lorsque l'objectif de comparaison n'est pas complètement circonscrit.

la population de répondants QVT d'une certaine population de référence (celle échantillonnée) qui elle-même serait proche de la population redressée de CT.

Ainsi, avant le calage sur marges, la correction de la non-réponse totale contribue à corriger la sous-représentation des ouvriers et la sur-représentation des cadres² parmi les répondants à QVT.

	B	CNR	C1	C2
Agriculteurs exploitants (%)	0,4	0,4	0,4	1,8
Artisans, commerçants et chefs d'entreprise (%)	3,7	3,8	3,7	6,3
Cadres et professions intellectuelles supérieures (%)	24,3	21,9	19,6	19,2
Professions intermédiaires (%)	29,8	28,5	27,5	24,8
Employés (%)	27,0	28,1	30,5	27,6
Ouvriers (%)	14,8	17,2	18,3	20,5

TABLE 4.1 – Exemple d'effets des corrections

B : pondérations brutes, CNR : après CNR totale, C1 : +calage sexe, âge, diplôme et nationalité, C2 : +calage sexe, âge, diplôme, nationalité et CS (niveau 1)

4.2.3 Utilisation du score de bien-être

Présentation du score de bien-être

Une variable que nous allons aussi intégrer dans le contrôle de la sélection est un score de bien-être. Ce score est imputé à partir de plusieurs questions posées dans CT et QVT. Ces questions et le score qui en découle ne sont pas des variables d'intérêt en tant que telles de l'enquête mais plutôt des covariables qui permettent d'expliquer les phénomènes étudiés.

Le score de bien être "WHO-5" a été développé par l'Organisation Mondiale de la Santé (Topp *et al.*, 2015). Il permet de capter, dans une certaine mesure, le bien-être émotionnel de l'individu à partir des réponses à une batterie de questions : l'individu doit dire à quelle fréquence (tout le temps, la plupart du temps, plus de la moitié du temps, moins de la moitié du temps, de temps en temps ou jamais), il a ressenti les affirmations suivantes au cours des deux dernières semaines :

- Je me suis senti bien et de bonne humeur ;
- Je me suis senti calme et tranquille ;
- Je me suis senti plein d'énergie et vigoureux ;
- Je me suis réveillé en me sentant frais et dispos ;
- Ma vie quotidienne a été remplie de choses intéressantes.

Pour calculer ce score, pour chacune des 5 questions, on impute 5 points lorsque la modalité choisie est "tout le temps", 4 points pour "la plupart du temps" ... et 0 point pour "jamais".

2. Par rapport à une distribution des CS issue de CT dont les marges sont utilisées dans le calage C2 du tableau 4.3. : parmi quelques autres différences constatées entre CT et QVT, on peut citer la sur-représentation des diplômés du supérieur dans QVT (47 % en pondération brute, 45 % après CNR contre 41 % dans CT) ou encore la sous-représentation des étrangers (3 % en pondération brute, 4 % après CNR contre 5 % dans CT)

Le total de points est ensuite multiplié par 4 pour obtenir un score entre 0 et 100. Ce score de bien-être est en moyenne, moins élevé de 5 points dans QVT (57,7 après calage sur marges sociodémographiques) que dans CT (62,7). Un écart qui reste malgré tout à relativiser au vu de la construction du score : deux personnes répondant de façon identique sauf à une question (donc 4 questions sur 5) où l'une choisit "tout le temps" et l'autre "la plupart du temps" auraient 4 points d'écart. Les paramètres de distribution du score de bien-être de QVT et CT sont donnés dans le tableau 4.1.

	QVT	CT
P10	24	32
P25	44	52
P50	60	68
P75	76	76
P90	80	84

TABLE 4.2 – Distribution du score de bien-être dans QVT et CT

Justification théorique de l'utilisation du score de bien-être

L'utilisation du score de bien-être dans le contrôle de la sélection semble pertinent pour plusieurs raisons. Une première tient au mode de collecte : cette variable est collectée de manière auto-administrée à la fois dans CT et dans QVT. Par ailleurs cette variable participe probablement au processus de sélection et semble également corrélée aux variables d'intérêt.

Rappelons en effet que l'objectif de la comparaison entre les données de CT et de QVT est de mesurer l'effet de mode intrinsèquement lié à la mesure. Pour cela, on cherche à rendre les populations les plus comparables possibles pour essayer de purger les différences liées à la sélection. Une partie de ces différences ont ainsi été corrigées par le traitement de la non-réponse totale et le calage sur marges sociodémographiques. Mais cela n'est sans doute pas suffisant comme contrôle de la sélection pour en déduire l'ampleur de l'effet de mode.

Une idée pour améliorer le contrôle de la sélection serait d'utiliser la partie auto-administrée du questionnaire de l'enquête CT. En effet, si tout le questionnement de QVT est auto-administré, c'est également le cas d'une petite partie de CT via un protocole sous casque afin d'obtenir les réponses les plus sincères possibles. Cette partie auto-administrée pose des questions subjectives sur les conditions de travail, interroge l'enquêté sur les relations avec les autres, les perspectives, les satisfactions et difficultés et enfin aborde l'état d'esprit, le bien-être dans la vie au quotidien. Si l'on considère que le questionnement auto-administré est similaire dans les deux enquêtes et donc qu'une personne répondrait de la même manière lors du questionnement CT que lors du questionnement QVT, alors les différences constatées dans les réponses aux questions communes et auto-administrées de CT et QVT relèveraient uniquement d'effets de sélection. Malheureusement, cette similarité entre les deux questionnements est assez théorique : les modes de collecte ne sont pas tout à fait les mêmes mais surtout, le contexte est différent. En effet pour CT, la partie du questionnaire auto-administrée vient après une longue phase de questionnement administré par un enquêteur alors que pour QVT, la partie ne se distingue pas du reste du questionnaire qui est totalement auto-administré (sur papier ou sur internet).

Même si l'hypothèse selon laquelle les différences entre CT et QVT sur la partie auto-administrée relèveraient uniquement d'effets de sélection n'est sans doute pas vérifiée, intégrer certaines des variables de cette partie dans le calage devrait contribuer à rapprocher les populations que l'on souhaite comparer. En effet, ces variables pourraient apporter une vraie plus-value dans le contrôle de la sélection par les thématiques abordées et le type de questionnement. Il reste donc à faire un choix parmi les variables de cette partie : certaines pouvant être considérées comme des variables d'intérêt à part entière, il n'est sans doute pas raisonnable de toutes les intégrer.

Dans la partie auto-administrée de CT, une bonne variable candidate pour intégrer le contrôle de la sélection est une variable qui participe au processus de sélection des individus dans les deux populations à comparer et qui est corrélée aux variables d'intérêt.

Lorsque le taux de réponse à une enquête est faible, il est possible que la population qui participe soit fortement sélectionnée : dans le cas de QVT, on peut craindre que les personnes qui répondent sont plutôt celles qui ont des choses à dire sur leurs conditions de travail parce qu'elles en sont mécontentes. Mais il n'est pas évident de disposer de variables permettant de corriger directement un tel effet.

Pour corriger les effets de sélection, des variables autour du rapport des individus aux enquêtes pourraient être mises à contribution, comme leur participation passée ou présente à des enquêtes de la statistique publique voire même à des enquêtes à caractère commercial (sondages d'opinion, panel de volontaires, etc.) : les individus ne sont pas tous égaux dans leur volonté intrinsèque de participer aux enquêtes et si cette volonté est liée aux variables d'intérêt, il faudrait, dans l'idéal pouvoir l'intégrer dans la correction.

Enfin, lorsque l'on suppose que la population fortement sélectionnée possède des caractéristiques spécifiques quant à son état d'esprit, ses opinions, ses attitudes, des questions assez généralistes qui synthétiseraient cela pourraient être utilisées dans la correction. Ainsi, si la population qui répond à une enquête est globalement peu satisfaite de sa vie au quotidien, cela pourrait être corrigé à partir de données de référence. Ces données de référence pourraient provenir d'une source externe, ou bien, comme ici, directement de l'enquête que l'on cherche à comparer, à savoir CT.

4.2.4 Une marge externe supplémentaire : la santé ressentie

La santé ressentie est une autre variable que l'on souhaite intégrer dans le calage pour contrôler la sélection. Elle se différencie assez nettement entre les deux enquêtes (voir tableau 4.2, la pondération utilisée pour QVT est celle après calage sur marges sociodémographiques) et semble assez pertinente en tant que variable pouvant à la fois participer au processus de sélection et corrélée aux variables d'intérêt. La limite de cette variable est qu'elle n'est pas collectée de la même manière dans les deux enquêtes (auto-administré dans QVT, avec enquêteur dans CT) et qu'il s'agit d'une variable particulièrement sensible aux conditions de collecte (Clark et Vicard, 2007) qui peut être potentiellement sujette aux effets de mode liés à la mesure. Cela peut sembler réducteur dans le cadre d'une utilisation de contrôle de la sélection pour mesurer des effets de mode liés à la mesure, mais, malgré, tout, nous choisissons d'intégrer cette variable, au moins pour analyser son apport potentiel dans le contrôle.

Une autre variable que l'on a souhaité dans un premier temps intégrer dans le contrôle de la

	QVT (%)	CT (%)
Très bon	16,1	26,3
Bon	44,2	50,4
Assez bon	31,1	18,8
Mauvais	7,5	4,0
Très mauvais	1,0	0,6

TABLE 4.3 – État de santé (ressentie) dans QVT et CT

sélection est la question relative au fait d'être syndiqué ou sympathisant d'un syndicat. En effet, il y a un peu plus d'adhérents et surtout de sympathisants dans QVT que dans CT. Le calage intégrant cette marge supplémentaire a été réalisé mais ne s'est pas avéré concluant. Nous ne présentons donc pas les résultats relatifs à cette variable supplémentaire.

4.3 Résultats des comparaisons entre CT et QVT

Plusieurs calages sur marges (Sautory, 1993) successifs ont ainsi été effectués. Pour rappel du principe général, étant donné une certaine variable d'intérêt Y et connaissant des totaux X_j sur l'ensemble de la population, on cherche à obtenir un estimateur de la forme

$$\hat{Y}_w = \sum_{k \in s} w_k y_k$$

où les poids w_k sont "proches" des pondérations initiales d_k (ici issues de l'étape de correction de la non-réponse) et vérifient des équations de calage de type

$$\forall j = 1 \dots J \quad \sum_{k \in s} w_k x_{jk} = X_j \quad (4.1)$$

On va alors chercher un système de poids w_k solution de

$$\min_{w_k} \sum_{k \in s} d_k \quad G\left(\frac{w_k}{d_k}\right)$$

où G est une fonction de distance, sous les contraintes des équations de calage (4.1).

(w_{k1}) tels que

$$\min_{w_{k1}} \sum_{k \in s} d_k \quad G\left(\frac{w_{k1}}{d_k}\right) \quad sc \quad \forall j = 1 \dots J \quad \sum_{k \in s} w_{k1} x_{jk} = X_j$$

(w_{k2}) tels que

$$\min_{w_{k2}} \sum_{k \in s} d_k \quad G\left(\frac{w_{k2}}{d_k}\right) \quad sc \quad \forall j = 1 \dots J \quad \sum_{k \in s} w_{k2} u_{jk} = U_j$$

(w_{k3}) tels que

$$\min_{w_{k3}} \sum_{k \in s} d_k \quad G\left(\frac{w_{k3}}{d_k}\right) \quad sc \quad \forall j = 1 \dots J \quad \sum_{k \in s} w_{k3} v_{jk} = V_j$$

(w_{k4}) tels que

$$\min_{w_{k4}} \sum_{k \in s} d_k G\left(\frac{w_{k4}}{d_k}\right) \quad \text{s.c.} \quad \forall j = 1 \dots J \quad \sum_{k \in s} w_{k4} Y_{jk} = Y_j$$

avec

- X =[sexe, âge, diplôme, nationalité],
- U =[X , CS (niveau 1)],
- V =[U , score de bien-être],
- Y =[V , santé ressentie].

Dans les tableaux suivants, la colonne B repose sur les pondérations brutes de tirage et la colonne CNR sur les pondérations issues de la correction de la non-réponse totale. Les calages C1, C2, C3 et C4 sont ceux liés aux marges respectives X , U , V , et Y .

Les calages sont effectués selon la méthode du raking ratio. Les rapports de poids (par rapport aux pondérations initiales que sont les poids après correction de la non-réponse totale) sont donnés dans le tableau 4.4 : il est normal que les poids se dispersent plus au fur et à mesure que l'on rajoute des marges de calage, mais cette dispersion semble plutôt raisonnable.

	C1	C2	C3	C4
P10	0,76	0,67	0,56	0,43
P25	0,79	0,75	0,70	0,56
P50	0,88	0,86	0,89	0,87
P75	1,12	1,09	1,14	1,22
P90	1,32	1,40	1,46	1,59

TABLE 4.4 – Rapports de poids après calages

C1 : +calage sexe, âge, diplôme et nationalité, C2 : + marge CS (niveau 1), C3 : + marge score de bien-être, C4 : + marge santé ressentie

4.3.1 Comparaisons détaillées par modules

Dans cette partie, on décline de manière linéaire les différents modules du questionnaire et on observe l'impact des différents calages sur les variables d'intérêt.

Divers : arrêts, maladie, accidents

Pour la question sur les arrêts maladie et sur le fait de venir au travail alors qu'on est malade, les réponses de CT et QVT, initialement pas trop éloignées, sont proches à l'issue du calage intégrant le score de bien-être. Et alors que la proportions sont légèrement plus élevées dans QVT que dans CT elles deviennent, après calage sur la santé plus faibles dans QVT que dans CT.

Pour ce qui est des accidents de travail, le calage sur la santé semble être pertinent pour rapprocher les estimations des deux enquêtes.

	B	CNR	C1	C2	C3	C4	CT
Divers							
Au moins un arrêt maladie (%)	29,5	29,8	30,3	29,5	28,0	25,8	27,1
Au moins un accident au travail (%)	9,2	10,1	10,6	10,7	10,1	9,1	8,5
Malade mais au travail (%)	45,6	45,7	45,8	45,6	42,3	39,2	41,1

TABLE 4.5 – Arrêts, maladie

B : pondérations brutes, CNR : après CNR totale, C1 : +calage sexe, âge, diplôme et nationalité, C2 : + marge CS (niveau 1), C3 : + marge score de bien-être, C4 : + marge santé ressentie, CT : enquête Conditions de travail après redressements

Temps de travail

	B	CNR	C1	C2	C3	C4	CT
Temps de travail							
Travaille à temps partiel (%)	17,7	17,7	18,4	18,0	18,3	18,3	19,0
Travaille la nuit (%)	20,0	20,7	20,6	21,1	20,9	20,5	16,4
Travaille le dimanche (%)	36,7	37,1	36,7	37,5	37,4	37,2	31,7
Temps de travail hebdomadaire (moyen)	37,7	37,5	37,2	37,6	37,4	37,4	38,0

TABLE 4.6 – Temps de travail

Il y a un peu plus de personnes dans QVT qui déclarent avoir des horaires atypiques : 37 % travaillent, habituellement ou occasionnellement, le dimanche (contre 32 % dans CT), 20 % travaillent la nuit (contre 16 % dans CT). Pour ces questions, l'effet des différents calages est très limité et ne rapproche pas vraiment les estimations. Pour le travail le dimanche, une possibilité est que la question n'est pas tout à fait comprise de la même manière : par exemple, dans CT, une consigne de la question sur le dimanche est que cela exclut le travail ramené à la maison. Les proportions de travailleurs à temps partiel sont très proches entre CT et QVT et ne sont pas trop modifiées au cours des différents calages. Il en est de même pour le temps de travail hebdomadaire moyen.

	B	CNR	C1	C2	C3	C4	CT
Conciliation							
Très bien (%)	20,6	20,6	20,7	20,4	22,4	23,9	35,3
Bien (%)	52,9	52,2	52,1	51,8	52,3	52,0	44,9
Pas très bien (%)	21,3	21,8	21,7	22,2	20,6	19,6	14,3
Pas bien du tout (%)	5,1	5,4	5,4	5,5	4,7	4,4	5,4

TABLE 4.7 – Conciliation vie professionnelle / vie personnelle

Dans cette partie du questionnaire, la question où les différences sont le plus marquées est celle qui porte sur la conciliation entre vie familiale et professionnelle. La question posée est "En général, vos horaires de travail s'accordent-ils avec vos engagements sociaux et familiaux en dehors de votre travail ?". Ainsi après les différents calages qui réduisent légèrement les écarts,

la proportion de personnes pour lesquelles vie professionnelle et vie personnelle se concilient très bien est de 24 % dans QVT contre 35 % dans CT. Il n'est pas impossible que le fait de parvenir à fixer un rendez-vous avec un enquêteur (pour CT), marque en soi et dans une certaine mesure, le fait d'arriver à concilier vie professionnelle et vie personnelle (plus que de répondre à QVT en tout cas). Notons tout de même que les écarts sont bien moindres si on cumule les modalités "Très bien" et "Bien" (76 % dans QVT contre 80 % dans CT).

	B	CNR	C1	C2	C3	C4	CT
Ressenti sur la paye							
Très bien payé (%)	1,1	1,0	0,9	1,0	1,1	1,2	2,0
Bien payé (%)	11,8	11,1	10,6	10,6	11,4	12,1	16,5
Normalement payé (%)	47,9	47,0	46,9	46,7	48,4	48,5	43,3
Plutôt mal payé (%)	33,9	35,0	35,4	35,1	33,4	32,8	31,4
Très mal payé (%)	5,3	5,9	6,1	6,4	5,7	5,4	6,3

TABLE 4.8 – Ressenti sur la paye

Sur la question du ressenti sur la paye, les personnes sont plus nombreuses à répondre qu'elles sont bien ou très bien payées dans CT que dans QVT et l'effet des différents calages est assez limité pour rapprocher les estimations.

Contraintes physiques

	B	CNR	C1	C2	C3	C4	CT
Travail impose							
rester longtemps debout (%)	42,8	45,4	47,3	48,5	48,4	47,4	50,8
autre posture pénible (%)	27,7	29,6	30,9	32,1	30,5	28,5	35,1
porter charges lourdes (%)	31,1	33,9	35,9	37,4	36,7	35,5	41,9
respirer fumées ou poussières (%)	21,7	24,1	25,2	26,6	25,8	24,7	31,4
contact produits dangereux (%)	19,2	20,6	21,2	22,0	21,4	20,6	29,6

TABLE 4.9 – Contraintes physiques

Pour les questions objectives sur les contraintes physiques de posture et de port de charges au travail³, le calage sur marges sociodémographiques tend à rapprocher les estimations de QVT avec celles de CT même si des écarts subsistent.

Les écarts sont assez marqués pour la question du contact avec les produits dangereux et celle sur le fait de respirer des fumées ou des poussières : pour rapprocher les estimations, il aurait été sans doute judicieux de pouvoir caler sur le secteur d'activité si une telle variable avait été

3. Il est possible qu'il existe un réel effet de mode sur la question de l'"autre posture pénible" : si l'on se restreint à la population spécifique des ouvriers du bâtiment, il y a un écart assez conséquent d'environ 10 points entre CT et QVT (après calages sur marges sociodémographiques) sur cette question alors que les écarts sont moindres pour la posture debout ou le port de charges lourdes. Il faut néanmoins nuancer ce constat en raison de la faiblesse des effectifs (une soixantaine d'observations dans QVT, environ le double pour CT).

disponible⁴.

Pour les différentes questions de ce bloc de questions, les écarts sont toujours dans le même sens : les contraintes physiques sont plus fréquentes dans CT que dans QVT. L'expérimentation QVT parvient sans doute moins que CT à faire participer les personnes dont le travail est physiquement le plus pénible mais le calage sur marges sociodémographiques parvient dans une certaine mesure à réduire cet écueil⁵. En revanche, les calages supplémentaires (santé et bien-être) ont plutôt tendance à augmenter les écarts : pour ces questions objectives sur les conditions de travail, les marges sociodémographiques sont les plus pertinentes et pourraient encore être affinées pour rapprocher les estimations de CT et QVT.

Rythme de travail

	B	CNR	C1	C2	C3	C4	CT
Le rythme est imposé par							
le déplacement d'une pièce (%)	7,4	8,2	8,7	9,1	8,7	8,2	7,5
la cadence d'une machine (%)	5,4	5,9	6,3	6,8	6,4	6,1	6,9
d'autres contraintes techniques (%)	18,4	19,1	19,2	19,6	18,7	18,0	17,4
la dépendance vis-à-vis de collègues (%)	33,4	33,7	33,6	33,3	31,6	30,8	28,2
des normes de productions, délais (%)	21,4	22,4	23,0	23,4	22,3	21,4	26,9
une demande extérieure (%)	60,1	60,0	59,7	59,4	58,1	57,6	59,4
une contrôle informatisé (%)	37,6	37,2	37,3	36,2	35,1	34,4	34,4

TABLE 4.10 – Rythme de travail

Sur les questions de dépendance du rythme de travail, les estimations de QVT et CT sont globalement assez proches, à l'exception de la question sur les normes de production. Les effets des différents calages sont relativement limités : le schéma général pour QVT est que les calages sociodémographiques tendent à augmenter les proportions en question alors que les calages ultérieurs les diminuent.

	B	CNR	C1	C2	C3	C4	CT
Rythme de travail							
Travaille à la chaîne (%)	2,7	3,1	3,4	3,7	3,5	3,4	3,7
Doit fréquemment interrompre tâche (%)	58,4	57,8	57,1	56,4	54,6	53,9	64,1
Objectifs précis à atteindre (%)	32,9	32,4	31,9	32,1	31,1	30,4	31,4
En contact avec le public (%)	73,2	73,3	73,3	72,5	72,8	72,7	73,1

TABLE 4.11 – Rythme de travail

4. Si le questionnement de QVT aborde l'activité de l'établissement, cela n'est pas suffisant pour coder la NAF à un niveau suffisamment détaillé pour pouvoir l'utiliser pour ce type de correction.

5. Le caractère auto-administré de l'expérimentation peut écarter, par sa nature, les personnes les moins à l'aise avec l'écrit. Intégrer des variables de diplôme et de nationalité permet dans une certaine mesure de tenir compte de cette sélection sur la littératie. Mais cela n'est peut-être pas suffisant et pourrait éventuellement expliquer les écarts constatés entre CT et QVT sur les contraintes physiques si les personnes les moins à l'aise avec l'écrit sont plutôt celles qui ont les métiers les plus pénibles physiquement.

Les proportions de personnes qui travaillent à la chaîne, de celles qui ont des objectifs précis à atteindre et de celles qui sont en contact avec le public sont très proches dans QVT et CT. Les effets des calages sont assez limités sur ces questions. En revanche, l'écart est assez marqué pour ceux qui déclarent devoir fréquemment interrompre leur tâche qui sont plus nombreux dans CT que dans QVT.

	B	CNR	C1	C2	C3	C4	CT
Obligé de vous dépêcher ?							
Toujours (%)	19,6	19,6	19,7	19,8	17,9	17,0	17,5
Souvent (%)	40,2	39,5	39,0	39,2	38,4	38,1	29,8
Parfois (%)	35,3	35,7	36,0	35,7	37,8	38,6	35,6
Jamais (%)	4,9	5,2	5,3	5,3	5,9	6,3	17,1

TABLE 4.12 – Obligé de vous dépêcher ?

De très gros écarts sont constatés dans les modalités de réponse choisies à la question "Êtes-vous obligé de vous dépêcher?". Ainsi, après calages, ils sont 38 % à choisir "souvent" dans QVT, contre 30 % dans CT et ils sont 6 % à choisir "jamais" dans QVT contre 17 % dans CT. Sur cette question, l'écart est tel qu'on ne peut l'imputer au seul effet de sélection non contrôlé. Un élément possible d'explication est que l'enchaînement des questions diffère quelque peu dans CT et QVT. Ainsi, dans QVT vient d'abord la question sur l'interruption de tâche puis celle sur les objectifs et enfin celle sur le fait de se dépêcher alors que pour CT la première question posée est celle sur le fait de se dépêcher avant d'aborder les interruptions puis les objectifs. Il s'avère en effet que le lien entre la question sur les objectifs et celle sur le fait de se dépêcher est un peu plus fort dans QVT que dans CT : les mesures d'association entre variables nominales que sont le coefficient de contingence C et le V de Cramer⁶ sont respectivement de 0,201 et 0,206 dans QVT contre 0,190 et 0,187 dans CT⁷.

Autonomie

	B	CNR	C1	C2	C3	C4	CT
Autonomie							
Peut interrompre travail momentanément (%)	67,7	66,6	65,6	66,2	67,2	67,9	71,2
Travail répète même gestes (%)	31,6	34,1	36,5	36,8	35,9	34,6	40,3
Travail apprend choses nouvelles (%)	78,7	77,3	76,1	76,0	77,9	78,4	78,9

TABLE 4.13 – Autonomie

Les proportions de personnes qui peuvent interrompre momentanément leur travail sont assez proches dans CT et QVT. C'est également le cas pour les proportions de personnes dont

6. Ces mesures neutralisent l'influence de la taille d'échantillon de la statistique du χ^2 . Pour rappel, $C = \sqrt{\chi^2 / (N + \chi^2)}$ et $V = \sqrt{\chi^2 / (N(\min(r - 1, c - 1)))}$ avec N la taille d'échantillon, r et c le nombre de modalités des 2 variables analysées.

7. De même, si l'on considère que les deux variables sont ordinales, les coefficients de corrélation de Pearson, Spearman ou encore le τ de Kendall, etc. sont plus élevés dans QVT que dans CT.

le travail permet d'apprendre des choses nouvelles. En revanche, la proportion de personnes pour lesquelles le travail consiste en une répétition de gestes est plus importante dans CT que dans QVT.

Ambiance

	B	CNR	C1	C2	C3	C4	CT
Collègues							
Aidé par collègues (%)	73,7	73,6	73,8	72,7	73,0	73,2	76,5
Désaccords avec collègues (%)	64,7	64,1	63,3	61,8	60,4	59,7	51,8
Tensions avec collègues (%)	25,4	24,8	24,6	24,0	22,1	21,2	21,3

TABLE 4.14 – Collègues

Sur les différentes questions autour de l'ambiance avec les collègues de travail, la principale différence entre CT et QVT est que les personnes sont plus nombreuses à être en désaccord avec leurs collègues sur la façon de bien faire le travail dans QVT que dans CT. L'écart est d'environ 13 points avant redressements mais se réduit en partie avec les différents calages puisque la proportion après calages est de 60 % dans QVT contre 52 % dans CT.

	B	CNR	C1	C2	C3	C4	CT
Au travail ...							
Contact avec personnes en détresse (%)	44,3	43,1	42,6	41,3	40,5	39,6	44,7
Doit calmer gens (%)	55,0	53,9	53,2	51,6	50,7	49,9	52,8

TABLE 4.15 – Calme, détresse

Si l'on se contente du calage sur marges sociodémographiques, les proportions d'individus qui sont en contact avec des personnes en détresse et celles qui doivent calmer des gens sont assez proches dans CT et QVT. Les calages additionnels ont plutôt tendance à dégrader la situation.

Évolutions

	B	CNR	C1	C2	C3	C4	CT
Évolutions							
Craintes pour emploi (%)	24,4	25,7	26,2	26,4	24,4	23,5	24,9
Pense changer qualif/métier (%)	25,1	26,5	26,9	26,2	24,6	23,8	26,9
Si perte, facile de trouver emploi (%)	24,2	25,6	26,4	26,4	27,3	27,8	36,9
Capable jusqu'à retraite (%)	51,1	49,3	48,3	48,1	51,0	53,8	60,7
Souhait jusqu'à retraite (%)	46,8	44,9	44,5	44,8	47,5	48,9	56,1

TABLE 4.16 – Évolutions

Les réponses aux questions sur les évolutions sont un peu plus pessimistes dans QVT que dans CT. Les proportions de personnes qui pensent qu'il sera facile de retrouver un emploi si elles le perdent, qui se pensent capables de faire le même travail jusqu'à la retraite ou qui souhaitent faire le même travail jusqu'à la retraite, sont plus faibles dans QVT que dans CT. Les différents calages et en particulier ceux intégrant le score de bien-être tendent à légèrement réduire les écarts.

Sur les questions de crainte quant à la perte de l'emploi ou le besoin de changer de qualification, les proportions sont sensiblement du même ordre dans CT et QVT après calages sur marges sociodémographiques, les calages additionnels ayant même plutôt tendance à fournir un redressement trop pessimiste de ces questions.

Notons par ailleurs que si l'écart est d'environ 10 points entre QVT et CT sur facilité à retrouver un emploi en cas de perte, il est encore beaucoup plus marqué sur les personnes dont le statut est plutôt précaire (intérim, contrats aidés, contrat à durée déterminée ou sans contrat de travail) : alors que dans QVT, 27 % de ces personnes déclarent pouvoir retrouver un emploi en cas de perte (soit une proportion proche des personnes en CDI ou titulaires de la fonction publique), ils sont 44 % dans CT, preuve sans doute que la question n'est pas comprise de la même manière.

Mais sur ces questions du module sur les évolutions, un élément est à prendre en compte : il s'agit de la seule partie du questionnaire commun à CT et QVT où la modalité "Ne sais pas" est proposée explicitement. Cette modalité est assez fréquemment choisie mais beaucoup plus dans QVT que dans CT : ainsi pour les questions de craintes pour l'emploi, d'évolution de métier, de retrouver un emploi en cas de perte, de tenir jusqu'à la retraite et de le souhaiter, les proportions de modalités de "Ne sais pas" sont respectivement de 14 %, 30 %, 33 %, 17 % et 17 % dans QVT contre 4 %, 13 %, 17 %, 4 % et 3 % dans CT. Alors que pour le reste des résultats de cette partie, le différentiel de "Ne sais pas" entre QVT et CT n'est pas tel qu'il remette en cause les résultats, on peut souhaiter, pour ce module, exprimer les résultats en excluant les "Ne sais pas". On obtient alors des résultats beaucoup plus proches entre les deux enquêtes, à savoir, si l'on reprend l'ordre des questions du tableau les proportions sont respectivement de 27 %, 34 %, 42 %, 65 % et 59 % dans QVT (après calage C4) et de 26 %, 31 %, 44 %, 64 % et 58 % dans CT.

Partie auto-administrée du questionnaire à la fois dans CT et QVT

Relations avec les autres au travail

Globalement, les réponses fournies dans QVT sont plus pessimistes que dans CT (table 4.17). Comme dans les deux cas, le questionnaire est auto-administré, il est possible que cela tienne plus des effets de sélection non contrôlée que de l'effet de mode pur même si le contexte du questionnaire et les modes de collecte ne sont pas tout à fait similaires. Le calage sur marges tend à réduire les écarts avec une forte contribution du score de bien-être à la convergence des estimations. Néanmoins pour certaines questions, en particulier celles qui portent sur la reconnaissance du travail, les écarts restent non négligeables.

Les estimations sont globalement assez proches entre les réponses redressées de QVT et celles de CT (table 4.18), les différents calages, et en particulier l'utilisation du score de bien-être, contribuant à ce rapprochement.

	B	CNR	C1	C2	C3	C4	CT
Pas d'accord ou Pas du tout d'accord							
Mon supérieur prête attention à ce que je dis (%)	21,2	21,0	21,0	20,3	17,7	16,8	17,0
Mon supérieur m'aide à mener mes tâches à bien (%)	26,6	26,1	25,8	24,9	22,2	21,3	17,8
Travail bien connu par ceux qui évaluent (%)	32,2	31,5	31,1	29,9	27,5	26,6	22,7
Mon supérieur traite équitablement ... (%)	35,2	35,0	35,1	33,9	30,9	29,9	25,8
Les collègues avec qui je travaille m'aident ... (%)	11,8	11,6	11,5	11,1	10,0	9,6	9,0
Reconnaissance du travail (%)	36,8	37,0	37,2	36,4	32,6	31,4	26,9
Perspectives de promotion satisfaisantes (%)	54,8	54,7	54,6	53,0	49,7	48,6	43,7
Quantité de travail excessive (%)	44,9	45,3	45,8	44,4	46,4	47,7	50,6
Peut organiser travail comme il convient (%)	19,1	19,7	20,1	19,7	17,6	16,7	17,1

TABLE 4.17 – Relations au travail

	B	CNR	C1	C2	C3	C4	CT
Toujours / souvent / parfois							
Je travaille sous pression (%)	86,5	85,6	84,7	84,5	82,2	81,4	80,6
Je continue à penser à mon travail même quand ... (%)	90,4	89,6	89,1	89,2	88,2	87,9	84,8
Je sais à l'avance quelles tâches ... (%)	79,7	78,1	77,5	77,1	77,5	77,5	70,1
Je vis des changements imprévisibles (%)	76,3	75,7	75,0	75,0	72,6	71,5	68,0
J'ai la possibilité de faire des choses qui me plaisent (%)	91,9	91,1	90,6	90,6	91,8	92,0	92,8
Je suis fier de travailler dans cette entreprise (%)	91,7	91,0	90,8	90,8	92,4	92,6	91,3
Je dois cacher mes émotions (%)	75,0	74,3	74,3	73,6	70,8	69,6	69,3
Je dois éviter de donner mon avis (%)	66,3	65,5	65,1	64,0	61,2	59,9	58,0
Il m'arrive d'avoir peur au travail (%)	39,0	39,8	40,2	40,4	38,7	36,9	37,6
Je dois faire des choses que je désapprouve (%)	43,8	43,5	43,2	42,4	39,5	38,2	63,8

TABLE 4.18 – Relations au travail

Une première différence importante porte sur le fait de savoir à l'avance quelles tâches sont prévues le mois suivant : la proportion de personnes qui répondent par l'affirmative est beaucoup plus forte dans QVT (78 % après calages) que dans CT (70 %).

Mais surtout, et il s'agit là d'un des écarts les plus spectaculaires de l'ensemble du questionnaire, les personnes qui disent devoir faire des choses qu'elles désapprouvent (toujours, souvent ou parfois) sont beaucoup moins nombreuses dans QVT (38 % après calages) que dans CT (64 %). Le résultat est d'autant plus remarquable qu'il est quelque peu contre-intuitif. Plus précisément, 32 % des personnes font parfois des choses qu'elles désapprouvent dans QVT contre 54 % dans CT (table 4.19). Un tel écart pourrait difficilement s'expliquer en termes d'effet de mode pur : rappelons que cette partie de questionnaire est également auto-administrée dans CT. Il est fort probable que l'enchaînement des questions place l'individu répondant dans des contextes bien différents. En effet, dans QVT, cette question se situe à la fin du bloc alors que dans CT elle suit immédiatement la question "J'ai la possibilité de faire des choses qui me plaisent" ⁸.

8. Une autre piste d'explication aurait pu être que les répondants de QVT, et en particulier sur internet choisissent la modalité "Jamais" (y compris en revenant sur la question) pour éviter de devoir répondre à la question ouverte qui suit. Si tel était le cas, la proportion de "Jamais" devrait être beaucoup plus importante sur internet que sur papier, où il est beaucoup plus aisé de sauter une question, mais cela n'est pas vraiment le

	B	CNR	C1	C2	C3	C4	CT
Des choses que je désapprouve							
Toujours (%)	1,0	1,1	1,2	1,1	1,0	1,0	1,2
Souvent (%)	6,2	6,3	6,5	6,3	5,4	4,9	8,2
Parfois (%)	36,7	36,0	35,5	35,0	33,2	32,3	54,3
Jamais (%)	56,0	56,4	56,7	57,4	60,3	61,6	35,9

TABLE 4.19 – Des choses que je désapprouve

Satisfactions et difficultés au travail

	B	CNR	C1	C2	C3	C4	CT
A vécu situation au travail : qqun ...							
vous ignore (%)	21,6	21,8	21,8	21,3	19,1	18,3	19,5
vous empêche de vous exprimer (%)	11,4	11,5	11,6	11,4	9,9	9,2	11,4
vous ridiculise en public (%)	8,0	8,2	8,2	8,2	7,2	6,8	6,9
critique injustement votre travail (%)	24,2	24,7	24,7	24,3	22,0	21,0	20,1
vous charge de tâches inutiles (%)	8,7	9,2	9,4	9,1	7,9	7,4	8,4
sabote votre travail (%)	9,6	9,9	9,8	9,7	8,7	8,0	7,8
laisse entendre que vous êtes mentalement dérangé (%)	2,9	3,0	3,0	3,0	2,6	2,3	2,9
vous dit des choses obscènes (%)	4,0	4,3	4,4	4,4	3,8	3,5	4,4
vous fait des propositions à caractère sexuel (%)	0,8	0,8	0,9	0,8	0,7	0,7	1,0
vous fait des blagues blessantes (%)	8,0	8,5	8,7	8,6	7,4	7,0	6,5

TABLE 4.20 – Difficultés au travail

	B	CNR	C1	C2	C3	C4	CT
Au travail, victime							
d'une agression verbale - public (%)	22,2	22,2	22,2	21,4	20,3	19,9	18,0
d'une agression physique ou sexuelle - public (%)	2,1	2,2	2,3	2,2	2,0	2,0	2,5
d'une agression verbale - collègues / supérieurs (%)	17,1	16,8	16,7	16,3	14,6	13,8	11,9
d'une agression physique - collègues / supérieurs (%)	0,6	0,6	0,7	0,7	0,6	0,5	0,7

TABLE 4.21 – Agressions au travail

Ces blocs de questions interrogent l'individu sur des situations assez concrètes de difficultés vécues au travail (tables 4.20 et 4.21). Les estimations issues de QVT sont relativement proches, après calages, de celles de CT. Les calages, dans leur ensemble, contribuent assez fortement à la convergence des estimations même si les calages additionnels sont, dans certains cas, superflus.

cas : s'ils sont un peu moins nombreux à choisir cette modalité sur papier, ils sont tout de même 56 % à le faire.

Synthèse de la comparaison "linéaire"

Cette comparaison blocs par blocs, variables par variables fournit des éléments quant à la compréhension des écarts entre CT et QVT selon la nature des questions, que l'on peut essayer de synthétiser.

Certaines questions présentent des écarts entre CT et QVT, qui tiennent probablement, pour partie du moins, au positionnement dans le questionnaire plus que de l'effet de mode lié à la mesure. Par exemple, la question sur les choses que les personnes doivent faire mais désapprouvent au travail présente de tels écarts que, manifestement, la question n'est pas comprise de la même manière dans les deux enquêtes. Le positionnement de cette question dans CT est peut-être problématique : il est possible que l'enchaînement avec la question sur la possibilité de faire des choses qui plaisent peut perturber le répondant. Reste que de tels écarts ne sont pas facilement interprétables...

D'autres questions présentent des écarts qui tiennent plutôt au fait que les explications qui accompagnent la question ne sont pas les mêmes : il s'agit par exemple de la question sur le travail le dimanche.

On peut parler d'effet de mode lié à la mesure lorsque la modalité "Ne sais pas" explicitement présentée dans le questionnaire est beaucoup plus fréquemment choisie dans QVT que dans CT. C'est ce qui se produit dans le module "Évolutions" et que l'on peut interpréter comme du *satisficing* : en l'absence d'enquêteur, le répondant n'est pas toujours disposé à effectuer les efforts nécessaires pour reporter la réponse "optimale", d'où un choix plus fréquent de la modalité "Ne sais pas".

Des questions subjectives présentent des écarts que l'on pourrait interpréter en tant qu'effets de mode liés à la mesure et plus précisément à la désirabilité sociale. En effet, dans le cadre d'une interaction avec un enquêteur, il est plus difficile d'émettre des opinions qui dévient, même légèrement, de ce que l'on pense être la norme sociale que seul face à son questionnaire auto-administré. On peut fournir une interprétation de ce type pour la question sur la conciliation ou celle du ressenti sur la paye où les réponses de QVT sont plus pessimistes que dans CT. Il n'est pas impossible non plus que les écarts relèvent pour partie de la sélection non contrôlée : le ressenti des individus sur leurs conditions de travail serait plutôt pessimiste chez les répondants de QVT et cela serait en partie, mais pas totalement, corrigé par l'utilisation du score de bien-être dans le calage.

Certaines questions assez factuelles et objectives (table 4.9) présentent des écarts que l'on peut probablement imputer aux limites des calages mis en oeuvre pour QVT qui, par exemple, n'intègrent pas l'activité des établissements : QVT a probablement du mal à mobiliser les personnes qui ont les métiers les plus physiquement pénibles et le redressement ne parvient que partiellement à combler cet écueil.

Enfin, tout un ensemble de variables présentent des écarts que l'on peut qualifier de raisonnables à l'issue des différents calages : c'est le cas, comme attendu, de la plupart des variables auto-administrées du questionnaire mais aussi d'un bon nombre de variables dont le mode

de collecte diffère dans CT et QVT.

Notons enfin que les contributions des 4 calages réalisés à la réduction des écarts entre CT et QVT diffèrent selon le type de variable. Par exemple le calage intégrant le score de bien-être contribue à la réduction des écarts sur les questions subjectives mais peut, au contraire, accentuer l'écart pour les questions plus factuelles. Symétriquement, les calages sociodémographiques sont efficaces pour réduire les écarts sur les questions factuelles, mais peuvent être neutres voire légèrement dégrader la situation initiale pour les questions subjectives.

Une autre perspective, présentée en annexe C, utilise des indicateurs synthétiques produits à partir d'analyses des correspondances multiples. Cinq dimensions sont étudiées : l'autonomie, les exigences émotionnelles, l'intensité du travail, la reconnaissance et les rapports sociaux. Il apparaît que les différents calages rapprochent globalement, au final, l'estimation des moyennes des indicateurs synthétiques de CT et QVT. De fait, ce sont les calages sur le score de bien-être et la santé ressentie qui contribuent fortement à ce rapprochement. Ce constat est à rapprocher des observations faites lors de la comparaison entre variables de QVT et CT au fil du questionnaire : au niveau variable, des écarts peuvent être conséquents mais les différents calages réalisés contribuent quand même à rapprocher globalement les estimations issues de CT et QVT en parvenant à contrôler une partie des différences liées à la sélection.

Conclusion

Si le taux de réponse de l'expérimentation Qualité de la vie au travail est plutôt meilleur que certaines autres enquêtes expérimentales et auto-administrées réalisées à l'Insee (Razafindranovona, 2016), il reste néanmoins beaucoup plus faible que celui de l'enquête de référence à laquelle on la compare : son taux de collecte est d'environ 40 % alors que celui de l'enquête Conditions de travail est légèrement supérieur à 70 %. Et le gain, en termes de taux de réponse, à proposer la possibilité de répondre par papier n'est pas complètement anecdotique même s'il peut sembler un peu limité (45 % pour le protocole internet et papier contre 38 % pour le protocole uniquement internet).

Inférer sur la population cible à partir de tels taux de réponse n'est pas évident et nécessite d'émettre des hypothèses fortes lors des phases de correction et de disposer de variables pertinentes pour redresser l'échantillon de répondants.

En effet, le caractère auto-administré du protocole de QVT sélectionne fortement la population de répondants. Des ajustements par calage sur marges sont ainsi nécessaires pour rendre les populations de CT et QVT "comparables", en termes de principales caractéristiques sociodémographiques, et donc contrôler une partie des différences liées à la sélection.

La forte sélection des répondants peut être particulièrement problématique lorsqu'elle est liée, dans une certaine mesure, à la thématique de l'enquête. La sélection s'opère ainsi lors des phases de contact avec l'enquêté : comme le montre l'exploitation de l'expérimentation contrôlée sur les modèles de lettre-avis, les éléments de contact peuvent avoir un certain impact sur les motivations à participer.

En dehors de certaines questions très particulières, les ordres de grandeur des variables d'intérêt recueillies dans CT se retrouvent dans QVT. Il y a néanmoins plusieurs cas que l'on peut schématiquement distinguer :

- pour certaines questions assez factuelles, les fréquences brutes (avant tout traitement) de QVT sont proches de celles de CT ;
- pour d'autres questions factuelles, le calage sur marges sociodémographiques contribue fortement à réduire les écarts entre QVT et CT ;
- certaines questions plus subjectives présentent des écarts assez importants après les calages sociodémographiques mais plus mesurés après les calages sur le score de bien-être et la santé ressentie ;
- enfin les écarts restent significatifs pour un nombre non négligeable de questions pour des raisons qui tiennent soit à la sélection (calage sociodémographique pas assez fin pour certaines questions factuelles, problème de caractéristiques inobservables qui participent à la sélection pour les questions plus subjectives) soit à la mesure (pessimisme plus important dans QVT qui tiendrait à la désirabilité sociale moins forte en auto-administré).

Si l'on souhaite reconduire un dispositif auto-administré similaire pour inférer sur une population de personnes en emploi, une réflexion est probablement à mener sur le type de questions supplémentaires que l'on pourrait souhaiter poser afin de caler les réponses à des marges, y compris externes à l'enquête, pertinentes car corrélées aux principales variables d'intérêt et participant au processus de sélection des répondants : par exemple, pour étudier les conditions de travail, il serait pertinent de pouvoir disposer d'une variable qui caractérise assez finement l'activité de l'établissement où travaille la personne.

Par ailleurs une autre difficulté est liée au nombre de variables concernées dans ce type d'enquête. La repondération (si un mode de référence peut être identifié) ne règle pas les effets de mesure : la "correction" est alors uniforme pour toutes les variables, elle peut permettre de corriger certains écarts, mais pas tous, en particulier si les effets de mode sont distincts d'une variable à l'autre. Il est très compliqué, voire impossible de travailler variable par variable : les travaux sont alors très volumineux, peuvent impliquer la production de plusieurs pondérations, etc. Des techniques d'imputation peuvent être envisagées, au risque de rompre les corrélations entre variables. Analyser la structure des corrélations selon les modes est alors un préalable à la mise en oeuvre d'imputations. L'utilisation de méthodes d'analyse des données pourrait être une solution.

Reste que l'effort principal doit porter sur l'amélioration du taux de réponse. Il faut, dans la mesure du possible, éviter que les répondants soient essentiellement des personnes mobilisées très directement par la thématique de l'enquête. Ceci n'est pas toujours évident car des arguments convaincants et ciblés doivent être avancés pour éveiller l'intérêt de l'enquêté...

Bibliographie

- AMIEL, M.-H. et DENOYELLE, T. (2012). Enquêtes en ligne : comparaison de modes de questionnement sur le thème du logement. In *Journées de Méthodologie Statistique*. http://jms.insee.fr/files/documents/2012/945_4-JMS2012_S26-3_AMIEL-ACTE.PDF.
- BUÉ, J., COUTROT, T. et GUIGNON, N. (2008). L'évolution des conditions de travail. *Insee Références - L'emploi, nouveaux enjeux*.
- CLARK, A. E. et VICARD, A. (2007). Conditions de collecte et santé subjective : une analyse sur données européennes. *Économie et statistique*, 403(1):143–163.
- DAVIE, E. (2015). Méthode de construction d'indicateurs synthétiques de conditions de travail et de risques psychosociaux. *DGAFP, Note*.
- de PERETTI, G. et RAZAFINDRANOVONA, T. (2014). Les enquêtes multimode : attention aux effets de mode. *Statistique et société*, 2.
- GIVORD, P. (2010). Méthodes économétriques pour l'évaluation de politiques publiques. *INSEE, Document de travail de la DESE, G2010-08*. http://www.insee.fr/fr/publications-et-services/docs_doc_travail/G2010-08.pdf.
- GROVES, R. M. et LYBERG, L. (2010). Total survey error : Past, present, and future. *Public Opinion Quarterly*, 74(5):849–879.
- HAINMUELLER, J. (2012). Entropy balancing for causal effects : A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, 20(1):25–46.
- HECKMAN, J. J., ICHIMURA, H. et TODD, P. E. (1997). Matching as an econometric evaluation estimator : Evidence from evaluating a job training programme. *The review of economic studies*, 64(4):605–654.
- HECKMAN, J. J., ICHIMURA, H. et TODD, P. E. (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies*, 65(2):261–294.
- LUGTIG, P., LENSVELT-MULDERS, G. J., FRERICHS, R. et GREVEN, F. (2011). Estimating nonresponse bias and mode effects in a mixed mode survey. *International Journal of Market Research*, 53(5):669–686.
- LYBERG, L. (2012). La qualité des enquêtes. *Techniques d'enquête*, 38(2):115–142.
- MCCARTY, J. A. et SHRUM, L. J. (2000). The measurement of personal values in survey research : A test of alternative rating procedures. *Public Opinion Quarterly*, 64(3):271–298.
- RAZAFINDRANOVONA, T. (2015). La collecte multimode et le paradigme de l'erreur d'enquête totale. *INSEE, Document de travail de la DMCSI, M2015-01*. http://www.insee.fr/fr/publications-et-services/docs_doc_travail/la%20collecte%20multimode-b.pdf.

- RAZAFINDRANOVONA, T. (2016). Exploitation de l'enquête expérimentale vols, violence et sécurité. *INSEE, Document de travail de la DMCSI, M2015-03*. http://www.insee.fr/fr/publications-et-services/docs_doc_travail/Exploitation.pdf.
- RAZAFINDRANOVONA, T., de PERETTI, G. et BARRAU, A. (2013). The philosophy of French experiments on internet and mixed-mode data collection. In *UNECE Seminar on statistical data collection*.
- ROSENBAUM, P. R. et RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.
- RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.
- RUBIN, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational and Behavioral statistics*, 2(1):1–26.
- SAUTORY, O. (1993). La macro calmar. *Redressement d'un échantillon par calage sur marges, Série des documents de travail de la Direction des Statistiques Démographiques et Sociales*, 55.
- TOPP, C. W., ØSTERGAARD, S. D., SØNDERGAARD, S. et BECH, P. (2015). The who-5 well-being index : a systematic review of the literature. *Psychotherapy and psychosomatics*, 84(3):167–176.
- VANNIEUWENHUYZE, J., LOOSVELDT, G. et MOLENBERGHS, G. (2014). Evaluating mode effects in mixed-mode survey data using covariate adjustment models. *Journal of Official Statistics*, 30(1):1–21.

Annexe A

Quelques éléments sur la partie de l'expérimentation QVT réalisée à La Réunion

En fin de collecte, le nombre global de questionnaires collectés pour cette opération est de 1 100, ce qui correspond à un taux de collecte de 37 %. Pour environ 12 % des individus contactés, les courriers reviennent en NPAI¹. En excluant les NPAI, le taux de réponse des individus pour lesquels un contact, via la réception de la lettre-avis, a pu être établi s'élève ainsi à 41 %.

De manière encore plus marquée qu'en métropole, les répondants sur internet sont plus aisés que les répondants papier et que les personnes sélectionnées dans l'échantillon

	Répondants internet	Répondants papier	NPAI	Pas de réponse	Total
Protocole : internet+papier	117	0	16	2 867	3 000
Protocole : internet seul	116	0	15	2 869	3 000

TABLE A.1 – Retours avant la première relance (QVT - La Réunion)

	Répondants internet	Répondants papier	NPAI	Pas de réponse	Total
Protocole : internet+papier	262	61	69	2 608	3 000
Protocole : internet seul	279	0	36	2 685	3 000

TABLE A.2 – Retours avant la deuxième relance (QVT - La Réunion)

	Répondants internet	Répondants papier	NPAI	Pas de réponse	Total
Protocole : internet+papier	350	248	167	2 235	3 000
Protocole : internet seul	489	0	189	2 322	3 000

TABLE A.3 – Retours à la fin de la collecte (QVT - La Réunion)

1. N'habite pas à l'adresse indiquée.

	Échantillon	Éch hors NPAI	QVT internet	QVT papier
Age (moyenne)	43,2	43,9	43,3	46,3
Femmes (proportion)	50 %	50 %	51 %	54 %
Propriétaires (proportion)	56 %	59 %	62 %	62 %
Revenus (euros annuels, médiane)	16 400	16 600	22 100	16 800

TABLE A.4 – Caractéristiques des répondants (QVT - La Réunion)

Annexe B

Internet et papier : des différences entre ces modes auto-administrés ?

Cette partie vise à exploiter l'expérimentation QVT pour essayer de répondre à une question parfois éludée dans la littérature : lorsque les populations de répondants sont "comparables", les réponses collectées sont-elles similaires sur internet et sur papier ?

B.1 En théorie : pas trop de différences entre modes auto-administrés

La littérature qui traite des effets de mode fait essentiellement la distinction entre, d'un côté, les modes auto-administrés et de l'autre, les modes avec intermédiation d'enquêteurs (Razafindranovona, 2015). C'est pour cette raison que le choix a été fait dans l'exploitation principale de l'expérimentation QVT de ne pas différencier le traitement des réponses selon le mode de collecte (internet ou papier). Quelques investigations sur cette similarité sont proposées dans cette partie en utilisant une méthode de *kernel matching* pour contrôler les différences de sélection entre la population de répondants sur papier et celle de répondants sur internet.

B.2 Contrôle de la sélection par kernel matching sur score de propension

Le contexte est ici proche de celui de l'évaluation de l'effet causal d'un traitement où le traitement serait le fait de répondre par l'un des deux modes de collecte considérés. Pour un panorama complet des méthodes pour évaluer un effet causal, on pourra se référer à Givord (2010).

Rappelons tout d'abord brièvement le cadre canonique de Rubin (1974) de l'évaluation économétrique.

Le cadre canonique de Rubin

Dans le cadre de Rubin, le traitement est représenté par la variable D : si la personne est traitée, alors $D = 1$, sinon $D = 0$. Soit une variable d'intérêt Y . Chaque personne peut être dans deux états possibles, 0 ou 1, avec comme variables d'intérêt "potentielles" respectivement

Y_0 et Y_1 . Le gain provenant du traitement est $\Delta = Y_1 - Y_0$.

Le problème de l'évaluation est que l'on ne connaît Δ pour personne étant donné que l'on observe $Y = DY_1 + (1 - D)Y_0$, c'est-à-dire, soit Y_0 , soit Y_1 .

Ainsi, pour une personne ayant reçu le traitement, on observe Y_1 : mais quelle serait la valeur de Y si la personne n'avait pas reçu le traitement ? L'essence des méthodes d'évaluation consiste à estimer ces situations contrefactuelles.

Une mesure de l'impact d'un programme que l'on souhaite classiquement obtenir est l'effet moyen du traitement dans la population des individus traités (*average treatment effect on the treated*) à savoir :

$$E(Y_1 - Y_0 | D = 1)$$

Un estimateur naïf consisterait à comparer la variable d'intérêt des traités avec celle des non traités mais il existe un biais de sélection dès lors que Y_0 et D ne sont pas indépendantes.

Effet du traitement conditionnel aux caractéristiques observables

L'effet moyen du traitement dans la population des individus traités, conditionnellement aux caractéristiques observables X est :

$$E(Y_1 - Y_0 | D = 1, X)$$

La moyenne $E(Y_1 - Y_0 | D = 1, X)$ est estimée à partir des données des participants aux programmes. En revanche, des hypothèses doivent être formulées pour estimer la moyenne contrefactuelle $E(Y_0 | D = 1, X)$. En l'absence d'expérience contrôlée, $E(Y_0 | D = 0, X)$ est utilisée pour approcher $E(Y_0 | D = 1, X)$. Le biais de sélection provenant de cette approximation est :

$$B(X) = E(Y_0 | D = 1, X) - E(Y_0 | D = 0, X)$$

Pour construire le contrefactuel, l'hypothèse souvent invoquée est $Y_0 \perp D | X$. En fait, pour identifier l'effet du traitement, l'hypothèse est que le conditionnement par X élimine le biais ($B(X) = 0$) c'est-à-dire que :

$$E(Y_0 | D = 1, X) = E(Y_0 | D = 0, X)$$

Principes généraux du matching

Pour estimer l'effet du traitement conditionnellement aux caractéristiques observables X , une méthode d'évaluation très populaire est celle du matching. Le matching est basé sur l'idée intuitive de comparer la variable d'intérêt des traités avec la variable d'intérêt d'individus comparables (par leurs caractéristiques observables) qui eux ne sont pas traités. La différence entre ces variables d'intérêt est alors attribuée au traitement. Cette méthode fait donc l'hypothèse forte que le biais de sélection peut être purgé en conditionnant sur un ensemble de caractéristiques observables. Un avantage de cette méthode est qu'elle ne suppose pas de spécifier la forme fonctionnelle de la relation entre variable d'intérêt et caractéristiques observables.

Un point important concernant l'utilisation de ces méthodes est que pour pouvoir comparer un individu traité avec un non traité aux caractéristiques similaires X , il faut s'assurer de

l'existence de cet individu non traité mais similaire par ses caractéristiques X . Ce point se formalise à travers une hypothèse dite de support commun qui doit être vérifiée :

$$0 < P(D = 1|X) < 1$$

Nous notons dorénavant I_0 et I_1 l'ensemble des individus respectivement non-traités et traités et N_0 , N_1 leur nombre.

La méthode initialement proposée par Rubin (1977) consiste à appairer chaque individu traité i à un individu non traité comparable $\tilde{i}(i)$, au sens des caractéristiques observables X . Par exemple, on peut prendre pour contrefactuel l'individu non traité dont l'indice vérifie :

$$\tilde{i}(i) = \operatorname{argmin}_{j \in I_0} \|X_i - X_j\|$$

où $\| \cdot \|$ est une distance ; on peut prendre par exemple la distance de Mahalanobis.

L'estimateur de l'effet sur les traités est alors :

$$\frac{1}{N_1} \sum_{i \in I_1} [Y_i - Y_{\tilde{i}(i)}]$$

Dans cette version simple du matching, le contrefactuel de l'individu traité se fait à partir d'un seul individu comparable. Il est également possible d'utiliser plusieurs individus comparables et de pondérer alors selon la distance à l'individu traité selon les caractéristiques observables.

Ainsi, une forme plus générale de l'estimateur de matching est proposée par Heckman, Ichimura et Todd (1997 et 1998) :

$$\sum_{i \in I_1} \omega_{N_0, N_1}(i) [Y_{1i} - \sum_{j \in I_0} W_{N_0, N_1}(i, j) Y_{0j}]$$

$W_{N_0, N_1}(i, j)$ est une pondération définie telle que pour tout $i \in I_1$, $\sum_{j \in I_0} W_{N_0, N_1}(i, j) = 1$.

Le choix d'une fonction de pondération reflète le choix d'une distance particulière dans la méthode de matching et les poids sont basés sur les distances dans l'espace des X .

Les $\omega_{N_0, N_1}(i)$ servent à sélectionner différents domaines de l'espace des X ou à tenir compte de l'hétéroscédasticité dans l'échantillon des traités.

Ainsi les différentes méthodes de matching correspondent à des choix de différentes fonctions de pondération $\omega_{N_0, N_1}(i)$ et $W_{N_0, N_1}(i, j)$.

Soit le voisinage $C(X_i)$ pour une personne de l'échantillon des traités. Les voisins de i sont les personnes $j \in I_0$ pour lesquelles $X_j \in C(X_i)$. Les personnes appariées à i sont celles de l'ensemble A_i tel que $A_i = \{j \in I_0 | X_j \in C(X_i)\}$.

Par exemple l'estimateur de matching de Rubin, dit du plus proche voisin présenté précédemment, consiste à prendre $\omega_{N_0, N_1}(i) = 1/N_1$, $C(X_i) = \min_{j \in I_0} \|X_i - X_j\|$ et où le schéma de pondération consiste à assigner tout le poids au seul appariement, soit $W_{N_0, N_1}(i, j) = 1$ si $j \in A_i$ et $W_{N_0, N_1}(i, j) = 0$ sinon.

Score de propension et matching

Pour que l'hypothèse d'indépendance conditionnelle soit vérifiée, on peut souhaiter effectuer l'appariement sur un grand nombre de variables. Mais les proches voisins peuvent alors être difficiles à trouver. En fait, les estimateurs sont d'autant plus biaisés que le nombre de variables de conditionnement est élevé. Un moyen de résoudre ce problème de dimensionnalité est d'utiliser le score de propension.

Rosenbaum et Rubin (1983) ont montré que si la propriété d'indépendance conditionnelle $Y_0 \perp D|X$ est vérifiée alors, en notant $P(X) = P(D = 1|X)$, si on a également $0 < P(X) < 1$, on a l'indépendance $Y_0 \perp D|P(X)$. L'estimation de l'effet sur les traités peut alors s'effectuer en appariant sur le score de propension. En pratique, si après avoir conditionné par l'estimation du score de propension, la distribution des X dépend encore du traitement D , cela suggère une mauvaise spécification.

Précisons que si le problème de la dimensionnalité semble résolu, en pratique, le score de propension devra être estimé (car le score exact n'est que rarement connu) souvent par des spécifications *logit* ou *probit*. Alors qu'un des intérêts du matching est son caractère *a priori* non paramétrique, le matching sur score de propension va souvent reposer en partie sur des hypothèses de spécification.

Kernel matching

Les méthodes de matching sur les plus proches voisins n'utilisent qu'une partie de l'information disponible : on peut souhaiter utiliser pour chaque individu traité l'ensemble de l'échantillon des non traités. Les méthodes de *kernel matching* reposent ainsi sur le fait que chaque individu non traité participe au contrefactuel de l'individu i avec une importance qui varie selon la distance avec l'individu considéré.

En reprenant les notations précédentes, l'estimateur de *kernel matching* s'écrit avec :

$$\omega_{N_0, N_1}(i) = \frac{1}{N_1} \quad , \quad A_i = I_0 \quad \text{et} \quad W_{N_0, N_1}(i, j) = \frac{G_{ij}}{\sum_{k \in I_0} G_{ik}} \quad \text{avec} \quad G_{ik} = G\left(\frac{X_k - X_i}{a_{N_0}}\right)$$

G étant une fonction noyau (positive, symétrique par rapport à 0 et telle que $\int_{-\infty}^{+\infty} G(u)du = 1$) et a_{N_0} un paramètre de fenêtre.

Des exemples de noyaux souvent utilisés sont le noyau gaussien et le noyau d'Epanechnikov.

Par définition de la fonction noyau, plus l'individu du groupe de contrôle est "proche" de l'individu traité plus sa contribution est importante dans la construction du contrefactuel. Le choix de la fenêtre relève d'un arbitrage classique entre biais et variance : plus la fenêtre est grande, plus la variance diminue mais en contrepartie le biais est également plus élevé. Par ailleurs, avec ce type d'estimateur où tous les individus du groupe de non-traités sont utilisés pour construire le contrefactuel de l'individu traité, l'hypothèse de support commun est très importante et le praticien devra y prêter une certaine attention.

Dans la version où le matching s'effectue sur le score de propension, cet estimateur peut s'écrire avec $P_i = P(D = 1|X_i)$:

$$\frac{1}{N_1} \sum_{i \in I_1} [Y_{1i} - \frac{\sum_{j \in I_0} G(\frac{P_j - P_i}{a_{N_0}}) Y_{0j}}{\sum_{k \in I_0} G(\frac{P_k - P_i}{a_{N_0}})}]$$

B.3 Résultats

Les travaux présentés dans cette partie ont été réalisés en collaboration avec Alexis Barrau (Razafindranovona *et al.*, 2013), en stage à la division Recueil et traitement de l'information de l'Insee.

B.3.1 Modélisation du score de propension

Pour modéliser la propension de répondre via l'un des deux modes de collecte (Lutig *et al.*, 2011), le choix a donc été fait de se concentrer sur les variables sociodémographiques, *a priori* peu sensibles aux effets de mode, et présentant de plus l'avantage d'être présentes en début de questionnaire (avant donc qu'un éventuel effet de lassitude différentielle entre les deux modes n'ait pu jouer). Finalement, le choix s'est porté sur l'utilisation des variables suivantes : le sexe, le fait d'être en couple, le niveau de diplôme, l'âge, le fait d'être à temps partiel, les revenus et le groupe social. Les variables qui n'étaient pas significatives (comme la nationalité ou le fait d'avoir un enfant) ou trop corrélées avec le groupe social (type de contrat notamment) n'ont pas été conservées afin de ne pas perturber l'estimation de la probabilité, et de conserver une certaine lisibilité quant à l'interprétation des résultats.

Le graphique B.1 représente les distributions du score de propension au sein des populations de répondants. Ce graphique confirme l'existence d'un support commun, hypothèse fondamentale pour mettre en oeuvre les techniques de *matching*.

B.3.2 Exemple pour illustrer la correction du biais de sélection

Les variables intégrées dans la modélisation du score de propension semblent permettre de contrôler correctement la sélection entre les deux populations de répondants. Cela peut, par exemple, s'illustrer en observant les fréquences de modalités avant et après contrôle de la sélection pour la question "À votre emplacement de travail, êtes-vous amené(e) à respirer des fumées ou des poussières?". Au premier abord, cette question semble ne pas devoir être soumise à trop d'effets de mode, même s'il est évidemment impossible de le dire *a priori*, la question étant relativement factuelle, avec un nombre très limité de modalités possibles. Par ailleurs, compte tenu de la différence de composition des deux sous-populations de répondants, en particulier le plus grand nombre d'ouvriers parmi les répondants papier, une différence nette entre les deux sous-échantillons est attendue.

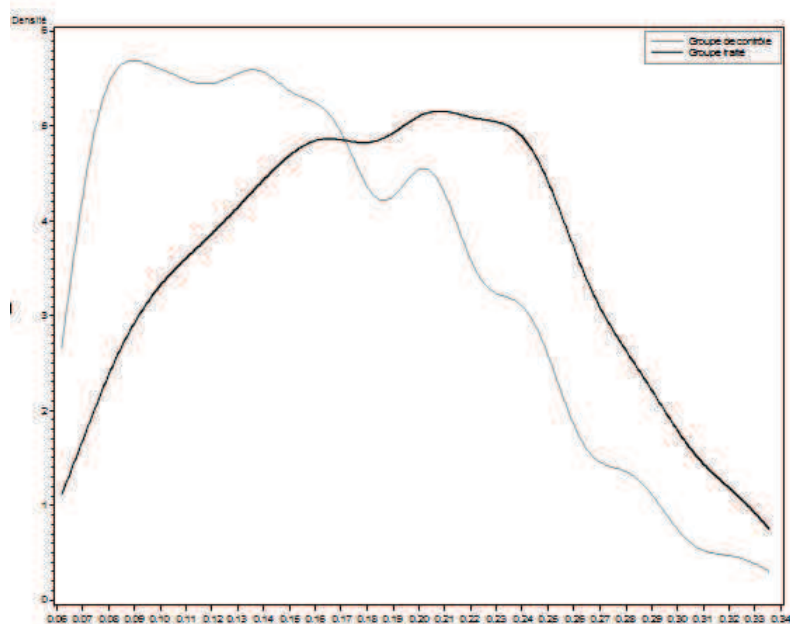


FIGURE B.1 – Score de propension et support commun

En effet, la comparaison brute des proportions de "oui" dans les deux sous-populations montre une différence assez marquée, de l'ordre de 5 points : 25,4 % sur papier contre 20,8 % sur internet. En revanche, après contrôle de la sélection par appariement, les différences sont pratiquement nulles puisque l'estimation de l'effet causal moyen est de l'ordre de 0,2 point, avec un intervalle de confiance à 5 %, calculé par bootstrap, qui est $[-1,3 ; 2,6]$.

B.3.3 Utilisation différente du spectre des réponses proposées

Les effets de mode peuvent prendre plusieurs formes, et peuvent notamment se manifester par une utilisation différente du spectre des réponses proposées, les individus cochant plus souvent les modalités extrêmes dans un mode que dans un autre par exemple.

Un indicateur a été mobilisé pour repérer des différences dans la différenciation des réponses entre les deux modes. Cet indice est proposé par McCarty et Shrum (McCarty et Shrum, 2000).

Si l'on considère un individu i , une batterie de D questions qui ont les mêmes N modalités (notées l), l'indice de différenciation ID_i est :

$$ID_i = 1 - \sum_{l=1}^N P_{li}^2$$

$$P_{li} = \frac{\sum_{k=1}^D I_{\{Y_{ki}=l\}}}{D}$$

Ainsi, on obtient un indice de différenciation qui sera à son maximum lorsque l'individu utilise avec la même fréquence toutes les modalités et à son minimum (zéro) lorsque l'individu coche la même modalité pour toutes les questions.

Le bloc de questions étudié porte sur les relations avec les autres au travail. La variable de différenciation sur ce bloc de questions présente une moyenne de 0,62 et un écart-type de 0,1. Surtout, la répartition est différente entre les deux modes puisque la moyenne est plus importante de 0,015 sur Internet par rapport au papier. Cette différence après contrôle de la sélection par appariement est du même ordre bien que légèrement moindre (0,01). Cet effet causal est certes limité, mais pas totalement négligeable car si on le rapporte à la moyenne pour l'ensemble de l'échantillon, cela représente une déviation d'environ 2 %, et cela correspond de plus à 10 % environ de l'écart-type. La différenciation des réponses semble donc un peu moindre sur papier ce qui peut s'expliquer par le processus cognitif : le report des réponses est plus "mécanique" sur papier et favorise des enchaînements de réponses typiques du *satisficing*.

Q72	Toujours	Souvent	Parfois	Jamais
Je travaille sous pression.....	☐ 1	☐ 2	☐ 3	☐ 4
Je continue à penser à mon travail même quand je n'y suis pas.....	☐ 1	☐ 2	☐ 3	☐ 4
Je sais à l'avance quelles tâches j'aurai le mois suivant.....	☐ 1	☐ 2	☐ 3	☐ 4
Je vis des changements imprévisibles ou mal préparés.....	☐ 1	☐ 2	☐ 3	☐ 4
J'ai la possibilité de faire des choses qui me plaisent.....	☐ 1	☐ 2	☐ 3	☐ 4
Je suis fier(fière) de travailler dans cette entreprise, organisation.....	☐ 1	☐ 2	☐ 3	☐ 4
Je dois cacher mes émotions ou faire semblant d'être de bonne humeur.....	☐ 1	☐ 2	☐ 3	☐ 4
Je dois éviter de donner mon avis, mon opinion.....	☐ 1	☐ 2	☐ 3	☐ 4
Il m'arrive d'avoir peur pendant mon travail, pour ma sécurité ou celle des autres.....	☐ 1	☐ 2	☐ 3	☐ 4
Là où je travaille, la santé et la sécurité des salariés passent avant la productivité et la performance.....	☐ 1	☐ 2	☐ 3	☐ 4
Je dois faire des choses que je désapprouve.....	☐ 1	☐ 2	☐ 3	☐ 4

→ Q74

FIGURE B.2 – Bloc de questions Q72

B.3.4 Score de bien-être

Le score de bien-être a déjà été présenté dans le chapitre 4. Il est imputé à partir d'un bloc de questions qui interroge l'individu sur son ressenti général, dans la vie de tous les jours. Rappelons que pour calculer ce score, pour chacune des 5 questions du bloc, on impute 5 points à une réponse "tout le temps", 4 points à une réponse "la plupart du temps" ... et 0 point à une réponse "jamais". Le total de points est ensuite multiplié par 4 pour obtenir un score entre 0 et 100.

Avant contrôle de la sélection, le score de bien-être moyen sur Internet est de 56,5 contre 59,2 pour les répondants papiers, soit une différence brute de 2,7 points. L'effet est de la même ampleur lorsqu'il est estimé avec contrôle par appariement : l'écart observé ne peut donc pas être expliqué par les seules différences de caractéristiques sociodémographiques des individus. Une explication possible est que les conditions de passation du questionnaire sur papier soient plus propices que sur internet à une ambiance apaisée qui induirait un score, en moyenne, légèrement plus élevé. Il est également possible que les variables de contrôle intégrées ici n'expliquent pas à elles seules les différences de sélection entre les deux sous-populations.

Q79	À quelle fréquence avez-vous ressenti les affirmations suivantes, au cours des deux dernières semaines dans <u>votre vie de tous les jours</u>, <u>au travail et en dehors</u> ?	Tout le temps	La plupart du temps	Plus de la moitié du temps	Moins de la moitié du temps	De temps en temps	Jamais
	Je me suis senti(e) bien et de bonne humeur	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
	Je me suis senti(e) calme et tranquille	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
	Je me suis senti(e) plein(e) d'énergie et vigoureux(se)	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
	Je me suis réveillé(e) en me sentant frais(fraiche) et dispos(e)	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
	Ma vie quotidienne a été remplie de choses intéressantes	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6

FIGURE B.3 – Bloc de questions Q79

Annexe C

Comparaisons sur indicateurs synthétiques

Pour aller au-delà d'une comparaison qui parcourt, de manière linéaire, le questionnaire, une solution peut être de mobiliser des indicateurs synthétiques. Pour cela, on s'inspirera très fortement des développements réalisés à la Dares et à la DGAFP et en particulier de la méthode dite *a priori* (Davie, 2015).

C.1 Indicateurs synthétiques issus de CT

L'objectif de la construction d'indicateurs synthétiques est de définir des indicateurs partagés, utiles à des fins d'études. Pour éviter l'écueil de l'arbitraire, les indicateurs s'appuient sur des critères statistique de maximisation de l'information des données d'enquête disponibles.

La méthode *a priori* repose sur une analyse de référence, issue d'un travail pluridisciplinaire reconnu, d'experts, mené par le collège d'expertise sur le suivi statistique des risques psychosociaux (RPS) au travail. Elle présente le grand avantage de pouvoir être reproduite sur différentes enquêtes ; les indicateurs produits à partir d'analyses des correspondances multiples (ACM) peuvent être comparés d'une enquête à l'autre sous réserve que le questionnaire ne soit pas trop modifié, et permettent d'étudier des évolutions. En revanche, des corrélations intéressantes entre des variables présentes dans les données d'enquête peuvent ne pas être mises en évidence.

La première étape consiste ici à déterminer des dimensions, pour lesquelles des indicateurs synthétiques sont souhaités et de spécifier les variables de l'enquête à prendre en compte dans chacune de ces dimensions. Neuf dimensions ont ainsi été choisies : les six dimensions définies par le collège d'expertise sur le suivi statistique des RPS au travail (l'intensité du travail, les exigences émotionnelles, l'autonomie, les rapports sociaux, les conflits de valeur et l'insécurité économique) ainsi qu'une dimension relative aux contraintes horaires et d'organisation du temps de travail, une dédiée aux contraintes physiques et environnementales et une concernant la reconnaissance, la rémunération et l'évaluation.

Pour chaque dimension, un certain nombre de variables de l'enquête sont sélectionnées. Une

ACM est ensuite réalisée sur ces variables. Elle permet de créer l'indicateur synthétique correspondant via les coordonnées des variables sur le premier axe.

C.2 Construction des indicateurs pour comparaison CT/QVT

Le questionnement n'étant pas le même dans CT et QVT, on doit se restreindre aux variables à la fois présentes dans CT et QVT : les dimensions ne sont donc pas tout à fait les mêmes et on va même abandonner certaines dimensions au vu du faible nombre de variables communes pour ces dimensions en question.

Les différents indicateurs et les variables qu'ils mobilisent, à la fois disponibles dans CT et QVT, sont présentés ci-dessous.

Autonomie

- Les supérieurs hiérarchiques disent comment faire ;
- Vous faites généralement appel à d'autres en cas d'incident ;
- Possibilité d'interrompre momentanément son travail quand on le souhaite ;
- Pouvoir organiser son travail de la manière qui convient le mieux (pas d'accord / pas du tout d'accord) ;
- Ennui ;
- Faire des choses qui me plaisent ;
- Possibilité d'apprendre des choses nouvelles ;
- Travailler à la chaîne ;
- Répéter continuellement une même série de gestes ou d'opérations ;

Exigences émotionnelles

- Cacher ses émotions ;
- Éviter de donner son avis ;
- Avoir peur pendant son travail pour sa sécurité ou celles des autres ;
- Être victime d'une agression verbale de la part du public ;
- Être victime d'une agression physique ou sexuelle de la part du public ;
- Être en contact avec des personnes en situation de détresse ;
- Devoir calmer des gens ;
- Vivre des tensions dans ses rapports avec le public ;
- Contact avec le public.

Intensité du travail

- Devoir se dépêcher pour faire son travail ;
- Devoir fréquemment abandonner une tâche pour une autre non prévue ;
- Devoir atteindre des objectifs chiffrés précis ;
- Avoir la possibilité de modifier ces objectifs ;
- Avoir du mal à atteindre les objectifs fixés ;
- Recevoir des ordres ou des indications contradictoires ;
- Être soumis à au moins 3 contraintes de rythme ;

- On me demande une quantité de travail excessive ;
- Travailler sous pression ;
- Penser à son travail même quand je n'y suis pas ;
- Au cours des 12 derniers mois, être allé travailler tout en pensant que vous auriez dû rester à la maison parce que vous étiez malade.

Reconnaissance

- Recevoir l'estime et le respect que mérite mon travail ;
- Mes perspectives de promotion sont satisfaisantes ;
- Être plutôt mal payé ou très mal payé ;
- Ma position professionnelle actuelle correspond à ma formation ;
- Les personnes qui évaluent mon travail le connaissent bien ;
- Mon supérieur traite équitablement les personnes qui travaillent sous ses ordres ;
- Fierté de travailler dans cette organisation ;
- Avoir au moins un entretien d'évaluation par an ;
- L'entretien porte sur des critères précis et mesurables.

Rapports sociaux

- En cas de travail délicat, être aidé par ses supérieurs hiérarchiques ;
- En cas de travail délicat, être aidé par les collègues ;
- Être plutôt mal payé ou très mal payé ;
- Depuis 1 an, la plupart des personnes avec qui je travaille ont changé ;
- Être en désaccord avec ses supérieurs sur la façon de bien faire son travail (toujours, souvent) ;
- Être en désaccord avec ses collègues sur la façon de bien faire son travail (toujours, souvent) ;
- Vivre des situations de tension dans ses rapports avec vos supérieurs hiérarchiques ;
- Vivre des situations de tension dans ses rapports avec ses collègues ;
- Vivre des situations de tension dans ses rapports avec les personnes encadrées ;
- Il m'arrive de superviser le travail d'autres salariés ;
- Mon supérieur prête attention à ce que je dis ;
- Mon supérieur m'aide à mener mes tâches à bien ;
- Mes collègues m'aident à mener mes tâches à bien ;
- Au cours des 12 derniers mois, une personne vous ignore, fait comme si vous n'étiez pas là ;
- (...) Vous empêche de vous exprimer ;
- (...) Vous ridiculise en public ;
- (...) Critique injustement votre travail ;
- (...) Vous charge de tâches inutiles ou dégradantes ;
- (...) Sabote votre travail, vous empêche de travailler correctement ;
- (...) Laisse entendre que vous êtes mentalement dérangé ;
- (...) Vous dit des choses obscènes ou dégradantes ;
- (...) Vous fait des propositions à caractère sexuel de façon insistante ;
- (...) Vous fait des blagues blessantes ou de mauvais goût, se moque de vous ;
- Avoir subi une agression verbale de la part des collègues ou des supérieurs ;
- Avoir subi une agression physique ou sexuelle de la part des collègues ou des supérieurs ;

- Avoir l'impression de faire partie d'une équipe.

C.3 Résultats

Les moyennes pondérées de chacun des indicateurs synthétiques au fur et à mesure des traitements post-collecte de QVT ainsi que la valeur de référence issue de CT (nulle, par construction) sont données dans le tableau C.1.

	B	CNR	C1	C2	C3	C4	CT
Indicateurs synthétiques							
Autonomie	0,27	0,37	0,46	0,44	0,29	0,24	0,00
Exigences émotionnelles	0,46	0,43	0,42	0,34	0,21	0,14	0,00
Intensité du travail	0,76	0,70	0,66	0,63	0,37	0,25	0,00
Reconnaissance	-0,13	-0,08	-0,07	0,03	-0,19	-0,26	0,00
Rapports sociaux	0,58	0,62	0,62	0,58	0,19	0,02	0,00

TABLE C.1 – Indicateurs synthétiques (bruts)

B : pondérations brutes, CNR : après CNR totale, C1 : +calage sexe, âge, diplôme et nationalité, C2 : + marge CS (niveau 1), C3 : + marge score de bien-être, C4 : + marge santé ressentie, CT : enquête Conditions de travail après redressements

Et pour mieux comprendre les valeurs des indicateurs, nous allons préciser rapidement la signification des dimensions, en fonction des coordonnées des variables sur les premiers axes des différentes. Pour schématiser, vont se projeter à droite des différents axes, les situations plutôt négatives, de difficultés au travail :

- à droite de la dimension **Autonomie** : peu ou pas d'autonomie, d'épanouissement dans le travail ;
- à droite de la dimension **Exigences émotionnelles** : fortes émotions et tensions dans le travail ;
- à droite de la dimension **Intensité du travail** : forte intensité et pression dans le travail ;
- à droite de la dimension **Reconnaissance** : peu ou pas de reconnaissance dans le travail ;
- à droite de la dimension **Rapports sociaux** : rapports conflictuels, tensions et difficultés dans le travail.

Il apparaît assez nettement que les différents calages rapprochent globalement, au final, l'estimation des moyennes des indicateurs synthétiques de CT et QVT ; en particulier, ce sont les calages sur le score de bien-être et la santé ressentie qui contribuent fortement à ce rapprochement. Le rapprochement est assez net en ce qui concerne les indicateurs sur les exigences émotionnelles, l'intensité du travail et les rapports sociaux. Le constat est beaucoup plus nuancé pour l'indicateur sur l'autonomie et surtout celui sur la reconnaissance.

La correction de la non-réponse et les calages sociodémographiques sont plutôt neutres en ce qui concerne les indicateurs sur les rapports sociaux et l'intensité du travail voire dégradent la situation pour l'indicateur sur l'autonomie. Pour l'indicateur sur les exigences émotionnelles,

chacun des calages contribue au rapprochement entre CT et QVT.

Ce constat est à rapprocher des observations faites lors de la comparaison entre variables de QVT et CT au fil du questionnaire : au niveau variable, des écarts peuvent être conséquents mais les différents calages réalisés contribuent quand même à rapprocher globalement les estimations issues de CT et QVT en parvenant à contrôler une partie des différences liées à la sélection.

Série des Documents de Travail « Méthodologie Statistique »

9601 : Une méthode synthétique, robuste et efficace pour réaliser des estimations locales de population.
G. DECAUDIN, J.-C. LABAT

9602 : Estimation de la précision d'un solde dans les enquêtes de conjoncture auprès des entreprises.
N. CARON, P. RAVALET, O. SAUTORY

9603 : La procédure FREQ de SAS - Tests d'indépendance et mesures d'association dans un tableau de contingence.
J. CONFAIS, Y. GRELET, M. LE GUEN

9604 : Les principales techniques de correction de la non-réponse et les modèles associés.
N. CARON

9605 : L'estimation du taux d'évolution des dépenses d'équipement dans l'enquête de conjoncture : analyse et voies d'amélioration.
P. RAVALET

9606 : L'économétrie et l'étude des comportements. Présentation et mise en œuvre de modèles de régression qualitatifs. Les modèles univariés à résidus logistiques ou normaux (LOGIT, PROBIT).
S. LOLLIVIER, M. MARPSAT, D. VERGER

9607 : Enquêtes régionales sur les déplacements des ménages : l'expérience de Rhône-Alpes.
N. CARON, D. LE BLANC

9701 : Une bonne petite enquête vaut-elle mieux qu'un mauvais recensement ?
J.-C. DEVILLE

9702 : Modèles univariés et modèles de durée sur données individuelles.
S. LOLLIVIER

9703 : Comparaison de deux estimateurs par le ratio stratifiés et application

aux enquêtes auprès des entreprises.

N. CARON, J.-C. DEVILLE

9704 : La faisabilité d'une enquête auprès des ménages.
1. au mois d'août.
2. à un rythme hebdomadaire
C. LAGARENNE, C. THIESSET

9705 : Méthodologie de l'enquête sur les déplacements dans l'agglomération toulousaine.
P. GIRARD.

9801 : Les logiciels de désaisonnalisation TRAMO & SEATS : philosophie, principes et mise en œuvre sous SAS.
K. ATTAL-TOUBERT, D. LADIRAY

9802 : Estimation de variance pour des statistiques complexes : technique des résidus et de linéarisation.
J.-C. DEVILLE

9803 : Pour essayer d'en finir avec l'individu Kish.
J.-C. DEVILLE

9804 : Une nouvelle (encore une !) méthode de tirage à probabilités inégales.
J.-C. DEVILLE

9805 : Variance et estimation de variance en cas d'erreurs de mesure non corrélées ou de l'intrusion d'un individu Kish.
J.-C. DEVILLE

9806 : Estimation de précision de données issues d'enquêtes : document méthodologique sur le logiciel POULPE.
N. CARON, J.-C. DEVILLE, O. SAUTORY

9807 : Estimation de données régionales à l'aide de techniques d'analyse multidimensionnelle.
K. ATTAL-TOUBERT, O. SAUTORY

9808 : Matrices de mobilité et calcul de la précision associée.
N. CARON, C. CHAMBAZ

9809 : Échantillonnage et stratification : une étude empirique des gains de précision.
J. LE GUENNEC

9810 : Le Kish : les problèmes de réalisation du tirage et de son extrapolation.
C. BERTHIER, N. CARON, B. NEROS

9901 : Perte de précision liée au tirage d'un ou plusieurs individus Kish.
N. CARON

9902 : Estimation de variance en présence de données imputées : un exemple à partir de l'enquête Panel Européen.
N. CARON

0001 : L'économétrie et l'étude des comportements. Présentation et mise en œuvre de modèles de régression qualitatifs. Les modèles univariés à résidus logistiques ou normaux (LOGIT, PROBIT) (version actualisée).
S. LOLLIVIER, M. MARPSAT, D. VERGER

0002 : Modèles structurels et variables explicatives endogènes.
J.-M. ROBIN

0003 : L'enquête 1997-1998 sur le devenir des personnes sorties du RMI - Une présentation de son déroulement.
D. ENEAU, D. GUILLEMOT

0004 : Plus d'amis, plus proches ? Essai de comparaison de deux enquêtes peu comparables.
O. GODECHOT

0005 : Estimation dans les enquêtes répétées : application à l'Enquête Emploi en Continu.
N. CARON, P. RAVALET

0006 : Non-parametric approach to the cost-of-living index.
F. MAGNIEN, J. POUGNARD

0101 : Diverses macros SAS : Analyse exploratoire des données, Analyse des séries temporelles.
D. LADIRAY

0102 : Économétrie linéaire des panels : une introduction.
T. MAGNAC

0201 : Application des méthodes de calages à l'enquête EAE-Commerce.
N. CARON

C 0201 : Comportement face au risque et à l'avenir et accumulation patrimoniale - Bilan d'une expérimentation.
L. ARRONDEL, A. MASSON, D. VERGER

C 0202 : Enquête Méthodologique Information et Vie Quotidienne - Tome 1 : bilan du test 1, novembre 2002.
J.-A. VALLET, G. BONNET, J.-C. EMIN, J. LEVASSEUR, T. ROCHER, P. VRIGNAUD, X. D'HAULTFOEUILLE, F. MURAT, D. VERGER, P. ZAMORA

0203 : General principles for data editing in business surveys and how to optimise it.
P. RIVIERE

0301 : Les modèles logit polytomiques non ordonnés : théories et applications.
C. AFSA ESSAFI

0401 : Enquête sur le patrimoine des ménages - Synthèse des entretiens monographiques.
V. COHEN, C. DEMMER

0402 : La macro SAS CUBE d'échantillonnage équilibré
S. ROUSSEAU, F. TARDIEU

0501 : Correction de la non-réponse et calage de l'enquête Santé 2002
N. CARON, S. ROUSSEAU

0502 : Correction de la non-réponse par répondération et par imputation
N. CARON

0503 : Introduction à la pratique des indices statistiques - notes de cours
J-P BERTHIER

0601 : La difficile mesure des pratiques dans le domaine du sport et de la culture - bilan d'une opération méthodologique
C. LANDRE, D. VERGER

0801 : Rapport du groupe de réflexion sur la qualité des enquêtes auprès des ménages
D. VERGER

M2013/01 : La régression quantile en pratique
P. GIVORD, X. D'HAULTFOEUILLE

M2014/01 : La microsimulation dynamique : principes généraux et exemples en langage R
D. BLANCHET

M2015/01 : la collecte multimode et le paradigme de l'erreur d'enquête totale
T. RAZAFINDROVONA

M2015/02 : Les méthodes de Pseudo-Panel
M. GUILLERM

M2015/03 : Les méthodes d'estimation de la précision pour les enquêtes ménages de l'Insee tirées dans Octopusse

E. GROS – K.MOUSSALAM

M2016/01 : Le modèle Logit
Théorie et application.
C. AFSA

M2016/02 : Les méthodes d'estimation de la précision de l'Enquête Emploi en Continu
E. GROS – K.MOUSSALAM

M2016/03 : Exploitation de l'enquête expérimentale Vols, violence et sécurité.
T. RAZAFINDROVONA

M2016/04 : Savoir compter, savoir coder. Bonnes pratiques du statisticien en programmation.
E. L'HOUE – R. LE SAOUT B. ROUPPERT

M2016/05 : Les modèles multiniveaux
P. GIVORD – M. GUILLERM

M2016/06 : Econométrie spatiale : une introduction pratique
P. GIVORD – R. LE SAOUT

M2016/07 : La gestion de la confidentialité pour les données individuelles
M. BERGEAT

M2016/08 : Exploitation de l'enquête expérimentale Logement internet-papier
T. RAZAFINDROVONA